# *Amateur Physics for the Amateur Pool Player*

***Third Edition***

*Ron Shepard*

Ron Shepard
Argonne Pool League
Argonne National Laboratory
Argonne, IL 60439
email: shepard@tcg.anl.gov

Copyright © Ron Shepard
3rd Edition 1997
2nd Edition 1996
1st Edition 1994

The author grants permission to individuals for personal use to make copies of this manuscript and to store and distribute this document in electronic form. All rights regarding the commercial use of this manuscript are reserved by the author.

## Introduction

The word *amateur* is based on the Latin words *amator* (a lover) and *amare* (to love). An amateur is someone who loves what he does, and pursues it for the pleasure of the act itself. These notes are intended for the pool player who enjoys playing the game, and who enjoys understanding how things work using the language of physics. There is probably very little pool playing technique discussed in this manuscript that will be new to the experienced pool player, and likewise, there is little physics that will be new to the experienced physicist. However, there will be hopefully new pool technique for the interested physicist and new physics for the interested pool player. The tone of the presentation is not directed necessarily toward either the pool student or the physics student, but rather toward the amateur who enjoys both. The physics that is used here is not derived from first principles; it is assumed that the reader is familiar with such ideas as Newton's laws of motion, center of mass transformations, moments of inertia, linear and angular acceleration, geometry, trigonometry, and vector notation. Reference to a calculus-based introductory college level physics textbook should be sufficient to understand fully any of the physics used or mentioned in this text. *The Feynman Lectures on Physics* (Vol. 1) is one such text that the reader will find enjoyable.

This discussion is divided into five sections. Section 1 discusses the equipment (balls, tables, cue sticks, cue tip, cloth) and some of its associated properties (various friction coefficients, forces, moments of inertia), section 2 discusses the concept of natural roll, section 3 discusses the cue tip and cue ball impact, section 4 discusses collisions between balls, and section 5 discusses the use of statistical methods. Each section includes some general discussion and specific problems (along with their solutions). Some exercises are also given along the way; it is intended for the reader to experiment on a pool table with some of the techniques that have been discussed.

## 1. Properties of the Equipment

Pool, billiard, and snooker balls are uniform spheres of, usually, a phenolic resin type of plastic. Older balls have been made of clay, ivory, wood, and other materials. On coin-operated tables, the cue ball is sometimes larger and heavier than the other balls; otherwise, all the balls in a set are the same size and weight. Standard pool balls are $2\frac{1}{4}$" in diameter, snooker balls are either of two sizes, $2\frac{1}{8}$" or $2\frac{1}{16}$", and carom billiard balls are one of three sizes, $2\frac{27}{64}$", $2\frac{3}{8}$", or $2\frac{7}{16}$". Tolerances in all cases are ±0.005". Pool balls weigh 5.5 to 6*oz*, snooker balls weigh 5 to 5.5*oz*, and billiard balls weigh 7 to 7.5*oz*.

**Problem 1.1:** What is the volume of a pool ball in terms of its radius *R*?
*Answer*: In spherical coordinates, the volume of a sphere is given by

$$V = \int_0^{2\pi}\int_0^{\pi}\int_0^{R} r^2 \sin\theta\, dr\, d\theta\, d\varphi = \left(\tfrac{1}{3}R^3\right)(2)(2\pi) = \tfrac{4}{3}\pi R^3$$

where *R* is the radius of the ball. Assuming that the ball is a perfect sphere, the minimum radius is $R_{min}$=1.1225" and the maximum radius is $R_{max}$=1.1275". The volume of a standard pool ball is between $\frac{4}{3}\pi R_{\min}^3 = 5.924 in^3 = 97.08 cm^3$ and $\frac{4}{3}\pi R_{\max}^3 = 6.004 in^3 = 98.39 cm^3$.

**Problem 1.2:** In order to satisfy the size and weight limits, what is the density range of the ball material in units of $g/cm^3$?
*Answer*: The density is the mass divided by the volume, $\rho = M/V$. The minimum mass is 5.5*oz*(28.35*g*/*oz*)=155.9*g*, and the maximum mass is 6.0*oz*(28.35*g*/*oz*)=170.1*g*. The minimum density is $\rho_{min} = M_{min}/V_{max} = 1.559g/98.39cm^3 = 1.58 g/cm^3$ and the maximum density is $\rho_{max} = M_{max}/V_{min} = 170.1g/97.08cm^3 = 1.75 g/cm^3$. For comparison, the density of water at room termperature is $0.997 g/cm^3$, a saturated sucrose (table sugar) solution is $1.44 g/cm^3$, a saturated cesium chloride solution is $1.89 g/cm^3$, and the density of mercury is $13.6 g/cm^3$, so a pool ball should easily sink in water, slowly sink in the sugar solution, barely float in the cesium chloride solution, and easily float in mercury.

The inertia tensor of a rigid body is defined as the elements of the 3 by 3 matrix

$$I_{ij} = \int_V \rho(\mathbf{r}) \left[\delta_{ij} \sum_k r_k^2 - r_i r_j\right] dv$$

where the components of the vector **r**=(*x*,*y*,*z*) are the cartesian coordinates. For a uniform sphere, $\rho(\mathbf{r}) = \rho$ is a constant for $r<R$ and is the density of the ball material. The mass of the ball is $M = \rho V = \frac{4}{3}\rho\pi R^3$.

**Problem 1.3:** Determine the inertia tensor for a ball in terms of *M* and *R*.

*Answer*: Taking the moment of inertia about the *x*-axis gives

$$I_{xx} = \int_V \rho(\mathbf{r})\left(z^2 + y^2\right)dv = S_{zz} + S_{yy} = 2S_{zz}$$

It is interesting to notice that the moment of inertia about the *x*-axis, for example as given above, depends only on how the mass of the object is distributed along the *z*- and *y*-axes. Some thoughtful reflection will reveal that, for the coordinate axes origin taken to be the center of the sphere, the $z^2$ integral $S_{zz}$ is the same as the $y^2$ integral $S_{yy}$, so only one integral really needs to be done as indicated in the last equality above. In fact, $S_{xx}=S_{yy}=S_{zz}$ since for a sphere, the choice of axis is completely arbitrary. Using $z = r\cos(\theta)$, $x = r\sin(\theta)\cos(\varphi)$, and $y = r\sin(\theta)\sin(\varphi)$ allows these integrals to be written in polar coordinates. Taking $S_{zz}$ for example gives

$$S_{zz} = \rho\int_0^R r^4 dr \int_0^\pi \cos^2\theta \sin\theta d\theta \int_0^{2\pi} d\varphi = \rho\left(\tfrac{1}{5}R^5\right)\left(\tfrac{2}{3}\right)(2\pi) = \tfrac{1}{5}MR^2$$

The moment of inertia about any axis is twice this value, giving $I_{xx} = I_{yy} = I_{zz} = \frac{2}{5}MR^2$. It may also be seen that the off-diagonal elements of the inertia tensor are all zero. This means that any choice of orthogonal coordinate axes is formally equivalent to any other, and any such choice corresponds to the *principle axes*. For other rigid bodies, the off-diagonal elements are generally nonzero, and only a special choice of the coordinate axes will result in a diagonal inertia tensor. Written as a matrix, the inertia tensor is

$$\mathbf{I} = \frac{2MR^2}{5}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

An important property of this inertia tensor is that its product with any vector ω is simply a scaling of that vector, the direction does not change: $\mathbf{I}\omega = \left(\frac{2}{5}MR^2\right)\omega$.

The kinetic energy of a ball consists of two parts, translational and rotational. The translational kinetic energy is given by $T_{(Trans)}={}^1\!/_2 MV^2$, where *V* is the velocity of the center of mass of the ball. The mass of the ball, *M*, is the proportionality constant between the velocity squared and the energy. The rotational kinetic energy about a principle axis is given by the similar equation $T_{(Rot)}={}^1\!/_2 I\omega^2$, where ω is the angular velocity, for example in radians per second. Therefore the moment of inertia, *I*, is the proportionality constant between the angular velocity squared and the rotational kinetic energy. The most general equation for the rotational energy of a rigid body is $T_{(Rot)}={}^1\!/_2 \omega\,\mathbf{I}\,\omega$, in which ω is the angular velocity about each axis, **I** is the 3 by 3 inertia tensor, and the dot implies the appropriate matrix-vector or vector-vector product. The quantity **L**=**I**ω is the rotational angular momentum about the center of mass, and the simple form for **I** given above means that for a pool ball the angular momentum is always aligned with the angular rotation. The rotational energy may then be written as $T_{(Rot)} = \left(\frac{1}{5}MR^2\right)\omega\ \omega = \left(\frac{1}{5}MR^2\right)|\omega|^2$. The freedom of axes choice for a uniform sphere will often allow the problem at hand to be simplified to only a single rotation axis, in which case the simple scalar equation may be used

When a force is applied to a rigid body, such as a ball, the velocity of the center of mass changes according to the equation $\mathbf{F} = M\dot{\mathbf{V}}$, and the angular velocity changes according to the equation $\mathbf{r} \times \mathbf{F} = \mathbf{I}\,\dot{\omega}$. When a single principle rotational axis is considered, the latter equation reduces to the simpler $r\sin(\theta)|F| = I\,\dot{\omega}$, where $\theta$ is the angle between the vectors **r** and **F**, with magnitudes $r$ and $|F|$ respectively. $\omega$ is in the direction perpendicular to the plane defined by the two vectors **r** and **F**, and aligned, by convention, with the right-hand-rule (*i.e.* when the fingers of the right hand curl in the direction that rotates **r** into **F**, then the thumb points along the direction of positive $\omega$; other analytic expressions for the vector cross product will also be used in this discussion, but the right-hand-rule provides a useful and intuitive defintion.) The vector **r** points from the center of mass of the ball to the point on the surface of the ball at which the force is applied. In these equations, $\dot{\mathbf{V}} \equiv \frac{d\mathbf{V}}{dt}$ is the linear acceleration along each coordinate axis and $\dot{\omega} \equiv \frac{d\omega}{dt}$ is the angular acceleration around each coordinate axis. The similarities in the relations between the force and the mass $M$ for the linear acceleration and between the force and the moment of inertia $I$ for the rotational acceleration are again seen. The $r\sin(\theta)$ factor shows how the angular acceleration depends on the direction of the force. When the force is applied directly toward the center of mass of the ball, then the $\sin(\theta)$ factor is zero and there is no angular acceleration; it is only when the force is applied in a direction askew from the center of the ball that angular acceleration occurs.

A force is required to rub two objects together. If the two objects are pressed together with a normal force $F_N$, and a sideways force of magnitude $F_f$ causes the two objects to slip against each other without acceleration, then the *coefficient of sliding friction* is defined as $\mu_{(sliding)}=F_f/F_N$. To a good approximation, the coefficient of friction between two surfaces is a constant, independent of the forces and independent of the speeds of the two sliding objects. A small coefficient of friction is associated with slippery object pairs, and a large coefficient of friction is associated with sticky object pairs. There is also a *static* coefficient of friction. Static friction is defined in a similar manner to sliding friction, but it applies to two surfaces that are at rest. For a given pair of surfaces, the static coefficient of friction is larger than the sliding coefficient, although for some surface pairs they are very close in value.

There are several frictional forces that are important in pool. The first is the sliding friction of a ball on the cloth, $F_s$. $F_s=\mu_{(sliding)}W$ where $W$ is the weight of the ball ($F_N=W=Mg$ where $g$ is the acceleration of gravity). Since the ball weight and the coefficient of friction are constants for a given ball and for a given table, the frictional force of a sliding ball is a constant. The magnitude of the frictional force does not depend on the velocity of the ball or upon ω for the ball as long as the ball is sliding on the cloth. The direction of this force does depend on the ball velocity and ω, and this will be examined in more detail in the following discussions. If the ball is not sliding on the cloth (e.g. the ball is at rest, or the ball is rolling smoothly without slipping on the cloth surface), then there is no sliding frictional force.

It is interesting to consider the nature of the cause of a sliding frictional force. At a microscopic level, the atoms in the molecules of one surface are attracted to those of the other surface. As the object slides forward, new interactions, or bonds, are formed in the forward direction, maintained momentarily, and then broken as the individual atoms are pulled apart. However, it is not directly these bonds that cause the friction. The reason is that the same kinetic energy is lost in forming the bond as is gained back again when it breaks, and there is no net change of energy due to the forming and breaking of these bonds as the surfaces slide across each other. But for the small amount of time that the individual atoms interact, vibrational energy of the surface molecules is transferred to the other molecules in the bulk of the objects. (Energy is also transferred in the opposite direction, but at a much smaller rate; the net energy flow is from the surface atoms to the bulk atoms, a consequence of the second law of thermodynamics.) The result of this energy transfer is that translational kinetic energy is transformed into vibrations of the molecules of the bulk materials, or in other words, into heat and sound. From this point of view of a physicist, it might be said that it is the heat and sound that cause the frictional force; this is somewhat the opposite of the layman’s point of view, namely, that friction causes the heat.

**Problem 1.4**: A block slides down an inclined plane without acceleration; what is the relation between μ and the angle of the slope of the plane?
*Answer*: The downward force is the weight of the object $W=Mg$. The component of this force normal to the plane surface is $F_N=W\cos(\alpha)$ where α is the angle of incline. The component of the downward force tangent to the surface of the plane is $F_t=W\sin(\alpha)$. This force is directed down the incline, accelerating the object, and it is opposed by the frictional force which is directed uphill. Since the object is sliding without acceleration, all of this tangential force is balanced exactly by the frictional force, $F_S=-F_t$. The coefficient of friction is then given by $\mu=F_t/F_N=\tan(\alpha)$. This relation between slope and the coefficient of friction is so fundamental that it is sometimes taken as a *de facto* defintion.

A sliding block provides a simple conceptual model for understanding several other aspects of sliding friction. Consider a sliding block of mass $M$ on a level surface with a sliding coefficient of friction μ. The downward force of the block is the weight of the block, $W=Mg$, and this force is exactly opposed by an upward force of the surface; this means that the block does not accelerate in the vertical direction. The horizontal force is constant in magnitude, $|F_s|=\mu W=\mu Mg$ and the direction of this force is opposite to the velocity which is taken to define the positive direction. This frictional force slows down the sliding block according to the equation $\dot{V}=-\mu g$ where the minus sign is due to the direction of the force. It is interesting that this equation does not depend on the block mass; several equations of motion in the following discussions will be similarly independent of the ball masses. Integration over time gives $V(t)=V_0-\mu gt$ where $V_0$ is the initial velocity at $t$=0. Of course, this equation is valid only as long as the block is sliding. Integration again over time gives the distance $x$ as a function of time as $x=V_0t-\frac{1}{2}\mu gt^2$ where the distance is measured from the starting point.

Since the block is slowing down, kinetic energy is not conserved in this process. This is a *dissipative* system, not a *conservative* system. How does the kinetic energy depend on time and distance? Substitution of $V(t)$ above gives

$$T = \tfrac{1}{2}MV^2 = \tfrac{1}{2}M(V_0^2 - 2V_0\mu gt + \mu^2 g^2 t^2)$$
$$= T_0 - \mu Mgx.$$

Kinetic energy is lost as a linear function of the distance and a quadratic function of time. When the block slides to rest, $T$=0, the initial energy and total sliding distance $d$ are simply related as $T_0=\mu Mgd$. If the initial energy of the block were doubled, then the distance that the block slides before coming to rest would also double. However, if the initial velocity were doubled, then the final distance would increase by a factor of four. Note also that for a given initial energy $T_0$, if the coefficient of friction were to increase, then the total sliding distance must decrease, and if the coefficient of friction were to decrease, then the total sliding distance must increase. A related quantity of interest is the *power dissipation*, defined as $\dot{T} \equiv \frac{dT}{dt}$. From the quadratic time function, or using the chain rule $\dot{T} = \frac{dT}{dx}\frac{dx}{dt}$, the power dissipation for a sliding block is seen to be $\dot{T}=-\mu MgV$. The treatment of frictional forces for a sliding block are relatively simple; the somewhat more complicated situations for a billiard ball sliding on a table and for two colliding billiard balls are treated in the following sections.

How can the coefficient of friction be measured? There are several possibilities, depending on the equipment available with which to make measurements or on the data available. (1) One method would be to attach a measuring scale to the block, and simply measure the force required to slide the block on the surface without acceleration; this force divided by the weight of the block would give directly the coefficient $\mu$. (2) If the surface can be held at an arbitrary slope, then $\mu$ can be determined as in P1.4. This may not be always practical (for example if the surface is a heavy billiard table). (3) If the velocity or the energy could be measured accurately at two points in a given trajectory, then the equation $T=T_0-\mu Mgx$ at these two points could be used to determine $T_0$ and the product $\mu Mg$. An independent determination of the weight $Mg$ would then allow $\mu$ to be determined. However, velocities are relatively difficult to measure, so this also may not be practical. (4) Suppose that the block slides a distance $d$ in time $t_d$ before coming to rest. Then the initial velocity was $V_0=\mu g t_d$. Substitution of this into the quadratic distance equation gives $\mu=d/(\frac{1}{2} g t_d^2)$. Of course, this is not an exhaustive list of possibilities, and many other schemes could be devised based on preparation of the initial velocity or trajectory measurements of various types.

A second force is the rolling resistance of a ball on the cloth. This is not, strictly speaking, a sliding frictional force since it does not invlove sliding surfaces, but the formal treatment of this force is similar to the above sliding frictional force. A detailed examination of the forces involved in this situation will be postponed until the next section. For the present discussion, this rolling resistance will be modeled as a ball rolling uphill on an inclined plane. This is a conservative model. The dissipative energy loss of an actual billiard ball is then considered to be analogous to the energy loss of the model ball in the conservative gravity field. Because this model is a conservative system, it is possible to determine the equations of motion of the ball without detailed consideration of the forces (which may not be intuitively obvious for this situation).

For an incline of slope α, the height above the starting point is given by $h=s \sin(\alpha)$, where $s$ is the distance up the incline from the starting point. The potential energy is then by given as a function of $s$ by $U(s)=Mgh=sMg\sin(\alpha)$. In this model it is assumed that there is no energy dissipation through heat. The total energy $E=T+U$ is a constant, so any kinetic energy lost by the ball is transferred to potential energy in the gravity field. This gives the relation $T(s)=T_0-\sin(\alpha)sMg$, where $T_0=E$ is the initial energy of the rolling ball at the bottom of the incline. It is now seen that the kinetic energy for a ball rolling on an incline obeys the same equation as for the sliding block, but with the incline slope, corresponding to sin(α), assuming the role of the sliding coefficient of friction of the block. However, in the case of a rolling ball, the kinetic energy expression is more complicated, and this, along with the examination of the associated forces, is discussed in more detail in the following section. Using the chain rule expression, the power dissipation for the ball rolling up an incline is given by $\dot{T} = \frac{dT}{ds}\frac{ds}{dt} = -\sin(\alpha)MgV$, where $V$ is determined by the speed parallel to the incline. If, for some reason, it were not possible to measure the slope of the incline, it could be determined indirectly by measuring the sin(α) factor in the above equations in the same manner that the sliding coefficient of friction μ can be measured for a sliding block.

The connection between an actual ball rolling on a level table and this model problem may be justified by considering the rolling ball at a microscopic level. The nature of the effective frictional force arises in part from the compression of the cloth fibers as the ball rolls past. Once compressed, they do not rebound immediately as the ball passes; if they did, then there would be no energy lost in this manner by the rolling ball. The energy lost by this irreversible compression of the fibers slows the rolling ball. Energy of the rolling ball is also lost to vibrations of the ball and table, and eventually to the increased temperature of the surroundings. As the ball rolls forward an infinitesimal amount, it rolls also uphill on the cloth, losing a small amount of kinetic energy. But the cloth cannot support the ball weight, so it compresses the fibers. This transfers the potential energy from the gravity field into the spring constants of these compressed fibers. As the ball continues to roll, the fibers remain compressed for a small time, and this time lag prevents the potential energy stored in the fibers from being returned to the ball kinetic energy. The horizontal distance that the ball rolls on the table can be measured, but the effective height that it would have risen if the cloth fibers had not compressed cannot be measured directly. Therefore, the effective slope sin($\alpha$), which may be associated with an effective rolling coefficient of friction $\mu^{eff}_{(rolling)}$, must be determined indirectly.

Consider a ball rolling a distance *d* on a table in time *t* before coming to a stop. At this time, an effective force is assumed of the form $F_r = \mu^{eff}_{(rolling)} Mg$ that opposes the rolling ball. Newton's equation $F_r = M\dot{V}$ may be rewritten as $\mu^{eff}_{(rolling)} g = -\dot{V}$. Integration over time results in $\mu^{eff}_{(rolling)} gt = V_0 - V$ where $V_0$ is the initial velocity. Integration over time again gives $\frac{1}{2}\mu^{eff}_{(rolling)} gt^2 = V_0 t - d$. The final velocity is zero when $V_0 = \mu^{eff}_{(rolling)} gt$ and this may be used to eliminate $V_0$ from the distance equation. The effective coefficient of friction for the rolling ball may then be determined from the equation

$$\mu^{eff}_{(rolling)} = \frac{d}{\frac{1}{2} g t^2}$$

The ball mass does not appear in this relation. The dimensionless quantity *table speed* is defined as $1/\mu^{eff}_{(rolling)}$ and is similarly independent of ball mass. With this definition of table speed, a very slow table is in the range of 50-70. Normal table speed is 80-100. A very fast pool table might have a speed higher than 120. The cloth on a billiard table is usually finer and smoother than that on a pool table, and a fast billiard table might have a speed over 150. The force due to rolling resistance is much smaller than that due to sliding friction.

The sliding frictional force and the rolling frictional force of a ball on a table are independent quantities. Consider for example a ball on a hard rubber surface; the sliding friction would be very large, while the rolling resistance would be relatively small. Alternatively, consider a ball on a Teflon surface with a soft backing; the sliding friction would be relatively very small, while the rolling resistance would be relatively large. The uniformity of billiard cloth material limits the range of extremes that are encountered in practice. The official BCA (Billiard Congress of America) rules specify a billiard cloth that is predominantly wool. The PBTA (Professional Billiard Tour Association) requirements are even more specific, and detail a brand and type of billiard cloth, namely Simonis 860; although this is partly a matter of sponsorship, it may be noted that this is a relatively fast pool table cloth that results typically in table speeds of 100 to 130 when newly installed.

**Problem 1.5**: A ball is lagged perfectly on a standard 9' pool table and it is observed that the ball travels from the foot cushion to the head cushion in 7.00 seconds. What is the table speed? What was the initial velocity of the ball as it left the last cushion?
*Answer*: The playing area of a standard 9' pool table is 50" by 100". After accounting for the ball width, the center of the ball travels (100"–2.25")=97.75" between cushions. The acceleration due to gravity is $g$=386 $in/s^2$. The table speed is

$$\text{TableSpeed} = \frac{1}{\mu^{eff}_{(rolling)}} = \frac{\frac{1}{2}gt^2}{d} = \frac{0.5\ \ 386\left(\frac{in}{s^2}\right)\ t^2}{97.75(in)} = 1.97\ \ t^2_{sec}$$

$$= 1.97\ \left(7.00^2\right) = 96.7$$

This is a fairly fast pool table. It is customary to approximate the $g/(2d)$=1.97 factor as 2.0 on a 9' table. The table speed may then be estimated simply as $2t^2$ where the time is measured in seconds. For playing purposes, it is usually unimportant to know the table speed to more than 2 significant figures. The velocity after the last cushion was

$$V_0 = \mu^{eff}_{(rolling)}gt = \frac{2d}{t} = \frac{2(97.75in)}{7.0s} = 27.9\left(\frac{in}{s}\right).$$

The initial velocity is seen to be twice the time-average velocity, which is given by $d/t$.

**Exercise 1.1**: Measure the table speed of some of the tables on which you play regularly. Rather than try to lag a ball perfectly, set up a ramp with cue sticks, and adjust the height of the ramp and initial ball placement so that the ball rebounds off the foot cushion and stops just before touching the head cushion. Disregard the small time it takes for the ball to achieve natural roll after impact with the foot cushion. Take the average time for several rolls in order to account for timing inaccuracies.

A third important frictional force is that between two colliding balls. The forces between two balls change during the collision. The collision time is very short, so these forces can be very large in order to transfer energy from one ball to another during a collision. The frictional forces act in a direction tangential to the surface of the ball at the point of contact between the balls. This is shown schematically in Fig. 1.1. The linear forces that accelerate the balls are directed between the ball centers. The resultant force on a ball is the sum of these two vector forces. That velocity component of a ball due to the tangential frictional forces is called either *collision induced throw* or *spin induced throw*, depending on the spinning condition of the balls and on the cut angles involved. When two balls slide against each other, both balls are accelerated by frictional forces. The frictional force vector that accelerates one ball is exactly opposite to that which accelerates the other ball. Note however that the angular acceleration due to the frictional forces has the same sign on both balls, due to the fact that the opposing forces are applied to the front of one ball but to the back of the other. As before, to a good approximation the frictional force is independent of the speed at which the two surfaces slide against each other. The force is constant unless the spinning balls "lock" against each other (as two interlocked gears), at which time the sliding frictional force vanishes.

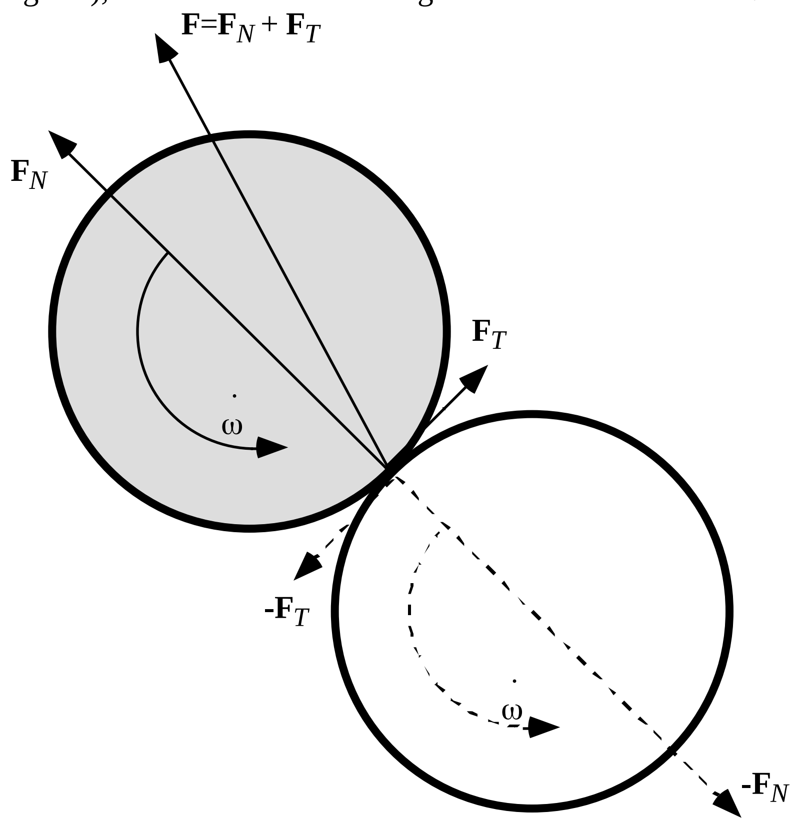



Fig. 1.1. The normal forces $\mathbf{F}_N$, tangential forces due to sliding friction $\mathbf{F}_T$, the resulting total force $\mathbf{F}$, and the angular acceleration $\dot{\omega}$ are shown schematically for two colliding balls. The magnitudes of the forces change during the collision, but the ratio of the tangential and normal forces are constant and are determined by the coefficient of friction. The magnitude of the tangential forces are shown greatly exaggerated. Note that although the tangential forces acting on the two balls exactly oppose each other, the resulting angular accelerations have the same sign.

**Problem 1.6**: Two object balls are frozen together and aligned straight toward the foot cushion exactly toward a marked spot. The nearest ball is 72" away from the cushion. The farthest ball from the cushion is hit at an angle with the cue ball. The object ball is observed to miss the point on the cushion by 4". Assuming that this collision induced throw is due to friction, what is the coefficient of friction for these two balls?

*Answer*: $\mathbf{F}_N$ is directed toward the marked spot, and $\mathbf{F}_T$ is perpendicular as in Fig. 1.1. The resultant velocity is parallel to the total force vector. The coefficient of friction is related to the angle of throw $\alpha$ by

$$\tan(\alpha) = \frac{F_T}{F_N} = \mu = \frac{V_T}{V_N} = \frac{D_T}{D_N}$$

Substitution of the appropriate distances gives the coefficient of friction as

$$\mu = \frac{4''}{72''} = 0.0556$$

**Exercise 1.2**: Measure the collision induced throw angle for several sets of balls at pool rooms where you play regularly. Generally, if the balls are worn or dirty, they will have a high coefficient of friction, and if they are new or polished, they will have a low coefficient of friction. Smear some chalk on the contact point between the frozen balls, and an increased coefficient of friction should be observed. Smear some talcum powder on the contact point, and a smaller coefficient of friction should be seen. Place a drop of water (or spit) on the contact point and the coefficient of friction will become essentially zero. Correcting for collision induced throw is one of the challenging aspects of playing with different sets of balls in tournaments, and of playing at different pool rooms.

A fourth frictional force is the static friction between the cue tip and the cue ball. The cue tip must not slide on the cue ball. If this occurs unintentionally, then a miscue results and the cue ball behaves unpredictably; if the cue tip slides intentionally against the cue ball, then an illegal "push shot" has occurred. The static frictional force is related to the normal force and to the static coefficient of friction by the relation $\mu_{static}=F_T/F_N$ where $F_T$ is the minimum force required to cause the cue tip to slide on the surface of the cue ball.

**Problem 1.7**: For a particular cue tip, it is observed that miscues begin to occur when the cue tip contacts the cue ball at a height halfway between the center and the top of the cue ball. What is the static coefficient of friction between the cue tip and the cue ball? If the static coefficient of friction is 1.0, what is the displacement at which miscues begin to occur?

*Answer*: Refer to Fig. 1.2. The slope of the cue ball at the point of contact. is determined by

$$\cot(\alpha) = \frac{\left(\frac{b}{R}\right)}{\sqrt{1-\left(\frac{b}{R}\right)^2}}$$

where $b$ is the displacement away from the center. When the force $F$ is applied to the cue ball in a horizontal direction, this may be written as a sum of the normal force toward the center of the cue ball $F_N=F\sin(\alpha)$, and the tangential frictional force with magnitude $F_T=F\cos(\alpha)$. The coefficient of friction and the maximum displacement are related by

$$\mu_{static} = \frac{F_T}{F_N} = \cot(\alpha) = \frac{\left(\frac{b_{max}}{R}\right)}{\sqrt{1-\left(\frac{b_{max}}{R}\right)^2}}$$

$$\left(\frac{b_{max}}{R}\right) = \frac{\mu_{static}}{\sqrt{1+\mu_{static}^2}}$$

For $b_{max}/R = ½$,

$$\mu_{static} = \frac{½}{\sqrt{1-¼}} = \frac{1}{\sqrt{3}} = .577$$

For $\mu_{static}=1.0$,

$$\left(\frac{b_{max}}{R}\right) = \frac{1}{\sqrt{2}} = .707$$

As seen for these two cases, a higher coefficient of friction allows the cue tip to contact the cue ball at larger displacements without miscuing.

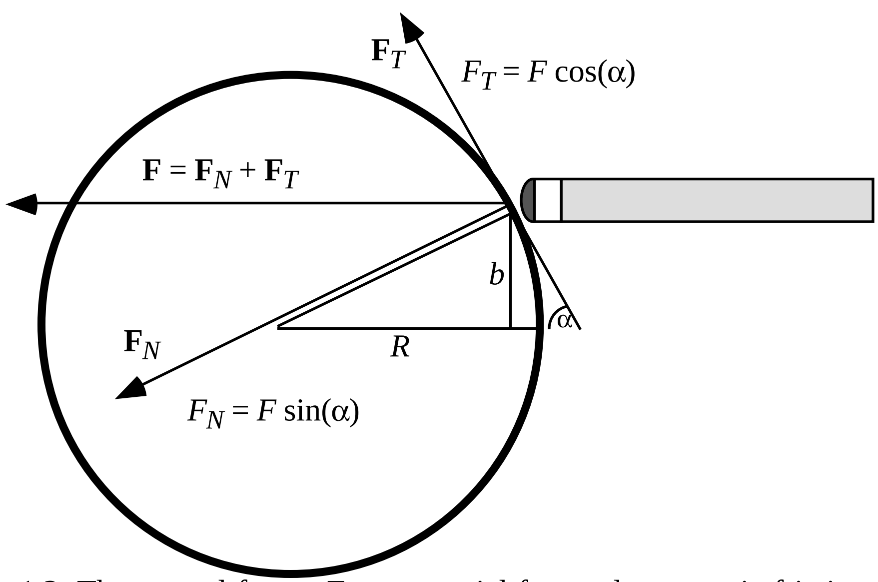



Fig. 1.2. The normal forces $F_N$, tangential forces due to static friction $F_T$, and the resulting total force $F$ for contact between the cue tip and cue ball are shown schematically. The magnitudes of the forces change during the collision, but the ratio of the tangential and normal forces are constant and are determined by the impact point and limited by the the static coefficient of friction.

**Exercise 1.3**: Determine the static coefficient of friction between your cue tip and a cue ball. Instead of determining the point of miscue (as in P1.7), hold a ball against a cushion and stand the cue shaft vertically on the ball. Estimate the distance away from the center ball, and use the equation in P1.7 to determine $\mu_{static}$. Wipe the cue tip clean, removing all chalk, and a smaller coefficient of friction should be observed. Experiment with different kinds of chalk and with different tip conditions. Note that it is the displacement of the actual contact point of the cue tip that should be measured, and not the displacement of the cue shaft edge.

## 2. Slide and Natural Roll

Suppose that at some time a ball is known to have some (center of mass) translational velocity and some spin (about the center of mass). For simplicity, assume that the spin axis is horizontal and is perpendicular to the translational velocity (*i.e.* the ball has straight topspin or draw; e.g. $\mathbf{V}=V\hat{\mathbf{i}}$ and $\omega=\omega\hat{\mathbf{j}}$ ). As the ball slides on the cloth on the table, the friction between the ball and cloth will cause both the translational and angular velocity to change. This force will act to accelerate the ball, that is, to increase or decrease the velocity, until an equilibrium situation occurs in which the translational and angular velocities "match" each other, at which time the sliding frictional force becomes zero. This is the *natural roll* (also called *normal roll*, *smooth roll,* or *rolling without slipping*) situation. Over a small time *dt*, the distance traveled by the ball will be *Vdt*, and the outside surface of the ball will roll a distance $R\omega dt$ relative to the ball center of mass. Therefore, this "matching" occurs when $V=R\omega$.

The natural roll condition is important to examine because the speed and spin of a sliding ball are always being forced toward the natural roll condition by the sliding friction, and once achieved, natural roll is maintained by the ball until it collides with another ball or cushion or rolls to a stop.

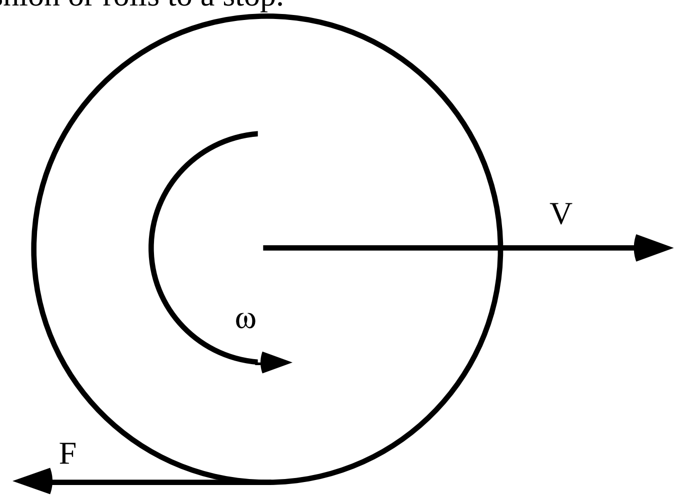



Fig. 2.1. The linear velocity *V*, angular velocity ω, and corresponding frictional force *F* are shown schematically for a backspin shot. *V* is positive, whereas *F* and ω are taken to be negative as shown.

Kinetic energy is not conserved during the equilibration period as the sliding ball approaches the natural roll condition. This is easy to see in the case in which the translational velocity and angular velocity oppose each other, as in a backspin shot depicted in Fig. 2.1. (Positive ω is taken to be in the clockwise direction in Fig. 2.1.) In a backspin shot, the initial frictional force acts to both slow down the ball and to decrease the magnitude of the spin, clearly decreasing simultaneously both types of kinetic energy.

A useful concept to introduce in this discussion is the spin/speed ratio $\omega/V$. In some situations, a more useful quantity is the dimensionless ratio $J=(R\omega/V)$; for the above backspin shot, this is the ratio of velocity at a point on a ball on the rotational equator that is due to the spin to the velocity of the center of mass of the ball. In situations in which several spin components are examined simultaneously, the dimensionless vector quantity $\mathbf{J}=(J_x,J_y,J_z)=R\omega/V$ is useful. As discussed above $J_y$=+1 corresponds to the natural roll condition when the velocity is directed along the *x*-axis.

The frictional force acts on the very bottom point of the ball, where the ball touches the cloth, and it points in a horizontal direction. The force acts to accelerate the ball according to the equation $\mathbf{F} = M\dot{\mathbf{V}}$. Integrated over some time period, this gives a change of momentum

$$\mathbf{F}t = M(\mathbf{V} - \mathbf{V}_0) .$$

in which $\mathbf{V}_0$ is the initial velocity vector. Note that since **F** and **V** point in opposite directions in a backspin shot; the ball slows down over time. When **F** and **V** point in the same direction, e.g. a ball over-spinning with topspin, the ball speeds up over time. In the case depicted in Fig. 2.1, this equation simplifies to

$$|F|t=-M(V-V_0)$$

or, after eliminating the mass from both side of the equation and introducing the ball-cloth sliding coefficient of friction,

$$\mu gt =-(V-V_0)$$

where the sign of the right hand sides results from the fact that the velocity and force vectors point in opposite directions. (In the general case for positive $V_0$, $F>0$ when $R\omega_0>V_0$, and $F<0$ when $R\omega_0<V_0$ or in other words, $F$ and $(J-1)$ have the same sign.)

The angular velocity of the sliding ball changes according to the equation $\mathbf{r} \times \mathbf{F} = \mathbf{I}\,\dot{\omega}$. For the backspin shot, $\mathbf{r}=-R\hat{\mathbf{k}}$, $\mathbf{F}=-|F|\,\hat{\mathbf{i}}$, and $\dot{\omega} = \dot{\omega}\hat{\mathbf{j}}$. In this situation, this equation simplifies to $R|F| = I\dot{\omega}$. Integrated over some time period, this gives

$$R|F|t = I(\omega - \omega_0).$$

Note in Fig. 2.1 that for a backspin shot the frictional force is acting to increase the angular velocity from an initial negative value to a final positive value. If the cue ball contacts an object ball while the angular velocity is still negative, this is called a *draw shot*. If all the spin is removed by the cloth friction and the ball is spinning neither forward nor backward upon impact with an object ball, this is called a *stun shot*. If forward roll, or in particular natural roll, is achieved prior to collision, this is called a *drag shot*. As shown in the above equation, it is the initial angular velocity, the sliding friction between the ball and the cloth, and the time before the collision that distinguishes these three shots.

**Problem 2.1:** What is the relation between linear and angular velocity for a sliding ball?
*Answer:* Eliminating the common $|F|t$ from the above two expressions gives

$$\frac{I}{R}(\omega - \omega_0) = -M(V - V_0) \quad .$$

Using the previous expression for $I$ for a ball results in

$$V = V_0 - \tfrac{2}{5}R(\omega - \omega_0) \quad .$$

This expression is valid at any time the ball is sliding on the cloth. Although derived specifically for the backspin shot, this expression is valid for any frictional force. Note that for the backspin shot, $V$ decreases as ω increases, and for the over-top-spin situation, $V$ increases as ω decreases. This shows that the relation between linear and angular velocity does not depend on the ball mass or on the ball-cloth sliding coefficient of friction.

**Problem 2.2:** Determine the final linear velocity of a ball after natural roll is achieved as a function of initial linear and angular velocities.
*Answer*: Natural roll is achieved when the linear and angular velocities equilibrate. Substituting $V=R\omega$ in the expression from P2.1 gives

$$V_{NR} = \frac{5}{7} V_0 + \frac{2}{7} R\omega_0$$

Note that if the initial angular velocity were zero, then the sliding ball would eventually slow down to $^5/_7$ of its initial velocity. If the initial angular velocity matched exactly the initial linear velocity, $V_0=R\omega_0$, then the linear velocity would remain unchanged. If the initial angular velocity is negative, as for a drag shot, then the final linear velocity is even less than $^5/_7$ of the initial velocity; for example, if the initial angular velocity is equal to natural roll, but in the opposite direction, $V_0=-R\omega_0$, then the final velocity is $^3/_7$ of the initial velocity. If the initial spin is very large and negative, then the final natural roll velocity will be negative; this can occur in masse shots, or in situations involving collisions with other balls. Note that the natural roll velocity does not depend on the ball-cloth friction or the ball mass.

**Exercise 2.1:** Experiment with the drag shot. Use a striped ball in place of the cue ball so that the spin is easily observed. Strike the "cue" ball below center. Observe how the ball initially spins backward. The cloth friction slows this backspin until at some point the ball is not rotating at all, but is simply sliding across the table. Beyond this point the ball begins rolling forward. At some point all sliding stops, and the ball achieves natural roll. During all of the time that the ball is sliding on the cloth, the speed of the ball is decreasing. If you have a video camera, record some of these shots and play them back in slow motion. The drag shot is useful when playing on dirty or unlevel tables, and a low-speed impact between the cue ball and object ball is required for position. The initial high speed of the cue ball reduces the effect of the unlevel table, and only at the very end after natural roll is achieved and the velocity is reduced to about $^3/_7$ of the initial velocity, does the impact occur. The average velocity of the cue ball is about $^5/_7$ of the initial velocity, which means that the effect of the unlevel table has been reduced by about $^2/_7$ or 29% from the case where natural roll is achieved immediately.

**Exercise 2.2**: Experiment with a stun shot. A stun shot is when the cue ball has zero angular velocity about the horizontal axis upon contact with an object ball or cushion.

Set up a straight-in shot with an object ball, and place the cue ball at various distances away from the object ball. (Use a striped ball in place of the cue ball so that the spin can be easily observed.) For a given distance and shot speed, shoot with just the right amount of backspin so that the cloth friction has time to remove the spin. The cue ball should stop exactly upon impact, and roll afterwards neither forward nor backward. For a fixed distance, the slower the shot speed, the more extreme will be the backspin required to achieve a stun shot impact. Experiment with stun shots on different tables. Sticky tables (high sliding friction between the cloth and ball) require more extreme backspin than slick tables to achieve stun. Stun shots are important for position play and, as discussed in later sections, for judging accurate carom angles.

**Problem 2.3:** What is the shape of the path taken by a sliding ball before natural roll is achieved? What is the shape of the path after natural roll is achieved?

*Answer*: Integration of $\mathbf{F} = M\dot{\mathbf{V}}$ twice gives

$$\tfrac{1}{2}\mathbf{F}t^2 = M(\mathbf{q} - \mathbf{q}_0 - \mathbf{V}_0 t)$$

$$\mathbf{q} = \mathbf{q}_0 + \mathbf{V}_0 t + \frac{1}{2M}\mathbf{F}t^2$$

Since the choice of coordinate axes is arbitrary, assume that the axes origin corresponds to $t$=0, and that the axes are oriented so that the $x$-component of the sliding force is zero. The coordinates of the path are then given by

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} V_{0x} \\ V_{0y} \end{pmatrix} t + \frac{1}{2M} t^2 \begin{pmatrix} 0 \\ F_{0y} \end{pmatrix} = \begin{pmatrix} V_{0x} \\ V_{0y} \end{pmatrix} t + \begin{pmatrix} 0 \\ \frac{1}{2}\mu g \end{pmatrix} t^2$$

Because of the choice of axes, the velocity in the $x$-direction remains unchanged over time. Using the relation $t=x/V_{0x}$ to eliminate $t$ from the $y$ part of this equation gives

$$y = \left(\frac{V_{0y}}{V_{0x}}\right) x + \left(\frac{\mu g}{2V_{0x}^2}\right) x^2$$

which may be recognized as an equation for a parabola. While the ball is sliding on the cloth, the path of the ball is a parabola, the shape of which is determined by the initial velocity and by the frictional force between the ball and the cloth. This path does not depend on the ball mass. This frictional force remains unchanged in both direction and magnitude as long as the ball is sliding. This applies to the paths taken by balls after collisions with cushions or with other balls, and also to the cue ball when struck with an elevated cue stick (*i.e.* masse or semi-masse shots). The ball is accelerated by the sliding force until natural roll is achieved. After natural roll is achieved, there is no sideways force exerted to further accelerate the ball, so the ball rolls in a straight line.

**Problem 2.4**: When a ball achieves natural roll, what fraction of its kinetic energy is translational and what fraction is rotational?

*Answer*: The total kinetic energy is

$$T = T_{(Trans)} + T_{(Rot)} = \tfrac{1}{2}MV^2 + \tfrac{1}{2}I\omega^2 = \tfrac{1}{2}MV^2 + \tfrac{1}{5}MV^2 = \tfrac{7}{10}MV^2.$$

This gives

$$\frac{T_{(Trans)}}{T} = \frac{5}{7}$$

$$\frac{T_{(Rot)}}{T} = \frac{2}{7}$$

Now that the total kinetic energy expression for a natural roll ball is known, the issue of rolling resistance can be examined in more detail. The previous conservative model of a ball rolling up an inclined plane will be used to understand the various forces involved. In the case of a ball rolling without slipping up an inclined plane, the result of these forces is known, namely that $R\omega = V$ is maintained as the ball slows down, but the forces themselves required to achieve this result are not obvious. In order to apply Newton's laws directly, these forces must be known beforehand. Therefore Lagrange's equations of motion will be used. The generalized coordinates will be taken to be the distance up the incline $s$, the angular rotation of the ball $\theta$, and the undetermined multiplier associated with the constraint equation, $\lambda$. The expressions for the kinetic energy, potential energy, and the constraint equation are

$$T = \tfrac{1}{2}MV_s^2 + \tfrac{1}{2}I\omega^2$$

$$U = sMg\sin(\alpha)$$

$$f(s,\theta) = R\theta - s = 0$$

The Lagrangian is $L=T-U+\lambda f$, and the equations of motion are determined from the equation, $\frac{\partial L}{\partial q_i} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}_i} = 0$, for the three coordinates $s$, $\theta$, and $\lambda$. Substitution gives the three equations

$$-Mg\sin(\alpha) - \lambda - M\dot{V}_s = 0$$

$$\lambda R - I\dot{\omega} = 0$$

$$R\theta - s = 0 \,.$$

Differentiating the last equation twice gives $R\dot{\omega} = \dot{V}_s$. Solving the second equation for the undetermined multiplier gives $\lambda = I\dot{V}_s/R^2$. Substitution into the first equation then gives

$$M\dot{V}_s = -\left(1 + \frac{I}{MR^2}\right)^{-1} Mg\sin(\alpha) = -\tfrac{5}{7}\,Mg\sin(\alpha)$$

$$= -Mg\sin(\alpha) + \tfrac{2}{7}Mg\sin(\alpha) \qquad \text{[\textit{rolling without slipping}]}$$

$$= F_{gravity} + F_{constraint}$$

If, instead of rolling without slipping, the ball were allowed to slide freely, then Newton's equation of motion in this coordinate system would have been simply

$$M\dot{V}_s = F_{gravity} = -Mg\sin(\alpha) \qquad \text{[\textit{with free slipping}]}$$

Therefore the sliding ball is seen to slow down faster than the rolling ball, all other things being the same. The effective force arising from the static coefficient of friction between the ball and the incline is seen to be $^2\!/_7 Mg\sin(\alpha)$, and this force is directed uphill, opposing the gravitational force. Because there is no sliding associated with this frictional force, there is no energy dissipation in this model system. The only kinetic energy lost is that associated with the corresponding increase in potential energy. As was done in the previous section for a sliding block, an association with the effective slope and a coefficient of friction is made, $\mu_{(rolling)} = \sin(\alpha)$. In the previous section, an equation of motion was assumed of the form $\mu^{eff}_{(rolling)} g = -\dot{V}_s$. It is now seen that this assumption was correct, with the association

$$\mu^{eff}_{(rolling)} = \left(1 + \frac{I}{MR^2}\right)^{-1} \sin(\alpha) = \left(1 + \frac{I}{MR^2}\right)^{-1} \mu_{(rolling)} = {}^5\!/_7 \mu_{(rolling)}$$

When should $\mu^{eff}_{(rolling)}$ be used, and when should $\mu_{(rolling)}$ be used? The answer is that for a rolling billiard ball, it doesn't matter which coefficient of friction is used, provided of course, that it is used with the corresponding equation of motion. The use of the equation of motion involving $\mu_{(rolling)}$ has the advantage that once it has been determined for one object, the same value can be used for other objects made of the same material but with different shapes, such as rolling cylinders, rolling tubes, rings, or hollow balls. The quantity $\mu_{(rolling)}$ is therefore, in some sense, more fundamental than is $\mu^{eff}_{(rolling)}$. The motion of these objects will of course be slightly different, due to the dependence on the moment of inertia of the equations of motion, as demonstrated in the following problem.

**Problem 2.5**: The table used in P1.5 is moved to the surface of the moon. The billiard ball is replaced with a cylinder made of the same material as a billiard ball. How long will it take for the cylinder to roll the length of the table?

*Answer*: First determine $\mu_{(rolling)}$ for the table from the previous data:

$$\mu_{(rolling)} = {}^7\!/_5 \mu^{eff}_{(rolling)} = \frac{7}{5\ \ 96.7} = 0.0145$$

For a solid cylinder, $I = MR^2/2$. $g_{moon} = 63.8 in/s^2$, about 1/6 the gravity of the earth. The equation of motion is

$$\dot{V} = -\left(1 + \frac{I}{MR^2}\right)^{-1} g_{moon} \mu_{(rolling)}$$

Integration twice over time, then solving for $t$ gives

$$t = \sqrt{\frac{2\left(1 + \frac{I}{MR^2}\right)d}{g_{moon}\mu_{rolling}}} = \sqrt{\frac{3\ \ 97.75 in}{63.8(in/s^2)\ \ 0.0145}} = 17.8s$$

Solving the same equation for a ball gives $t = 17.2s$, a result that may also be obtained simply by scaling the earth time, 7.00$s$ by the factor $\sqrt{g_{earth}/g_{moon}} = 2.46$. Therefore, most of the lag time difference is due to the different gravitational forces of the earth and moon, with a smaller difference due to the different moments of inertia of the cylinder and ball.

**Problem 2.6**: Taking into account both the sliding friction and the rolling resistance, what is the total distance traveled by a cue ball with $V_0$=0 as a function of the initial spin $R\omega_0$? (neglect collisions with other balls and cushions)

*Answer*: As discussed in more detail in the following sections, $V_0$=0 is the appropriate initial condition immediately after the cue ball collides head-on with an object ball; the object ball removes the velocity of the cue ball, but leaves its spin unchanged. According to P2.2, this spin then accelerates the cue ball to the natural roll velocity $V_{NR}=\frac{2}{7}R\omega_0$. The time required to achieve natural roll is given by

$$t_{NR} = \frac{2R\omega_0}{7\mu_s g}$$

where $\mu_s$ is the sliding friction coefficient and $\omega_0$ is taken to be positive. The distance covered by the sliding cue ball during this time is

$$d_{NR} = \tfrac{1}{2}\mu_s g t_{NR}^2 = \frac{2(R\omega_0)^2}{49\mu_s g}.$$

Upon achieving natural roll, the equation of motion is then determined by the rolling resistance. The total rolling time is given by

$$t_R = \frac{V_{NR}}{\mu_{(rolling)}^{eff} g}$$

and the total rolling distance is given by

$$d_R = V_{NR}t_R - \tfrac{1}{2}\mu_{(rolling)}^{eff} g t_R^2 = \frac{V_{NR}^2}{2\mu_{(rolling)}^{eff} g} = \frac{2(R\omega_0)^2}{49\mu_{(rolling)}^{eff} g}.$$

The total distance for both the slide and the roll is

$$d_{total} = d_{NR} + d_R = (R\omega_0)^2 \left(\frac{2}{49g}\right)\left(\frac{1}{\mu_s} + \frac{1}{\mu_{(rolling)}^{eff}}\right).$$

This equation holds for both topspin and draw shots. An important point to notice is that the total distance is proportional to the square of the initial spin. This explains why it is much easier to position the cue ball accurately on a stop shot than on a strong draw or force-follow shot; a small variation in the initial spin is magnified into a larger distance for large initial $R\omega_0$ than for a small initial $R\omega_0$.

**3. Cue Tip/Cue Ball Impact**

Consider the situation in which a level cue stick strikes the cue ball. The cue tip applies a force to the cue ball at some point on the surface of the ball. This contact time is not instantaneous, but it is very short. Unlike a ball-to-ball impact (characterized by small tangential frictional forces and therefore resulting in a force that is directed essentially between the centers of the balls), the cue tip does not slip on the cue ball (except of course in a miscue situation). With these assumptions, the force is directed along the direction of the cue shaft. The angular acceleration from this force is given by the equation $\mathbf{r} \times \mathbf{F} = \mathbf{I}\,\dot{\omega}$. When a level cue stick strikes the cue ball, the angular acceleration along the direction of force, **F**/|**F**|, is given by

$$\dot{\omega}\ \frac{\mathbf{F}}{|\mathbf{F}|} = \left(\mathbf{I}^{-1}(\mathbf{r}\times\mathbf{F})\right)\ \frac{\mathbf{F}}{|\mathbf{F}|} = 0 \quad .$$

There is no component of angular acceleration around the axis of the cue stick, so there is no sideways frictional force between the ball and the cloth; the cue ball slides in a straight line in the direction of the cue shaft, while rotating about either or both the vertical axis (*i.e.* sidespin) and the horizontal axis perpendicular to the cue shaft (*i.e.* topspin or draw). This results from the fact that the moment of inertia for a pool ball is proportional to the unit matrix. (If the inertia tensor of an object is not proportional to the unit matrix, e.g. if the ball has an embedded off-center weight, then it will in general curve as it slides or rolls instead of moving in a straight line.)

First consider the case in which the cue tip strikes the cue ball exactly in the center. In this situation $\mathbf{r}\times\mathbf{F} = 0 = \mathbf{I}\,\dot{\omega}$, and there is no angular velocity imparted directly to the cue ball. The only thing that occurs is a transfer of linear momentum and translational energy between the cue stick and the cue ball. It will be assumed that the contact time is so short that the hand/skin/cuestick effects can be ignored. That is, at the very beginning of the contact time, the cue stick slows down and starts moving slower than the hand, and the skin begins to tighten, but by the time any significant extra force is exerted on the cue stick, the cue ball has already departed and lost contact with the cue tip.

**Problem 3.1:** What is the relation between the cue stick energy and velocity, the length of the stroke, and the applied force? (Assume a constant force is applied by the hand to the cue stick during the stroke.)

*Answer*: Integration of the equation $F = M_S\dot{V}$ over time gives $Ft = M_S(V - V_0) = M_S V$ where $F$ is the force applied to the stick and $M_S$ is the mass of the cue stick. Integration again gives $\frac{1}{2}Ft^2 = M_S(x - x_0) = M_S d$ in which $d$ is the distance of the stroke. Solving the first equation for $t$ and substitution into the second gives for the kinetic energy

$$T = \tfrac{1}{2}M_S V^2 = Fd.$$

Solving for $V$ gives

$$V = \sqrt{\frac{2Fd}{M_s}}$$

The cue stick energy is proportional to the stroke length and to the applied force, and the cue stick velocity is proportional to the square root of the stroke length and of the applied force. It is important to note that in the expression $T=Fd$, the energy does not depend on the mass of the cue stick. This means that for a given force on the cue stick and a given stroke length, a light cue stick will acquire the same energy as a heavy cue stick.

**Problem 3.2:** What is the relation between the final cue ball velocity and initial and final cue stick velocity, and the mass of the cue stick?

*Answer*: Before the impact, only the cue stick has momentum $M_s V_0$ and energy $\frac{1}{2} M_s V_0^2$. After the collision, both the cue stick and the cue ball have energy and momentum. Conservation of momentum and energy, assuming a center-ball impact, give

$$M_s V_0 = M_s V_s + M_b V_b$$

$$\frac{1}{2} M_s V_0^2 = \frac{1}{2} M_s V_s^2 + \frac{1}{2} M_b V_b^2 \quad .$$

Solve the first equation for $V_s$, and substitute into the second equation to obtain

$$V_b = \frac{2M_s}{M_s + M_b} V_0$$

$$V_s = \frac{M_s - M_b}{M_s + M_b} V_0$$

$$\frac{V_b}{V_s} = \frac{2M_s}{M_s - M_b}$$

A typical cue stick weighs 18*oz*, or about three times the weight of a pool ball. In this case, $V_b = \frac{3}{2} V_0$, $V_s = \frac{1}{2} V_0$, and $V_b/V_s = 3$, so the cue ball is moving about 3 times faster than the cue stick immediately after impact. If the masses were exactly equal (a very light cue stick), then the final ball velocity would be equal to the initial stick velocity, and the final stick velocity would be zero; all of the energy would be transferred from the stick to the ball. If the stick mass were less than the ball mass, then the final stick velocity would be in the opposite direction to the initial stick velocity; that is, the stick would bounce back from the cue ball. Under no condition does $V_b = V_s$; that is, there does not exist a combination of cue stick mass and ball mass such that both are moving forward immediately after impact at the same velocity.

**Problem 3.3:** What is the fraction of energy that is transferred from the cue stick to the cue ball as a function of the stick and ball masses?

*Answer:* Using the final stick and ball velocities from P3.2 gives

$$T_b = \frac{1}{2} M_b V_b^2 = \frac{4 M_b M_s}{\left(M_s + M_b\right)^2} \left(\frac{1}{2} M_s V_0^2\right) = \frac{4 M_b M_s}{\left(M_s + M_b\right)^2} T_0$$

Let $\alpha_s = M_s/M_b$ be the stick to ball mass ratio. Then the ratio of energies is given by

$$\frac{T_b}{T_0} = \frac{4\alpha_s}{(1+\alpha_s)^2}$$

When $\alpha_s$=1, then this energy ratio is unity, in agreement with the conclusions in P3.2. When there is a mismatch of masses, this energy ratio is less than one and the efficiency of transfer of energy in the collision is reduced.

If a 6*oz* cue stick results in optimal transfer of energy, then why not use one? If it is not optimal, then what is? There are two separate components to the answer. First, it is not always the most efficient transfer of energy that is important, but rather control of the energy that is transferred to the cue ball. It is easier to control a heavier stick than an extremely light one, and the inherent inefficiency from the mass difference is a way to reduce errors in the speed of the cue ball. A possible exception to this is the break shot in open-break games such as 8-ball and 9-ball in which the maximization of cue ball energy is desired. This leads to the second component of the answer.

As the bicep contracts to accelerate the cue stick on the break stroke, both the mass of the forearm and cue stick mass are accelerated. To understand how this affects the final object ball energy in at least a qualitative manner, some simplifying assumptions may be imposed. Assume that the forearm is a thin rod of uniform mass. The moment of inertia of the forearm would be $M_f L^2/3$ where $M_f$ is the mass of the forearm and $L$ is the forearm length. The moment of inertia of the cue stick about the elbow is $M_s L^2$. As both the arm and stick are accelerated about the elbow by a constant force $F$ for an angle $\theta$, the total energy is given by $T=FL\theta$. For a given stroke length $L\theta$ and force $F$, the total kinetic energy is independent of the cue stick and forearm masses. Writing the two parts of the energy explicitly gives

$$T = T_0 + T_f = \tfrac{1}{2} M_s L\omega^2 + \tfrac{1}{6} M_f L\omega^2 = T_0\left(1 + \frac{M_f}{3M_s}\right)$$

where $T_0$ is the cue stick energy. Although $T$, the total kinetic energy of the arm and stick, is fixed by $T=FL\theta$, the fractional division of this energy between the stick and arm is seen to be determined by the mass ratio. It is interesting in this expression that the only important factor is the mass ratio of the forearm and stick; the length of the forearm does not matter, at least within the current set of simplifying assumptions. This means that the optimal cue stick weight will be the same for tall players as for short players, provided the forearm masses are the same. Some players pivot their arm from the shoulder rather than the elbow on the break shot. The above analysis indicates that the additional arm length is irrelevant, but with this technique the entire arm mass rather than simply the forearm mass must be included into the $M_f$ term. Whether this is beneficial or not depends also on the relative forces applied by the different muscle groups involved in the two stroke techniques.

The dilemma is now apparent from the above equation and P3.3. In order to achieve the highest transfer of energy from the cue stick to the cue ball, a very light 6*oz* cue stick would be necessary. But in order to maximize the cue stick energy $T_0$ for a fixed total energy $T$ during the stroke, a very large cue stick mass would be necessary. Consequently, maximization of the cue ball energy requires some kind of compromise between these two extremes.

The quantity $T_0$ is the cue stick energy at the end of the stroke, and P3.3 gives the relation between $T_0$ and the cue ball energy $T_b$. The combination of these relations gives

$$\frac{T_b}{T} = \frac{4M_bM_s^2}{\left(M_b + M_s\right)^2\left(M_s + \frac{1}{3}M_f\right)} = \frac{4\alpha_s^2}{\left(1+\alpha_s\right)^2\left(\alpha_s + \frac{1}{3}\alpha_f\right)}$$

In the last expression, $\alpha_s = M_s/M_b$ is the ratio of the stick mass to ball mass, and $\alpha_f = M_f/M_b$ is the forearm to ball mass ratio. For a given forearm mass, the optimum stick mass is determined by differentiating the above expression with respect to $\alpha_s$, setting the result to zero, and solving for $\alpha_s$ as a function of $\alpha_f$. The final expression is

$$\alpha_{s(opt)} = \frac{1}{2} + \sqrt{\frac{1}{4} + \frac{2}{3}\alpha_f}$$

which is an equation for a parabola. When $\alpha_f=0$, it is seen that $\alpha_{s(opt)}=1$, and the optimal cue stick mass would be 6*oz*, a result which agrees with the conclusions from P3.3. A light forearm mass might be 24*oz*, which corresponds to $\alpha_f=4$, $\alpha_{s(opt)}=2.2$, and an optimal cue stick mass of 13.2*oz*. A typical forearm mass might be 36*oz*, which corresponds to an optimal stick weight of 15.4*oz*. A heavy forearm mass might be 64*oz*, which corresponds to an optimal stick weight of 19.3*oz*. A person who breaks with his entire arm, pivoting at the shoulder rather than the elbow, might have an arm mass of 150*oz*, which corresponds to an optimal stick weight of 27.2*oz*.

In the last few years, many professional 9-ball players have switched from heavy break cues to lighter break cues. These players may still use a typical 19-20*oz* cue for their normal strokes in a game, but they break with a lighter 15-18*oz* break cue. Break cues of this weight are consistent with the above equations, elbow pivots rather than shoulder pivots, and slim to medium body types. The actual breaking technique used by these players is more complicated than that considered above, and involves pivots about both the shoulder and the elbow.

**Problem 3.4:** What is the spin/speed ratio of the cue ball immediately after contact as a function of the vertical cue tip contact point?

*Answer*: For simplicity assume that the contact point is in the vertical plane through the center of the cue ball. When the cue tip applies a force in an off-center hit, the force accelerates the center of mass, and the resulting momentum is $p=MV$. The linear momentum is given by the expression $p = \int_0^t F(t')dt'$ in which the force is not constant during the contact time and $t$ is the (very short) contact time between the cue tip and the cue ball. (An ideal *impulsive* force is one that integrates to a constant momentum change as the contact time decreases. A cue tip contacting a cue ball and a hammer driving a nail are two examples of nearly ideal impulsive forces.) Integrating the angular acceleration equation in the same way gives $pR\sin(\theta)=pb=I\omega$. The quantity $b=R\sin(\theta)$ is the impact parameter, and is the vertical offset away from a center-ball hit. $b$ is positive for an above-center hit, zero for a center ball hit, and negative for a below-center hit. Eliminating the linear momentum $p$ from these two sets of equations gives

$$MV = \frac{I\omega}{b} = \frac{2MR^2\omega}{5b}$$

$$J = \left(\frac{R\omega}{V}\right) = \frac{5}{2}\left(\frac{b}{R}\right)$$

If $b$=0, then the angular velocity $\omega$ is also zero, which means that there is no spin imparted with a center-ball hit of the cue tip. If the cue tip hits above center, then $b$ is positive and $\omega=\omega_y$ is positive, which means that the ball is rolling in the same direction as the velocity. If the cue tip hits below center, then $b$ is negative and $\omega$ is negative, which means that the cue ball is spinning in the opposite direction as in a draw or drag shot. Note that the above equations are valid only for $-R\le b\le R$, or else $b$ is meaningless; the cue tip would miss the cue ball. For practical reasons, $b$ is restricted even more due to the fact that contact points close to the edge of the cue ball result in miscues (see P1.7). Although determined above for angular velocity about the horizontal axis, the same equation applies to angular velocity about the vertical axis resulting from a horizontal impact parameter, or, in fact, to any arbitrary angular velocity axis.

**Problem 3.5:** At what vertical contact point $b_{NR}$ will the cue ball have natural roll?
*Answer*: Natural roll occurs when $V=R\omega_y$. Substitution into the above equation gives

$$b_{NR} = \tfrac{2}{5}R$$

Noting that the height above the cloth is given by $z=R+b$, this point may also be written

$$z_{NR} = \tfrac{7}{5}R = \tfrac{7}{10}D$$

where $D=2R$ is the height of the ball. This point is actually rather high on the cue ball, and it is risky to attempt to hit higher than this due to the possibility of miscuing (see P1.7). Sidespin that is imparted to the cue ball with a level stick has no effect on natural roll, so the set of points on the cue ball for which natural roll is achieved immediately with no sliding are along the horizontal line at a height $^7\!/_{10}$D above the table surface.

**Exercise 3.1**: Experiment with shots involving natural roll impact points. Use a striped object ball in place of the cue ball. Orient the ball so that the plane defined by the stripe center is tilted at various angles away from vertical. The cue stick should be held as level as possible and should be within the plane defined by the stripe. The cue tip contact point should be exactly in the center of the stripe at a height $^7\!/_{10}$D above the table. When executed correctly, the stripe will appear “stationary” as the ball rolls.  A small error in the contact point, or in the ball setup, will result in a small wobble of the stripe on the rolling ball.

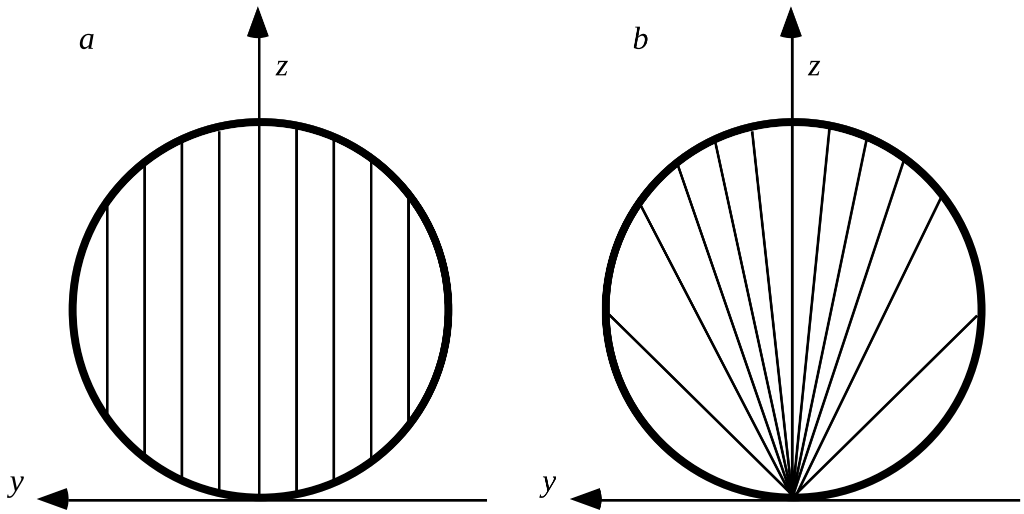



Fig. 3.1. The cue tip contact points corresponding to various arbitrary sidespin/speed ratios are denoted by the thin lines as viewed from the rear of the cue ball.  Figure *a* denotes constant spin/speed ratios immediately after contact with the cue tip; these are vertical straight lines.  Figure *b* denotes constant spin/speed ratios after natural roll is achieved; these are straight lines that all intersect at the point $(y,z)=(0,0)$.  In both cases, the larger offsets from the center are associated with higher spin/speed ratios.

**Problem 3.6:** Which cue tip contact points will result in the same sidespin/speed ratios immediately after contact with the cue tip?  Which contact points will result in the same sidespin/speed ratios after the cue ball achieves natural roll?

*Answer*: Consider the coordinate axes in Fig. 3.1.  The *z*-coordinate is the height above the cloth, and the *y*-coordinate is the distance away from the vertical plane through the center of the ball.  $b_y = y$ is the horizontal impact parameter, and $b_z = (z-R)$ is the vertical impact parameter.  Denote the point of contact with coordinates $(y,z)$.  In terms of the linear momentum $p$, the initial forward velocity and forward rotation are given by

$$V_0 = \frac{p}{M}$$
$$\omega_{0y} = \frac{p(z-R)}{I} = \frac{5p(z-R)}{2MR^2} \quad .$$

The forward rotation depends only on the height of the cue tip contact point $z$ and not on the sideways displacement $y$.  Upon achieving natural roll, the final forward velocity (see P2.2) is given by

$$V_{NR} = \tfrac{5}{7} V_0 + \tfrac{2}{7} R\omega_{0y} = \frac{5p}{7M} + \frac{5p(z-R)}{7MR} = \frac{5p}{7M}\left(\frac{z}{R}\right) \quad .$$

The sidespin (*i.e.* the angular velocity about the vertical axis) is assumed to be unchanged by the frictional forces of the sliding ball. From P3.4, the initial, and final, sidespin about the $z$-axis is given by

$$\omega_z = \left(\frac{5y}{2R^2}\right)V_0 = \frac{5yp}{2R^2 M} \quad .$$

The sidespin depends only on the horizontal displacement, $y$. The sidespin/speed ratio for the initial velocity is given by

$$J_z = \frac{R\omega_z}{V_0} = \frac{5y}{2R}$$

This ratio depends only on the horizontal impact parameter $y$, and is independent of the ball speed $V_0$ and vertical contact point $z$. The same ratio would occur with a soft hit as with a very hard hit.

Taking the ratio of the sidespin and final natural roll velocity gives

$$J_{z,NR} = \frac{R\omega_z}{V_{NR}} = \frac{7}{2}\left(\frac{y}{z}\right)$$

where $J_{z,NR}$ is the desired spin/speed ratio. The set of points ($y$,$z$) that correspond to the same $J_{z,NR}$ are given by the straight line defined by

$$z = \left(\frac{7}{2J_{z,NR}}\right)y$$

The lines corresponding to several $J_{z,NR}$ are shown in Fig. 3.1. It is interesting that exactly the same effect may be obtained by striking the cue ball at any point on a given straight line, provided the cue ball has sufficient time to achieve natural roll through sliding friction. For a desired final velocity, a higher initial velocity is required for small-$z$ contact points in order to overcome the drag. Note that higher sidespin/speed ratios (larger $J_{z,NR}$) are associated with straight lines closer to horizontal, and smaller ratios (smaller $J_{z,NR}$) are associated with more vertical slopes.

**Problem 3.7:** Of the set of points ($y$,$z$). that correspond to a constant natural-roll spin/speed ratio $J_{z,NR}$, which point ($y_0$,$z_0$) corresponds to the smallest displacement from center ball?

*Answer*: Consider Fig. 3.2. All the points a given distance from center ball will form a circle. The smallest circle that touches the desired straight line, as determined in P3.6, will define the smallest displacement that gives the desired spin/speed ratio. The point at which this smallest circle touches the appropriate straight line is denoted ($y_0$,$z_0$). At this point, the curve defining the circle and the straight line will be tangent, and the three points (0,0), (0,R), and ($y_0$,$z_0$) will form a right triangle. Let $\alpha$ be the angle away from vertical as indicated in Fig. 3.2. The tangent of this angle is given by $\tan(\alpha)=y_0/z_0$, and also by $\tan(\alpha)=(R-z_0)/y_0$. Equating these two expressions gives

$$y_0^2 = z_0(R - z_0).$$

Completing the square on the right hand side of this equation and rearranging gives

$$y_0^2 + \left(z_0 - \tfrac{1}{2}R\right)^2 = \left(\tfrac{1}{2}R\right)^2 \ .$$

This is recognized as the equation for a circle of radius $\frac{1}{2}R$ centered at the point $(0, \frac{1}{2}R)$. Contacting these points with the cue tip is called *aiming on the small circle*. When a player aims on the small circle, and the cue ball subsequently achieves natural roll, the desired spin/speed ratio $J_{z,NR}$ is achieved with the minimal displacement from center ball. It is possible to achieve much higher spin/speed ratios when the cue ball is allowed to achieve natural roll than the ratios that can be obtained immediately after cut tip contact as demonstrated in the following problem.

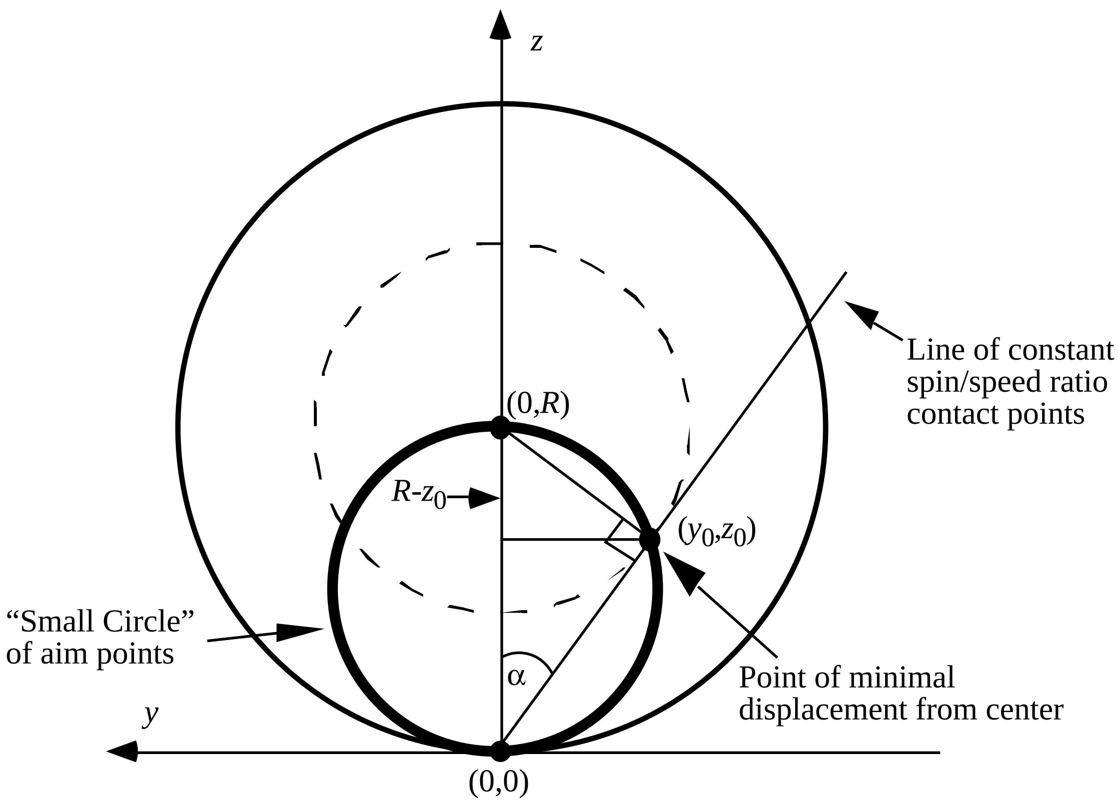



Fig. 3.2. The set of points that correspond to the minimal displacements from center ball for various spin/speed ratios after natural roll is achieved fall on a small circle of radius $R/2$ that touches the bottom point of the cue ball.

**Problem 3.8:** What is the natural roll sidespin/speed ratio, $R\omega_z/V_{NR}$, for the equatorial cue tip contact point $P_1=(y_1,z_1)=\left(\frac{1}{\sqrt{2}}R, R\right)$? What is the natural roll sidespin/speed ratio for the contact point $P_2=(y_2,z_2)=\left(\frac{1}{2}R, \frac{1}{2}R\right)$? At what contact points $P_3=(y_3,z_3)$ would the initial spin/speed ratio, $R\omega_z/V_o$, be the same as the natural roll spin/speed ratio of $P_2$?
*Answer*: From P3.6 the natural roll spin/speed ratio for $P_1$ is given by

$$\left(\frac{R\omega_z}{V_{NR}}\right) = \frac{7}{2}\left(\frac{y_1}{z_1}\right) = \frac{7}{2\sqrt{2}} = 2.475$$

The natural roll spin/speed ratio for $P_2$ is

$$\left(\frac{R\omega_z}{V_{NR}}\right) = \frac{7}{2}\left(\frac{y_1}{z_1}\right) = \frac{7}{2} = 3.5$$

Although the displacements away from center of these two points are the same, namely $R/\sqrt{2}$ , the sidespin/speed ratio for the second point is over 41% larger than the first point. The second point $P_2$ is on the "small circle" and therefore results in the maximal natural roll sidespin/speed ratio for this displacement distance.

In order to achieve a comparable initial sidespin/speed ratio

$$\frac{7}{2} = \frac{R\omega_z}{V_0} = \frac{5}{2}\left(\frac{y}{R}\right)$$

$$P_3 = (y_3, z_3) = \left(\tfrac{7}{5}R, z\right)$$

However, the set of points $P_3$ are not on the cue ball. Therefore, it is impossible to achive such a large sidespin/speed ratio without taking advantage of the drag to reduce the ball velocity. For practical purposes, a sidespin/speed ratio of 3.5 is about as large as can be attained with a cue tip impact with a level cue stick. Larger ratios can be achieved only with elevated cue stick strokes (masse) or with collisions involving other balls.

It is sometimes convenient to think of the cue ball spin and velocity at any moment in time for a sliding ball in terms of an "effective cue tip contact point". That is, for a given linear and angular velocity of a cue ball, there exists a contact point on the cue ball at which, if the cue tip where to strike a stationary ball at that point, with the correct velocity, the result would be to match, or to reproduce, exactly the same spin and speed. Because the linear and angular velocities change as the ball slides, the effective contact point is time dependent. From P3.6, the horizontal and vertical components of the spin are related to the vertical and horizontal components of the impact parameter of the cue tip contact point according to

$$\frac{5}{2}\left(\frac{b_y^{eff}}{R}\right) = \frac{R\omega_z}{V} = \frac{R\omega_{0z}}{\left(V_0 - \mu g t\right)}$$

$$\frac{5}{2}\left(\frac{b_z^{eff}}{R}\right) = \frac{R\omega_y}{V} = \frac{\left(R\omega_{0y} + \frac{5}{2}\mu g t\right)}{\left(V_0 - \mu g t\right)}$$

where the time dependence of the angular and linear velocities due to the cloth friction on the sliding ball from Section 2 have been used. The origin $t$=0 is taken in the above equations to be the time at which the cue tip strikes the ball.

**Problem 3.9**: Show that the set of effective contact points corresponding to $b_y^{eff}(t)$ and $b_z^{eff}(t)$ for a sliding ball lie on a straight line passing through the coordinate points $(y,z)=(0,0)$ and $(y,z)=(b_y^{eff}(0),R+b_z^{eff}(0))$.

*Answer*: Let $b_z^{eff}$ be considered as a function of $b_y^{eff}$ and defined parametrically through the time variable $t$. Solve the first equation above for $t$ in terms of $b_y^{eff}$, and substitute into the second to give

$$\frac{\left(R + b_z^{eff}(t)\right)}{b_y^{eff}(t)} = \frac{\left(R + b_z^{eff}(0)\right)}{b_y^{eff}(0)}$$

The right hand side of this equation is time independent. Therefore, the slope of the curve defined by the points $(y,z)=(b_y^{eff}(t),R+b_z^{eff}(t))$ is a constant, independent of time, and the set of time-dependent effective contact points lie on a straight line. The distance $(R+b_z^{eff}(t))$ is the height of the tip contact point above the cloth as seen for example in Fig. 3.2, and the distance $b_y^{eff}(t)$ is the horizontal tip displacement. Therefore, the line passing through the point (0,0) at the bottom of the ball to the initial point $(b_y^{eff}(0),R+b_z^{eff}(0))$ has the same slope as the rest of the line. The line segment of effective contact points ends when $b_z^{eff}(t)=\frac{2}{5}R$, at which time the ball achieves natural roll.

The result of P3.9 allows the player to compensate accurately for the effects of table friction on the spin axis with the following approach. First determine the desired spin axis at the eventual position of the cue ball. A stun shot for example, which is a frequent goal, would have a vertical spin axis at the time the cue ball collides with the object ball. This spin axis corresponds to some effective contact point $(b_y^{eff}(t),R+b_z^{eff}(t))$. In the case of a stun shot, this point would have coordinates $(b_y^{eff}(t),R)$ and correspond to pure sidespin. The player must then estimate, based on shot speed and the cloth friction, the required vertical offset below center in order to achieve a stun shot. Let this vertical distance be denoted $\delta$. The player then draws an imaginary line from the point $(b_y^{eff}(t),R)$, corresponding to the desired target spin state of the cue ball, to the point (0,0). The point on that imaginary line that corresponds to $(b_y^{eff}(0),R-\delta)$ is the desired contact point. Other final spin states would be estimated in the same manner. The straight line is always drawn from the final effective contact point to the origin (0,0), and the player works backward in time, so to speak, from the final spin state of the cue ball to the initial tip/ball contact time. If, during this process, the actual contact point $(b_y^{eff}(0),b_z^{eff}(0))$ is judged to be outside the boundary at which miscues begin to occur (see P1.7), then the desired shot is not possible, and the player should seek other alternatives.

**Problem 3.10**: What is the relation between the cue stick velocity immediately before contact, the cue ball velocity immediately after contact, and the impact parameter $b$? (assume that the total kinetic energy is conserved)

*Answer*: Conservation of linear momentum and kinetic energy give

$$M_s V_0 = M_s V_s + M_b V_b$$

$$\tfrac{1}{2} M_s V_0^2 = \tfrac{1}{2} M_s V_s^2 + \tfrac{1}{2} M_b V_b^2 + \tfrac{1}{2} I \,\omega_b^2$$

$$= \tfrac{1}{2} M_s V_s^2 + \left(\tfrac{1}{2} + \tfrac{5}{4}\left(\frac{b}{R}\right)^2\right) M_b V_b^2$$

Solve the first equation for $V_s$, and substitute into the second equation to obtain

$$V_b = \frac{2V_0}{1 + \dfrac{M_b}{M_s} + \tfrac{5}{2}\left(\dfrac{b}{R}\right)^2}$$

It may be verified that this expression agrees with that of P3.2 when $b$=0. It may now be understood why it is desirable to avoid spin on the cue ball during the break shot. For a given cue stick energy, or velocity $V_0$, any spin corresponding to nonzero $b$ has the effect of reducing the cue ball velocity and the translational kinetic energy; the maximum cue ball speed is achieved with a centerball $b$=0 contact point. The ratio $V_b/V_0$ is plotted as a function of impact parameter for some selected ball/stick mass ratios in Fig. 3.3.

**Problem 3.11**: What is the vertical impact parameter that maximizes the ratio $V_{NR}/V_0$ where $V_{NR}$ is the cue ball natural roll velocity and $V_0$ is the before-collision cue stick velocity?

*Answer*: From P2.2, P3.6, and P3.10, the natural roll velocity is given by

$$V_{NR} = \tfrac{5}{7} V_b + \tfrac{2}{7} R\omega_b = \left(\tfrac{10}{7}\right)\left(\frac{1 + \left(\dfrac{b}{R}\right)}{1 + \dfrac{M_b}{M_s} + \tfrac{5}{2}\left(\dfrac{b}{R}\right)^2}\right) V_0$$

Solving for the velocity ratio, differentiating with respect to $b$, setting the result to zero, and simplifying gives

$$\left(\frac{b}{R}\right)_{\max V_{NR}} = -1 + \sqrt{\frac{7}{5} + \frac{2}{5}\left(\frac{M_b}{M_s}\right)}$$

For a 6*oz* ball and an 18*oz* stick, the optimal impact point is given by

$$\left(\frac{b}{R}\right)_{\max V_{NR}} = 0.238 \qquad [M_s/M_b=3]$$

and for a 24 *oz* stick the optimal impact point is

$$\left(\frac{b}{R}\right)_{\max V_{NR}} = 0.225 \qquad [M_s/M_b=4]$$

This range includes most common stick weights and shows that the optimal impact point is only weakly dependent on the stick weight in this range. In both cases, the impact point is between centerball $b$=0 and the natural roll height $b=\tfrac{2}{5}R$. The initial cue ball velocity is maximized at $b$=0, but $\tfrac{2}{7}$ of this velocity is lost upon achieving natural roll to sliding friction; at $b = {}^2\!/_5 R$ there is no velocity loss due to sliding friction, but the initial velocity is relatively small due to the energy and momentum transfer conditions between the stick and ball. The above contact point is the optimal compromise between these two extremes. Maximization of the natural roll velocity is the same as maximizing the natural roll energy, and is the same as maximizing the distance that the ball rolls before stopping due to rolling resistance. Because this distance is maximized, this also means that the distance is relatively insensitive to small deviations of the contact point away from this optimal value. This is most useful when cue ball placement is of utmost importance such as, for example, during the lag shot at the beginning of a match. The ratio $V_{NR}/V_0$ is plotted as a function of impact parameter for some selected ball/stick mass ratios in Fig. 3.4.

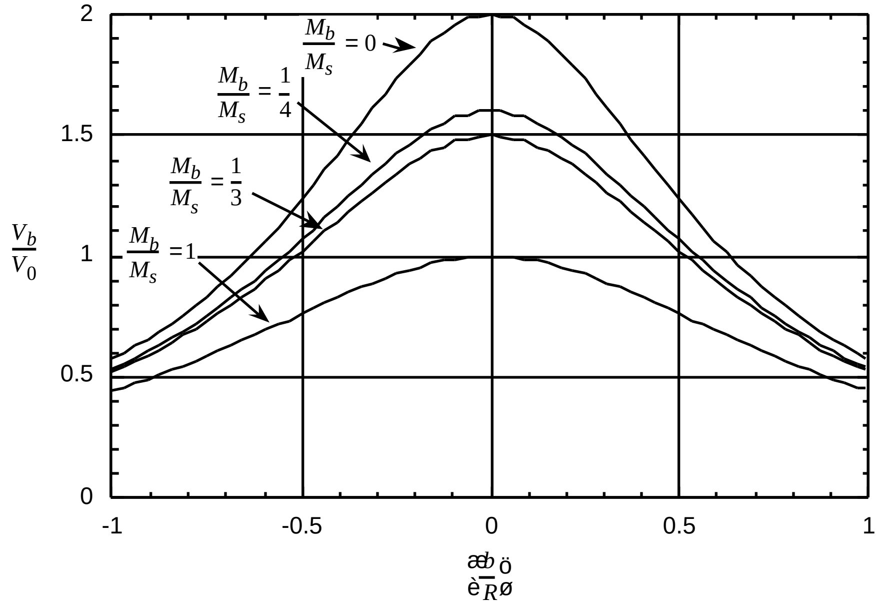



Fig 3.3. The ratio of the cue ball velocity $V_b$ to the before-collision cue stick velocity $V_0$ is shown as a function of the vertical impact parameter $(b/R)$ for some selected ball/stick mass ratios.

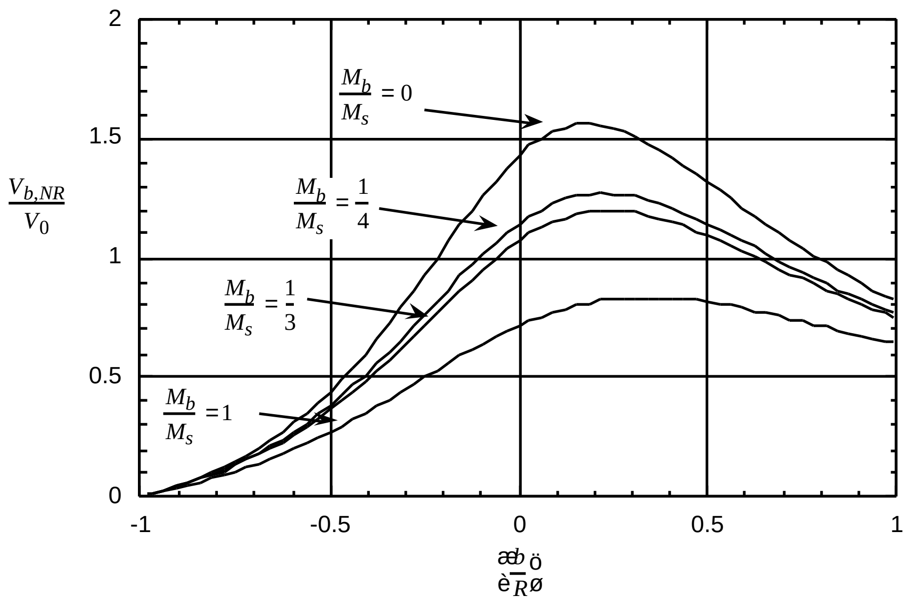



Fig. 3.4. The ratio of the final natural roll cue ball velocity $V_{b,NR}$ to the before-collision cue stick velocity $V_0$ is shown as a function of the vertical impact parameter ($b/R$) for some selected ball/stick mass ratios. For a given ball/stick mass ratio, the optimal contact point for a lag shot is determined by the flat region near the curve maximum.

## 4. Collisions Between Balls

Consider the motions of two colliding balls. One ball is assumed to be moving before the collision, and both balls are assumed to be moving afterwards. For this discussion, assume that the initially moving ball is the cue ball, and the initially stationary ball is an object ball. As the two balls collide in an off-center hit, the frictional forces acting tangential to the surfaces are relatively small (e.g. compared to the frictional forces between a ball and the cue tip). All of the remaining force is directed along the line between the centers of the balls.

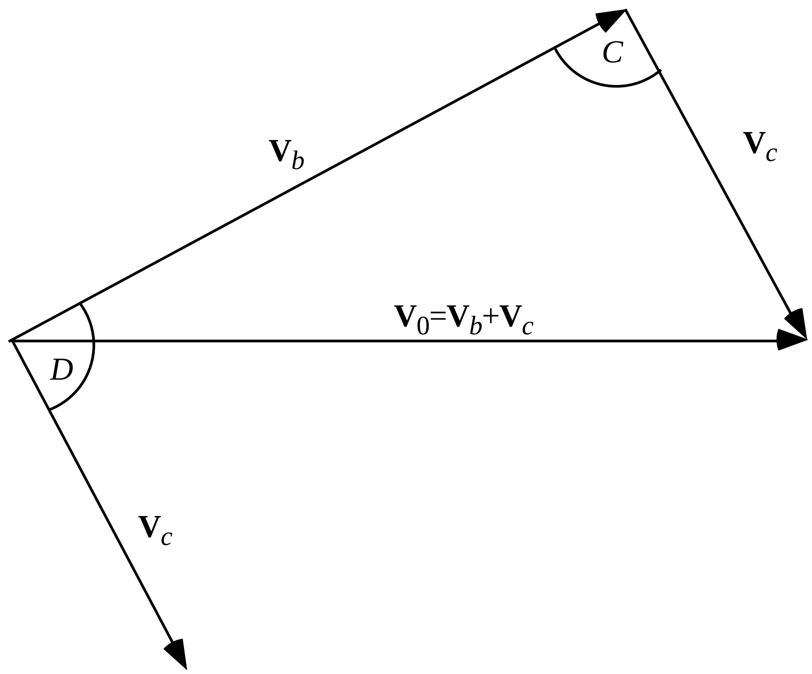



Fig. 4.1. Pictorial representation of the conservation of momentum vector relation $\mathbf{V}_0=\mathbf{V}_b+\mathbf{V}_c$. The angles C and D are supplementary and satisfy the relation $C+D=\pi$.

Consider first the ball motions just before the collision and just after the collision; in this situation, the friction between the cloth and the sliding/rolling balls has not had time to affect the ball trajectories. Linear momentum ($\mathbf{p}=M\mathbf{V}$) is conserved in both the *x*- and *y*-coordinate directions. Represented with vectors, the vector sum of the final momentum of the two balls is equal to the initial momentum of the cue ball. Eliminating the mass $M$ of the balls, results in the vector relation $\mathbf{V}_0=\mathbf{V}_b+\mathbf{V}_c$ between the initial and final velocities. This relation is shown pictorially in Fig. 4.1. The final velocity of the cue ball $\mathbf{V}_c$ has been drawn twice: once with its base common to that of the $\mathbf{V}_b$ vector, which is consistent with both balls departing from the same collision point on the table, and again with its base at the end of the $\mathbf{V}_b$ vector to show pictorially that $\mathbf{V}_0=\mathbf{V}_b+\mathbf{V}_c$. The angles $D$ and $C$ are supplementary and are related by (in radians) $C+D=\pi$, and consequently, $\cos(C)=-\cos(D)$.

In addition to momentum, energy is also conserved in this collision to a good approximation. The relatively small amount of energy that is lost is turned into sound or heat within the balls. An *elastic* collision is one in which energy is assumed to be conserved, so this energy loss will be denoted $E_{inelastic}$. As discussed in the previous sections, there are two kinds of kinetic energy, translational and rotational, associated with each ball. Equating the energy before and after the collision gives

$$T_{0(Trans)} + T_{0(Rot)} = T_{c(Trans)} + T_{c(Rot)} + T_{b(Trans)} + T_{b(Rot)} + E_{inelastic}$$

Collecting all the $T_{(Rot)}$ terms together, and multiplying by 2/M gives the relation

$$V_0^2 = V_b^2 + V_c^2 + \Delta_{elastic} + \Delta_{inelastic} = V_b^2 + V_c^2 + \Delta_{total}$$

with

$$\Delta_{elastic} = \frac{2}{M}\left(T_{c(Rot)} + T_{b(Rot)} - T_{0(Rot)}\right)$$

$$\Delta_{inelastic} = \frac{2}{M} E_{inelastic}$$

The term $\Delta_{elastic}$ depends on the total change of rotational energy. The contribution $\Delta_{elastic}$ may be positive, zero, or negative, but the term $\Delta_{inelastic}$ is always positive, since it represents an energy loss in the collision process. There are two types of contributions to $E_{inelastic}$, the first type of energy loss is due to the frictional forces of the sliding balls. These frictional forces result in the exchange of energy between the various translational and rotational components. Just as in the case of the simple sliding block, the frictional forces are intimately related to the inelastic energy loss; without this inelastic energy loss, there would be no sliding friction. As will be seen in the following discussions, this inelastic energy loss can be determined by analysis of the resulting momentum exchange between the balls. Other contributions to the inelastic energy loss involve the imperfect transfer of energy between the balls. For example, the sound made by the colliding balls represents a transfer of kinetic energy from the collision process to the surroundings. This energy loss would occur even in the absence of sliding frictional forces. In the present discussion, this latter type of energy loss will not be considered quantitatively in the analysis. With this simplification, both the elastic and inelastic contributions to $\Delta_{total}$ are assumed to be associated with the tangential forces of sliding friction.

The law of cosines for an arbitrary triangle with sides *a, b,* and *c* with corresponding angles *A*, *B*, and *C* is

$$c^2 = a^2 + b^2 - 2ab\cos(C)$$

This allows the angles of a triangle to be related to the lengths of the three sides. In particular, the sides of the triangle resulting from the pictorial representation of the conservation of momentum relation may be related to the departure angle. Comparing the law of cosines with the above velocity equation gives the relation

$$\cos(C) = \frac{-\Delta_{total}}{2|V_b||V_c|} = -\cos(D)$$

between the angles $C$ and $D$ and the change of translational energy term $\Delta_{total}$. If there is no translational kinetic energy loss during the collision of the balls, then $\Delta_{total}=0$, $\cos(C)=0$, and $C=\pi/2$ is a right angle (*i.e.* 90 degrees). In this case, the law of cosines reduces to the familiar theorem of Pythagoras. If $C=\pi/2$, then $D=\pi/2$ and the two balls depart at exactly a right angle. In this initial discussion it will be assumed that the balls are rotating about the vertical axes only; the more general situation is examined later. If there is no rotational energy change during the collision, then $\Delta_{elastic}=0$. There are three situations in which there will be no total rotational energy change during a collision. First, if there is no friction between the balls, then there will be no tangential forces acting at the point of contact. This is, of course, an approximation, but for many shots such an approximation is sufficient, and in any case it defines a convenient reference point. The second situation in which no spin change occurs is when the cue ball has just the right amount of outside spin so that the ball surfaces are not moving relative to each other during the (very short) collision time. In this case the cue ball spin is unchanged, and the object ball acquires no spin during the collision. The third situation in which no total rotational energy change occurs is when the cue ball has just the right amount of inside spin so that all of the cue ball spin is transferred to the object ball, and the cue ball departs with no spin. The first situation is an ideal, and occurs only with no friction between the colliding balls; $\Delta_{total}=0$ in this case for all collision situations. The second situation is independent of the ball friction, but depends on matching exactly the outside spin and the cut angle; $\Delta_{total}=0$ for this situation since both components vanish when there is no friction. The third situation depends on matching the amount of inside spin with the friction between the balls and the cut angle; since there are accelerations associated with the frictional forces, there is a nonzero $\Delta_{inelastic}$ component, $\Delta_{total}\neq 0$, and therefore the departure angle will differ from $\pi/2$.

To appreciate the importance of spin transfer, consider a cut shot, with ball friction, when the cue ball has no spin initially. In this case, the $T_{0(Rot)}$ term will be zero, but both $T_{c(Rot)}$ and $T_{b(Rot)}$ will be nonzero. The cue ball acquires some sidespin by rubbing against the object ball, and the initially motionless object ball acquires some sidespin by rubbing against the cue ball. In this case, both $\Delta_{elastic}>0$ and $\Delta_{inelastic}>0$, the angle $C$ will be larger than $\pi/2$, and the angle of departure $D$ will be smaller than a right angle. In actual practice this is a small effect, in the neighborhood of 2-4 degrees depending on how sticky are the pair of colliding balls, but a 4 degree angle, over 8 feet results in a deviation of 6.7", or about half a diamond on a 9' table ($\tan(\alpha)=d/L$ with deviation angle $\alpha$, distance $L$, and deviation distance $d$). When referring to the resulting object ball deviations, this effect is called *collision-induced throw*, and clearly this must be accounted for, to some extent, on any but the most trivial of shots.

**Problem 4.1**: What are the conditions in which $\Delta_{elastic}$ will be positive, zero, and negative? (assume all spins are about the vertical axes)

*Answer*: Substituting the rotational energy expression gives

$$\Delta_{elastic} = \tfrac{2}{5} R^2 \left( \omega_c^2 + \omega_b^2 - \omega_0^2 \right)$$

where all angular velocities are relative to the vertical axes of each ball. However, any change of angular velocity in the cue ball must be compensated exactly by a corresponding change in the object ball angular velocity, since the frictional forces on each ball are equal but opposite in direction.

$$\omega_0 = \omega_c - \omega_b \ .$$

Substitution of this relation gives

$$\Delta_{elastic} = \tfrac{4}{5} R^2 \omega_b \omega_c \ .$$

When the final spins of both balls are in the same direction (*i.e.* both are clockwise when looking down on the table from above, or both are counterclockwise), then $\Delta_{elastic}$ will be positive, cos($D$) will be positive, and the angle of departure of the two balls will be $<\pi/2$. When the final spin of either the cue ball or the object ball is zero, then $\Delta_{elastic}$ will be zero, and the departure angle will be $\leq\pi/2$, and the magnitude will depend entirely on $\Delta_{inelastic}$ which is always nonnegative. These are the only situations that result in $\Delta_{elastic}=0$. When the final spins of the two balls are in opposite directions (*i.e.* one clockwise and the other counterclockwise), then $\Delta_{elastic}$ will be negative, and the departure angle will depend on the relative magnitudes of the two components $\Delta_{elastic}$ and $\Delta_{inelastic}$. Note that cos($D$) depends on the final spin/speed ratios of the balls, so within the current set of simplifying approximations, the contribution of $\Delta_{elastic}$ to the departure angle is independent of the overall shot speed.

The above qualitative analysis did not require a detailed examination of the forces during the collision process. These forces and the resulting ball trajectories are now examined in more detail. For this pupose, it is useful to define two coordinate systems as shown in Fig. 4.2. The first coordinate system, denoted ($x'$,$y'$,$z'$), is appropriate for the initial cue ball velocity before the collision; the second, denoted ($x$,$y$,$z$) is the natural coordinate system to describe the trajectories after the collision. Unit vectors along these two coordinate axes satisfy the transformation relation

$$\begin{pmatrix} \hat{\mathbf{i}}' \\ \hat{\mathbf{j}}' \\ \hat{\mathbf{k}}' \end{pmatrix} = \begin{pmatrix} \cos(\alpha) & \sin(\alpha) & 0 \\ -\sin(\alpha) & \cos(\alpha) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \hat{\mathbf{i}} \\ \hat{\mathbf{j}} \\ \hat{\mathbf{k}} \end{pmatrix}$$

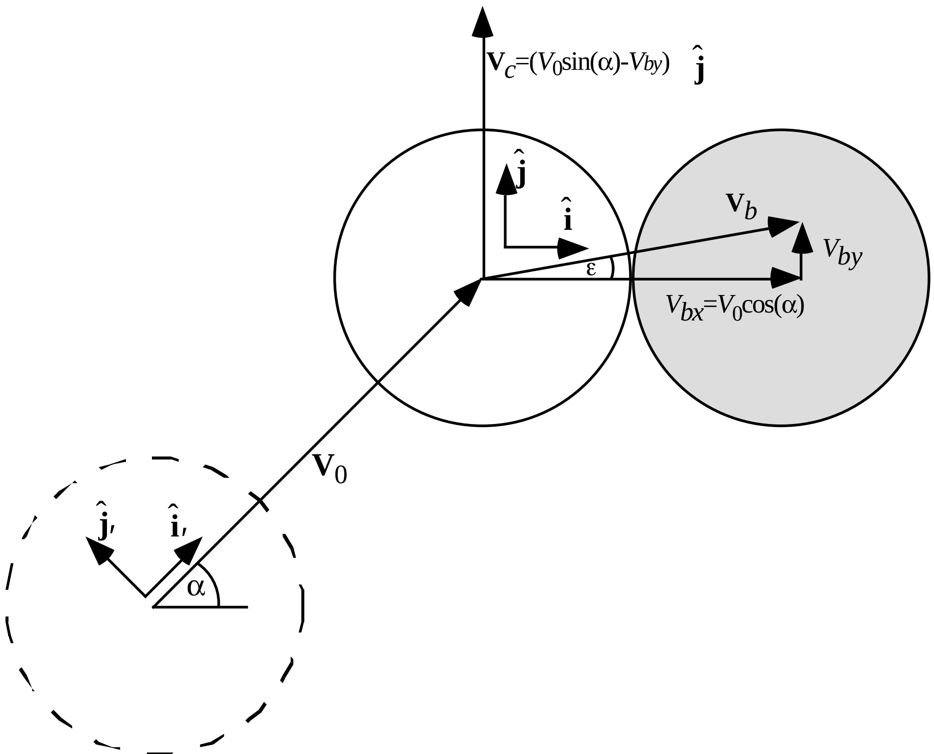



Fig. 4.2. The effects of the sliding frictional forces on the object ball and cue ball are shown in detail on the after-collision velocity vectors. Two coordinate systems are used in the analysis of object ball throw. The first is relative to the initial cue ball velocity $\mathbf{V}_0$, the second is appropriate to describe the after-collision velocities. The vertical $z$-coordinate is not shown, but is directed out of the plane of the figure. The angle $\alpha$ would be the object ball cut angle if there were no friction.

It is convenient to take the origin of the $(x,y,z)$ coordinate system to be the cue ball center at the moment of contact with the object ball. With this choice, the contact point of the cue ball and object ball lies on the $x$-axis. In the absense of friction, the object ball would depart along the $x$-axis and the cue ball would depart along the $y$-axis. The frictional forces are tangential to the point of contact, and therefore lie in the $yz$ plane. The direction of the frictional force is determined by the velocity of the contact point of the cue ball at the moment of contact. The contact point velocity is the sum of the linear velocity

$$\mathbf{V}_0 = V_0\hat{\mathbf{i}}' = V_0\left(\cos(\alpha)\hat{\mathbf{i}} + \sin(\alpha)\hat{\mathbf{j}}\right)$$

and the angular velocity

$$\omega_0 \times \mathbf{r}_{cp} = R\,\omega_0 \times \hat{\mathbf{i}} = \begin{vmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ \omega_{0x} & \omega_{0y} & \omega_{0z} \\ R & 0 & 0 \end{vmatrix} = R\left(\omega_{0z}\hat{\mathbf{j}} - \omega_{0y}\hat{\mathbf{k}}\right)$$

If the cue ball is struck with a level cue stick (*i.e.* no masse), then the cue ball rotation may be written as

$$\omega_0 = \omega'_{0y}\hat{\mathbf{j}}' + \omega'_{0z}\hat{\mathbf{k}}' = -\omega'_{0y}\sin(\alpha)\hat{\mathbf{i}} + \omega'_{0y}\cos(\alpha)\hat{\mathbf{j}} + \omega_{0z}\hat{\mathbf{k}}$$

$\omega'_{0y}<0$ for backspin, $\omega'_{0y}=0$ for a stun shot, and $\omega'_{0y}>0$ for topspin. $\omega'_{0z} = \omega_{0z}$ corresponds to sidespin. The resulting contact point velocity is

$$\mathbf{V}_{cp} = V_0\cos(\alpha)\hat{\mathbf{i}} + \left(V_0\sin(\alpha) + R\omega_{0z}\right)\hat{\mathbf{j}} - R\omega'_{0y}\cos(\alpha)\hat{\mathbf{k}}$$
$$= V_{cpx}\hat{\mathbf{i}} + V_{cpy}\hat{\mathbf{j}} + V_{cpz}\hat{\mathbf{k}}$$

It is the sign of $V_{cpy}$ that determines the direction of throw of the object ball. $V_{cpy}>0$ results in throwing the object ball in the $+\hat{\mathbf{j}}$ direction, $V_{cpy}<0$ results in $-\hat{\mathbf{j}}$ throw, and $V_{cpy}=0$ results in no throw. It is interesting that, for a given angle $\alpha$, $V_{cpy}$ depends only on the cue ball sidespin $\omega_{0z}$. Cue ball topspin or draw does not change the direction of throw, but it does change the magnitude of the throw.

The $V_{cpx}$ component of the contact point velocity is directed exactly along the object ball center of mass. As the balls collide, the momentum component $p_x=MV_{cpx}$ is transferred entirely from the cue ball to the object ball. This momentum is transferred during the very short collision time $t$ according to the equation $p_{bx} = \int_0^t F_x(t')dt'$. If there are any tangential components of the contact point velocity, then at any time during the collision there is a tangential frictional force with magnitude given by $F(t) = \mu_{bb}F_x(t)$ where $\mu_{bb}$ is the ball-ball sliding coefficient of friction. The direction of this tangential force is determined by the tangential components of the contact point velocity. A unit vector in this tangential direction may be defined as

$$\hat{\mathbf{e}} = \frac{\mathbf{V}_{cp}}{\left|\mathbf{V}_{cp}\right|} = \frac{V_{cpy}\hat{\mathbf{j}} - V_{cpz}\hat{\mathbf{k}}}{\left|V_{cpy}\hat{\mathbf{j}} - V_{cpz}\hat{\mathbf{k}}\right|}$$
$$= \frac{\left(V_0\sin(\alpha) + R\omega_{0z}\right)\hat{\mathbf{j}} - R\omega'_{0y}\cos(\alpha)\hat{\mathbf{k}}}{\left[\left(V_0\sin(\alpha) + R\omega_{0z}\right)^2 + \left(R\omega'_{0y}\cos(\alpha)\right)^2\right]^{1/2}}$$
$$= \cos(\gamma)\hat{\mathbf{j}} + \sin(\gamma)\hat{\mathbf{k}}$$

with obvious definitions for the horizontal component $\cos(\gamma)$ and vertical component $\sin(\gamma)$. The vertical component of this force direction $\sin(\gamma)$ either works in conjuction or opposition to the weight of the ball; it does not affect the *direction* of the cue ball or object ball velocities in the plane of the table after the collision. However, the horizontal component of the force $\cos(\gamma)$ does affect the object ball direction. It is this horizontal component of the force that results in the object ball throw. Fig. 4.3 shows the possible combinations of directions for the unit vector $\hat{\mathbf{e}}$ and the geometrical meaning of the components $\cos(\gamma)$ and $\sin(\gamma)$. The factor $\cos(\gamma)$ may be thought of as a *geometrical efficiency factor* in converting the frictional forces into throw velocities.

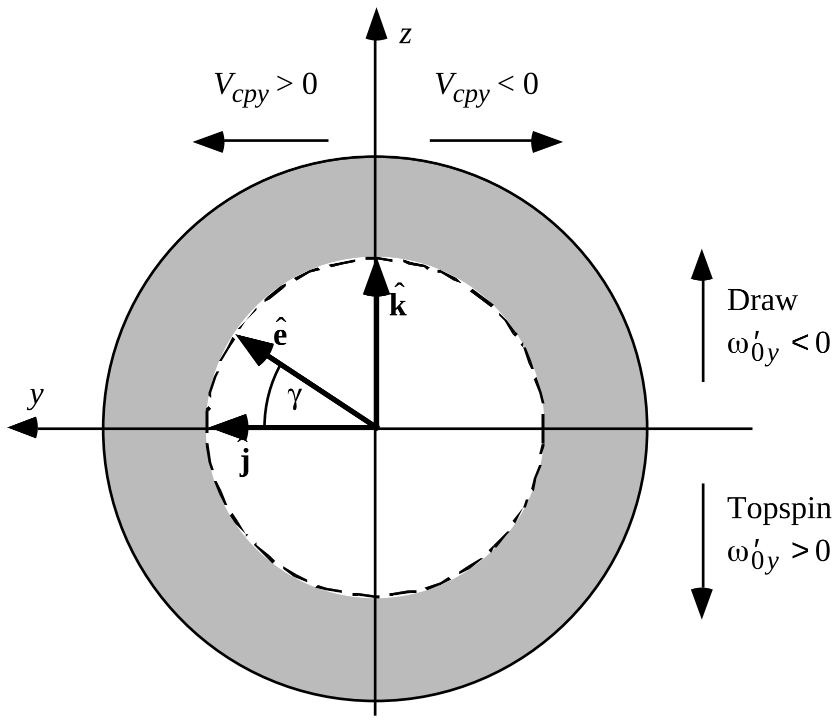



Fig. 4.3. The unit vector $\hat{\mathbf{e}}$, parallel to the direction of the sliding frictional force on the object ball, is decomposed into the horizontal and vertical components characterized by the angle γ. This force is applied to the object ball at the contact point, and an opposing force is applied to the cue ball. This force is tangential to the ball surfaces and lies in the *yz*-plane. The direction of the unit vector depends on the cut angle and the spin axis of the cue ball at the moment of the collision. The object ball throw is proportional to the horizontal component of the frictional force.

The object ball throw is determined by the *y*-component of the frictional force. Substitution of the above decomposition of $\hat{\mathbf{e}}$ gives the relations

$$p_{by} = \int_0^t F_y(t')dt' = \cos(\gamma)\mu_{bb}\int_0^t F_x(t')dt' = \cos(\gamma)\mu_{bb}p_{bx}$$

$$V_{by} = \cos(\gamma)\mu_{bb}V_{bx}$$

The horizontal component of the tangential frictional force results in the throw velocity $V_{by}$ being added to the object ball velocity, and the opposing frictional force acts to subtract exactly this velocity from the post-collision cue ball velocity. Because the factor cos(γ) depends on several parameters, it is useful to consider some special cases.

**Problem 4.2**: How does the throw angle ε defined by $\tan(\varepsilon)=V_{by}/V_{bx}$, depend on overall shot speed?

*Answer*: Rewriting the cos(γ) expression in terms of spin/speed ratios gives

$$\cos(\gamma) = \frac{\left(\sin(\alpha) + \frac{R\omega_{0z}}{V_0}\right)}{\left(\left(\sin(\alpha) + \frac{R\omega_{0z}}{V_0}\right)^2 + \left(\frac{R\omega'_{0y}}{V_0}\cos(\alpha)\right)^2\right)^{1/2}}$$

The geometrical factor cos(γ) is seen to depend entirely on spin/speed ratios, not overall shot speed. The throw angle is $\varepsilon = \arctan(V_{by}/V_{bx}) = \arctan(\mu_{bb}\cos(\gamma))$. The velocity ratio, and therefore the throw angle ε is independent of the shot speed. In practice, this result is not entirely true; the throw angle decreases slightly for very hard shots. This change of throw angle with shot speed is due to a slight speed-dependence of $\mu_{bb}$. Fig. 4.4 shows the dependence of the object ball throw factor cos(γ) as a function of the sidespin/speed ratio ($R\omega_{0z}/V_0$) for a specific cut angle of π/6 (a half-ball cut) for several values of the topspin/speed ratio.

**Problem 4.3**: For a stun shot, $\omega'_{0y}=0$, how does the throw velocity depend on the cue ball cut angle α?

*Answer*: For a stun shot, the cos(γ) factor reduces to the form

$$\cos(\gamma) = \frac{V_{cpy}}{|V_{cpy}|} = \frac{\left(V_0\sin(\alpha) + R\omega_{0z}\right)}{|V_0\sin(\alpha) + R\omega_{0z}|} = \pm 1 \qquad [\text{for } \omega'_{0y}=0]$$

The sign of the cos(γ) factor is determined by the initial velocity component, the cut angle α, and the sidespin $\omega_{0z}$. The throw velocity is then given by

$$V_{by} = \pm\mu_{bb} V_{bx}$$

If the cue ball has no sidespin, then cos(γ)=+1, and $V_{by} = \mu_{bb}V_{bx}$ for the shot angle in Fig. 4.2. This result was *assumed* in P1.6, as a way to determine $\mu_{bb}$, but it is now seen with a careful analysis that this assumption was indeed correct [provided the frozen object ball acts the same as a stun-shot collision]. The only dependence of the throw velocity on the cut angle is in the *direction* of the frictional force. Fig. 4.4 shows the dependence of the object ball throw factor cos(γ) as a function of the sidespin/speed ratio ($R\omega_{0z}/V_0$) for a stun shot. There is an abrupt change in value as $V_{cpy}$ changes sign.

**Problem 4.4**: For a natural roll cue ball, $R\omega'_{0y}=V_0$ (or a reverse natural roll cue ball, $R\omega'_{0y}=-V_0$) how does the throw angle depend on the cue ball cut angle α?

*Answer*: For a natural roll cue ball, the cos(γ) factor reduces to the form

$$\cos(\gamma) = \frac{\left(\sin(\alpha) + \frac{R\omega_{0z}}{V_0}\right)}{\left(\left(\sin(\alpha) + \frac{R\omega_{0z}}{V_0}\right)^2 + \cos(\alpha)^2\right)^{1/2}} \qquad [\text{NR or RNR}]$$

In Fig. 4.4, this factor is plotted as a function of sidespin/speed ratio for a specific cut angle $\alpha=\pi/6$. The throw angle is determined by $\varepsilon=\arctan(\mu_{bb}\cos(\gamma))$. Although the slope is steepest in the region near $V_{cpy}=0$, the slope is not as steep in this region as that for smaller values of $|R\omega'_{0y}/V_0|$.

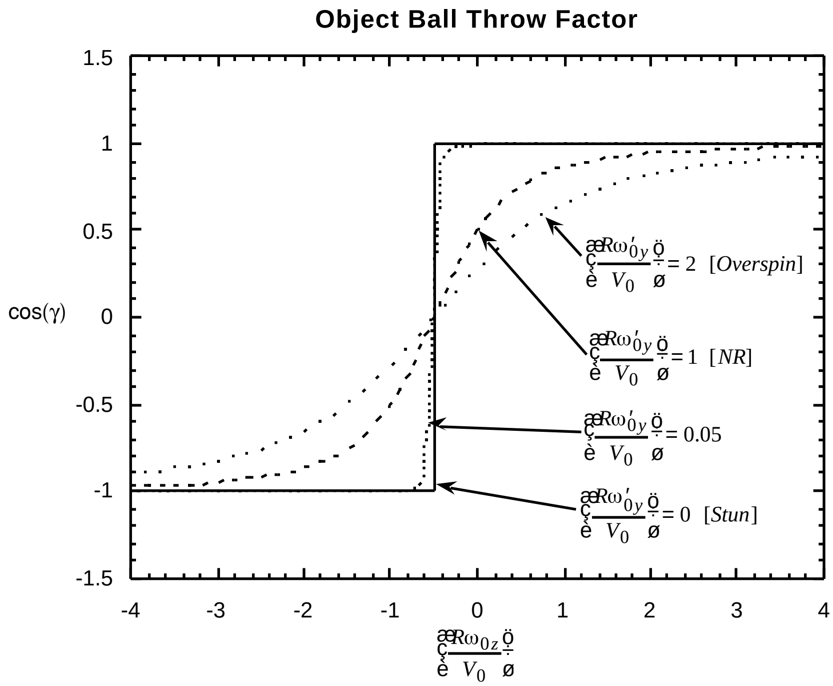



Fig. 4.4. The object ball throw factor $\cos(\gamma)$ is shown as a function of the cue ball sidespin to speed ratio ($R\omega_{0z}/V_0$) for selected values of of cue ball topspin/draw. The slope of the given curve determines how sensitive is the object ball throw to small variations in the sidespin.

In practice, it is impossible to achieve an exact stun shot. There will always be some small value of $\omega'_{0y}$. Similarly, the quantity $V_{cpy}=(V_0\sin(\alpha)+R\omega_{0z})$ will never be exactly zero; it may be very small, but it will never be exactly zero. This leads to the question of how the throw angle depends on small variations from these limiting conditions. The answer is that the direction of the unit vector $\hat{\mathbf{e}}$ becomes very sensitive, rotating wildly even with very small changes in the cue ball spin. Both the numerator and the denominator of the components become small, but without a definite limit. Therefore, the $\cos(\gamma)$ factor can vary between –1 and +1, and the throw velocity can vary anywhere between $-\mu_{bb}V_{bx}$. and $+\mu_{bb}V_{bx}$. For small values of $\omega'_{0y}$, the slope of the $\cos(\gamma)$ curve becomes very steep; this steepness reflects the sensitivity of the object ball throw to the sidespin. This correlation of steepness of slope with small $\omega'_{0y}$ values may be seen in Fig. 4.4. This slope reflects the sensitivity of the throw factor $\cos(\gamma)$ with respect to changes in the sidespin. The sensitivity of the throw factor with respect to changes in the topspin is related to the derivative of cos(γ) with respect to the other spin/speed ratio ($R\omega'_{0y}/V_0$).

**Problem 4.5**: What is the sensitivity of the object ball throw with respect to both components of the cue ball spin?

*Answer*: It is conventient to characterize the sensitivity in terms of the spin/speed ratios $J_{0z}=(R\omega_{0z}/V_0)$ and $J_{0y}=(R\omega'_{0y}/V_0)$. The sensitivity of the throw factor to the cue ball spin is characterized by the derivatives

$$\frac{d\cos(\gamma)}{dJ_{0y}} = \frac{-\left(\sin(\alpha)+J_{0z}\right)J_{0y}\cos^2(\alpha)}{\left(\left(\sin(\alpha)+J_{0z}\right)^2+\left(J_{0y}\cos(\alpha)\right)^2\right)^{3/2}}$$

$$\frac{d\cos(\gamma)}{dJ_{0z}} = \frac{\left(J_{0y}\cos(\alpha)\right)^2}{\left(\left(\sin(\alpha)+J_{0z}\right)^2+\left(J_{0y}\cos(\alpha)\right)^2\right)^{3/2}}$$

The first equation gives the sensitivity of the throw with respect to changes in the topspin or backspin of the cue ball, the second equation gives the sensitivity with respect to changes in the sidespin. When $J_{0y}$ is small, then the slope of the cos(γ) factor is approximately

$$\frac{d\cos(\gamma)}{dJ_{0z}} \approx \frac{\left(J_{0y}\cos(\alpha)\right)^2}{\left|\sin(\alpha)+J_{0z}\right|^3} \qquad \text{[for small } J_{0y}\text{]}$$

This shows why the slope of the cos(γ) curve becomes essentially vertical in Fig. 4.4 as the sidespin $J_{0z}$ passes through the zero point of $V_{cpy}$ and the denominator of this component of the sensitivity vanishes.

A combined measure of the sensitivity of the object ball throw to the cue ball spin may be defined as

$$F(\mathbf{J}_0) = \sqrt{\left(\frac{d\cos(\gamma)}{dJ_{0y}}\right)^2 + \left(\frac{d\cos(\gamma)}{dJ_{0z}}\right)^2}$$

For values of $\mathbf{J}_0$ that correspond to small $F(\mathbf{J}_0)$, the player is allowed larger margins of error in shot execution (e.g. in the accuracy of the cue tip contact point) and in judgement (e.g. in estimating, and compensating for, the object ball throw). Regions with large $F(\mathbf{J}_0)$ are those where very small spin variations result in large changes in the object ball throw; these are the regions that the player should try to avoid. Fig. 4.5 shows a contour plot of the sensitivity $F$ as a function of the two components of the cue ball spin, $J_{0z}$ and $J_{0y}$, for the same cut angle as was used in Fig. 4.4, namely $\alpha=\pi/6$ (a half-ball cut). It may be observed that the regions of least sensitivity are those with small $J_{0y}$ (*i.e.* close to being a stun shot), and large sidespin $|J_{0z}|$ (*i.e.* corresponding to extreme underspin or overspin). Regions of high sensitivity are seen to correspond to $V_{cpy}$=0 (*i.e.* to $J_{0z}$=-sin(α)=-½). The highest sensitivity contours correspond to the region near the point $V_{cpy}$=0 and $J_{0y}$=0; a magnified view of this region is shown in the inset in Fig. 4.5. The sensitivity of the throw angle becomes enormous in this region. Ironically, the spin combinations that result in the smallest object ball throw sometimes correspond also to the largest sensitivity, and the spin combinations that result in the largest throw sometimes correspond also to the smallest sensitivity.

With this sensitivity in mind, it is possibly a wise tactic to avoid these conditions so as to avoid the large uncertainty in the throw angle. That is, stun shots with outside spin should be avoided, according to this argument, when the effects of throw might be critical to the success of the shot. This uncertainty may be avoided in practice by ensuring that the numerator or the denominator (or both) are significantly different from zero at the moment of collision of the cue ball with the object ball. This may be done for a given shot either by avoiding stun-shot spin (*i.e.* ensuring $\omega'_{0y} \neq 0$ thereby reducing the magnitude of the cos(γ) factor), or by avoiding the $V_{cpy}$=0 condition (thereby producing a predictable, although nonzero throw), or by avoiding both simultaneously.

It should be pointed out that this recommendation is somewhat contrary to that given by some other players, teachers, and authors. Their argument is that minimizing the $V_{cpy}$ factor will minimize the throw. As seen in Fig. 4.5, this is only true if $|\omega'_{0y}|$ differs from zero and is large compared to $|V_{cpy}|$. In practice for some types of shots, it may be easier to avoid the $V_{cpy}$=0 combinations of speed and sidespin by intentionally overspinning or underspinning the cue ball, and to account explicitly for the throw by adjusting the aim point. This approach might be preferable in situations where stun-shot spin is necessary for position. Examples of this compensation are described in the following problems. Another complicating factor is the seemingly random phenomenon called *skid* (also called *cling* or *kick*). Skid occurs when a small piece of chalk or dust is trapped between the contact point of the balls, increasing dramatically the coefficient of friction for that particular shot. When this occurs, the amount of throw associated with nonzero $V_{cpy}$ is very unpredictable.

**Problem 4.6**: For a natural roll cue ball (or reverse natural roll cue ball) with no sidespin, $\omega_{0z}$=0, how does the throw angle depend on the cue ball cut angle α?
*Answer*: From P4.4, the cos(γ) factor reduces to the form

$$\cos(\gamma) = \frac{(\sin(\alpha))}{\left(\sin(\alpha)^2 + \cos(\alpha)^2\right)^{1/2}} = \sin(\alpha) \qquad \text{[NR or RNR with } \omega_{0z}=0]$$

The throw angle is determined by $\varepsilon=\arctan(\mu_{bb}\cos(\gamma))=\arctan(\mu_{bb}\sin(\alpha))$. The throw depends only on the cut angle α. It is 0 for a straight in shot (α=0), and increases to a maximum value for very thin cuts (α≈π/2). The impact parameter for the cue ball/object ball collision is $b_{bb}=R\sin(\alpha)$. This allows the factor $\cos(\gamma)=b_{bb}/R$ to be easily determined geometrically for any given cut shot with natural roll and no sidespin.

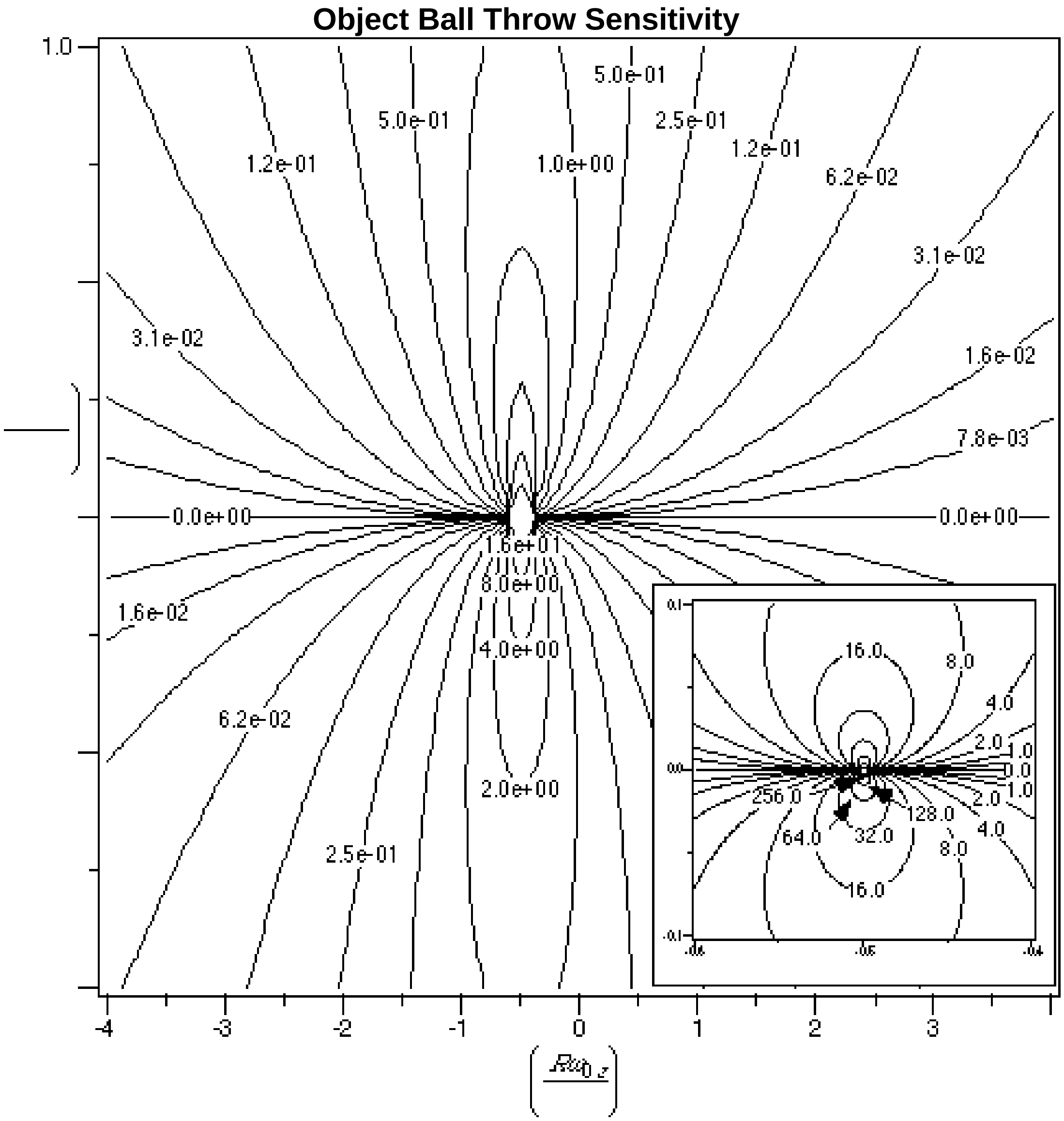



Fig. 4.5. A contour plot of the sensitivity of the object ball throw factor $\cos(\gamma)$ is shown as a function of the cue ball sidespin to speed ratios $J_{0z}=(R\omega_{0z}/V_0)$ and the topspin-draw spin to speed ratio $J_{0y}=(R\omega'_{0y}/V_0)$. Adjacent contours differ by a factor of two in the sensitivity function $F(\mathbf{J}_0)$. The inset figure is an expanded view of the small region near $J_{0y}=0$ and $V_{cpy}=0$.

Based on these considerations, the following procedure may be used to adjust for object ball throw for natural roll shots with no sidespin. (1) Determine $\mu_{bb}$ using the procedure in P1.6. This only needs to be done once for a given set of balls. (2) For the particular shot of interest, estimate the distance $D$ from the object ball to the pocket; the corresponding maximum throw distance will be $\mu_{bb}D$. (3) For the zero-friction cut angle for the particular shot of interest, estimate the impact parameter for the ball-ball collision, and the ratio $b_{bb}/R$. (4) Multiply the maximum throw distance $\mu_{bb}D$ by the impact parameter ratio $b_{bb}/R$, and call the result *s*. (5) Imagine a point that is displaced by the distance *s* from the pocket target, and aim for this offset point as if there were no throw.

For an example of this procedure, assume that $\mu_{bb}$ has been determined for the set of balls as in P1.6 to be 4/72. For the shot of interest, the distance from the object ball to the pocket is 36". The maximum throw distance for this shot is (4/72)*36"=2"; that is, half the reference shot distance results in half the maximum throw distance. Suppose that the shot of interest is almost straight-in, a slight cut to the left, with $b_{bb}/R$=¼. The offset distance is given by s=¼*2"=½". Now a displaced point ½" to the left of the pocket center is used as a corrected aim point. This aim point is valid for either natural roll or reverse natural roll. With a little bit of practice, these estimations become second nature and may be done almost instantaneously. For other $\omega'_{0y}$ spin combinations, the offset point will be displaced from the target pocket somewhere between the maximum value of 2" (appropriate for a stun shot) and the natural roll value of ½", but the offset aim point will always be on the "overcut" side of the pocket center. Experienced pool players know to "cut 'em thin to win" when the balls are sticky, and the above procedure quantifies just "how thin" is "thin" to achieve the most consistent results.

The use of sidespin also requires further adjustments to the above procedure, but this requires even more judgement on the part of the shooter. One way to adjust for sidespin is to estimate mentally the cos(γ) factor by imagining how the cue ball will be spinning at the time of contact. Replacing the cue ball with a striped ball, and practicing various combinations of topspin, draw, stun, and sidespin will help the player develop this estimation skill. In general, the offset point will always be displaced less than the maximum value determined by $\mu_{bb}D$. Of course, small $\mu_{bb}$ values mean that any errors made in the estimation of the cos(γ) factor result in smaller errors in the object ball trajectory. Sticky balls with large $\mu_{bb}$ are very challenging. One of the challenges faced by tournament players is the accurate adjustment to different sets of balls, each with different $\mu_{bb}$, as they move from table to table in the tournament matches.

**Problem 4.7**: What is the resulting object ball spin $\omega_b$ due to the frictional force $\mathbf{F}(t)$?
*Answer*: The angular acceleration is given by the equation $\mathbf{r}\times\mathbf{F}=\mathbf{I}\,\dot{\omega}$. Integration of the force over the contact time gives

$$\omega_b=\frac{-R}{I}\int_0^t\hat{\mathbf{i}}\times\mathbf{F}(t')dt'=\frac{-R\mu_{bb}}{I}\hat{\mathbf{i}}\times\left(\cos(\gamma)\hat{\mathbf{j}}+\sin(\gamma)\hat{\mathbf{k}}\right)\int_0^t F_x(t')\,dt'$$

$$=\frac{-5\mu_{bb}V_{bx}}{2R}\left(-\sin(\gamma)\hat{\mathbf{j}}+\cos(\gamma)\hat{\mathbf{k}}\right)$$

**Problem 4.8**: What is the relation between the natural roll spin axis and the object ball throw angle? (For simplicity, ignore the effects of the vertical friction components during the cue ball and object ball collision.)

*Answer*: Since both the object ball sidespin and the object ball throw angle are caused by the same frictional force, the magnitudes of these two effects are closely related. Immediately after the collision with the cue ball (and ignoring the object ball spin due to the vertical friction components), the object ball linear velocity and angular velocity vectors are given by

$$\mathbf{V}_b = V_{bx}\hat{\mathbf{i}} + V_{by}\hat{\mathbf{j}} = V_{bx}\hat{\mathbf{i}} + \mu_{bb}\cos(\gamma)V_{bx}\hat{\mathbf{j}} = V_b\hat{\mathbf{e}}_b$$

$$\omega_b = \omega_{bz}\hat{\mathbf{k}} = -\left(\frac{5\mu_{bb}\cos(\gamma)V_{bx}}{2R}\right)\hat{\mathbf{k}}$$

with $\tan(\varepsilon)=\mu_{bb}\cos(\gamma)$. After achieving natural roll, the object ball linear and angular velocity vectors are

$$\mathbf{V}_{bNR} = \tfrac{5}{7}V_b\hat{\mathbf{e}}_b$$

$$\omega_{bNR} = \omega_b + \frac{5V_b}{2R}\hat{\mathbf{e}}_b$$

where the unit vector $\hat{\mathbf{e}}_b = \hat{\mathbf{k}}\times\hat{\mathbf{e}}_b$ is the horizontal vector perpendicular to the object ball velocity. The angle of the natural roll spin axis is related to the components of the spin axes according to

$$\tan(\beta) = \frac{\omega_z}{\omega_z} = -\frac{5V_{by}}{2V_{bNR}} = -\frac{7\tan(\varepsilon)}{2\sqrt{1+\tan^2(\varepsilon)}} = -\frac{7}{2}\sin(\varepsilon)$$

For the typically small object ball throw angles, the approximate relations

$$\beta \approx -\tfrac{7}{2}\varepsilon \approx -\tfrac{7}{2}\mu_{bb}\cos(\gamma) \qquad \text{[for small } \varepsilon\text{]}$$

show that the natural roll spin axis tilt angle is about 3½ times larger in magnitude than the corresponding object ball throw angle, and that both angles are approximately linear with respect to the ball-ball friction coefficient. This axis tilt is most easily observed by viewing the rolling object ball from directly behind its path and by noting the equivalent tilt of the stationary rotational equator. The relation between the spin axis and the rotational equator is shown in Fig. 4.6. This axis tilt may be used to give the player additional feedback in adjusting the compensation for object ball throw on cut shots.

P4.7 gives the resulting object ball spin if the frictional force acts on the ball without opposition. During the collision, in order for a horizontal component of angular acceleration to occur, the ball-ball friction must act simultaneously with the ball-cloth friction. It will be assumed hereafter that the ball-cloth friction is insignificant during the collision time, and its affects will be ignored. The practical accuracy of this approximation may be estimated by the following considerations. A typical collision time is $t$=0.0001$s$, and a typical object ball velocity is $V_{bx}$=100$in/s$. The average impact force is then given by $F_{avg}=MV_{bx}/t$. The sliding frictional force of the ball on the cloth is given by $F_s=\mu_s Mg$. The ratio is given by $F_s/F_{avg}=\mu_s gt/V_{bx}$. Assuming a ball-cloth

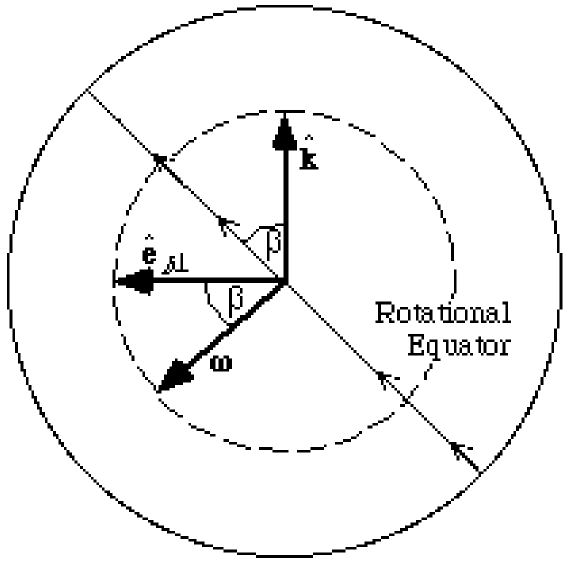



Fig. 4.6. The relation between the tilt of the spin axis ω and the rotational equator is shown pictorially as viewed from the rear of a naturally rolling ball. For an object ball, the angle β of the tilt of this axis is approximately 7⁄2 times larger in magnitude than the object ball throw angle.

sliding coefficient of friction $\mu_s$=0.1, this ratio is $F_s/F_{avg}$=0.0000386. Therefore, the ball-ball frictional forces do indeed dominate the ball-cloth frictional forces during the collision.

The treatment of the vertical acceleration due to the vertical component of the frictional force is somewhat complicated. The table surface prevents any vertical acceleration in the downward direction. The weight of the ball opposes any upward frictional force, but it doesn't prevent upward acceleration. Therefore, during the contact period, if the ball is on the table surface and ($F_z$–$Mg$) is negative, resulting in a downward net force, there is no acceleration at that instant. But if ($F_z$–$Mg$) is positive, then that upward force results in vertical acceleration of the ball off the table surface. If $Mg$ is neglible compared to a large positive $F_z$ then the maximum vertical velocity immediately after the collision would be the same as the maximum throw velocity; the maximum angle that the object ball departs from the table surface would be the same as the maximum horizontal object ball throw angle. With the average impact force given by $F_{avg}=MV_{bx}/t$ and the downward force of gravity given by $F_{grav}=Mg$, then the ratio is given by $F_{grav}/F_{avg}=gt/V_{bx}$. For the typical shot considered in the previous paragraph, the numerical value of this ratio is $F_{grav}/F_{avg}$=0.000386. Therefore, the ball-ball frictional forces also dominate the gravitational forces during the collision.

**Problem 4.9**: A cue ball with backspin strikes an object ball straight on. Assume the gravitational force on the ball is negligible during the collision, a shot speed of 36"/$s$, and $\mu_{bb}$=4/72 as in P1.6. What height does the object ball achieve over the table, and how far away from the starting point does it land?

*Answer*: The vertical velocity is given by

$$V_{bz} = \mu_{bb}\sin(\gamma)\, V_{bx}$$

For a straight on shot with backspin, sin(γ)=+1 and the entire frictional force is directed upward. $V_{bz}$ =(4/72)36"/$s$=2"/$s$. The height of the ball trajectory above the table is given by

$$z = V_{bz}\, t - \tfrac{1}{2} g t^2 = (2"/s)t - \tfrac{1}{2}(386"/s^2)t^2$$

The maximum height is achieved when $dz/dt$=0. This occurs at $t_{max} = V_{bz}/g = \mu_{bb} V_{bx}/g$. The time to achieve maximum height is linear in the coefficient of friction $\mu_{bb}$ and in the shot speed $V_{bx}$.

$$t_{max} = V_{bz}/g = 2/386\ s = 0.00518\ s$$

The height achieved at this time is

$$z_{max} = V_{bz}\, t_{max} - \tfrac{1}{2} g t_{max}^{\ 2} = \frac{V_{bz}^2}{2g} = \frac{\mu_{bb}^2 V_{bx}^2}{2g}$$

$$= (2/386)" = 0.00518"$$

The maximum height achieved is proportional to the square of the coefficient of friction and to the square of the shot speed. The ball returns to the table at the time ($2t_{max}$). At this time, the horizontal distance traveled by the ball while airborne is

$$x = V_{bx}(2t_{max}) = \frac{2\mu_{bb} V_{bx}^2}{g}$$

$$= 36"(2)(2/386) = 0.373"$$

The horizontal distance of the jump is proportional to the coefficient of friction and to the square of the shot speed. Due to the very short times and small distances that the object ball is airborne, this jumping effect can be neglected, for the most part, during play.

One point to notice in P4.9 is that while the object ball has a vertical momentum immediately after the collision, the cue ball is constrained to the table surface. If the cue ball strikes the object ball with topspin, then it is the cue ball that leaves the table and the object ball that is constrained to the table surface. In either case, the vertical component of the linear momentum is not conserved by the balls during the collision. The reaction of the downward-directed ball is absorbed by the table. If the table had been considered to be part of the system, then linear momentum would have been conserved in the analysis. In this respect, the nonconservation of linear momentum in the vertical direction is an artifact of the formal separation between the “system” and the “surroundings” in this simple analysis.

**Problem 4.10**: Using the velocity and spin results from P4.2-P4.7, compute the total kinetic energy before and after the collision. Determine $E_{inelastic}$. (For simplicity, ignore the velocity and spin resulting from the vertical components of the frictional force.)

*Answer*: The total kinetic energy immediately before the collision is

$$T_0 = T_{0(Trans)} + T_{0(Rot)} = \tfrac{1}{2} M V_0^2 + \tfrac{1}{2} I \omega_0^2$$

The kinetic energy immediately after the collision is

$$T_f = \tfrac{1}{2} M\left(V_c^2 + V_b^2\right) + \tfrac{1}{2} I\left(\omega_c^2 + \omega_b^2\right)$$

Writing all of the friction-dependent contributions in terms of $V_{by}$ gives

$$V_{by} = \mu_{bb}\cos(\gamma)V_{bx} = \mu_{bb}\cos(\gamma)V_0\cos(\alpha)$$

$$\mathbf{V}_b = V_0\cos(\alpha)\hat{\mathbf{i}} + V_{by}\hat{\mathbf{j}}$$

$$\mathbf{V}_c = \left(V_0\sin(\alpha) - V_{by}\right)\hat{\mathbf{j}}$$

$$\omega_b = \frac{-5V_{by}}{2R}\hat{\mathbf{k}}$$

$$\omega_c = \omega_0 + \omega_b = -\omega'_{0y}\sin(\alpha)\hat{\mathbf{i}} + \omega'_{0y}\cos(\alpha)\hat{\mathbf{j}} + \left(\omega_{0z} - \frac{5V_{by}}{2R}\right)\hat{\mathbf{k}}$$

Substitution into the kinetic energy expression gives

$$T_f = T_0 + M\left(-V_{by}\left(V_0\sin(\alpha) + R\omega_{0z}\right) + \tfrac{7}{2}V_{by}^2\right)$$

$$= T_0 + M\left(-V_{by}V_{cpy} + \tfrac{7}{2}V_{by}^2\right)$$

The kinetic energy change $E_{inelastic}$ is given by

$$E_{inelastic} = T_0 - T_f = M\left(V_{by}\left(V_0\sin(\alpha) + R\omega_{0z}\right) - \tfrac{7}{2}V_{by}^2\right)$$

$$= M\left(V_{by}V_{cpy} - \tfrac{7}{2}V_{by}^2\right)$$

The friction allows for transfer of energy between the translational and rotational degrees of freedom, but only at a cost. This is consistent with the effect of ball-cloth friction on the kinetic energy as discussed previously. In the expressions above, $V_{cpy}$ is the horizontal tangential component of the contact point velocity of the cue ball at the instant of collision. $V_{cpy}$ determines the direction of the frictional force on the object ball and therefore has the same sign as $V_{by}$. The lowest order term in $\mu_{bb}$ in the loss of energy due to friction, $MV_{by}V_{cpy}$, is positive. The second term, which is second order in $\mu_{bb}$ and therefore in general much smaller in magnitude, is always negative.

**Problem 4.11**: Determine $\Delta_{elastic}$, $\Delta_{inelastic}$, and $\Delta_{total}$ in terms of $V_{by}$. What are these quantities when $V_{cpy}=0$?

*Answer*: From P4.10, $\Delta_{inelastic}$ is given by

$$\Delta_{inelastic} = \tfrac{2}{M}E_{inelastic} = 2V_{by}V_{cpy} - 7V_{by}^2$$

Generalizing the approach of P4.1 for arbitrary cue ball spin $\omega_0$,

$$\Delta_{elastic} = \tfrac{2}{5}R^2\left(\omega_c^2 + \omega_b^2 - \omega_0^2\right) = \tfrac{2}{5}R^2\left(\omega_c^2 + \omega_b^2 - (\omega_c - \omega_b)\ (\omega_c - \omega_b)\right)$$

$$= \tfrac{4}{5}R^2\ \omega_c\ \ \omega_b$$

$$= -2R\omega_{0z}V_{by} + 5V_{by}^2$$

$$\Delta_{total} = \Delta_{elastic} + \Delta_{inelastic} = 2V_{by}V_0\sin(\alpha) - 2V_{by}^2$$

In general $\Delta_{total}$ is a quadratic function of the ball-ball sliding coefficient of friction $\mu_{bb}$.

In the special case of $V_{cpy}$ =0, then also $V_{by}$=0 and $\Delta_{total}$ vanishes, indicating that the departure angle of the cue ball and object ball is exactly a right angle.

The initial velocity of the cue ball immediately after collision is given by $\mathbf{V}_c = \left(V_0 \sin(\alpha) - V_{by}\right)\hat{\mathbf{j}}$. The *magnitude* of this velocity depends on the object ball throw, but its *direction* is independent of any frictional forces. If the cue ball has no spin about the horizontal axis (*i.e.* only sidespin, no backspin or topspin), then this initial direction is unchanged by the sliding friction of the cloth. The cue ball will slow down upon achieving natural roll, but the velocity direction will remain unchanged. In this sense, the trajectory of the cue ball after the collision is less dependent on the ball-ball coefficient of friction $\mu_{bb}$ than the object ball trajectory. This observation is useful in judging and executing accurate stun shot caroms.

**Exercise 4.1**: Experiment with stun shot caroms. Begin by placing the cue ball a few inches away from the object ball, and cueing exactly in the center. The cue ball should not curve after the collision. Mark the position of the cue ball center at the collision point and the two contact points where the balls touch the cushions. Measure the angle and determine how close is the departure angle to a right angle. Include shots with sidespin to determine the effects of $\Delta_{total}$ on the departure angle. With some practice, stun shot caroms can be executed very accurately. Stun shot caroms are particularly useful in 9-ball.

**Problem 4.12**: Determine the total angular momentum immediately before and after the collision relative to the point that corresponds to the cue ball center at the moment of collision. Is angular momentum conserved? (ignore the linear velocity components due to the vertical frictional forces)

*Answer*: There are two contributions to the total angular momentum. One is the rotational contributions of the balls spinning about their centers, $\mathbf{L}^{spin} = \mathbf{I}\,\omega$, and the other is the orbital contribution of the centers of mass moving about the point of origin, $\mathbf{L}^{orbit} = \mathbf{r} \times \mathbf{p}$. Before the collision, these contributions are

$$\mathbf{L}_0^{orbit} = \mathbf{r}_0(t) \times \mathbf{p}_0(t) = \left(\mathbf{V}_0 t\right) \times \left(M\mathbf{V}_0\right) = 0$$

$$\mathbf{L}_0^{spin} = \mathbf{I}\,\omega_0$$

$$\mathbf{L}_0 = \mathbf{L}_0^{orbit} + \mathbf{L}_0^{spin} = \mathbf{I}\,\omega_0$$

After the collision the contributions are

$$\begin{aligned}\mathbf{L}_b^{orbit} &= \mathbf{r}_b(t) \times \mathbf{p}_b(t) = \left(2R\hat{\mathbf{i}} + \mathbf{V}_b t\right) \times \left(M\mathbf{V}_b\right) \\ &= 2MR\left(\hat{\mathbf{i}} \times \mathbf{V}_b\right) = 2MR\left(\hat{\mathbf{i}} \times \left(V_{bx}\hat{\mathbf{i}} + V_{by}\hat{\mathbf{j}}\right)\right) = 2MRV_{by}\hat{\mathbf{k}} \\ &= -2I\omega_{bz}\hat{\mathbf{k}}\end{aligned}$$

$$\mathbf{L}_b^{spin} = \mathbf{I}\,\omega_b$$
$$\mathbf{L}_c^{orbit} = \mathbf{r}_c(t) \times \mathbf{p}_c(t) = \left(\mathbf{V}_b t\right) \times \left(M\mathbf{V}_b\right) = 0$$
$$\mathbf{L}_c^{spin} = \mathbf{I}\,\omega_c = \mathbf{I}\left(\omega_0 + \omega_b\right)$$
$$\mathbf{L} = \mathbf{L}_c^{orbit} + \mathbf{L}_c^{spin} + \mathbf{L}_b^{orbit} + \mathbf{L}_b^{spin} = \mathbf{I}\,\omega_0 + 2I\omega_{by}\hat{\mathbf{j}}$$

The total angular momentum difference before and after the collision is then

$$\mathbf{L} - \mathbf{L}_0 = 2I\omega_{by}\hat{\mathbf{j}}$$

The total angular momentum is always conserved except for the horizontal component about the *y*-axis, which is conserved only when $\omega_{by}$=0. This component arises from the vertical frictional force during the collision, and vanishes only when $\omega'_{0y}$=0 (*i.e.* for stun shot collisions). The vertical component of angular momentum is always conserved, as is the other horizontal component about the *x*-axis; the orbital angular momentum arising from the object ball throw compensates exactly for the change in the spin angular momentum. This compensation cannot occur for the vertical frictional force because of the constraint of the table surface. In the above equations, the vertical linear acceleration was neglected, but even if it had been included for the jumped ball (as determined in P4.9), the corresponding contribution from the nonjumped ball during the collision is eliminated by the table surface. Indeed, as discussed previously, because the vertical components of linear momentum are not conserved in the collision, it should not be expected that the angular momentum components due to these same frictional forces could be conserved using the same simple analysis.

In the previous few problems, various aspects of object ball throw have been examined. The object ball throw affects the trajectories of the balls immediately after the collision. The behavior of the balls after the collision is determined by both the initial post-collision conditions of the balls and by the action of the cloth friction on the sliding balls which was discussed in some detail in the previous sections. The results of the present section heretofore, involving ball-ball interaction will now be combined with the results of the previous sections to examine the behavior of the sliding balls as a function of the collision conditions, and eventually, as a function of the tip-ball contact point. In the following discussions, object ball throw will be largely ignored in order to simplify the derivations. In most cases, the effects of object ball throw may be included, at the cost of some additional complexity, but this adds relatively little to the basic understanding of the situations. The first situation to be considered is the behavior of a natural roll cue ball after collision with an object ball. This special case is particularly central to pool and billiards because of the special importance of natural roll.

**Problem 4.13:** What is the angle of deflection of a natural roll cue ball as a function of the object ball cut angle after the collision and after natural roll is achieved by both balls?

(ignore friction between the balls)

*Answer*: With no ball-ball friction, the initial cue ball deflection direction is $\pi/2$ (90 degrees) from the object ball cut angle. In terms of unit vectors $\hat{\mathbf{i}}'$ and $\hat{\mathbf{j}}'$ in the $x'$ and $y'$ coordinate directions respectively in Fig. 4.2, the initial velocity vectors immediately after collision are given by

$$\mathbf{V}_b = V_0 \cos(\alpha)\left(\cos(\alpha)\hat{\mathbf{i}}' - \sin(\alpha)\hat{\mathbf{j}}'\right)$$

$$\mathbf{V}_c = V_0 \sin(\alpha)\left(\sin(\alpha)\hat{\mathbf{i}}' + \cos(\alpha)\hat{\mathbf{j}}'\right) \quad .$$

The cut angle $\alpha$ is the angle between vectors $\mathbf{V}_b$ and $\mathbf{V}_0$. There is no initial object ball angular velocity immediately after the collision, so only the speed changes and not the direction upon achieving natural roll. The final natural roll velocity is given by

$$\mathbf{V}_{b,NR} = \tfrac{5}{7}\mathbf{V}_b = \tfrac{5}{7}V_0 \cos(\alpha)\left(\cos(\alpha)\hat{\mathbf{i}}' + \sin(\alpha)\hat{\mathbf{j}}'\right) \; .$$

The situation is somewhat different for the cue ball. The cue ball has natural roll before the collision, $V_0 = R\omega'_{0y}$, and this angular velocity is unchanged by the collision with the object ball. The ball-cloth friction from this initial angular velocity creates a force component in the $\hat{\mathbf{i}}'$ direction only. The final velocity vector for the cue ball is

$$\mathbf{V}_{c,NR} = \tfrac{5}{7}\mathbf{V}_c + \tfrac{2}{7}V_0\hat{\mathbf{i}}' = \left(\tfrac{5}{7}V_0 \sin^2(\alpha) + \tfrac{2}{7}V_0\right)\hat{\mathbf{i}}' + \left(\tfrac{5}{7}V_0 \sin(\alpha)\cos(\alpha)\right)\hat{\mathbf{j}}' \, .$$

The cue ball deflection angle $\theta$, relative to the velocity vector $\mathbf{V}_0$, after natural roll is achieved, is determined by

$$\tan(\theta) = \frac{\sin(\alpha)\cos(\alpha)}{\sin^2(\alpha) + \frac{2}{5}}$$

Immediately after the collision, the cue ball path is a parabola as determined in P2.3. The frictional force accelerates the cue ball until natural roll is achieved. At the point that natural roll is achieved, the cue ball rolls in a straight line with no acceleration. The angle between this straight line and the initial velocity direction $\mathbf{V}_0$ is the deflection angle $\theta$ which satisfies the above equation.

**Problem 4.14:** Show that $\tan(\alpha + \theta) = \tfrac{7}{2}\tan(\alpha)$

*Answer*: Using the tangent addition relation $\tan(\alpha + \theta) = \dfrac{\tan(\alpha) + \tan(\theta)}{1 - \tan(\alpha)\tan(\theta)}$ with $\tan(\alpha) = \dfrac{\sin(\alpha)}{\cos(\alpha)}$ and $\tan(\theta) = \dfrac{\sin(\alpha)\cos(\alpha)}{\sin^2(\alpha) + \frac{2}{5}}$ gives

$$\tan(\alpha + \theta) = \frac{\sin(\alpha)\left(\sin^2(\alpha) + \cos^2(\alpha) + \frac{2}{5}\right)}{\frac{2}{5}\cos(\alpha)} = \tfrac{7}{2}\tan(\alpha)$$

**Problem 4.15:** What cut angle $\alpha$ maximizes the natural roll deflection angle $\theta$?

*Answer*: Rewrite the above expression as $\theta = \arctan\left(\tfrac{7}{2}\tan(\alpha)\right) - \alpha$. Differentiate with respect to α to obtain

$$\frac{d\theta}{d\alpha} = \frac{14}{4+45\sin^2(\alpha)} - 1.$$

Setting the derivative to zero and solving for α gives

$$\alpha_{(\theta_{max})} = \arcsin\left(\frac{\sqrt{2}}{3}\right) = 0.49088 = \frac{\pi}{6.3999} \quad [= 28.125 \text{ deg}]$$

Note that this is just a bit thicker than a half-ball hit, which is a π/6 or a 30 degree cut angle (neglecting collision induced throw).

**Problem 4.16:** What is the maximum deflection angle θ for a natural-roll cue ball collision?

*Answer*: Substitution of $\alpha_{(\theta_{max})}$ gives

$$\theta_{max} = \arctan\left(\tfrac{7}{2}\tan\left(\alpha_{(\theta_{max})}\right)\right) - \alpha_{(\theta_{max})}$$

$$= \frac{\pi}{2} - 2\alpha_{(\theta_{max})} = 0.58903 = \frac{\pi}{5.3335} \quad [= 33.749 \text{ deg}]$$

This is very useful to know because a natural-roll cue ball carom at this angle is intrinsically more accurate than a cut shot with the same cut angle as demonstrated in the following problem.

**Problem 4.17:** If the object ball is cut about 2 degrees away from that corresponding to the maximum deflection angle as determined in P4.15, what is the change in the cue ball deflection angle?

*Answer*: If the cut angle is 2 degrees less, then

$$\theta = \arctan\left(\tfrac{7}{2}\tan(26\text{deg})\right) - 26\text{deg} = 33.64\text{deg}$$

which is 0.11 degrees away from the maximal value as determined in P4.16. If the cut angle is 2 degrees more, corresponding to a half-ball hit of 30 degrees, then

$$\theta = \arctan\left(\tfrac{7}{2}\tan(30\text{deg})\right) - 30\text{deg} = 33.67\text{deg}$$

which is 0.08 degrees away from the maximal value. In both cases, the cue ball deflection angle is much more stable to small deviations than the object ball cut angle.

**Problem 4.18**: What is the relation between the cut angle α and the natural roll deflection angle θ for small cut angles α?

Answer: For small angles (measured in radians), $\tan(x) \approx x$. The relation, $\tan(\alpha + \theta) = \tfrac{7}{2}\tan(\alpha)$, from P4.14 then gives

$$\theta \approx \tfrac{5}{2}\alpha \qquad \text{[for small } \alpha\text{]}.$$

This relation is useful to know when playing position using natural roll on nearly straight-in shots. It is difficult to achieve a larger amount of topspin than $V_0=R\omega_0$ with a direct cue-tip/cue-ball shot due to the risk of miscue (see P1.7). However, higher spin/speed ratios can be achieved with carom shots. A higher spin/speed ratio would result in a smaller factor than that in the above equation.

**Problem 4.19**: What is the cut angle α at which exactly half of the kinetic energy of a natural-roll cue ball is transferred to the object ball? What is the corresponding natural roll deflection angle θ? At this angle, what are the final kinetic energies of both balls?

*Answer*: When the cue ball has natural roll, $V_0=R\omega_0$, the total kinetic energy is

$$T = \tfrac{1}{2} MV_0^2 + \tfrac{1}{2} I\omega_0^2 = \tfrac{7}{10} MV_0^2$$

The energy of the object ball immediately after collision is

$$T_b = \tfrac{1}{2} MV_b^2 = \tfrac{1}{2} MV_0^2 \cos^2(\alpha)$$

Setting $T_b=\tfrac{1}{2}T$ and simplifying gives

$$\alpha_{(\frac{1}{2}T)} = \arccos\left(\sqrt{\tfrac{7}{10}}\right) = 0.57964 = \frac{\pi}{5.4199} \quad [= 33.211 \text{ deg}]$$

This angle is unchanged as the object ball achieves natural roll. The corresponding deflection angle after natural roll of the cue ball is achieved is

$$\theta_{(\frac{1}{2}T)} = \arctan\left(\tfrac{7}{2}\tan(\alpha_{(\frac{1}{2}T)})\right) - \alpha_{(\frac{1}{2}T)} = 2\alpha_{(\frac{1}{2}T)} - \alpha_{(\frac{1}{2}T)} = \alpha_{(\frac{1}{2}T)}$$

The relation $\tfrac{7}{2}\tan(\alpha_{(\frac{1}{2}T)}) = \tan\left(2\alpha_{(\frac{1}{2}T)}\right)$, used to simplify the above expression, may be verified using the tangent addition formula in P4.14. Therefore, when the final deflection angles are equal for both balls, then each ball has the same kinetic energy immediately after the collision. Note that the cut angle at which this occurs is just a bit thinner than that for a half-ball hit (which would be 30 degrees, neglecting collision induced throw).

The final object ball and cue ball kinetic energies, using $V_{b,NR}$ and $V_{c,NR}$ from P4.13 are

$$T_{b,NR} = \tfrac{1}{2} MV_{b,NR}^2 = T_0\left(\tfrac{25}{49}\cos^2(\alpha)\right)$$

$$T_{c,NR} = \tfrac{1}{2} MV_{c,NR}^2 = T_0\left(\tfrac{25}{49}\sin^4(\alpha) + \tfrac{20}{49}\sin^2(\alpha) + \tfrac{4}{49} + \tfrac{25}{49}\sin^4(\alpha)\cos^2(\alpha)\right)$$

where $T_0$ is the initial cue ball translational energy. These relations are satisfied for any cut angle α. Substitution of $\cos^2(\alpha_{(\frac{1}{2}T)})=7/10$ and $\sin^2(\alpha_{(\frac{1}{2}T)})=3/10$ for the specific half-energy cut angle results in

$$T_{b,NR} = T_{c,NR} = \tfrac{5}{14} T_0 \ .$$

Not only is the energy divided equally between the two balls upon collision with a cut angle of $\alpha_{(\frac{1}{2}T)}$, but the final energies of the two balls are equal after both balls achieve natural roll. The distance that a ball rolls after achieving natural roll, neglecting subsequent cushion and ball collisions, is directly proportional to the natural roll kinetic energy. This relation is useful in situations in which it is necessary that both the object ball and the cue ball roll the same distance, and as a point of reference when unequal distances are required.

In Fig. 4.7 the deflection angle $\theta$ of a natural roll cue ball, as determined in P4.13 and P4.14, is plotted as a function of the object ball cut angle $\alpha$. Also shown on the same graph is the derivative curve $\left(\frac{d\theta_{NR}}{d\alpha}\right)$ as determined in P4.14. The points on this curve corresponding to a half-ball hit, the maximum deflection angle $\alpha_{(\theta_{NR:\max})}$ from P4.14, and the deflection angle corresponding to splitting the kinetic energy as determined in P4.19, are also plotted. The derivative curve is monotonic in the range shown in Fig. 4.7 (in general, it is an even function, symmetric about $\alpha$=0). The derivative curve starts with a value of $5/2$ at $\alpha$=0 (see P4.15), decreases to the value of zero at $\alpha_{(\theta_{\max})}$, and then approaches its asymptotic value of $-5/7$ as the cut angle approaches $\pi/2$. Another point of interest shown in Fig. 4.7 is the value of the cut angle $\alpha$ at which the slope $\left(\frac{d\theta_{NR}}{d\alpha}\right)$ has a value of one. This occurs at $\alpha_{crit}=\arcsin\left(\sqrt{1/15}\right)$=.26116 [=14.963 deg]. For cut angles less than $\alpha_{crit}$, $\left|\frac{d\theta_{NR}}{d\alpha}\right|>1$ and the natural roll cue ball trajectory is more sensitive than the object ball trajectory to small variations in the cut angle. However for the rest of the range of cut angles, $\left|\frac{d\theta_{NR}}{d\alpha}\right|<1$ and the cue ball trajectory is less sensitive than the object ball trajectory. Less sensitivity means that it is easier for the shooter to control, and this may be used to advantage, for example, in placing the cue ball more precisely in position and safety play.

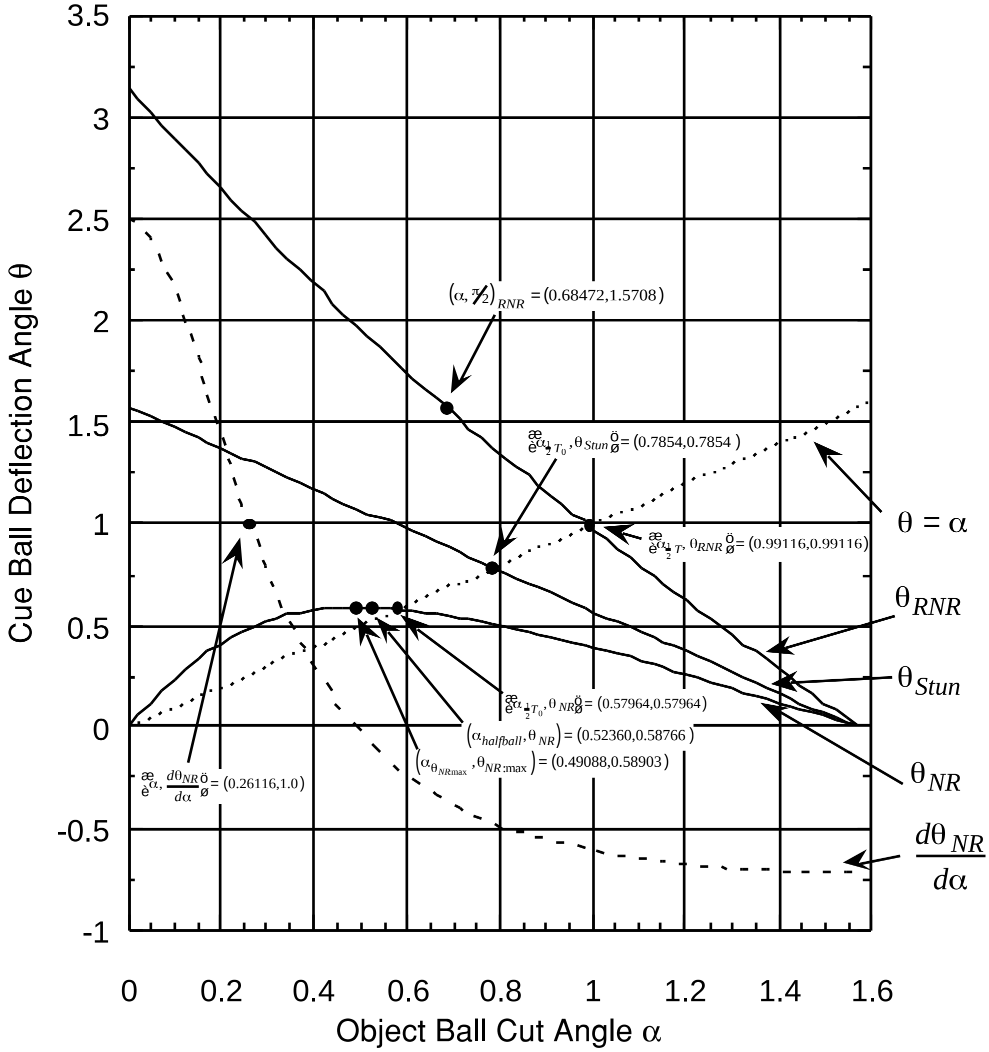



Fig. 4.7. The post-collision natural roll cue ball deflection angle is shown as a function of the object ball cut angle. The $\theta_{NR}$ curve is applicable when the cue ball has natural roll before the collision. $\theta_{Stun}$ is when the cue ball has no spin before the collision. $\theta_{RNR}$ is when the cue ball has reverse natural roll before the collision. The straight line $\theta=\alpha$ corresponds to an equal splitting of the kinetic energy after both balls achieve natural roll. Also shown is the dashed curve defined by $\left(\frac{d\theta_{NR}}{d\alpha}\right)$. Several important individual points on each of these curves are also shown as discussed in the text.

**Problem 4.20**: If the cue ball is not rotating upon impact with the object ball (a stun shot), at what cut angle α is half of the kinetic energy transferred? What are the final energies of the balls? (neglect any frictional forces between the balls)

*Answer*: Taking the velocities immediately after the collision from P4.13, the initial kinetic energies are

$$T_b = \tfrac{1}{2} M V_b^2 = \tfrac{1}{2} M V_0^2 \cos^2(\alpha)$$

$$T_c = \tfrac{1}{2} M V_c^2 = \tfrac{1}{2} M V_0^2 \sin^2(\alpha)$$

Equating these two energies gives

$$\tan^2(\alpha) = 1$$

$$\alpha = \arctan(1) = \frac{\pi}{4} \quad [= 45 \text{ deg}]$$

Each ball has initially after the collision an energy of $\tfrac{1}{2}T_0$. Since neither ball has any angular velocity immediately after the collision, both balls slow down upon achieving natural roll by $\tfrac{5}{7}$ of the initial ball velocities. There is no change of angle, since the velocity directions of the balls do not change. The natural roll kinetic energy of each ball is then $(\tfrac{1}{2})(\tfrac{5}{7})^2 T_0 = (\tfrac{25}{98}) T_0$. Compared to the results of P4.19 involving natural roll of the cue ball, it is seen that the cut angle is thinner and that the final energies of both balls are smaller relative to $T_0$ with a stun shot than with natural roll. This half-energy cut angle point for stun shots is shown on the $\theta_{Stun}$ curve in Fig. 4.7. The $\theta_{Stun}$ curve is a straight line that ranges from the limiting values of $\theta_{Stun}=\pi/2$, at cut angle $\alpha=0$, to $\theta_{Stun}=0$, at $\alpha=\pi/2$.

**Problem 4.21**: What is the natural roll cue ball deflection angle as a function of the cue ball spin $\omega_{0y}$ at the moment of collision and the object ball cut angle?

*Answer*: Generalizing the results of P4.13, it is convenient to write the natural roll cue ball velocity in terms of the spin/speed ratio $J_{0y} = (R\omega_{0y}/V_0)$.

$$\mathbf{V}_{c,NR} = \tfrac{5}{7}\mathbf{V}_c + \tfrac{2}{7} V_0 J_{0y}\hat{\mathbf{i}}' = \tfrac{5}{7} V_0\left(\sin^2(\alpha) + \tfrac{2}{5} J_{0y}\right)\hat{\mathbf{i}}' + \tfrac{5}{7} V_0\left(\sin(\alpha)\cos(\alpha)\right)\hat{\mathbf{j}}'$$

The cue ball deflection angle is determined by the ratio of the two components.

$$\tan(\theta) = \frac{\sin(\alpha)\cos(\alpha)}{\sin^2(\alpha) + \tfrac{2}{5} J_{0y}}$$

Using the tangent addition relation, this may be written as

$$\tan(\alpha + \theta) = \left(\frac{1 + \tfrac{2}{5} J_{0y}}{\tfrac{2}{5} J_{0y}}\right)\tan(\alpha)$$

For the natural roll condition, $J_{0y}=+1$, these results all agree with those of P4.13-P4.14.

**Problem 4.22**: In P4.19 and P4.20 it is seen that a particular cut angle splits evenly both the initial kinetic energy and the natural roll kinetic energies of the two balls. Under what conditions will a cut angle split both energies? (assume $\omega_z$=0)

*Answer*: Half of the initial kinetic energy is transferred when $T_b = \frac{1}{2} T_0$. This occurs when

$$\cos^2(\alpha) = \left(\tfrac{1}{2} + \tfrac{1}{5} J_{0y}^2\right)$$

where $J_{0y}$ is the spin/speed ratio ($R\omega_{0y}/V_0$). The natural roll kinetic energy is split evenly when $T_{b,NR}=T_{c,NR}$. Using the previous natural roll conditions, this occurs when

$$\cos^2(\alpha_{NR}) = \left(\tfrac{1}{2} + \tfrac{1}{5} J_{0y}\right)$$

The angles $\alpha$ and $\alpha_{NR}$ are equal only when

$$J_{0y}(J_{0y} - 1) = 0$$

There are only two possible solutions to this equation: $J_{0y}$ =1, the natural roll situation discussed in P4.19, and $J_{0y}$ =0, the stun shot condition discussed in P4.20. For other spin/speed ratios, there will be one angle $\alpha$ that splits the initial kinetic energy, and a separate angle $\alpha_{NR}$ that splits evenly the natural roll kinetic energies.

**Problem 4.23**: If the cue ball has reverse natural roll (RNR), $V_0$=- $R\omega'_{0y}$, what is the relation between the cut angle $\alpha$ and the natural roll deflection angle $\theta$?

*Answer*: For reverse natural roll, $J_{0y}$ =-1. Referring to the result in P4.21,

$$\tan(\alpha + \theta) = -\tfrac{3}{2}\tan(\alpha)$$

The sign factor in this equation indicates that ($\alpha$+$\theta$) is in a different quadrant than $\alpha$. Specifically, $0 \le \alpha \le \pi/2$ is always in the first quadrant, and $\pi/2 \le (\theta+\alpha) \le \pi$ is always in the second quadrant. Taking the appropriate quadrant for $\theta$ gives the relation

$$\theta = \arctan\left(-\tfrac{3}{2}\tan(\alpha)\right) - \alpha + \pi \quad .$$

For small cut angle $\alpha$, it is seen that

$$\theta \approx \pi - \tfrac{5}{2}\alpha \qquad \text{[for small } \alpha\text{]}$$

The same factor of $\frac{5}{2}$ is seen for the RNR draw shot as for the (topspin) natural roll shot in P4.15. However, in the case of a draw shot the deviation is away from the reverse direction $\pi$ (or 180 degrees), rather than the forward direction. As in the case with topspin, it is difficult to achieve a larger amount of draw than $V_0$=–$R\omega_0$ with a normal direct cue-tip/cue-ball shot due to the risk of miscue (see P1.7). However, higher spin/speed ratios can be achieved with carom and masse shots.

**Problem 4.24**: In P4.19 and P4.20 it is seen that the kinetic energy of the cue ball and object ball is split evenly when the cut angle is equal to the cue ball deflection angle for $J_{0y}$=1 and $J_{0y}$=0. Show that this condition is true for arbitrary $J_{0y}$. What is the cut angle that splits the natural roll energy of a reverse natural roll collision? How does this angle compare to the natural roll angle from P4.19.

*Answer*: From P4.22, the post-collision natural roll kinetic energy is split evenly when $\cos^2(\alpha) = \left(\tfrac{1}{2} + \tfrac{1}{5} J_{0y}\right)$ and $\sin^2(\alpha) = \left(\tfrac{1}{2} - \tfrac{1}{5} J_{0y}\right)$. Substitution of these relations into the general deflection angle equation of P4.21 gives

$$\tan(\theta) = \frac{\sin(\alpha)\cos(\alpha)}{\sin^2(\alpha) + \frac{2}{5}J_{0y}} = \frac{\sin(\alpha)\cos(\alpha)}{\cos^2(\alpha)}$$

$$= \tan(\alpha)$$

or in general $\theta = \alpha$ when the natural roll kinetic energy is split evenly. This line is shown in Fig. 4.7. The above equation for the cut angle may be written as

$$\alpha = \theta = \arcsin\left(\sqrt{\frac{1}{2} - \frac{1}{5}J_{0y}}\right)$$

In particular, for reverse natural roll, $J_{0y}$=-1, the half-energy cut angle is given by

$$\alpha_{\frac{1}{2}T,RNR} = \arcsin\left(\sqrt{\frac{7}{10}}\right) = 0.99116 = \frac{\pi}{3.1696} \quad [= 56.789\text{deg}]$$

From comparison with P4.19, it is seen that $\alpha_{\frac{1}{2}T,RNR} + \alpha_{\frac{1}{2}T,NR} = \frac{\pi}{2}$. This is an example of the general relation

$$\alpha_{\frac{1}{2}T,J_{0y}} + \alpha_{\frac{1}{2}T,-J_{0y}} = \frac{\pi}{2}$$

which follows from the relation, $\cos^2(\alpha_{\frac{1}{2}T,J_{0y}}) = \sin^2(\alpha_{\frac{1}{2}T,-J_{0y}})$

The reverse natural roll deflection angle is shown as a function of the object ball cut angle in Fig. 4.7. Considering $\theta_{RNR}$ as a function of cut angle $\alpha$, it is seen that $\theta_{RNR}$ ranges from zero, for very thin cuts, to $\pi$, for very thick cuts. In contrast $\theta_{NR}$ from P4.15 only ranged from zero to a bit over $\pi/6$. Since natural roll topspin and reverse natural roll backspin represent the practical extremes of cue ball spin (neglecting collision effects and masse), the area between the $\theta_{NR}$ and $\theta_{RNR}$ curves in Fig. 4.7 respresents all possible practically allowed shots. The area between the $\theta_{Stun}$ curve and the $\theta_{RNR}$ curve represents all possible draw shots, and the area between the $\theta_{Stun}$ and $\theta_{NR}$ curves represents all possible topspin shots. Inspection shows that the area associated with draw shots is much larger than that associated with topspin shots. This means that there is much more flexibility with respect to carom angles with draw than with topspin, and correspondingly, that topspin shots are usually less sensitive than draw shots to variations in the cut angle or amount of spin. It may be seen in Fig. 4.7 that $\theta_{RNR}$ is almost a straight line, with an average slope of about twice that of $\theta_{Stun}$. Since $\theta_{Stun}$ is relatively easy to determine, this allows in turn $\theta_{RNR}$ to be estimated for any cut angle simply by multiplying $\theta_{Stun}$ by 2. Inspection of Fig. 4.7 shows that this simple factor will always overestimate the actual deflection angle. The following problem demonstrates the magnitude of error of this approximation.

**Problem 4.25**: At what cut angle does a reverse natural roll cue ball deflect at exactly a right angle?

*Answer*: From P4.23, the desired cut angle satisfies the relation

$$\tan(\alpha \quad + \frac{\pi}{2}) = -\frac{3}{2}\tan(\alpha \quad)$$

Using the identity $\tan(\alpha + \frac{\pi}{2}) = -1/\tan(\alpha)$, $\alpha$ may be determined to be

$$\alpha = \arctan\left(\sqrt{\tfrac{2}{3}}\right) = 0.68472 = \frac{\pi}{4.5881} \quad [=39.232\text{deg}]$$

This point is plotted on the $\theta_{RNR}$ curve in Fig. 4.7. The simple "factor of 2" estimate from the stun-shot curve would have predicted this angle to be π/4 (or 45 degrees), which would have been about 12% in error. The correct cut angle $\alpha$ is about midway between a half-ball cut angle and the π/4 angle.

**Problem 4.26**: For a given cut angle α, what sidespin/speed ratio will result in no horizontal tangential frictional forces?

*Answer*: The surfaces of the balls must not slide against each other in order for the frictional forces to vanish during the collision. The velocity of the cue ball contact point just before the collision is the sum of the linear velocity $\mathbf{V}_0$ and the instantaneous velocity due to the angular velocity about the vertical axis $\omega \times \mathbf{r}$. The contact point velocity is given by

$$\begin{aligned}\mathbf{V}_{cp} &= V_0\cos(\alpha)\hat{\mathbf{i}} + \left(V_0\sin(\alpha) + R\omega_{0z}\right)\hat{\mathbf{j}} - R\omega'_{0y}\cos(\alpha)\hat{\mathbf{k}} \\ &= V_{cpx}\hat{\mathbf{i}} + V_{cpy}\hat{\mathbf{j}} + V_{cpz}\hat{\mathbf{k}}\end{aligned}$$

When $V_{cpy}$=0, then the horizontal frictional forces vanish. Solving for the ratio $R\omega_{0z}/V_0$ gives

$$J_{0z} = \frac{R\omega_{0z}}{V_0} = -\sin(\alpha)$$

**Problem 4.27**: Using the initial spin/speed ratio and the final natural roll spin/speed ratio from P3.6, and the $V_{cpy}$=0 relation from P4.26, what cue tip contact points will result in no horizontal tangential frictional forces between the two colliding balls with a cut angle α?

*Answer*: For the spin/speed ratio immediately after cue tip contact, the contact points are given by the vertical line satisfying

$$\sin(\alpha) = \frac{5y'_{tip}}{2R}$$

Note that the object ball contact point satisfies the relation, $y'_{cp} = -R\sin(\alpha)$. This gives the relation between $y'_{tip}$ and $y'_{cp}$ as

$$y'_{tip} = -\tfrac{2}{5}y'_{cp}$$

The sign difference means that the cue tip impact parameter is in the opposite hemisphere from the object ball contact point. Note that in the limit of an extreme cut shot of angle π/2, this result agrees with that of P3.5; that is, "sideways natural roll" is achieved with a horizontal impact parameter of $^2/_5R$. This relation is useful when the object ball collision occurs very soon after the cue tip contact, before the friction between the ball and cloth has time to change the velocity.

When the cue ball is allowed to achieve natural roll before colliding with the object ball, the desired cue tip contact points satisfy

$$\sin(\alpha) = \frac{7}{2}\left(\frac{y'_{tip}}{z_{tip}}\right)$$

$$y'_{tip} = -\left(\frac{2y'_{cp}}{7R}\right)z_{tip}$$

For a given cut angle $\alpha$, this is a straight line that passes through the origin (0,0). An easy way to estimate the sets of points defined by this straight line is as follows. Refer to Fig. 4.8. Determine the correct contact point at height $z_{tip}=\frac{7}{5}R$. At this contact point, natural roll would be achieved immediately (see P3.5), and the natural roll horizontal offset is the same as the initial horizontal offset determined above, namely the contact point would be $(y,'z)=(-\frac{2}{5}y'_{cp},\frac{7}{5}R)$. The set of desired points is then given by drawing a straight line between this particular contact point and the point at the very bottom of the ball (0,0). In particular, the point on this straight line that is the minimum distance from the center is on the small circle as shown in P3.7.

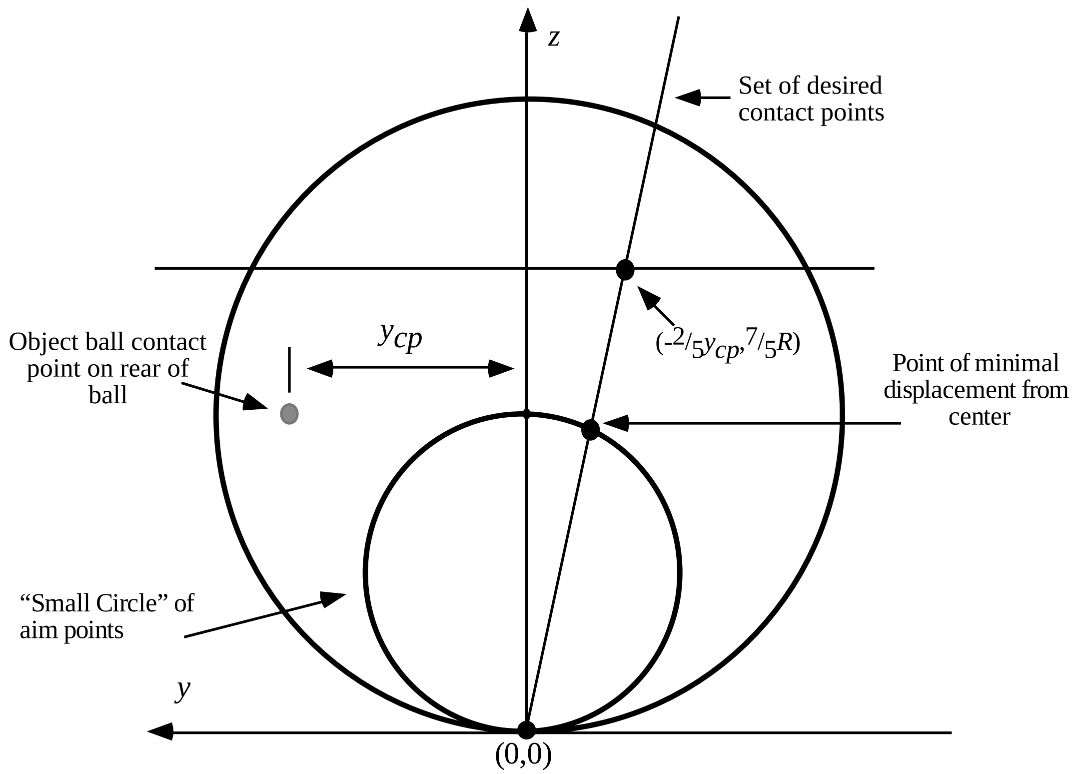



Fig. 4.8. The set of cue tip contacts points that correspond to no (horizontal) frictional forces when the cue ball achieves natural roll prior to collision with the object ball fall on a straight line. The object ball contact point depends on the cut angle. The slope of the line depends on the object ball contact point $y'_{cp}$ as indicated.

**5. Statistics**

The mathematical fields of statistical analysis, combinatorial analysis, stochastic analysis, and game theory are all useful in both physics and pool, and they are all interesting fields of study for the amateur. Statistical methods, which is used here in a general way to include all of these fields, can be used to assess performance, to judge a technique or strategy, and to predict future outcomes based on previous and perhaps incomplete information. These and other uses of statistics will be examined in this section. First some elementary background material and notation will be introduced.

The *average*, or *arithmetic mean*, of a set of values $\{x_i\}$, called a *population*, is

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$N$ is the number of values and the index $i$ runs over the members of the set. There can be repetitions among the values $x_i$, and it may be more convenient to sum over the distinct values $\{y_i\}$, weighted by their repetitions $\{n_i\}$, rather than the individual members of the sample space. In this case the mean may be written as

$$\bar{x} = \sum_{i=1}^{N_{val}} n_i y_i \Big/ \sum_{i=1}^{N_{val}} n_i \quad \text{where } N = \sum_{i=1}^{N_{val}} n_i$$

The probability for each distinct value is

$$p_i = \frac{n_i}{N}$$

and these form a set of nonnegative numbers $\{p_i\}$; this gives another useful expression for the mean.

$$\bar{x} = \sum_{i=1}^{N_{val}} p_i y_i$$

Note that the mean does not necessarily correspond to a member of the sample set.

The set of probabilities $\{p_i\}$ and the corresponding distinct values $\{y_i\}$ defines the probability distribution. For many purposes, it is convenient to consider the probability as a function of the value, $p(y)$. For an ordered set of values $\{y_i\}$, say with $y_i < y_{i+1}$, and corresponding probabilities $\{p_i\}$, there is a cumulative probability defined by

$$p_m^{cum} = \sum_{i=1}^{m} p_i = p_{m-1}^{cum} + p_m$$

The cumulative probability increases monotonically to its maximum value of 1. It is sometimes useful to study properties of various subsets of the population, and the cumulative probability is often used to pick out, for example, the bottom third, or the middle third, or the top quartile, or the top 5%.

Another useful property of a distribution is the *median*. Suppose that the individual members of the sample space $x_i$ are ordered by value. The median of the sample is the value of the INT((*N*+1)/2) element in the ordered list, where INT() implies truncation to an integer value. (There are several conventions used to handle the situation in which there is an even number of members, and the two middle members have different values; for simplicity, this situation will not be considered in this section.) In terms of the ordered probability distribution, the median is determined by the smallest value *m* which satisfies

$$p_m^{cum} \geq \frac{1}{2}$$

The median of the set $\{x_i\}$ is denoted $\tilde{x}$. If the members of the sample space are chosen randomly, then it is just as likely that a value less than the median will be picked as a value that is greater than the median.

The *distribution maximum* or *mode* is the value corresponding to the largest probability value. For a given set, a maximum may not exist, or it may not be unique. If the distribution is symmetric and centrally peaked, then the mode, the median, and the mean will all be the same. If the probability distribution is symmetric, but not necessarily peaked, then the median and the mean are the same but the mode may be different. If the distribution is skewed, meaning that it is not peaked about a central value, then the mean, the median, and the mode will generally have all different values.

Another important value that characterizes a sample set is the standard deviation, which, like the mean, may be computed in various ways in terms of the sample elements, repetition counts, and probability distributions.

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2} = \sqrt{\frac{1}{N}\sum_{i=1}^{N_{val}} n_i\left(y_i - \bar{x}\right)^2} = \sqrt{\sum_{i=1}^{N_{val}} p_i\left(y_i - \bar{x}\right)^2}$$

$$= \sqrt{\left(\frac{1}{N}\sum_{i=1}^{N} x_i^2\right) - \bar{x}^2} = \sqrt{\left(\frac{1}{N}\sum_{i=1}^{N_{val}} n_i y_i^2\right) - \bar{x}^2} = \sqrt{\left(\sum_{i=1}^{N_{val}} p_i y_i^2\right) - \bar{x}^2}$$

If the sample values are very tightly clustered about the mean value, then σ will be small, and if the sample values are broadly spread apart then σ will be large. The *variance* is defined as the quantity $\sigma^2$.

**Problem 5.1**: Given the set $\{x\}=\{0,1,1,4,5\}$, compute the mean using all three methods, the median, the mode, and the standard deviation. What are these same quantities for the set $\{x\}=\{0,1,1,8,9\}$.

*Answer*: For the first set, the distinct values and corresponding probabilities are $\{y_i\}=\{0,1,4,5\}$ and $\{p_i\}=\{1/5, 2/5, 1/5, 1/5\}$. The mean may be written as

$$\bar{x} = \frac{0+1+1+4+5}{5} = \frac{0+2\cdot 1+4+5}{1+2+1+1} = \tfrac{1}{5}\cdot 0 + \tfrac{2}{5}\cdot 1 + \tfrac{1}{5}\cdot 4 + \tfrac{1}{5}\cdot 5 = \tfrac{11}{5}$$

The median is the value of the third element (i.e. (5+1)/2) in this ordered list, $\tilde{x}=x_3=1$. The largest distribution value is $p_2=2/5$, so the mode is $y_2=1$. In this case the mode and the median happen to have the same value, but they both differ from the mean. The standard deviation of the first set is

$$\sigma = \sqrt{\frac{0^2 + 2\cdot 1^2 + 4^2 + 5^2}{5} - \left(\frac{11}{5}\right)^2} = \sqrt{\frac{94}{25}} = 1.939$$

For the second set the mean is

$$\bar{x} = \tfrac{1}{5}\cdot 0 + \tfrac{2}{5}\cdot 1 + \tfrac{1}{5}\cdot 8 + \tfrac{1}{5}\cdot 9 = \tfrac{19}{5}$$

and the standard deviation is σ=Sqrt(374/25)=3.868. For this set, the median is still $\tilde{x}=x_3=1$, and the distribution maximum $p_2=2/5$ still occurs for $y_2=1$, the same as for the first set. For both sets, the mean value does not correspond to a set member. The standard deviation is larger for the second set than for the first set, reflecting the wider range of values.

It is sometimes useful to merge various subsets of values into one large set. If the subset size, mean, and standard deviation is known for each of the subsets, then it is possible to compute the size, mean, and standard deviation of the combined set without knowing the individual values. The parameters of the combined set are given by

$$N = \sum_i N_i$$

$$\bar{x} = \frac{1}{N}\sum_i N_i \bar{x}_i$$

$$\sigma^2 = \frac{1}{N}\sum_i N_i \sigma_i^2 + \frac{1}{N}\sum_i N_i \left(\bar{x}_i - \bar{x}\right)^2$$

The summations in these equations are over the subsets, not the individual elements. The combined average is simply the weighted average of the subset averages. The variance of the combined set contains two contributions, the first is the weighted mean of the subset variances, and the second is the weighted variance of the subset means.

**Problem 5.2**: Compute the mean and variance for the combined set {0,0,1,1,1,1,4,5,8,9}={0,1,1,4,5} $\cup$ {0,1,1,8,9} using the results from P5.1.

*Answer*: $N$=5+5=10

$\bar{x}=(5(11/5)+5(19/5))/10=3$

$\sigma^2=(5(94/25)+5(374/25)+5(11/5-3)^2+5(19/5-3)^2)/10=10$

It may be verified that these values agree with those computed using the individual elements of the combined set.

In some situations the *sample space* is only a subset of a larger *population space*. The sample space may be used to estimate the statistical parameters (mean, median, mode, standard deviation, etc.) of the population space, or the population space statistics may be used to predict possible subspace statistics. In some cases the population space may be too large to handle, or may even be infinite in size, in which case only a smaller sample space is available. There are two conceptual ways of constructing a sample space. One way is by randomly choosing elements from the population space, and setting aside the member once it has been chosen so that it cannot be drawn again; the other way is to replace the elements as they are chosen so that they may be chosen again. Some care must be taken with this choice to ensure that the sample space gives the best possible representation of the population space.

Suppose a population space consists of *N* distinguishable objects (e.g. numbered slips of paper). If one member of this set is chosen, and if the probability for all the members is the same (e.g. the slips are the same size and mixed well before selection), then there are *N* possible, equally likely, outcomes. Now consider choosing two members of the set, without replacement. What is the number of possible outcomes? The act of drawing two objects can be thought of conceptually in two steps: drawing one object, setting it aside, and then drawing the second object. There are *N* possible outcomes after the first draw, and (*N*-1) possible outcomes for the second draw, so it would appear that there might be *N*(*N*-1) possible outcomes. However, if the order of drawing the two objects is unimportant, then this overcounts the outcomes by a factor of two, and the correct answer would be *N*(*N*-1)/2. In the general case, what is the number of possible outcomes for choosing *m* distinguishable objects where the order that they might be drawn is unimportant? The answer is the binomial coefficient which is written

$$\binom{N}{m} = \frac{N!}{m!(N-m)!} = \frac{N(N-1)(N-2)\cdots(N-m+1)}{1\cdot 2\cdot 3\cdots m}$$

The numerator in the last expression is the number of ways to select the *m* objects one at a time, without replacement, from the population space, and the denominator is the number of permutations of these objects to account for the fact that their order is irrelevant. The binomial coefficient $\binom{N}{m}$ is often pronounced "*N* choose *m*" to stress this important relationship. Binomial coefficients satisfy the recursion

$$\binom{N}{m} = \binom{N-1}{m-1} + \binom{N-1}{m}$$

with the boundary conditions $\binom{N}{0} = \binom{N}{N} = 1$. This leads to "Pascal's Triangle"

```
                  1
                1   1
              1   2   1
            1   3   3   1
          1   4   6   4   1
        1   5   10  10  5   1
      1   6   15  20  15  6   1
    …             …             …
```

in which the row is determined by *N* and the element within the row corresponds to *m*. In the triangle, each element is the sum of the two nearby elements in the row above it, a result of the two-term recursion. The name “binomial coefficient” comes from the fact that these numbers are the coefficients of the individual elements in the term-by-term expansion of $(p+q)^n$.

$$(p+q)^n = \sum_{m=0}^{n} \binom{n}{m} p^m q^{n-m} = q^n + npq^{n-1} + \ldots + np^{n-1}q + p^n$$

**Problem 5.3**: Given the sample set {1,2,3,4}, enumerate all of the ways of choosing zero, one, two, three, and four elements without replacement.
*Answer*: There is 1 way to choose zero elements: {}; there are 4 different ways to choose one element: {1}, {2}, {3}, and {4}; there are 6 ways to choose two elements: {1,2}, {1,3}, {1,4}, {2,3}, {2,4}, and {3,4}; there are 4 ways to choose three elements: {1,2,3}, {1,2,4}, {1,3,4}, and {2,3,4}; there is 1 way to choose four elements: {1,2,3,4}. These numbers, 1, 4, 6, 4, and 1, agree with the $n$=4 row of Pascal’s triangle and with the closed-form expression for the binomial coefficients.

Suppose that the probability for a “successful” event to occur is $p$. The probability for a failure is $q=(1-p)$. If the sample space is infinite, or if the space is finite and the sampling is done with replacements, then the probability for success does not change upon repetition and the probability for two consecutive successes is $p^2$. The probability for a single success and a single failure is $2pq$, because there are two ways to arrive at this result, each of which has probability $pq$. The probability of two failures is $q^2$. In the general case, the probability of obtaining $m$ successes and $n$ failures after $N=n+m$ attempts is given by $P(p;m,n)=\binom{m+n}{m} p^m q^n$. Comparison with the binomial expansion shows that this probability is the $m^{\text{th}}$ term in the expansion of $(p+q)^{(m+n)}$; consequently such distributions are called *binomial distributions*.

**Problem 5.4**: Two players are playing 9-ball, and the probability that player-1 will win an individual game is $p=2/3$. What is the probability that after 4 games the score will be 3:1? Enumerate all the possible ways of arriving at this game score.
*Answer*: Using the above equation the probability is

$$P(2/3;3,1)=\binom{4}{1}\left(\tfrac{2}{3}\right)^3\left(\tfrac{1}{3}\right)^1 = 4\left(\tfrac{8}{27}\right)\left(\tfrac{1}{3}\right) = \tfrac{32}{81} = 0.395$$

There are four ways of arriving at this game score: LWWW, WLWW, WWLW, and WWWL. The probability of each of these individual ways occurring is $p^3q^1 = \left(\tfrac{2}{3}\right)^3\left(\tfrac{1}{3}\right)^1 = 0.0988$.

**Problem 5.5**: Two players are playing 9-ball, and the probability that player-1 will win an individual game is $p=2/3$. The match is handicapped at 3:2, meaning that player-1 must win 3 games whereas player-2 must win only 2 games in order to win the match. What is the probability that player-1 will win this match? If the match is handicapped at $N_1$:$N_2$, what is the general expression that player-1 will win?

*Answer*: There are two ways that player-1 can win the match handicapped at 3:2, namely 3:0, and 3:1. In order to arrive at a 3:0 score, player-1 must win the last game from a 2:0 score; the probability for this to occur is $pP(p;2,0)=\left(\frac{2}{3}\right)\binom{2}{0}\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right)^0=\frac{8}{27}=0.296$. In order to arrive at a 3:1 score, player-1 must win the last game from a 2:1 score; the probability for this to occur is $pP(p;2,1)=\left(\frac{2}{3}\right)\binom{3}{1}\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right)^1=\frac{8}{27}=0.296$. The probability of player-1 winning the match is the sum of these two terms, $W$=16/27=0.593.

In the general case, there are $N_2$ ways for player-1 to win the match: $N_1$:0, $N_1$:1, ..., $N_1$:($N_2$-1). The probability for the individual $N_1$:$m$ case is $pP(p;N_1-1,m)$. The probability that player-1 will win the match is the sum over all of the individual probabilities

$$W(p;N_1,N_2)=\sum_{m=0}^{N_2-1} pP\left(p;N_1-1,m\right)=\sum_{m=0}^{N_2-1}\binom{N_1+m-1}{m}p^{N_1}q^m$$

In the following two problems, the match probability $W$ is examined, first with the game probability $p$ fixed and varying the matchup $N_1$ and $N_2$, and then with $N_1$ and $N_2$ fixed and varying $p$.

**Problem 5.6**: Write a computer program to compute a table of values containing $P(p;m,n)$ for $0\le m\le 9$ and $0\le n\le 9$. From this table, compute the corresponding $W(p;m,n)$ tables for $1\le m\le 10$ and $1\le n\le 10$. Compute these tables for $p=\frac{1}{2}$, for which the players are equally likely to win an individual game, and for $p=\frac{2}{3}$, for which player-1 is twice as likely to win an individual game as player-2.

*Answer*: For this purpose, it is better to formulate the $P(p;m,n)$ table construction using the following recursion approach (which is similar to computing Pascal's triangle).

$P(p;0,0)=1$

$P(p;0,n)=qP(p;0,n-1)$ ; for $n$=1,2,...9

$P(p;m,0)=pP(p;m-1,0)$ ; for $m$=1,2,...9

$P(p;m,n)=qP(p;m,n-1)+pP(p;m-1,n)$ ; for $m$=2,...,9 and for $n$=2,...,9

Basically, this recognizes the fact that to arrive at a score of $m$:$n$, either player-1 must win the last game from a score of ($m$-1):$n$, which occurs with probability $p$, or player-2 must win from a score of $m$:($n$-1), which occurs with a probability $q$.

The $W(p;m,n)$ table is then constructed in a similar manner.

$W(p;m,1)=pP(p;m-1,0)$ ; for $m$=1,...,10

$W(p;m,n)=W(p;m,n-1)+pP(p;m-1,n-1)$ ; for $m$=1,...,10 and for $n$=2,...,10

These tables are included below for the two specified values. Note that the $P(\frac{1}{2};m,n)$ table is symmetric (i.e. $P(\frac{1}{2};m,n)=P(\frac{1}{2};n,m)$), as would be expected for two equally matched players. Note also that the $W(\frac{2}{3};3,2)$ entry agrees with the hand-calculated value from P5.5.

Such a program may be easily written in almost any programming language or spreadsheet. It is sometimes handy to have such a program available when directing tournaments, or even for personal use, in order to determine fair handicapped matchups between players of varying strengths.

**Problem 5.7**. Compare the $W(p;n,n)$ and the $W(p;2n,n)$ match probabilities for $1 \le n \le 10$ numerically as a function of the game probability $p$.

*Answer*: Using the computer program from P5.6, the appropriate elements of the $W$ table may be determined as a function of $p$. These match probabilities are shown in Fig. 5.1. In general, it may be observed that each curve of $W(p;m,n)$ is an increasing function of the game probability $p$. It is seen that $W(\frac{1}{2};n,n)=\frac{1}{2}$ for all matches. This means that if player-1 is the stronger player, $p>\frac{1}{2}$, it is to his advantage to play a longer even matchup rather than a shorter match, but if player-1 is the weaker player, $p<\frac{1}{2}$, then it is to his advantage to play a shorter match. A beginner might be able to win a game (i.e. a 1:1 match) against a professional, but it is most unlikely that he would win a longer 10:10 match.

There is no single common point of exact intersection for the $W(p;2n,n)$ curves; these curves cross at slightly different values of $p$. If a match probability of $W=\frac{1}{2}$ is defined as “fair”, then it is clear in Fig. 5.1 that player-1 must have a larger game probability $p$ to survive a 2:1 match than a 4:2 match. An interesting region occurs for the 2:1 and 4:2 curves after they intersect ($W(.641;2,1)=W(.641;4,2)=.411$) but before the point corresponding to $W(.686;4,2)=\frac{1}{2}$. In this domain, $.641<p<.686$, $W<\frac{1}{2}$ for both curves, so player-1 is expected to lose both matches, yet it is still to his advantage to play the longer match. This handicapped situation is in contrast to the even-matchup situation in which the expected winner always benefits from the longer match. Such a domain exists for the other pairs of $2n{:}n$ matchup curves, but it becomes much smaller because the curves are steeper for longer matches. Furthermore, in the domain $.5<p<.641$, before the 2:1 and 4:2 curves intersect, player-1 is the stronger player but his best chances of winning are with the shorter 2:1 match. Again, this is in contrast to the even-matchup situation in which the stronger player always benefited the most with longer matches.

$P(1/2;m,n)$

| m\n | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1.000 | 0.500 | 0.250 | 0.125 | 0.063 | 0.031 | 0.016 | 0.008 | 0.004 | 0.002 |
| 1 | 0.500 | 0.500 | 0.375 | 0.250 | 0.156 | 0.094 | 0.055 | 0.031 | 0.018 | 0.010 |
| 2 | 0.250 | 0.375 | 0.375 | 0.313 | 0.234 | 0.164 | 0.109 | 0.070 | 0.044 | 0.027 |
| 3 | 0.125 | 0.250 | 0.313 | 0.313 | 0.273 | 0.219 | 0.164 | 0.117 | 0.081 | 0.054 |
| 4 | 0.063 | 0.156 | 0.234 | 0.273 | 0.273 | 0.246 | 0.205 | 0.161 | 0.121 | 0.087 |
| 5 | 0.031 | 0.094 | 0.164 | 0.219 | 0.246 | 0.246 | 0.226 | 0.193 | 0.157 | 0.122 |
| 6 | 0.016 | 0.055 | 0.109 | 0.164 | 0.205 | 0.226 | 0.226 | 0.209 | 0.183 | 0.153 |
| 7 | 0.008 | 0.031 | 0.070 | 0.117 | 0.161 | 0.193 | 0.209 | 0.209 | 0.196 | 0.175 |
| 8 | 0.004 | 0.018 | 0.044 | 0.081 | 0.121 | 0.157 | 0.183 | 0.196 | 0.196 | 0.185 |
| 9 | 0.002 | 0.010 | 0.027 | 0.054 | 0.087 | 0.122 | 0.153 | 0.175 | 0.185 | 0.185 |

$W(1/2;m,n)$

| m\n | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.500 | 0.750 | 0.875 | 0.938 | 0.969 | 0.984 | 0.992 | 0.996 | 0.998 | 0.999 |
| 2 | 0.250 | 0.500 | 0.688 | 0.813 | 0.891 | 0.938 | 0.965 | 0.980 | 0.989 | 0.994 |
| 3 | 0.125 | 0.313 | 0.500 | 0.656 | 0.773 | 0.855 | 0.910 | 0.945 | 0.967 | 0.981 |
| 4 | 0.063 | 0.188 | 0.344 | 0.500 | 0.637 | 0.746 | 0.828 | 0.887 | 0.927 | 0.954 |
| 5 | 0.031 | 0.109 | 0.227 | 0.363 | 0.500 | 0.623 | 0.726 | 0.806 | 0.867 | 0.910 |
| 6 | 0.016 | 0.063 | 0.145 | 0.254 | 0.377 | 0.500 | 0.613 | 0.709 | 0.788 | 0.849 |
| 7 | 0.008 | 0.035 | 0.090 | 0.172 | 0.274 | 0.387 | 0.500 | 0.605 | 0.696 | 0.773 |
| 8 | 0.004 | 0.020 | 0.055 | 0.113 | 0.194 | 0.291 | 0.395 | 0.500 | 0.598 | 0.685 |
| 9 | 0.002 | 0.011 | 0.033 | 0.073 | 0.133 | 0.212 | 0.304 | 0.402 | 0.500 | 0.593 |
| 10 | 0.001 | 0.006 | 0.019 | 0.046 | 0.090 | 0.151 | 0.227 | 0.315 | 0.407 | 0.500 |

$P(2/3;m,n)$

| m\n | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1.000 | 0.333 | 0.111 | 0.037 | 0.012 | 0.004 | 0.001 | 5E-04 | 2E-04 | 5E-05 |
| 1 | 0.667 | 0.444 | 0.222 | 0.099 | 0.041 | 0.016 | 0.006 | 0.002 | 9E-04 | 3E-04 |
| 2 | 0.444 | 0.444 | 0.296 | 0.165 | 0.082 | 0.038 | 0.017 | 0.007 | 0.003 | 0.001 |
| 3 | 0.296 | 0.395 | 0.329 | 0.219 | 0.128 | 0.068 | 0.034 | 0.016 | 0.007 | 0.003 |
| 4 | 0.198 | 0.329 | 0.329 | 0.256 | 0.171 | 0.102 | 0.057 | 0.030 | 0.015 | 0.007 |
| 5 | 0.132 | 0.263 | 0.307 | 0.273 | 0.205 | 0.137 | 0.083 | 0.048 | 0.026 | 0.013 |
| 6 | 0.088 | 0.205 | 0.273 | 0.273 | 0.228 | 0.167 | 0.111 | 0.069 | 0.040 | 0.022 |
| 7 | 0.059 | 0.156 | 0.234 | 0.260 | 0.238 | 0.191 | 0.138 | 0.092 | 0.057 | 0.034 |
| 8 | 0.039 | 0.117 | 0.195 | 0.238 | 0.238 | 0.207 | 0.161 | 0.115 | 0.077 | 0.048 |
| 9 | 0.026 | 0.087 | 0.159 | 0.212 | 0.230 | 0.214 | 0.179 | 0.136 | 0.096 | 0.064 |

$W(2/3;m,n)$

| m\n | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.667 | 0.889 | 0.963 | 0.988 | 0.996 | 0.999 | 1.000 | 1.000 | 1.000 | 1.000 |
| 2 | 0.444 | 0.741 | 0.889 | 0.955 | 0.982 | 0.993 | 0.997 | 0.999 | 1.000 | 1.000 |
| 3 | 0.296 | 0.593 | 0.790 | 0.900 | 0.955 | 0.980 | 0.992 | 0.997 | 0.999 | 0.999 |
| 4 | 0.198 | 0.461 | 0.680 | 0.827 | 0.912 | 0.958 | 0.980 | 0.991 | 0.996 | 0.998 |
| 5 | 0.132 | 0.351 | 0.571 | 0.741 | 0.855 | 0.923 | 0.961 | 0.981 | 0.991 | 0.996 |
| 6 | 0.088 | 0.263 | 0.468 | 0.650 | 0.787 | 0.878 | 0.934 | 0.965 | 0.983 | 0.991 |
| 7 | 0.059 | 0.195 | 0.377 | 0.559 | 0.711 | 0.822 | 0.896 | 0.942 | 0.969 | 0.984 |
| 8 | 0.039 | 0.143 | 0.299 | 0.473 | 0.632 | 0.759 | 0.851 | 0.912 | 0.950 | 0.973 |
| 9 | 0.026 | 0.104 | 0.234 | 0.393 | 0.552 | 0.690 | 0.797 | 0.873 | 0.925 | 0.957 |
| 10 | 0.017 | 0.075 | 0.181 | 0.322 | 0.476 | 0.618 | 0.737 | 0.828 | 0.892 | 0.935 |

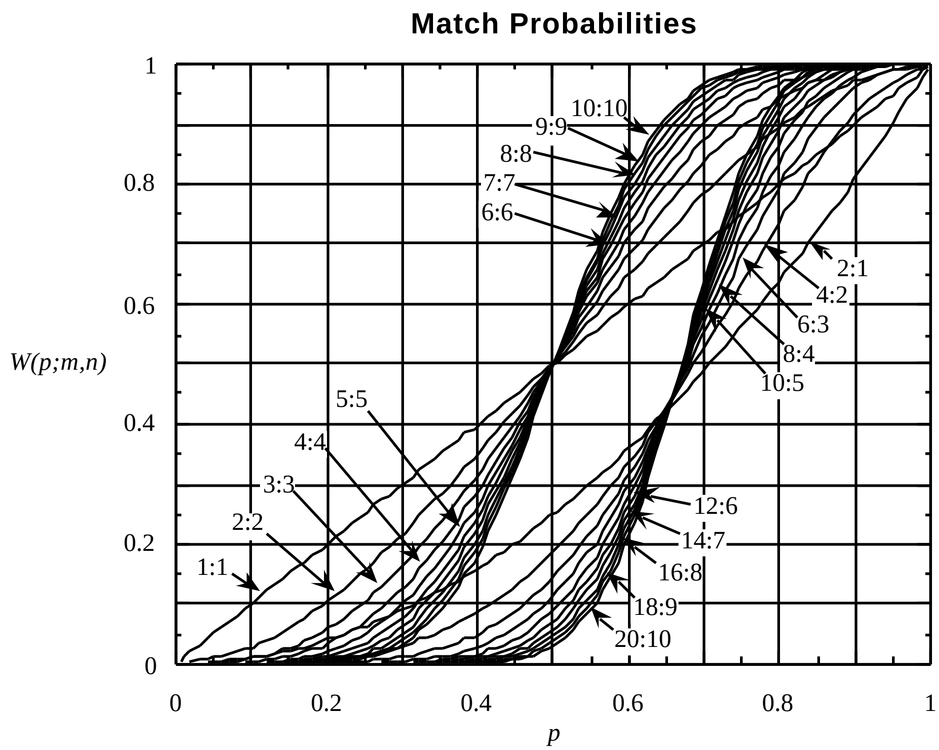



Fig. 5.1. The match probability *W* as a function of the individual game probability *p* for even *n:n* matchups and uneven 2*n:n* matchups of various lengths. For all of the individual curves, the match probability is an increasing function of the game probability. The steepness of a curve is related to how sensitive is the match outcome to the game probability.

**Problem 5.8**. A strong player is negotiating a matchup with a weaker opponent and he knows that his game probability against this opponent is $p=2/3$. He is offered a choice between a single long match of 9:5, and a 3:1 match of sets where each set is handicapped at 3:3. Which option is best for player-1?

*Answer*: At first this seems very complicated, so it is best to break the problem down into smaller pieces that are easier to understand. Player-1 will win the long match with a probability of $W(2/3;9,5)=.552$ according to the table in P5.6. He will win a 3:3 set with a probability of $W(2/3;3,3)=.790$, also according to the table in P5.6. The match probability for the second option is given by $W(.790;3,1)$. That is, the statistical analysis for winning multiple-set matches is the same as that for winning multiple-game matches, but with the variable $p$ being the set probability instead of the game probability. This may be computed using the program in P5.6, or from the polynomial expression from P5.5: $W(p;3,1)=p^3$. In either case, the result is seen to be $W(.790;3,1)=.493$. Player-1 would have a small 5.2% advantage over player-2 in the long-match format, but he would have a slight 0.7% disadvantage with this particular set format.

In either case, player-1 must win 9 games total in order to win the match. In the long match format, player-2 needs to win 5 games to win the match, whereas in the set format he needs only to win 3 games, provided they are all in the same set. In the set format, player-2 can win as many as 6 games and still lose the match, provided they are split evenly with 2 games in each set. There are apparently no shortcuts, based simply on the total games required by each player, that will give the correct choice in these negotiations. The actual statistical analysis is required to correctly assess each possible option.

In the general case of multiple-set matches, the player-1 match probability is given by the general expression

$$W^{match} = W\left(W\left(p; N_1^{set}, N_2^{set}\right); N_1^{match}, N_2^{match}\right)$$

in which the required games per set and sets per match are indicated and in which $p$ is the individual game probability for player-1.

For a given $N$ with $N=m+n$, binomial distributions characterized by the probabilities $P(p;m,n)$ are centrally peaked for $p\approx\frac{1}{2}$ (i.e. the peak occurs near $m\approx N/2$), the peak is shifted toward large $m$ values for large $p\approx 1$, and the peak is shifted toward small $m$ values for small $p\approx 0$. Because binomial distributions are so common, it is useful to characterize the peak $\tilde{x}$, the mean $\bar{x}$, and the standard deviation σ in a general way.

**Problem 5.9**: Compute the mode, the mean, and the standard deviation of a binomial distribution in terms of $N$ and $p$.

*Answer*: The possible values of a binomial distribution correspond to the integers $\{m\ ; m=0,\ldots N\}$ and the corresponding probabilities are given by $P(p;m,N\text{-}m)$. The mode, or distribution peak, occurs for the smallest value of $m$ for which $P(p;m+1,N\text{-}m\text{-}1)<P(p;m,N\text{-}m)$. The peak of a binomial distribution is given by

$$\tilde{x} = m_{small} = Ceiling(\ Np - q\ )$$

where $Ceiling(x)$ denotes the smallest integer that is greater than or equal to $x$. The mean is given by

$$\begin{aligned}
\bar{x} &= \sum_{m=0}^{N} mP(p;m,N-m) = \sum_{m=0}^{N} m\frac{N!}{m!(N-m)!}p^m q^{N-m} \\
&= \sum_{m=0}^{N-1} \frac{N(N-1)!}{m!(N-1-m)!}p^{m+1}q^{N-1-m} = Np\sum_{m=0}^{N-1}\frac{(N-1)!}{m!(N-1-m)!}p^m q^{N-1-m} \\
&= Np(p+q)^{N-1} = Np
\end{aligned}$$

The mean and the mode of a binomial distribution differ by, at most, one. A similar sequence of operations gives the standard deviation of a binomial distribution.

$$\sigma = \sqrt{Npq}$$

An important property of the binomial distribution is that for large *N*, it approaches the *normal distribution* defined by

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(x-\bar{x})^2/\sigma^2}$$

This distribution is symmetric about the mean and is peaked at the mean. It is often useful to shift and scale the domain of the distribution using the equation $z = (x - \bar{x})/\sigma$. In terms of these dimensionless transformed values, called *standard units*, the normal distribution takes the simple form

$$P(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

In this *standard form*, the normal distribution is peaked at $z$=0 and has a standard deviation of σ=1. Areas under the normal distribution correspond to various cumulative probabilities. However, the form of the normal distribution does not allow for a simple closed-form expression of the antiderivative, so integrals must be computed numerically or interpolated from tables. One form for these tables is in terms of the symmetric integral $P^{cum}(z_c) = \int_{-z_c}^{z_c} P(z)\,dz$. The following short table gives some of the more commonly used cumulative probabilities and their corresponding critical values $z_c$.

Table 5.1. Normal Distribution Critical Values

| $P^{cum}(z_c)$ | .9973 | .99 | .98 | .96 | .9545 | .95 | .90 | .80 | .6827 | .50 |
|---|---|---|---|---|---|---|---|---|---|---|
| $z_c$ | 3 | 2.58 | 2.33 | 2.05 | 2 | 1.96 | 1.645 | 1.28 | 1 | .6745 |

**Problem 5.10**: When two players play 9-ball the probability that player-1 will win any particular game is 0.52. These players play 120 games. What is the expected mean score for player-1 and the expected variation about this mean score? What is the range of scores that would be expected to occur 95% of the time? What is the range expected to occur 50% of the time?

*Answer*: The possible game scores form a binomial distribution. The mean score for player-1 is $\bar{x}=Np$=120(0.52)=62.4. The standard deviation is $\sigma=\sqrt{Npq} = \sqrt{120(.52)(.48)}$ =5.47. For 120 games, the binomial distribution can be approximated by a normal distribution. The critical value corresponding to 95% is $z_c$=1.96. $z_c\sigma$=(1.96)(5.47)=10.7, so there is approximately a 95% probability that the final game score for player-1 will be between $\bar{x}-z_c\sigma$=51 and $\bar{x}+z_c\sigma$=73. There is only about a 5% chance that the final score will be outside of this range. For the 50% range, $z_c\sigma$=(.6745)(5.47)=3.69, so there is a 50% chance that the player-1 game score will be between 59 and 66 using the normal approximation to the binomial distribution. The exact probability, using the exact binomial statistics as in P5.6, for this range of scores is 53.5%, which shows that the normal distribution approximation is quite reliable.

Suppose that the probability $p$ corresponds to some average probability of successfully executing a shot, and runlength statistics are of interest, where "runlength" means the number of consecutive successful shots. The chances of success on the first shot is $p$, and for two consecutive successful shots is $p^2$, and so on. The probability of running $n$ balls or greater is $p^n$. What is the probability of running exactly $n$ balls and then missing? The answer is $r_n=p^n q$ where $q=(1-p)$. The set $\{r_n\}$ then defines a probability distribution for a population of infinite size.

**Problem 5.11**: What is the mode, mean, and median runlength for a probability distribution defined by $r_n=p^n q$ as a function of $p$?

*Answer*: The ratio of two successive runlength probabilities, $r_{n+1}/r_n=p<1$ shows that the distribution is monotonically decreasing, and therefore the maximum of the distribution occurs always at $n$=0, regardless of $p$. This shows that the mode is not particularly useful for predicting typical outcomes if the distribution is severely skewed. The mean runlength is

$$\bar{r} = \sum_{n=0}^{\infty} n r_n = (1-p)\sum_{n=1}^{\infty} n p^n$$

That is, a run of length $n$ occurs with probability $r_n$. It may be verified by straightforward division that

$$\frac{1}{(1-p)} = \sum_{n=0}^{\infty} p^n = 1 + p + p^2 + \ldots + p^k + \ldots$$

Differentiating both sides with respect to $p$ , followed by multiplication by $p$, gives the identity

$$\frac{p}{(1-p)^2} = \sum_{n=1}^{\infty} n p^n = p + 2p^2 + 3p^3 + \ldots + kp^k + \ldots$$

Substitution of this relation into the expression for the mean runlength gives

$$\bar{r} = \frac{p}{(1-p)} = \frac{p}{q}$$

$$p = \frac{\bar{r}}{(\bar{r}+1)}$$

The first equation gives the average runlength in terms of the individual shot probability, whereas the second gives the individual shot probability as a function of the average runlength.

The median runlength is the smallest value $m$ that satisfies the equation

$$\tfrac{1}{2} \le r_m^{cum} = \sum_{n=0}^{m} r_n = (1-p)\sum_{n=0}^{m} p^n = (1-p)\left(\frac{1-p^{m+1}}{1-p}\right) = 1 - p^{m+1}$$

The summation identity is easily verified from the above expansion of 1/(1-*p*). Some rearrangements then give the result that the median runlength corresponds to the smallest integer *m* that satisfies the relation

$$m \ge -\left(\frac{\log(2)}{\log(p)} + 1\right)$$

It is interesting that the median runlength $\tilde{r} = m$ is always less than the mean runlength $\bar{r} = p/(1-p)$, as demonstrated in the following table.

Table 5.2. Runlength statistics for selected shot probabilities *p*.

| $p$ | $\bar{r} = p/(1-p)$ | -(1+log(2)/log(p)) | $\tilde{r}$ | $\tilde{r}/\bar{r}$ |
|---|---|---|---|---|
| 0.5 | 1.0 | 0.0 | 0 | 0.000 |
| 0.6 | 1.5 | 0.4 | 1 | 0.667 |
| 0.7 | 2.3 | 0.9 | 1 | 0.429 |
| 0.8 | 4.0 | 2.1 | 3 | 0.750 |
| 0.9 | 9.0 | 5.6 | 6 | 0.667 |
| 0.91 | 10.1 | 6.3 | 7 | 0.692 |
| 0.92 | 11.5 | 7.3 | 8 | 0.696 |
| 0.93 | 13.3 | 8.6 | 9 | 0.677 |
| 0.94 | 15.7 | 10.2 | 11 | 0.702 |
| 0.95 | 19.0 | 12.5 | 13 | 0.684 |
| 0.96 | 24.0 | 15.9 | 16 | 0.667 |
| 0.97 | 32.3 | 21.8 | 22 | 0.680 |
| 0.98 | 49.0 | 33.3 | 34 | 0.694 |
| 0.99 | 99.0 | 67.9 | 68 | 0.687 |

**Problem 5.12**: What is an approximate relation between the median and the mean runlength for the $r_n$ distribution?

*Answer*: Using natural logarithms, the median runlength may be written

$$\tilde{r} \approx -\left(\frac{\ln(2)}{\ln(p)} + 1\right) = -\left(\frac{\ln(2)}{\ln\left(\frac{\bar{r}}{1+\bar{r}}\right)} + 1\right)$$

For reasonably large $\bar{r}$, the denominator simplifies using the approximations

$$\frac{\bar{r}}{1+\bar{r}} = 1 - \frac{1}{\bar{r}} + \left(\frac{1}{\bar{r}}\right)^2 - \ldots \approx 1 - \frac{1}{\bar{r}}$$

$$\ln\left(\frac{\bar{r}}{1+\bar{r}}\right) \approx \ln\left(1 - \frac{1}{\bar{r}}\right) = -\left(\frac{1}{\bar{r}}\right) + \tfrac{1}{2}\left(\frac{1}{\bar{r}}\right)^2 - \tfrac{1}{3}\left(\frac{1}{\bar{r}}\right)^3 + \ldots \approx -\left(\frac{1}{\bar{r}}\right)$$

This gives the approximate relation

$$\frac{\tilde{r}}{\bar{r}} \approx \ln(2) = 0.693$$

The last column of Table 5.2 shows the actual median to mean ratios for some selected values of *p*. This approximation is seen to be accurate to within a few percent for mean runlengths $\bar{r}$ of about 5.0 or larger.

**Problem 5.13**: An experienced 14.1 player knows that his mean runlength is 24.0 balls. What is the shot probability using the $r_n$ estimate of the statistical distribution? What is the probability that this player will run between 50 and 75 balls? What is the probability of a run 100 or larger?

*Answer*: The individual shot probability for this player is $p$=24.0/25.0=.96. The probability of a run between 50 and 75, inclusive, is $r_{75}^{cum} - r_{49}^{cum}$=(1-$p^{76}$)-(1-$p^{50}$)= $p^{50}$-$p^{76}$=0.085. That is, the player should expect a run between 50 and 75 to occur in 8.5% of the attempts. The probability of a run of 100 or over is $1 - r_{99}^{cum}$=$p^{100}$=0.017; such a run will occur in 1.7% of the attempts.

**Problem 5.14**: This same 14.1 player is offered a friendly wager that for the rest of the day, every run over 20 balls he will win the wager amount, and every run of 20 balls or less he will lose the wager amount. Using the above statistical runlength model, is this a good proposition for the player?

*Answer*: Since his 24.0 mean is over 20 balls, it might seem at first that it would be a good proposition. However, upon closer inspection, the wager is really a matter of the median runlength, not the mean runlength. The individual shot probability for this player is $p$=24.0/25.0=.96, and this corresponds to a median runlength of $\tilde{r}$=16 according to Table 5.2. The approximation from P5.12 gives $\tilde{r}$≈(0.693)(24.0)=16.6, so even if the player did not have the benefit of the table or a calculator to compute the exact median, he should expect to run 20 balls less than half of the time. Computation of the exact cumulative probability for $m$=20 gives $r_{20}^{cum} = 1 - 0.96^{21} = 0.576$, which means that he should expect to lose the wager 57.6% of the time, and win it only 42.4% of the time using the simple statistical model.

As shown in the following problem, safety play between opponents in an actual game situation skews the differences between the mean and the median runlengths even more than that predicted by this simple statistical model.

**Problem 5.15**: Assume that due to safety play by the opponent, the first shot of a player's inning has a success probability of only $\alpha p$, with $0<\alpha\leq 1$, and each subsequent shot then has a success probability of $p$. What is the runlength probability distribution, the mean, the median, and the approximate ratio of the median to the mean as a function of $\alpha$ and $p$?

*Answer*: $r(\alpha)_0$=(1-$\alpha p$), $r(\alpha)_1$=$\alpha pq$, and, in general, $r(\alpha)_n$=$\alpha p^n q$ for $n$>0. Using the same approach as in P5.11, the mean runlength is found to be

$$\bar{r}(\alpha) = \frac{\alpha p}{1-p}$$

The cumulative probabilities are given by

$$r(\alpha)_m^{cum} = 1 - \alpha p^{m+1}$$

and the median runlength is determined by the smallest integer *m* that satisfies the equations

$$r(\alpha)_m^{cum} = 1 - \alpha p^{m+1} \geq \tfrac{1}{2}$$

$$m \geq -\left(\frac{\log(2\alpha)}{\log(p)} + 1\right)$$

The approximate ratio of the median to the mean, using the same approximations as in P5.12, is given by

$$\frac{\tilde{r}(\alpha)}{\bar{r}(\alpha)} \approx \frac{\ln(2\alpha)}{\alpha}$$

A few sample values for this ratio are shown in the following table.

| $\alpha$ | 1.0 | 0.9 | 0.8 | 0.7 | 0.6 | 0.5 |
|---|---|---|---|---|---|---|
| $\tilde{r}(\alpha)/\bar{r}(\alpha)$ | 0.69 | 0.65 | 0.59 | 0.48 | 0.30 | 0.0 |

Only in the best possible case, $\alpha$=1, is this ratio as good as that predicted in P5.12; in the other cases, this ratio becomes progressively worse with more aggressive safety play. (Note that in all of the equations above, setting $\alpha$=1 produces agreement with the previous results.) This shows that even though the mean runlength is strongly dependent on safety play, the median runlength is even more sensitive.

The previous discussion concerned runlengths in which there were *n* successes followed by a single miss. In a game situation, this would apply to a single inning of a longer game. What is the runlength distribution after several innings? Using a similar approach as before, it is seen that the probability of accumulating *n* successful shots and *m* misses out of $N=m+n$ total shots is the binomial expansion term $P(p;n,m)$. The probability of a runlength score of exactly *n* after *m* innings (neglecting penalty points that might apply to the misses in the game) is given by

$$R_{nm} = qP(p;n,m-1) = \binom{n+m-1}{m-1} p^n q^m$$

That is, the first ($m$-1) misses can occur anywhere during the first ($n+m$-1) shots, but the last miss must occur on the last shot. It may be verified that $R_{n1}=r_n$ for all *n*, which is the single-inning runlength distribution that has been previously examined.

**Problem 5.16**: What is the mean score after *m* innings, using the $R_{nm}$ distribution, as a function of *p*? What is the standard deviation of these scores?
*Answer*: The mean score is

$$\bar{R}_m = \sum_{n=0}^{\infty} nR_{nm} = \sum_{n=0}^{\infty} n\binom{n+m-1}{m-1}p^n q^m = q^m \sum_{n=1}^{\infty} \frac{(n+m-1)!}{(n-1)!(m-1)!}\, p^n$$

$$= mpq^m \sum_{n=0}^{\infty} \frac{(n+m)!}{(n)!(m)!}\, p^n = mpq^m \sum_{n=0}^{\infty} \binom{n+m}{m} p^n = \frac{mpq^m}{(1-p)^{m+1}}$$

$$= m\frac{p}{(1-p)} = m\bar{r}$$

The summation identity used in the above sequence may be verified using induction and repeated differentiation of the 1/(1-*p*) expansion as in P5.11. This result says simply that if a player has a mean, single-inning, runlength of $\bar{r}$, then after *m* innings, his mean score will be $m\bar{r}$.

The variance and standard deviation of the scores are

$$\sigma_m^2 = \left(\sum_{n=0}^{\infty} n^2 R_{nm}\right) - \bar{R}_m^2 = -\bar{R}_m^2 + \sum_{n=0}^{\infty} \big(n(n-1)+n\big)R_{nm}$$

$$= \bar{R}_m - \bar{R}_m^2 + \sum_{n=0}^{\infty} n(n-1)R_{nm} = \bar{R}_m - \bar{R}_m^2 + m(m+1)p^2 q^m \sum_{n=0}^{\infty} \frac{(n+m+1)!}{n!(m+1)!}\, p^n$$

$$= \bar{R}_m - \bar{R}_m^2 + \frac{m(m+1)p^2}{q^2} = \frac{mp}{q^2} = \frac{mp}{(1-p)^2}$$

$$\sigma_m = \sqrt{\sigma_m^2} = \frac{\sqrt{mp}}{q} = \frac{\sqrt{mp}}{1-p}$$

There are significant qualitative differences in the $R_{nm}$ distributions (for a given inning count *m*) and the single-inning $r_n$ distribution, particularly for large *m*. The most obvious of these is that the mode (or peak) of $R_{nm}$ may occur for a nonzero value, as may be verified by inspection of a few examples; in contrast, the single-inning $r_n$ distribution always peaks at *n*=0. The ratio of two successive probability values is

$$\frac{R_{n+1,m}}{R_{n,m}} = \left(\frac{n+m}{n+1}\right)p$$

The first factor is nonincreasing and approaches one from above as *n* increases, the second factor *p* is less than one, and the product may be either larger than one, indicating that the distribution is increasing towards a peak, or less than one, indicating that the distribution is decreasing after the peak. The mode is determined by the smallest nonnegative value of *n* for which the ratio is less than one, or zero if the ratio is always less than one. Solving the above ratio for this value gives the relation

$$Mode = Ceiling\left(\frac{mp-1}{1-p}\right) = Ceiling\left(\bar{R}_m - \frac{1}{q}\right)$$

The closed-form expression for the multiple-inning median involves special function evaluations (namely, the *incomplete beta function*), but it is easily determined by inspection of the numerical cumulative distributions for a given shot probability $p$ and for a given value of the inning count $m$. The following table gives some examples of these statistical parameters for selected values of $m$ and $p$.

Table 5.3. Multiple-Inning Runlength Statistics

| | Mode | | | | | $\bar{R}_m$ | | | | | $\tilde{R}_m$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $p\backslash m$ | 1 | 2 | 4 | 8 | 16 | 1 | 2 | 4 | 8 | 16 | 1 | 2 | 4 | 8 | 16 |
| .5 | 0 | 0 | 2 | 6 | 14 | 1.0 | 2.0 | 4.0 | 8.0 | 16.0 | 0 | 1 | 3 | 7 | 15 |
| .6 | 0 | 1 | 4 | 10 | 22 | 1.5 | 3.0 | 6.0 | 12.0 | 24.0 | 1 | 2 | 5 | 11 | 23 |
| .7 | 0 | 2 | 6 | 16 | 34 | 2.3 | 4.7 | 9.3 | 18.7 | 37.3 | 1 | 4 | 8 | 18 | 36 |
| .8 | 0 | 3 | 11 | 27 | 59 | 4.0 | 8.0 | 16.0 | 32.0 | 64.0 | 3 | 7 | 15 | 31 | 63 |
| .9 | 0 | 8 | 26 | 62 | 134 | 9.0 | 18.0 | 36.0 | 72.0 | 144.0 | 6 | 15 | 33 | 69 | 141 |

Because the median has no simple closed-form expression, as do the mode and the mean, a useful empirical approximation of the median is given by the weighted average

$$\tilde{R}_m = \tfrac{1}{3}\left(2\bar{R}_m + Mode\right) \approx \bar{R}_m - 1/(3q)$$

Comparison of this estimate with the above exact values in Table 5.3 shows that it is fairly accurate for $m$=4 or greater. Note that apart from the integer truncation in the mode and median evaluations, the differences between the mean, median, and the mode depend only on the individual shot probability $p$ and are independent of the inning count $m$. This is why the three statistics appear to merge together in a relative sense in Table 5.3; they all increase with the inning count $m$ but with constant differences.

When the multiple-inning runlength mean and the mode are relatively close to each other, indicating little skew, and the distribution has a single peak that is not close to zero, then the distribution is well approximated by a normal distribution with mean and standard deviation as determined in P5.16. That is, even though the single-inning distribution appears very different than a normal distribution, the multiple-inning distributions approach nonetheless a normal form as the inning count $m$ increases. A perhaps simpler example of this common phenomenon is the point totals for two fair dice; each die individually has a flat distribution of point values from one to six, but when the totals are added for two dice, the probability of totaling to seven ($p_7$=$1/6$), which is the mean, is six times larger than rolling a two ($p_2$=$1/36$), with the other possible totals having intermediate probabilities. The results from P5.16 show that although the standard deviation of the score distributions are increasing with the inning count $m$, the deviation increases only as $\sqrt{m}$, whereas the mean score increases linearly as $m$. This means that the *relative deviations* (also called the *relative dispersions*), given by $\sigma_m / \bar{R}_m = 1/\sqrt{mp}$, decrease with respect to increasing inning count. When viewed in terms of percentages, the score will appear to more tightly cluster with larger inning counts, but when viewed in terms of actual score points, the score will appear to disperse more with larger inning counts; this is demonstrated in the following problem.

**Problem 5.17**: The 14.1 player from P5.13 plays 5 innings. What is his mean score, most likely score, and his median score? What is the standard deviation of the expected score distribution? What are these parameters after 50 innings?

*Answer*: This player's mean single-inning runlength is $\bar{r}$ =24.0, so his shot probability is $p$=0.96. After 5 innings, the mean score is given by

$$\bar{R}_5 = m\bar{r} = 5(24.0) = 120.0$$

His most likely score is the distribution peak, or the mode, given by

$$Mode = Ceiling\left(\bar{R}_m - 1/q\right) = Ceiling\left(120.0 - 25.0\right) = 95$$

The median is estimated as

$$\tilde{R}_5 = \tfrac{1}{3}\left(2\bar{R}_5 + Mode\right) \approx \bar{R}_5 - 1/\left(3q\right) = \bar{R}_5 - 25.0/3 = 112$$

This shows that the distribution of scores after 5 innings still is skewed significantly to the right. The standard deviation after 5 innings is

$$\sigma_5 = \sqrt{mp}/q = \sqrt{5(.96)}/.04 = 54.8$$

and the relative deviation is $\sigma_5/\bar{R}_5 = 54.8/120.0 = 0.457$, which is quite large.

If the player were to wager a fixed amount per point on the score after 5 innings, then the 120 point score demarks the fair betting point; those betting against the player at a lower score should expect to lose. But if someone were to wager in a betting pool of all possible scores, then a score of 95 is the most likely winner, and the scores close to 95 would be the best alternatives. And finally if the player were to wager simply whether the score is beyond a certain value after 5 innings (as in P5.14), then the 112 point score demarks the fair betting point; those betting against the player at a lower score should expect to lose.

For 50 innings the statistical parameters are: $\bar{R}_{50}$ =1200, *Mode*=1175, $\tilde{R}_{50}$ =1192, $\sigma_{50}$=173, and $\sigma_{50}/\bar{R}_{50}$ =0.144. This distribution is only slightly skewed to the right, and although the standard deviation is larger for 50 innings than for 5 innings, the relative deviation compared to the mean is much smaller.

In most of the previous discussion, the population distribution and statistical parameters have been assumed to be known, and the questions have been about the properties of various samples of this population. The reverse situation is now examined, namely how to predict the total population statistics from known sample statistics. For this purpose, consider a population from which all possible subsets of a particular size are formed. Each of these subsets has a mean, and the questions of interest are how reliable of an estimate for the total population mean is one of these subset means, how does this estimate depend on the standard deviation of the population, and how does this estimate improve with increasing sample size.

Assume that there is some population $\{x_i\}$ that has a mean $\bar{x}_p$ and a standard deviation $\sigma_p$. To simplify the following steps, it is assumed that the population is finite of size $N$, but the final results will also hold for infinite populations. Now suppose that $m$-element subsets are drawn with replacement from the population. There are $N^m$ possible $m$-element subsets, and $\beta$ is used as an index symbol to enumerate them. Such $m$-element sets are called *cartesian product* sets. The mean of each of these subsets is

$$\bar{x}_\beta = \frac{1}{m} \sum_{i=1}^{m} x_{\tau(i,\beta)}$$

where $\tau(i,\beta)$ is the population index of the *i-th* element of the $\beta$-*th* subset. The mean and variance of these $m$-element subset means is given by

$$\bar{x}_{[m]} = \frac{1}{N^m} \sum_{\beta=1}^{N^m} \bar{x}_\beta$$

$$\sigma[\bar{x}]_m^2 = \frac{1}{N^m} \sum_{\beta=1}^{N^m} \bar{x}_\beta^2 - \bar{x}_{[m]}^2$$

For $m$=1, there are $N$ 1-element subsets, and the mean of each of these subsets is simply the value of that element, $\bar{x}_\beta = x_{\tau(i,\beta)} = x_i$, and the mean of the subset means is the same as the population mean.

$$\bar{x}_{[1]} = \frac{1}{N} \sum_{\beta=1}^{N} \bar{x}_\beta = \bar{x}_p$$

Similarly, the variance of these subset means is the same as the population variance.

$$\sigma[\bar{x}]_1^2 = \frac{1}{N} \sum_{\beta=1}^{N} \bar{x}_\beta^2 - \bar{x}_{[1]}^2 = \sigma_p^2$$

Now consider the situation in which the $m$-element subset parameters $\bar{x}_{[m]}$ and $\sigma[\bar{x}]_m^2$ are assumed to be available. The $(m+1)$-element subsets are constructed by forming the cartesian products $\{x_i \quad x_\beta ; i = 1 \quad N, \beta = 1 \quad N^m\}$. That is, each of the new set members is formed by combining the $N$ population set elements with all possible $N^m$ $m$-element subsets. The mean of each of these new subsets is

$$\bar{x}_{\beta'} = \bar{x}_{i\beta} = \frac{x_i + m\bar{x}_\beta}{m+1}$$

The mean of these subset means is

$$\bar{x}_{[m+1]} = \frac{1}{N^{m+1}} \sum_{\beta'=1}^{N^{m+1}} \bar{x}_{\beta'} = \frac{1}{(m+1)N^{m+1}} \sum_{\beta=1}^{N^m} \sum_{i=1}^{N} \left( x_i + m\bar{x}_\beta \right) = \frac{\bar{x}_p + m\bar{x}_{[m]}}{(m+1)}$$

This relation gives the results $\bar{x}_{[2]} = \bar{x}_p$, $\bar{x}_{[3]} = \bar{x}_p$, and, in general,

$$\bar{x}_{[m]} = \bar{x}_p$$

for all subset sizes *m*. The variance of the subset means is

$$\begin{aligned}
\sigma[\bar{x}]_{m+1}^2 &= \frac{1}{N^{m+1}} \sum_{\beta=1}^{N^m} \sum_{i=1}^{N} \bar{x}_{i\beta}^2 - \bar{x}_{[m+1]}^2 \\
&= \frac{1}{N^{m+1}} \sum_{\beta=1}^{N^m} \sum_{i=1}^{N} \frac{1}{(m+1)^2} \left( x_i^2 + 2m x_i \bar{x}_\beta + m^2 x_\beta^2 \right) - \bar{x}_p^2 \\
&= \frac{1}{(m+1)^2} \left( \left( \frac{1}{N} \sum_{i=1}^{N} x_i^2 - \bar{x}_p^2 \right) + m^2 \left( \frac{1}{N^m} \sum_{\beta=1}^{N^m} \bar{x}_\beta^2 - \bar{x}_p^2 \right) \right) \\
&= \frac{\sigma_p^2 + m^2 \sigma[\bar{x}]_m^2}{(m+1)^2}
\end{aligned}$$

This relation gives the results $\sigma[\bar{x}]_2{}^2=\sigma_p{}^2/2$, $\sigma[\bar{x}]_3{}^2=\sigma_p{}^2/3$, and, in general, $\sigma[\bar{x}]_m{}^2=\sigma_p{}^2/m$. The standard deviation of the subset means is then given by

$$\sigma[\bar{x}]_m = \frac{\sigma_p}{\sqrt{m}}$$

That is, the distribution of subset means becomes more narrowly peaked about the population mean as the size of the subset becomes larger. This relation also says that if the population standard deviation is small, then the mean estimates obtained from the subsets will be similarly sharp. In practice, the population standard deviation is usually not known, so it must be estimated from the *m*-element subspace statistics, along with the estimate of the mean.

When the subsets are formed by selection from the population without replacement, then the mean of the subset means is also given by

$$\bar{x}_{[m]} = \bar{x}_p$$

and the standard deviation of the subset means is given by

$$\sigma[\bar{x}]_m = \frac{\sigma_p}{\sqrt{m}} \sqrt{\frac{N-m}{N-1}}$$

The standard deviation of the means of the subsets formed without replacement are always smaller than those with replacement, and in particular $\sigma[\bar{x}]_m=0$ when $m=N$. When $N>>m$, then the standard deviation of the mean is essentially the same for both types of subsets, and in the limit of an infinite population, both expressions are seen to be formally equivalent.

When the standard deviation is computed for an *m*-element sample space, then it is customary to use the factor (*m*-1) rather than *m* in the denominator; this has the effect of making the estimate for the population standard deviation slightly larger, but for reasonably large sample sizes the difference is unimportant. There are also other corrections that are sometimes applied when estimating population statistics from sample spaces. With the knowledge that such corrections can lead to slightly better estimations, they will not be used in the following examples for the sake of simplicity.

**Problem 5.18**: A population space has the values {1,2,4,5} which occur with equal probabilities. Compute the mean and standard deviation of the population set. Compute the mean and the standard deviation of the mean of the 2-element cartesian product set.
*Answer*: The population space mean is 3.0, and the population standard deviation is $\sigma_p$=sqrt(5/2)=1.5811. The 2-element cartesian product set is the same as the 2-element subsets drawn with replacement from the original set. There are 16 of these 2-element subsets, all with equal probability,

| | | | |
|---|---|---|---|
| {1,1} | {1,2} | {1,4} | {1,5} |
| {2,1} | {2,2} | {2,4} | {2,5} |
| {4,1} | {4,2} | {4,4} | {4,5} |
| {5,1} | {5,2} | {5,4} | {5,5} |

and with the corresponding means

| | | | |
|---|---|---|---|
| 1.0 | 1.5 | 2.5 | 3.0 |
| 1.5 | 2.0 | 3.0 | 3.5 |
| 2.5 | 3.0 | 4.0 | 4.5 |
| 3.0 | 3.5 | 4.5 | 5.0 |

The mean of these 2-element subset means is $\bar{x}_{[2]}$3.0, which demonstrates the general relation $\bar{x}_{[m]}=\bar{x}_p$. The standard deviation of these subspace means is $\sigma[\bar{x}]_2$=1.1180, which agrees with the equation $\sigma[\bar{x}]_2=\sigma_p$/sqrt(2). Note that these computations apply to a 4-element population space, or to a larger finite population space with the appropriate repetitions, or to an infinite population space with the appropriate probabilities.

**Problem 5.19**: A player has a practice routine that involves a particular sequence of shots. He keeps track of his numerical score for this routine for 10 weeks with the following results: {50, 44, 46, 52, 47, 51, 49, 45, 48, 50}. What is his mean score? Assuming a normal distribution of scores, what is the range of scores for which there is an 80% confidence level that the range includes the player's true mean score for this practice drill? The player experiments with a new technique (e.g. a different stroke technique) and scores a 53 on this drill, his highest score ever. Can the player be 95% certain that this is due to the technique change rather than to random chance? Can he be 90% certain that the score is due to the technique change? How does the sample size affect these assessments?
*Answer*: The mean score for the 10 weeks is 48.2. The standard deviation of the sample set is $\sigma$=2.52, which is taken as an approximation of the population standard deviation. The standard deviation of the mean is estimated as $\sigma[\bar{x}]_{10}=\sigma$/Sqrt(10)=0.797. Using the normal distribution approximation, Table 5.1 gives the critical value for an 80% confidence level as $z_c$=1.28. There is an 80% chance that the true mean score, which

would be the long-term mean of the player's score for this drill, is between $\bar{x} - z_c\sigma[\bar{x}]$=47.2 and $\bar{x} + z_c\sigma[\bar{x}]$=49.2.

The score of 53 is 4.8 points higher than the mean, or 4.8/2.52=1.90 standard units. The critical value for 95% confidence is $z_c$=1.96, so such a score would be expected to occur even with no change in stroke technique due to random chance within the 95% confidence predicted by a normal distribution assumption. For a 90% confidence $z_c$=1.645; such a score would not be expected to occur at this confidence level due simply to random chance. If the estimates for the mean and standard deviation were reliable (e.g. if the sample were much larger), then the player could say that he is 90% certain that the stroke technique change improved his score, but he could not say that he is 95% certain. However, this is a fairly small difference based on such a small sample. There is, after all, a good chance that the true mean is as high as 49.2, so a score of 53 is only 3.8 points, or 3.8/2.52=1.51 standard units, above the mean, and such a score can be expected to occur about 87% of the time due to random chance; that is, the player could say only that he is 13% confident that the score is due to the stroke change. Additional scores with the original technique would allow for a more accurate estimation of the population mean, and therefore a more accurate estimation of the effect of the stroke change in terms of standard units. This is one reason why players should establish practice routines and record their numerical scores over long periods of time; not only does it allow the player to track his progress, but it also allows for accurate statistical assessments of technique and equipment changes.

More exact determinations of the confidence can be achieved with additional data, including those obtained using the new stroke technique, by comparing the means and standard deviations of the different data sets (original stroke technique vs. the new one). When the means are sufficiently different, and when the standard deviations are sufficiently narrow, then there is a high confidence that the technique change is responsible for the score difference rather than simply the expected random fluctuations in the score. In general, the confidence is determined by the overlap regions of the two distribution tails. This type of comparative analysis can be quantified further with *chi-squared* tests (to compare expected and measured distribution statistics), the *Student's t-test* (to determine if two samples with the same variance have different means), the *F-Test* (to compare sample variances), and with *Analysis of Variance* techniques (to determine if two different samples actually are drawn from the same population). These methods are all outside the scope of this section, but they are mentioned in case the interested reader wishes to follow up on this interesting topic.

Statistical analysis can be used as a basis of choosing from among a set of possible tactics. This requires estimations of individual shot outcomes. A simple example of using probabilities to assess tactical options is a simple "one-ball" game, which occurs in actual game situations when, for example, both players are shooting at the 9-ball in a game of 9-ball, or when both players are shooting at the 8-ball in a game of 8-ball, or when both players are shooting at the black in a game of snooker, or when both players are shooting at the last ball in a game of one-pocket. In the one-ball game, a player is faced with a particular shot at a single ball. If he succeeds, then he wins immediately, and if he fails then the outcome depends on the outcome of the opponent's shot. As a way of keeping track of the details, such game situations may be represented with a diagram. The diagram corresponding to the one-ball game is shown in Fig. 5.2.

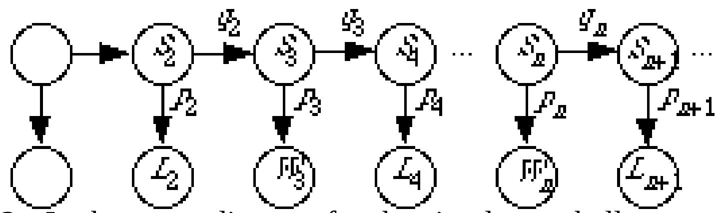


Fig. 5.2. In the game diagram for the simple one-ball game, there are an infinite number of nodes. Only the first few are shown explicitly. Player-1 wins at the terminal nodes *Wn* and he loses at the terminal nodes *Ln* . The transition probabilities are next to the connecting arcs.

In this diagram, the various *states* (or game situations) are called *nodes* and are shown by the circles. The possible state transitions are shown by the lines the connecting the nodes, and these lines are called *arcs*. The states labeled by *Sn* are where the winning ball has not been pocketed by the $n^{th}$ inning, the states labeled by *Wn* are those where player-1 has won on the $n^{th}$ inning, and the states labeled by *Ln* are those where player-1 has lost on the $n^{th}$ inning. Each arc is associated with a particular transition probability. There are probability *weights* associated with each state. *S*1 is called the *graph head*, and $P_{S1}$=1 means that the *S*1 node is the starting point of the game. The weight of any other state is given by the sum of the weights of the previously connected state multiplied by the transition probability associated with the connecting arc. That is, the probability of arriving at a particular state in the game is the summation of the probability of arriving at all previous states, times the probability of making a transition from these previous states to the current state. The states labeled *Wn* and *Ln* are called *terminal nodes*, or probability *sinks*, or *tails*, since there are no arcs leaving these nodes; the game is over when the destination is one of these nodes. These game diagrams are a pictorial way of enumerating all possible paths as the probability density flows from the sources, through the transient states, to the probability sinks.

In the simple one-ball game depicted in Fig. 5.2, there is only one probability source *S*1, an infinite number of transient states *Sn*, and an infinite number of probability sinks *Wn* and *Ln*. There are only two arcs leaving each of the *Sn* nodes; in more complicated game situations, there may be several arcs leaving a node, each depicting a transition to a new possible state or to a previous state. The sum of all of the arc transition probabilities from a node is 1. In Fig. 5.2, the successful shots by either player are labeled $p_n$ and the unsuccessful shots are labeled $q_n$ with $q_n$=(1-$p_n$). This general idea of assigning probability weights to nodes, and to computing these weights from transition probabilities has already been used in the recursive algorithm for computing game score probabilities in P5.6.

In the general one-ball game, all of the individual $p_n$ values will be different. It is interesting to consider some simpler cases in which the shot success probabilities are assumed to have special relations.

**Problem 5.20**: Assume that all of the shots taken by player-1 in the game depicted in Fig. 5.2 have a success probability of $p_1$, and all of the shots taken by player-2 have a success probability of $p_2$. In terms of these two parameters, what is player-1's total probability of winning, assuming that an infinite number of shots is allowed in the game? What combinations of $p_1$ and $p_2$ lead to a game probability of $W=½$?

*Answer*: The total chance of winning is the summation of the node weights $P_{W1}$, $P_{W3}$,…. By multiplying the appropriate arc weights to get the node weights, the probabilities are given by $P_{W1}=p_1$, $P_{W3}=q_1q_2p_1$, $P_{W5}=(q_1q_2)^2p_1$, and so on. This will be called the two-parameter infinite-look-ahead approximation to the general one-ball game. The summation is

$$W^{[\infty]} = P_{W1} + P_{W3} + P_{W5} + \ldots = p_1 + q_1q_2p_1 + (q_1q_2)^2 p_1 + \ldots$$

$$= \sum_{i=1}^{\infty} P_{W(2i-1)} = \sum_{i=0}^{\infty} (q_1q_2)^i p_1 = \frac{p_1}{1-q_1q_2} = \frac{1}{1+p_2\dfrac{(1-p_1)}{p_1}}$$

Setting $W^{[\infty]}=½$ and solving for $p_2$ in terms of $p_1$ gives

$$p_2^{crit} = \frac{p_1}{(1-p_1)}$$

When the actual value of $p_2$ is larger than this critical value, player-2 is expected to win, and when $p_2$ is smaller than this critical value then player-1 is expected to win.

A contour plot of $W^{[\infty]}$ as a function of the two parameters $p_1$ and $p_2$ is shown in Fig. 5.3. The region of the contour plot corresponding to small $p_1$ and large $p_2$ is the "sell-out" region; shots in this region should usually be avoided and other shots should be considered. The area of the contour plot corresponding to large $p_1$ and small $p_2$ is the "2-way shot" region; it is a great tactical advantage when these shots are available, as indicated by the large $W^{[\infty]}$ values. The area of the contour plot with small $p_1$ and small $p_2$ corresponds to defensive safety shots; the primary purpose is to keep the opponent from winning immediately, and to exploit any small advantage in probability over several innings. It is interesting to note how sensitive is the game probability estimate $W^{[\infty]}$ to small changes in the shot probabilities $p_1$ and $p_2$ in this region.

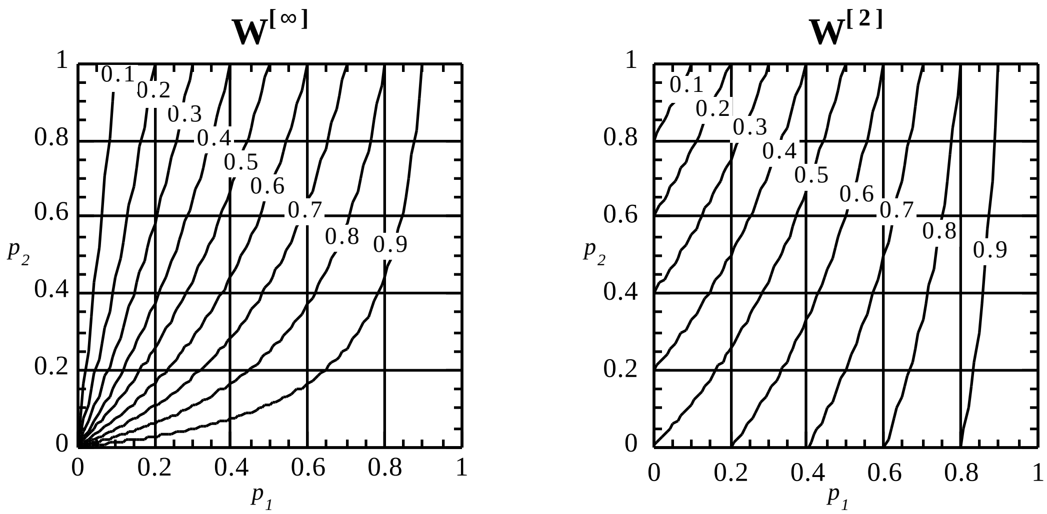



Fig. 5.3. Contour plots of the $W^{[\infty]}$ and $W^{[2]}$ approximations to the general one-ball game as a function of the two independent parameters $p_1$ and $p_2$. The $W$=½ contours are the same for both approximations.

A particularly important set of values corresponds to $W^{[\infty]}=½$. The $p_1$ and $p_2$ combinations for which $W^{[\infty]}>½$ are those in which player-1 is expected to win the simple one-ball game, and those values for which $W^{[\infty]}<½$ are those in which player-1 is expected to lose. In this simple game model, $p_2$ will be large either when player-2 is a very good shotmaker, or when the balls end up consistently in easy positions after a miss by player-1. It is clear from this graph that $W^{[\infty]}>½$ for all values of $p_2$ when $p_1>½$; this means that no matter how good of a shotmaker the opponent is, or how easy of a shot is left after each miss, player-1 is the expected winner when $p_1>½$. This is supported also by the $p_2^{crit}$ expression given in P5.20. This advantage is afforded player-1 because he gets the first shot in the game. However, it is still useful to compare two possible strategies, even when both of them result in favorable outcomes for player-1. When $p_1<½$, then the expected outcome clearly depends on $p_2$; when $p_2$ is sufficiently small, then player-1 is still the expected winner, but when $p_2$ is large, then player-1 is expected to lose.

**Problem 5.21**: Assume that player-1 in the one-ball game takes his first shot with a success probability of $p_1$, and that a good estimate of the value of $p_2$ is known, but after these first two shots both players are assumed to have an even chance of winning the game. In terms of these two parameters, what is player-1's total probability of winning? What combinations of $p_1$ and $p_2$ lead to a game probability of $W$=½?

*Answer*: The game diagram for this approximation to the general one-ball game is shown in Fig. 5.4. This will be called the two-parameter two-shot-look-ahead approximation. There are now only two nodes in Fig. 5.4 that correspond to wins for player-1. The game probability is the sum of the weights for these two nodes.

$$W^{[2]} = P_{W1} + P_W = p_1 + \frac{1}{2} q_1 q_2$$

A contour plot of $W^{[2]}$ is shown in Fig. 5.3. When compared to the $W^{[\infty]}$ contour plot, it is seen that the two approximations give similar, but not exactly equivalent, predictions of game probabilities.

Rearranging the $W^{[2]}=\frac{1}{2}$ equation to solve for $p_2$ as a function of $p_1$ gives

$$p_2^{crit} = \frac{p_1}{(1 - p_1)}$$

which is the same curve of critical values as determined previously for $W^{[\infty]}$.

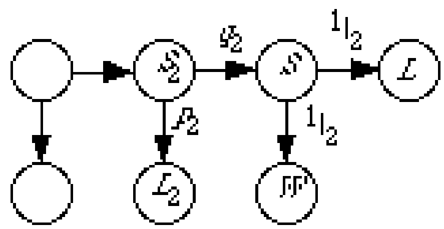


Fig. 5.4. The two-shot-look-ahead approximation to the general one-ball game depends on only two independent parameters $p_1$ and $p_2$ that characterize the shot success for the first two innings. After the second inning, the game outcome probability is split equally between the two players.

If three independent shot parameters are known, then this leads to a three-shot-look-ahead approximation, and in general there are $n$-shot-look-ahead approximations involving $n$ independent probability parameters. It should be stressed that both the two-parameter infinite-look-ahead and the two-parameter two-shot-look-ahead equations are approximations to the general one-ball game; one of these should not be regarded as an approximation to the other. In some situations, the infinite-look-ahead assumption may be more appropriate, while in other situations the two-shot-look-ahead assumption might be best. For example, if player-1 is a weaker player than player-2, then from a relatively neutral position the infinite-look-ahead model with a large $p_2$ value would provide the most reliable estimates of game outcomes; but if both players are roughly equal in ability, and if player-1 has a decided positional advantage for his first shot (e.g. a strong 2-way shot corresponding to a large $p_1$ and a small $p_2$), then the two-shot-look-ahead model would provide the most reliable estimates of game outcomes.

If two players are playing a multigame 9-ball match (or, for example, 8-ball or one-pocket), then the opening break shot is usually regarded as an advantage. Matches are sometimes played in which the winner of each game is rewarded by being allowed the opening break in the next game (*winner-breaks*), or they may be played where the loser of one game breaks in the next game (*loser-breaks*), or the players may *alternate breaks* from game to game, or one of the players may break all of the games (e.g. *player-1* *breaks*; this is usually regarded as a handicap advantage for the breaking player to compensate for some difference in skills). This last situation is interesting in the context of the above analysis of the one-ball game. Suppose that player-1 breaks each game, and that he wins each of these games on his first inning with a probability of $p_1$. In the second inning, player-2 wins with a probability of $p_2$, and so on. The game diagram for this situation is the same as for the one-ball game. The nodes of the diagram correspond to inning counts rather than individual shots, and the winning probabilities are with respect to games rather than individual shot successes, but the mathematical structure is the same for both situations. The infinite-look-ahead and the two-shot-look-ahead approximations, and the discussions of these two parameters in P5.20 and P5.21 apply also, in a perhaps more approximate way, to this multigame match situation. With these approximations, the contour plots in Fig. 5.3 show that player-1 would be expected to have an advantage over player-2 by virtue of playing the first inning, and this advantage becomes more significant for larger values $p_1$.

**Problem 5.7**: Using the game probability estimates $W^{[\infty]}$ and $W^{[2]}$, compute the probability for player-1 to win when $(p_1,p_2)$ have the values: (0.1,0.9), (0.9,0.1), (0.9,0.9), (0.5,0.5), (0.25,0.33), (0.4,0.8), and (0.3,0.3).

*Answer*: The game probability estimates are given in the following table

| $p_1$ | $p_2$ | $W^{[\infty]}$ | $W^{[2]}$ |
|---|---|---|---|
| 0.1 | 0.9 | 0.11 | 0.15 |
| 0.9 | 0.1 | 0.99 | 0.95 |
| 0.9 | 0.9 | 0.90 | 0.90 |
| 0.5 | 0.5 | 0.67 | 0.62 |
| 0.25 | 0.33 | 0.50 | 0.50 |
| 0.4 | 0.8 | 0.45 | 0.46 |
| 0.3 | 0.3 | 0.59 | 0.55 |

The $(p_1,p_2)$=(0.1,0.9) shot is a sell-out shot. Player-1 is expected to lose this game, even with the first-shot advantage; he should consider another choice of shot. The (0.9,0.1) situation is a strong 2-way shot; player-1 is the favorite in this game. For (0.9,0.9), player-1 is again the favorite. Even though $p_1$ for this case is the same as the previous one, it is clear that it is better to plan to leave a low-percentage shot for the opponent than a high-percentage one (i.e., $W^{[\infty]}$=0.99 is better than $W^{[\infty]}$=0.90). For the (0.5,0.5) shot, player-1 is the expected game winner using both estimates, even though he and his opponent are evenly matched with equally difficult shots; this is due to the first-shot advantage. The (0.25,0.33) shot corresponds to a $(p_1,p_2^{crit})$ pair, so each player has equal probability of winning according to both estimates.

Player-1 has a disadvantage at (0.4,0.8) and a fairly significant advantage at (0.3,0.3) using both estimates of the game probability. It is interesting to compare these last two situations, since it appears to be a paradox to many inexperienced pool players. In both cases, the individual shot probability $p_1$ is relatively small. In fact, $p_1$ is smaller for the second (favorable) game outcome than for the first (unfavorable) game outcome. In the first case, he leaves a high-probability shot for his opponent, while in the second case he leaves a low-probability shot. Inexperienced players often choose shots based only on their estimate of the first-shot success probability, that is only on $p_1$. This is an example of how the down-side consequences (what occurs after the miss) outweigh the up-side reward (which shot has the higher $p_1$). In other words, it is sometimes more important not to “sell out” than it is to try to succeed with a spectacular shot. In “tactic-rich” games involving relatively difficult shots, such as one-pocket, this kind of decision is part of the routine shot-selection process. The simple one-ball game mathematical model used here provides an approximate way to quantify the relative importance of the up-side reward and the down-side consequences for these more complex situations in actual pool games.

In physical simulations, processes that may be characterized by probabilities are called *stochastic* systems, and an important class of stochastic systems, called *Markov* processes, are those in which the probability of making a transition from one state to another depends only on the initial and final states, and not upon a history of the previous states. The game diagrams described above are examples of Markov processes. One way to analyze these types of diagrams is to consider them as a “time dependent” process. In some situations, it is the transient short-time behavior that is of importance, and at other times it is the long-time steady-state behavior that is most interesting. In the above game diagrams, the “time” parameter corresponds to the inning count, or to the shot count, or, as will be discussed below, to a game count. An initial probability distribution is assigned to the nodes of the graph, and this probability density flows through the graph as individual time steps are taken. The information that is most important in the pool-game situation is how much of this probability density ends up in the various terminal nodes. In the above examples of game diagrams, it was possible to answer this question by recognizing relatively simple algebraic simplifications that allowed closed-form expressions to be obtained. But in more complicated situations, such closed-form expressions may not be apparent, or they may not even exist. In these situations, it is still possible to extract the long-time steady state probability densities numerically, and this general procedure is now discussed.

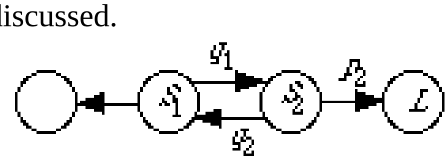


Fig. 5.5. The game diagram for the two-parameter infinite look-ahead approximation to the general one-ball game consists of four nodes: two terminal nodes and two transition nodes.

This procedure will be applied first to the one-ball game so that the results can be compared to those previously found. The two-parameter infinite look-ahead game probability will be examined. For this purpose, it is convenient to use a simpler game diagram, shown in Fig. 5.5, that has only a finite number of nodes (four in this case). In this diagram, all of the wins for player-1 are treated equivalently, with a single terminal node, and likewise all loses for player-1 are treated with a single terminal node. The two transient states are related simply to which player is shooting the shot. The important quantity in Markov analysis is the *probability transition matrix*. The rows and columns of this matrix correspond to the states of the system, and therefore to the nodes of the game diagram. The element $M_{ij}$ corresponds to the arc weight of the arc that connects node $j$ to node $i$; that is $M_{ij}$ is the probability of making a transition from the state corresponding to node $j$ to the state corresponding to node $i$. Nodes that are not connected correspond to zero $M_{ij}$ values. Terminal nodes are assumed to make transitions to themselves each time step with unit probability. The transition matrix corresponding to the two-parameter infinite-look-ahead approximation to the one-ball game is

$$\mathbf{M} = \begin{pmatrix} 1 & p_1 & 0 & 0 \\ 0 & 0 & q_2 & 0 \\ 0 & q_1 & 0 & 0 \\ 0 & 0 & p_2 & 1 \end{pmatrix}$$

The rows and columns of the matrix correspond to the nodes *W*, *S1*, *S2*, and *L*, respectively. The sum of the elements in a column of **M** is one, which reflects the fact that probability density is not destroyed by taking a time step. The vector-matrix product relation $(1,1,1,1)\mathbf{M}=(1,1,1,1)$ is a consequence of this, and such a relation is always satisfied for a Markov transition matrix. This means that there exists at least one left eigenvector of **M** that corresponds to an eigenvalue of one, and therefore there also must exist a right eigenvector with this same eigenvalue; the existence of this unit eigenvalue is important in this analysis. Let $\kappa=(1,1,1,1)$ be this left eigenvector and $\mathbf{v}^0$ be an arbitrary column vector, then the dot product $\kappa\mathbf{v}^0$ is equal to the sum of the elements of $\mathbf{v}^0$. The vectors of interest correspond to probability densities, and such vectors contain only nonnegative elements that sum to one. A single time-propagation step from an initial vector $\mathbf{v}^0$ is given by the matrix-vector product $\mathbf{v}^1=\mathbf{M}\mathbf{v}^0$. Operating on the left of this equation with $\kappa$ gives the result $\kappa\mathbf{v}^1=\kappa\mathbf{v}^0=1$, which means that the sum of the probability density after the time step is the same as the sum before the time step. After two steps, the density is given by $\mathbf{v}^2=\mathbf{M}\mathbf{v}^1=\mathbf{M}^2\mathbf{v}^0$, and after $n$ steps the density is given by $\mathbf{v}^n=\mathbf{M}\mathbf{v}^{(n-1)}=\mathbf{M}^n\mathbf{v}^0$. Operating on the left by $\kappa$ on any of these relations shows that the total density is conserved always by the propagation operations.

What does the vector $\mathbf{v}^n$ look like after a large number of steps? The answer depends on the eigenvalues of the matrix **M**. In general the right eigenvectors of **M** satisfy the equation $\mathbf{MR}=\mathbf{R}\lambda$ in which the right eigenvectors form the columns of **R**, and the diagonal matrix $\lambda$ contains the corresponding eigenvalues. This allows the matrix **M** to be written as $\mathbf{M}=\mathbf{R}\lambda\mathbf{R}^{-1}$. The matrix $\mathbf{M}^2$ is given by $\mathbf{M}^2=\mathbf{R}\lambda\mathbf{R}^{-1}\mathbf{R}\lambda\mathbf{R}^{-1}=\mathbf{R}\lambda^2\mathbf{R}^{-1}$, and in general $\mathbf{M}^n=\mathbf{R}\lambda^n\mathbf{R}^{-1}$ with

$$\lambda^n = \begin{pmatrix} \lambda_1^n & 0 & 0 & 0 \\ 0 & \lambda_2^n & 0 & 0 \\ 0 & 0 & \lambda_3^n & 0 \\ 0 & 0 & 0 & \lambda_4^n \end{pmatrix}$$

This expression allows the probability distribution after an arbitrary number of time steps to be determined with relatively little effort, compared to the straightforward approach using repeated multiplications. The rows of the matrix $\mathbf{L}=\mathbf{R}^{-1}$ are the left eigenvectors, and, with the appropriate choice of normalization, one of these rows is $\kappa=\mathbf{L}^{[1]}$. If $\lambda_i=1$, then $\lambda_i^n=1$ for all $n$, and if $|\lambda_i|<1$, then $\lambda_i^n \to 0$ as $n \to \infty$. For the transition matrices associated with game diagrams, the eigenvalues are $-1<\lambda_i\leq 1$. This allows the vector limit after an infinite number of time steps, called the *stochastic limit*, to be written as

$$\mathbf{v}^\infty = \mathbf{M}^\infty\mathbf{v}^0 = \mathbf{R}\lambda^\infty\mathbf{R}^{-1}\mathbf{v}^0 = \mathbf{R}\lambda^\infty\mathbf{L}\mathbf{v}^0 = \sum_i^{(\lambda_i=1)} \mathbf{R}_i\left(\mathbf{L}_i\mathbf{v}^0\right)$$

in which the summation includes only the right and left eigenvector pairs that correspond to eigenvalues of unit magnitude. In many cases, there is only a single eigenvalue in this summation, and in this case $\mathbf{v}^\infty=\mathbf{R}_1(\kappa\mathbf{v}^0)=\mathbf{R}_1$, where $\mathbf{R}_1$ is the right eigenvector associated with the single eigenvalue of unit magnitude (and scaled appropriately to conserve density). In this case there is a single stochastic limit $\mathbf{v}^\infty$ that is approached for any arbitrary starting density $\mathbf{v}^0$. In other cases, there may be more than one such vector in the summation, in which case the final stable probability distribution depends on the starting distribution $\mathbf{v}^0$. It may be verified that $\mathbf{M}^\infty\mathbf{M}^\infty=\mathbf{M}^\infty$, and therefore $\mathbf{M}^\infty$ is a projection operator; it operates upon an arbitrary probability density distribution and projects this vector onto the subspace of the stable state distribution(s).

In the specific case of the two-parameter infinite-look-ahead approximation to the one-ball game, the eigenvalues of $\mathbf{M}$ are $(1,1,\sqrt{q_1q_2},-\sqrt{q_1q_2})$. There are two eigenvalues of unit magnitude. The matrix $\mathbf{M}^\infty$ is given by

$$\mathbf{M}^\infty = \begin{pmatrix} 1 & \frac{p_1}{1-q_1q_2} & \frac{p_1q_2}{1-q_1q_2} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \frac{q_1p_2}{1-q_1q_2} & \frac{p_2}{1-q_1q_2} & 1 \end{pmatrix}$$

**Problem 5.23**: Given an initial probability distribution of $\mathbf{v}^0=(0,1,0,0)^T$, compute $\mathbf{v}^1$, $\mathbf{v}^2$, $\mathbf{v}^3$, $\mathbf{v}^4$, $\mathbf{v}^5$, and $\mathbf{v}^\infty$. What are these same vectors for $\mathbf{w}^0=(0,0,1,0)^T$? What is the stochastic limit for the vector $(0,\frac{1}{2},\frac{1}{2},0)^T$? What is the meaning of these three limits?
*Answer*: For $\mathbf{v}^0=(0,1,0,0)^T$, the first few vectors are

$$\mathbf{v}^0=\begin{pmatrix}0\\1\\0\\0\end{pmatrix},\ \mathbf{v}^1=\begin{pmatrix}p_1\\0\\q_1\\0\end{pmatrix},\ \mathbf{v}^2=\begin{pmatrix}p_1\\q_1q_2\\0\\q_1p_2\end{pmatrix},\ \mathbf{v}^3=\begin{pmatrix}p_1\left(1+q_1q_2\right)\\0\\q_1^2q_2\\q_1p_2\end{pmatrix},\ \mathbf{v}^4=\begin{pmatrix}p_1\left(1+q_1q_2\right)\\\left(q_1q_2\right)^2\\0\\q_1p_2\left(1+q_1q_2\right)\end{pmatrix},$$

$$\mathbf{v}^5=\begin{pmatrix}p_1\left(1+q_1q_2+\left(q_1q_2\right)^2\right)\\0\\q_1^3q_2^2\\q_1p_2\left(1+q_1q_2\right)\end{pmatrix},\text{ and }\mathbf{v}^\infty=\mathbf{M}^\infty\mathbf{v}^0=\begin{pmatrix}\dfrac{p_1}{1-q_1q_2}\\0\\0\\\dfrac{q_1p_2}{1-q_1q_2}\end{pmatrix}.$$

The first element of these vectors, which corresponds to the probability that player-1 will win the game after the appropriate number of innings, is seen to be the same as the series of cumulative probabilities computed in P5.20, and the corresponding element of the $\mathbf{v}^\infty$ vector agrees also with that from P5.20. Although the 4-node game diagram in Fig. 5.5 seems simpler than the infinite-node diagram in Fig. 5.2, the step-by-step propagation of the 4-node density vector gives the same information as the more complicated game diagram. It may also be noted in this example that the sum of the densities for the four nodes always adds up to 1.

For $\mathbf{w}^0=(0,0,1,0)^T$, the first few vectors are

$$\mathbf{w}^0=\begin{pmatrix}0\\0\\1\\0\end{pmatrix},\ \mathbf{w}^1=\begin{pmatrix}0\\q_2\\0\\p_2\end{pmatrix},\ \mathbf{w}^2=\begin{pmatrix}p_1q_2\\0\\q_1q_2\\p_2\end{pmatrix},\ \mathbf{w}^3=\begin{pmatrix}p_1q_2\\q_1q_2^2\\0\\p_2\left(1+q_1q_2\right)\end{pmatrix},\ \mathbf{w}^4=\begin{pmatrix}p_1q_2\left(1+q_1q_2\right)\\0\\\left(q_1q_2\right)^2\\p_2\left(1+q_1q_2\right)\end{pmatrix},$$

$$\mathbf{w}^5=\begin{pmatrix}p_1q_2\left(1+q_1q_2\right)\\q_1^2q_2^3\\0\\p_2\left(1+q_1q_2+\left(q_1q_2\right)^2\right)\end{pmatrix},\text{ and }\mathbf{w}^\infty=\mathbf{M}^\infty\mathbf{w}^0=\begin{pmatrix}\dfrac{p_1q_2}{1-q_1q_2}\\0\\0\\\dfrac{p_2}{1-q_1q_2}\end{pmatrix}.$$

For the $(0,\tfrac{1}{2},\tfrac{1}{2},0)^T$ vector, the stochastic limit is

$$\mathbf{M}^\infty\begin{pmatrix}0\\\tfrac{1}{2}\\\tfrac{1}{2}\\0\end{pmatrix}=\tfrac{1}{2}\mathbf{M}^\infty\left(\mathbf{v}^0+\mathbf{w}^0\right)=\tfrac{1}{2}\begin{pmatrix}\dfrac{p_1\left(1+q_2\right)}{1-q_1q_2}\\0\\0\\\dfrac{p_2\left(1+q_1\right)}{1-q_1q_2}\end{pmatrix}$$

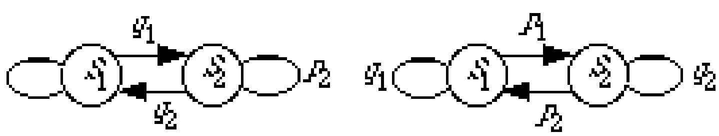
For the $\mathbf{v}^0$ case, player-1 is given the first shot of the game, and consequently the first chance to win, whereas in the $\mathbf{w}^0$ case player-2 is given the first shot. For this game diagram, there are two eigenvalues equal to one, so there are two possible independent, asymptotic stable solutions for these two initial conditions, $\mathbf{v}^\infty$ and $\mathbf{w}^\infty$. The same game diagram, and the same Markov analysis, covers both situations. The $(0,\tfrac{1}{2},\tfrac{1}{2},0)^T$ initial probability distribution corresponds to the alternating break situation, or perhaps to some other situation in which each player has the first shot an equal number of times.

In this simple approximation to the one-ball game, closed-form expressions could be found for all of the eigenvalues and eigenvectors, but in more complicated game situations this may not be true. If only a numerical solution is possible, then the asymptotic stable solutions can be found for any initial density vector by computing the eigenvalues and eigenvectors of the transition matrix numerically.

In the winner-breaks match situation, a player with a strong break advantage has the opportunity to break and win several consecutive games in a row. This may occur because the player breaks and runs, never allowing the opponent to have a shot, or it may be because the player is a good tactical player and he never allows his opponent an open shot on a winnable table. The winner-breaks match situation amplifies the break advantage in this case through a positive feedback situation. In a loser-breaks match, the player with the break advantage cannot exploit it because when he wins one game, his opponent gets to break, nullifying the break advantage. In this case the break advantage is damped through a negative feedback situation. The following problem shows how this feedback situation can be quantified.

**Problem 5.24**: Two players are playing a series of 9-ball games, and the winner of one game breaks in the subsequent game. When player-1 breaks, player-1 wins with a probability of $p_1$. When player-2 breaks, player-2 wins with a probability of $p_2$. What fraction of the total games will player-1 win if a large number of games are played? If the loser of one game breaks in the next game, what fraction of the total games will player-1 win? What is the expected outcome in the alternating break situation?

*Answer*: The game diagrams for the winner-breaks (WB) and for the loser-breaks (LB) situations are:



There are two states of interest, *S*1 is when player-1 breaks and *S*2 is when player-2 breaks. The corresponding transition matrices are:

$$\mathbf{M}^{WB} = \begin{pmatrix} p_1 & q_2 \\ q_1 & p_2 \end{pmatrix}, \quad \mathbf{M}^{LB} = \begin{pmatrix} q_1 & p_2 \\ p_1 & q_2 \end{pmatrix}$$

For both of these situations, closed-form solutions can be found for the eigenvalues and eigenvectors. The eigenvalues for the two cases are $(1, p_1+p_2-1)$, and $(1, 1-p_1-p_2)$, respectively. In both cases, there is only a single eigenvalue of unit magnitude, so each situation has a single asymptotic distribution given by $\mathbf{v}^{\infty}=\mathbf{R}_1$ for any choice of initial density $\mathbf{v}^0$. This means that it does not matter which player has the initial break in the match; the long-run winner is determined only by the two probability parameters $p_1$ and

$p_2$ and by the match type, WB, LB, or AB. For the WB and LB cases, the $\mathbf{R}_1$ eigenvectors are, respectively,

$$\mathbf{R}_1^{WB} = \begin{pmatrix} \frac{q_2}{q_1+q_2} \\ \frac{q_1}{q_1+q_2} \end{pmatrix}, \text{ and } \mathbf{R}_1^{LB} = \begin{pmatrix} \frac{p_2}{p_1+p_2} \\ \frac{p_1}{p_1+p_2} \end{pmatrix}$$

In the WB case, player-1 wins a fraction of games corresponding to $W^{WB}=R_{11}{}^{WB}= q_2/(q_1+q_2)$; in the LB case, the fraction of games won by player-1 is $W^{LB}=R_{21}{}^{LB}= p_1/(p_1+p_2)$.

In the alternating-break situation, player-1 breaks half of the games, and proceeds to win the fraction $p_1$ of these, and player-2 breaks the other half of the games, and player-1 wins the fraction $q_2$ of these games. The total fraction of games won by player-1 is therefore $(p_1+q_2)/2$ in the alternating break situation.

Contour plots of the winning probabilities $W^{WB}$, $W^{LB}$ and $W^{AB}$ determined from P5.24 are shown as a function of $p_1$ and $p_2$ in Fig. 5.6. It is surprising how different these plots appear. It may be verified that the critical values of $p_1$ and $p_2$ that correspond to $W=½$ are the same in the WB, LB, and AB matches, namely $W=½$ when $p_1=p_2$ in all three cases; this means that the break choice is not expected to change the eventual winner, provided a large number of games are played in the match; however the margin by which the winner is expected to win can depend in a significant way on the match format due to the interplay between the positive and negative feedback effects. This means that in a match that is handicapped by payout stakes can depend in a significant way on the break choice. When $p_1=q_2$, then the player-1 game probability does not depend on which player breaks, and there is no break advantage or disadvantage; this relation is equivalent to $p_1+p_2=1$, and it may be verified that $W^{WB}=W^{LB}=W^{AB}=p_1$ in all three match situations when this condition is satisfied.

A contour plot of the difference probability, $W^{diff}=W^{WB}-W^{LB}$, is also shown in Fig. 5.6. The solid positive contour lines correspond to the situations in which player-1 has the best chance in a WB match, and the dashed negative contour lines correspond to the situations in which player-1 has the best chance in a LB match. It is clear in this figure that there can be significant differences in the outcomes of the WB and LB matches. In most situations, the individual probabilities are expected to be close to 0.5 for both players, and the difference contour plot in Fig. 5.6 shows that the break choice makes only a small difference in the outcome in these situations. However, the most drastic differences occur when one of the players has a strong break advantage or disadvantage. Table 5.4 gives the player-1 winning probabilities for a few selected values of $p_1$ and $p_2$ that demonstrate these trends.

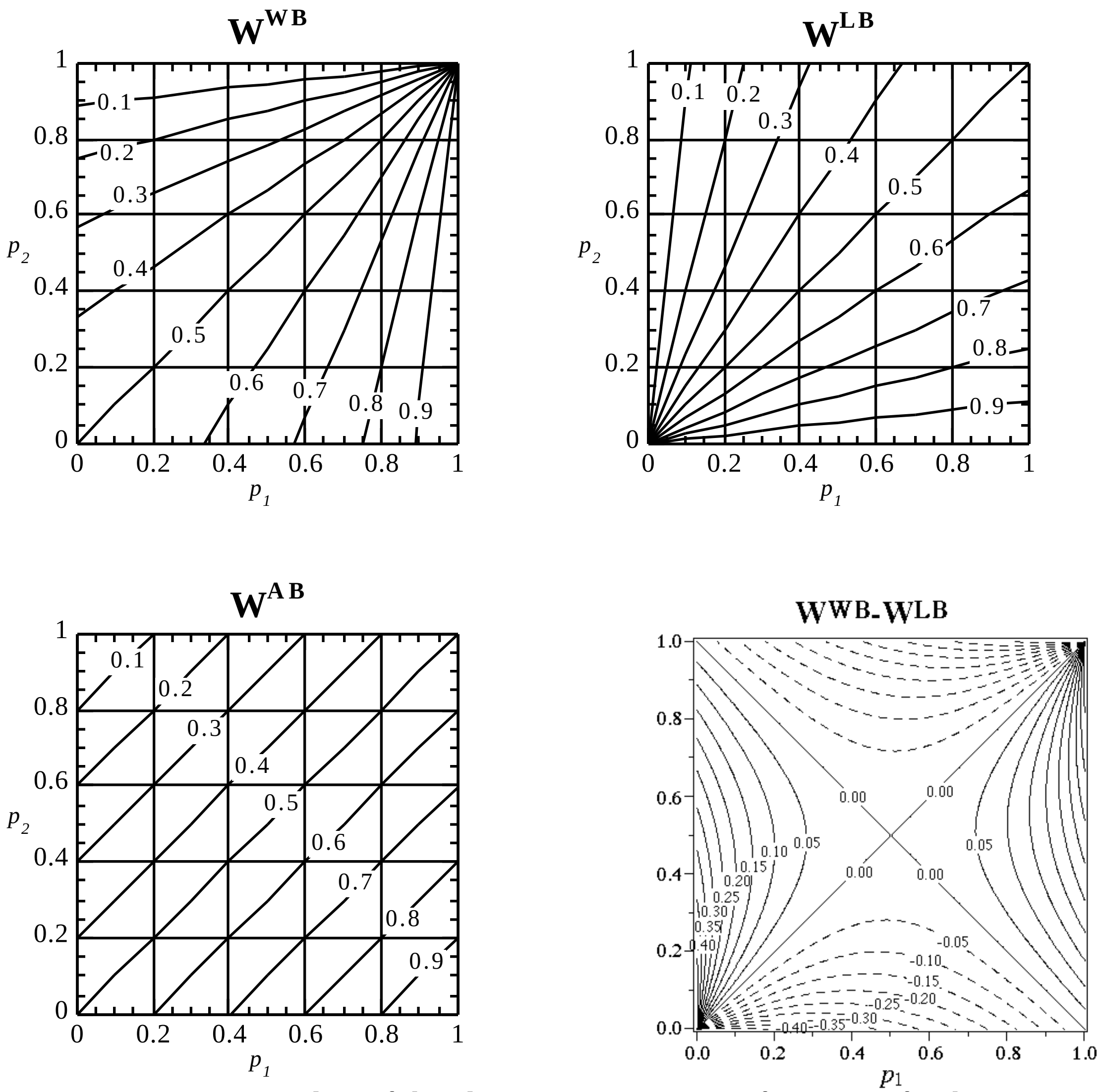



Fig. 5.6. Contour plots of the player-1 winning game fraction *W* for three types of matches: winner-breaks, loser-breaks, and alternating breaks. The $p_1$ parameter is the fraction of games that player-1 wins when player-1 breaks, and $p_2$ is the fraction of games that player-2 wins when player-2 breaks. When the contour lines are closely spaced, then the winning fraction *W* is very sensitive to small changes in the parameters $p_1$ and $p_2$.

The first three rows of Table 5.4 correspond to equally matched players. In the first row, both players might be weak runout players, or weak tactical players, who tend to lose most of the games that they break, and both players have a break disadvantage; in the second row both players win the same percentage of games that they lose when they break, and neither player has a break advantage or a break disadvantage; in the third row, both players might be strong runout players, or strong tactical players who tend to control the table once they get a shot, and the break is an advantage that can be exploited by both players. Matches between equal-strength players are expected to be even in WB, LB, and

AB situations, as seen in the first three rows. The next two rows show the expected result from a strong mismatch; in the fourth row player-1 is the underdog, and in the fifth row he is the strong favorite. Since $p_1+p_2=1$ in these two cases, there is no break advantage and the winning expectations are equal for all three match situations.

Table 5.4. Comparison of WB, LB, and AB winning probabilities.

| *Row* | $p_1$ | $p_2$ | $W^{WB}$ | $W^{LB}$ | $W^{AB}$ |
|---|---|---|---|---|---|
| 1 | .1 | .1 | .5 | .5 | .5 |
| 2 | .5 | .5 | .5 | .5 | .5 |
| 3 | .9 | .9 | .5 | .5 | .5 |
| 4 | .1 | .9 | .1 | .1 | .1 |
| 5 | .9 | .1 | .9 | .9 | .9 |
| 6 | .5 | .1 | .64 | <u>.83</u> | .7 |
| 7 | .5 | .9 | .17 | <u>.36</u> | .3 |
| 8 | .9 | .5 | <u>.83</u> | .64 | .7 |
| 9 | .1 | .5 | <u>.36</u> | .17 | .3 |
| 10 | .95 | .94 | <u>.54</u> | .50 | .505 |

Rows six and seven are two cases in which moderate mismatches occur and in which player-1 is the medium-strength player. In row six, he is favored to win over a weak opponent in both WB and LB match situations, but he is expected to win a much higher fraction of games in the LB match than in the WB match. This is because in the LB situation, player-1 can win one game and then take advantage of his opponent's breakshot weakness immediately in the next game by forcing him to break. The AB win fraction is between those of the WB and LB, and this trend holds for all combinations of the parameters $p_1$ and $p_2$. In row seven, player-1 is a medium strength player playing against a strong player; he is expected to lose in all three types of match situations, but his game percentage is about twice as large in the LB match as in the WB match. This is because he breaks more often than his stronger opponent in the LB situation, and although he does not benefit particularly from his own breaks, he keeps his opponent from exploiting his break advantage. The last two rows show the same types of mismatches as rows six and seven, but with the assumption that player-1 is the strong player (row eight) or the weak player (row nine) against a medium strength player; in both cases, a WB match situation is most beneficial to player-1. In row eight, player-1 benefits in the WB situation by exploiting his break advantage. In row nine, player-1 is actually penalized by being forced to break, and he breaks fewer times in the WB situation than in the LB situation which helps him limit his loses.

Row ten shows the expected results for two strong players who are closely, but not exactly, matched. Player-1 has a very slight 1% win-while-breaking game probability advantage over that of his opponent. It is interesting that in the WB match, this small advantage is magnified into a 4% difference in the expected game fraction, whereas in the LB match, the effect of this small advantage is almost eliminated in the expected game fraction. In the AB match, the 1% $p$ advantages gets diluted to a 0.5% $W$ advantage. The amplification of small differences of WB matches is a consequence of the clustering of the contour lines in the upper right corner of the $W^{WB}$ graph in Fig. 5.6. Similarly, the damping out of such differences in the LB situation is a consequence of the wide spacing of the contour lines in the upper right corner of the $W^{LB}$ graph. When player-1 has a slight breakshot advantage over his strong opponent, then he should prefer the WB situation, but when he has a slight breakshot disadvantage compared to his strong opponent, then he should prefer the LB situation.

**Problem 5.25**. Two players play a stakes-handicapped match in which player-1 wins 1.0 points for each game that he wins and he loses 2.0 points for each game that he loses. The win-while-break percentages for the two opponents are $p_1$=0.9 and $p_2$=0.5. What is the expected outcome for WB, LB, and AB matches?
*Answer*: The expectation of return $R$ by player-1 for each game is given by

$$R = W\ Z_W - L\ Z_L$$

where $W$ is the probability of winning each game, $Z_W$ is the number of points that he wins for each of these games, $L$=(1-$W$) is the losing game probability and $Z_L$ is the number of points that he loses. Using the results from Table 5.4, it is seen that
$R^{WB}$=(.83)(1.0)-(.17)(2.0)=0.49, $R^{LB}$=(.64)(1.0)-(.36)(2.0)=-0.08, and
$R^{AB}$=(.70)(1.0)-(.30)(2.0)=0.10. Player-1 is expected to win in the WB match and AB match situation, but he is expected to lose in the LB match situation.

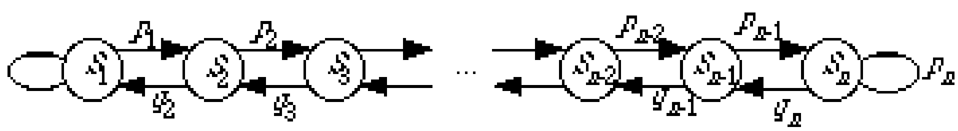


Fig. 5.7. The general game diagram for a progressive drill has $n$ nodes and $n$ independent probability parameters.

A progressive practice drill has a difficulty parameter that can be increased or decreased. A difficulty parameter might be a shot angle, or a shot distance, or some cue ball position goal, the number of object balls, or some combination of such parameters. In a progressive practice drill, when the player succeeds at one level of difficulty, then he is rewarded by being allowed to attempt the next level of difficulty; when the player fails at a level of difficulty, then he is penalized by being forced back to the previous level. Suppose that there are $n$ levels of difficulty, numbered 1…$n$. Failure at the first level means that the player attempts that level again, and success at the $n^{th}$ level means that level-$n$ is attempted again. Suppose that the probability of success at the $i^{th}$ level is denoted $p_i$, and the failure probability is therefore $q_i$=1-$p_i$. The game diagram for a general progressive practice drill is shown in Fig. 5.7. The Markov transition matrix for such a progressive drill has the form

$$\mathbf{M} = \begin{pmatrix} q_1 & q_2 & 0 & \cdots & 0 \\ p_1 & 0 & \ddots & \cdots & \vdots \\ 0 & p_2 & \ddots & q_{n-1} & 0 \\ \vdots & \cdots & \ddots & 0 & q_n \\ 0 & \cdots & 0 & p_{n-1} & p_n \end{pmatrix}$$

This matrix form is called a tridiagonal matrix because the only nonzero elements occur in the diagonal or in the elements adjacent to the diagonal. A property of such a tridiagonal matrix is that for $0<p_i<1$, there are no repeated eigenvalues; in particular, there is a single eigenvalue of unity, and therefore there is a single stable probability distribution for a progressive drill. If a practice drill is performed for a large number of steps, then this unique distribution will be approached in the stochastic limit, and the statistical parameters associated with this distribution can be used to assess the player's performance at the drill.

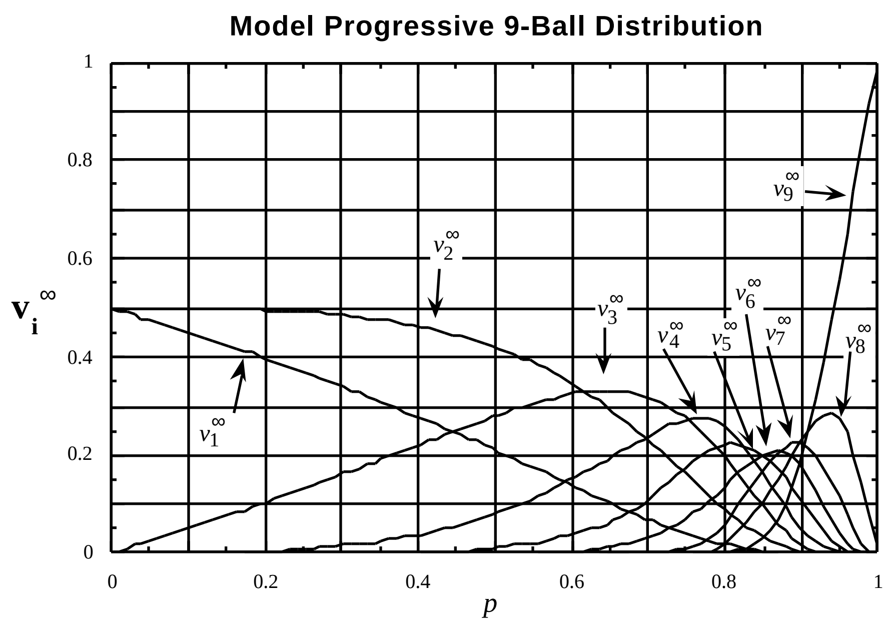



Fig. 5.8. The components of the stochastic limit distribution $\mathbf{v}^\infty$ are plotted as a function of the shot-success parameter $p$ for the model progressive 9-ball drill with $p_i=p^{(i-1)}$. The distribution changes significantly as a function of $p$, and this means that the distribution is a sensitive measure of performance.

**Problem 5.26**: In the progressive 9-ball drill, the player starts by throwing the 9-ball randomly on the table, taking the cue ball in hand, and shooting the 9-ball. Upon success, the 8-ball and the 9-ball are thrown on the table and the player attempts to run both balls from ball in hand. In general, a successful run of $i$ balls means that a run of $i$+1 balls is attempted, and a failure at a run of *i* balls means that, on the next turn, a run of *i*-1 balls is attempted. Assume that the probability of success for *i* balls is $p_i=p^{(i-1)}$ where *p* is an average probability of making an individual shot. What is the expected distribution for *p* equal to 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, and 0.99? What are the mean, median, mode, and standard deviations for each of these distributions?

*Answer*: The expected distributions are the stochastic limit $\mathbf{v}^\infty$ which is determined from the right eigenvector $\mathbf{R}_1$ of the matrix $\mathbf{M}^\infty$ associated with the eigenvalue $\lambda=1$, and scaled so that the elements total to unit probability. The coefficients of $\mathbf{R}_1$ are plotted as a function of the shot success parameter *p* in Fig. 5.8. These probability distributions for the specific values of *p* are given in Table 5.5, along with the associated statistical parameters. Because the distribution changes significantly with small changes in *p*, it provides a sensitive assessment of performance.

Table 5.5. Model Progressive 9-Ball Drill Statistics

| $p$ | $d_1$ | $d_2$ | $d_3$ | $d_4$ | $d_5$ | $d_6$ | $d_7$ | $d_8$ | $d_9$ | $\bar{x}$ | *Mode* | $\tilde{x}$ | $\sigma$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| .5 | .210 | .419 | .280 | .080 | .011 | .001 | .000 | .000 | .000 | 2.264 | 2 | 2 | .920 |
| .6 | .138 | .346 | .324 | .149 | .037 | .005 | .000 | .000 | .000 | 2.618 | 2 | 3 | 1.051 |
| .7 | .069 | .231 | .318 | .237 | .107 | .031 | .006 | .001 | .000 | 3.201 | 3 | 3 | 1.236 |
| .8 | .018 | .088 | .195 | .256 | .222 | .135 | .060 | .020 | .005 | 4.354 | 4 | 4 | 1.526 |
| .9 | .000 | .004 | .017 | .052 | .110 | .177 | .223 | .227 | .190 | 6.964 | 8 | 7 | 1.567 |
| .95 | .000 | .000 | .000 | .002 | .011 | .038 | .112 | .272 | .565 | 8.333 | 9 | 9 | .928 |
| .99 | .000 | .000 | .000 | .000 | .000 | .000 | .005 | .076 | .918 | 8.912 | 9 | 9 | .306 |

**Exercise 5.1**. Practice the progressive 9-ball drill over an extended period of time and accumulate the data for the number of successes and number of failures at each level. From these data, an empirical value of the success probability $p_i$ at each level can be estimated. Use these empirical values and determine the corresponding stochastic limit distribution. Compare this hypothetical distribution to the actual distribution, which consists of the total attempts (successes+failures) at each level. If there are significant differences, then this shows where the most significant improvements in performance are possible. For example, if there seem to be too many small-*i* attempts, then additional focus may be needed for these “easy” cases, or there may be some intimidation on the long run attempts.

Almost any kind of shot or game situation may be turned into a progressive drill and subjected to this kind of stochastic analysis. For example, for 8-ball, the player might throw out an equal number of stripes and solids along with the 8-ball, take ball in hand, and attempt to run out. Upon success, one more ball of each type is thrown out at the beginning, until all 15 balls are initially on the table.

In the National Pool League (NPL) handicap system (see, for example, `http://www.accessone.com/~mavlon/handicap.html`), each player has a numerical skill rating estimate. If $R_1$ and $R_2$ are the skill ratings for two opponents, then the probability $p$ that player-1 will win an individual game is assumed to be given by

$$p = \frac{1}{1 + 2^{-(R_1 - R_2)/30}}$$

or, equivalently, the rating difference between two opponents satisfies the relation

$$R_1 - R_2 = \frac{30}{\log 2} \log\left(\frac{p}{1-p}\right)$$

Skill ratings range from about 20 for beginners to around 80 for experienced amateur players to over 130 for professional-level players. Each additional rating difference of 30 points results in another factor of two in the ratio of game probabilities $p/q$. Matchups are chosen based on the analysis in P5.5. In general, for a given $p_1$, The match probability $W(p;m,n)$ is determined for values of $m+n$ that are reasonable for tournament play, and the combination that gives the match probability closest to $W$=0.5 is chosen. The following table contains four sets of matchups. Chart-8 is used for short matches when the time for each match needs to be minimized, Chart-10 is used for regular length matches, and Chart-12 is used when longer matches can be played. In some situations, short charts are used for lower-rated players and longer charts are used for higher-rated players. Longer and shorter charts than those shown here may also be used in particular league or tournament situations. Chart-20 is a very long match chart and is included for comparison purposes. When a player wins a match in the NPL system, his skill rating increases by a point, and when a player loses a match his skill rating decreases. Because of this adjustment, the skill rating estimate tends to fluctuate somewhat about a mean value that reflects the player's true skill rating. The skill rating value may be used to label the states in a game diagram, and because transitions are allowed only between nearby states, the game diagram for the NPL handicap system is the same as for a progressive drill as shown in Fig. 5.7, and the corresponding Markov transition matrix is tridiagonal.

In order to perform a stochastic simulation of the NPL handicap system, it is useful to introduce a few simplifying approximations. It is assumed that a particular player of interest, player-1, has a true skill that corresponds to a skill rating of $R^{Actual}$. He plays against an infinite number of opponents, all of whom have skill ratings that also correspond to $R^{Opponent}=R^{Actual}$. As player-1 plays against these opponents, his skill rating estimate will fluctuate about $R^{Actual}$. At any particular time player-1's apparent skill rating will be denoted $R^{Apparent}$. It is $R^{Apparent}$ that is used to determine the game matchup, using the charts in Table 5.6, but the actual game probability is determined by $R^{Actual}$-$R^{Opponent}$=0. For example, suppose that a tournament is using Chart-10, and the apparent rating difference is 5 points, which means that the matchup is 5:5. $W(0.5;5,5)$=0.5 then defines the transition probability for player-1 to advance to the next higher skill rating, and (1-$W(0.5;5,5)$)=0.5 defines the probability for the player to fall back to the next lower skill rating. If player-1 wins this match, the apparent skill rating difference for his next match will be 6 points and the next matchup will be 5:4. From P5.6 it is seen that $W(0.5;5,4)=0.363$. Player-1 is now overrated and is more likely to lose this match than to win it. Similarly, when player-1 loses enough matches his apparent rating will be 6 points too low, the match probability will be $W(0.5;4,5)=0.637$. At this point, player-1 is underrated and is more likely to win than to lose. According to this mechanism, the player has a tendency to fluctuate about his true skill rating; if his apparent rating gets too low there is a tendency for him to start winning a majority of his matches and for his rating to adjust up back to its correct level, and if his apparent rating gets too high there is a tendency to lose a majority of his matches and for his rating to adjust back down to its correct level.

Table 5.6. Examples of four charts used in the NPL handicap system.

**Chart-10**

| Rating Difference | Match Games |
|---|---|
| 0-5 | 5:5 |
| 6-14 | 5:4 |
| 15-21 | 6:4 |
| 22-28 | 5:3 |
| 29-36 | 6:3 |
| 37-46 | 7:3 |
| 47-56 | 6:2 |
| 57-up | 7:2 |

**Chart-12**

| Rating Difference | Match Games |
|---|---|
| 0-4 | 6:6 |
| 5-11 | 6:5 |
| 12-17 | 7:5 |
| 18-22 | 6:4 |
| 23-28 | 7:4 |
| 29-35 | 8:4 |
| 36-42 | 7:3 |
| 43-48 | 8:3 |
| 49-58 | 9:3 |
| 59-68 | 8:2 |
| 69-up | 9:2 |

**Chart-8**

| Rating Difference | Match Games |
|---|---|
| 0-6 | 4:4 |
| 7-18 | 4:3 |
| 19-29 | 5:3 |
| 30-39 | 4:2 |
| 40-48 | 5:2 |
| 49-up | 6:2 |

**Chart-20**

| Rating Difference | Match Games |
|---|---|
| 0-2 | 10:10 |
| 3-7 | 10:9 |
| 8-12 | 10:8 |
| 13-17 | 11:8 |
| 18-22 | 11:7 |
| 23-27 | 12:7 |
| 28-33 | 12:6 |
| 34-36 | 13:6 |
| 37-40 | 14:6 |
| 41-45 | 13:5 |
| 46-51 | 14:5 |
| 52-59 | 14:4 |
| 60-68 | 16:4 |
| 69-75 | 15:3 |
| 76-77 | 16:3 |
| 78-87 | 17:3 |
| 88-97 | 16:2 |
| 98-100 | 17:2 |
| 101-up | 18:2 |

Fig. 5.9 shows the stochastic distributions $\mathbf{v}^{\infty}$ for the four charts in Table 5.6 with the above assumptions. These are called the natural distributions because they depend only on the granularity introduced by the matchups. In general, it is seen that there are

two components to the widths of a given distribution. One component is the flat region at the top which is due to the W=0.5 transition probability for near-zero apparent rating differences. This flat region is wider for the shorter-match charts than for the longer-match charts. The other component of the width is the falloff that is induced by the matchup differences. This component would occur even for longer matches than shown in Fig. 5.9, but in general the falloff is more rapid for longer matches than for shorter matches, as discussed in P5.7. The Chart-20 distribution displays both characteristics of the long-matches: a narrow flat region and a rapid falloff.

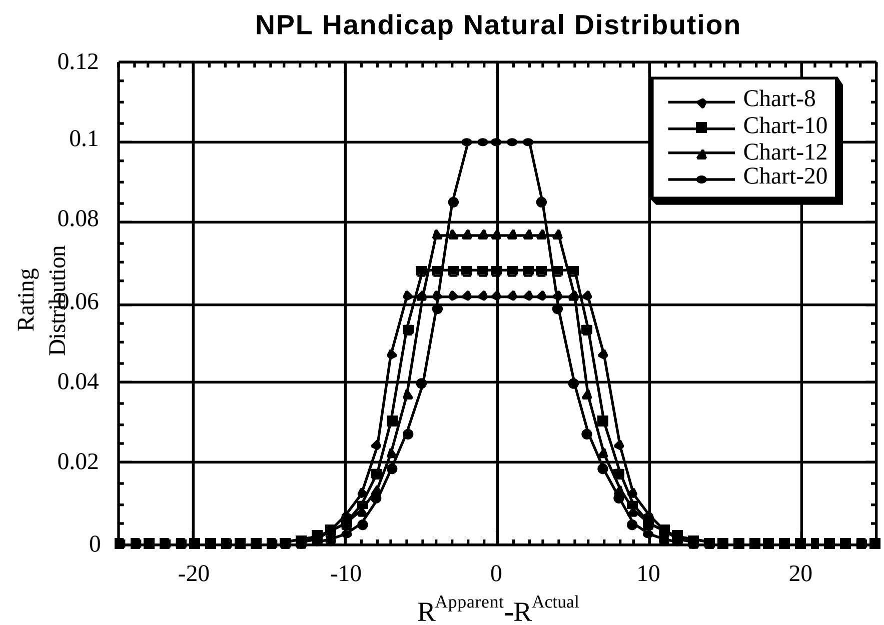



Fig. 5.9. The natural distributions $\mathbf{v}^\infty$ of rating variations are shown for four NPL charts. The variations away from the correct rating are relatively small in all cases, but the peak is sharper generally for longer matches than for shorter matches.

An improved simulation can be achieved by relaxing some of the simplifying approximations used in the preceding stochastic analysis. One approximation is that the opponents' skill ratings are estimated exactly. It can be assumed that there is a probability distribution $\{d_j\}$ of actual skill ratings for the opponents. Player-1 has an actual skill rating of $R^{Actual}$ and an apparent skill rating of $R_i$ ; the opponent has an actual skill rating of $R_j$ , occurring with probability $d_j$, and an apparent skill rating of $R^{Actual}$. The probability for player-1 to win a match is then given by the expression

$$W_i = \sum_j d_j W\left(p_j; m_i, n_i\right)$$

in which $p_j$ is determined by the actual skill difference ($R^{Actual}$-$R_j$) and the matchup $m_i$:$n_i$ is determined by the apparent skill difference ($R_i$-$R^{Actual}$). This kind of expression involving summations of probability distributions is called a *convolution*. The question then arises as to what opponent distribution $\{d_j\}$ should be used. An obvious answer is to use the same distribution for both the opponents and for player-1. This distribution is determined in a self-consistent manner. Some reasonable approximation for $\{d_j\}$is assumed, and the corresponding stochastic distribution $\mathbf{v}^{\infty}$ for player-1 is determined from the eigenvalue analysis of the transition matrix. This stochastic distribution then defines a new $\{d_j\}$, which then results in a new transition matrix, which then results in a new stochastic distribution. After a few cycles of this process, the input distribution $\{d_j\}$ converges to the same as the output stochastic distribution $\mathbf{v}^{\infty}$, and self-consistency is achieved. This process in which the stochastic distribution depends on itself is called *autocorrelation*. In the case of the NPL handicap distributions, this has a very small effect, too small to notice the difference when plotted as in Fig. 5.9. The standard deviation for the Chart-8 distribution widens from 4.963 for the natural distribution to 4.978 with autocorrelation, for Chart-10 it widens from 4.634 to 4.654, for Chart-12 it widens from 4.269 to 4.291, and for Chart-20 it widens from 3.668 to 3.697. Further improvements in the simulation require additional assumptions about the distribution of actual and apparent skill ratings for the opponents, and about the day-to-day and match-to-match fluctuations of actual skill that all players display. In general, all of these effects tend to smooth and widen the stochastic distributions compared to the natural distributions shown in Fig. 5.9 and to the autocorrelated distributions described above. For the efficient numerical treatment of the convolution of several distribution variables, methods based on Fourier transforms are usually employed.

In addition to the matches of limited length that have been analyzed previously in this section, another common type of match is the *n-ahead* match. The players keep playing games until one of them manages to get *n* games ahead of the other player, and this terminates the match. It is also possible to handicap such a match, so that one of the players needs *m* games ahead to win, while the other player needs *n* games. The game diagram for a general handicapped *n*-ahead match is shown in Fig. 5.10. For a match of this type handicapped at *m:n*, the game graph has ($m$+$n$+1) nodes, *m* of which are on one side of the starting node *S*0, and *n* of which are on the other. It is assumed that the probability of winning an individual game is independent of the score, and for simplicity it is assumed that the breaker of each game does not affect the game probability, although it is straightforward to incorporate differing probabilities for these situations if such data is available. The Markov transition matrix for the *n*-ahead game always has two eigenvalues of unity; this may be verified by expanding the secular equation in cofactors and minors first along the first column (corresponding to the losing node *L*), and then along the last column (corresponding to the winning node *W*), exposing two ($1-\lambda$) factors in the characteristic polynomial. The following problem shows these general features for a specific game diagram, but the general approach can be applied to any *n*-ahead type match situation.

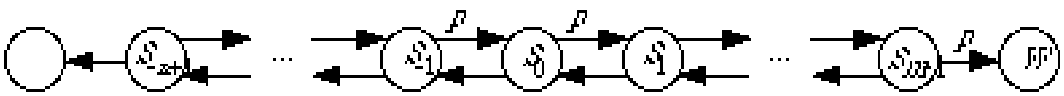


Fig. 5.10. The game diagram for a general handicapped $n$-ahead type match is shown. From the starting node $S_0$, player-1 needs $m$ games to win the match and player-2 needs $n$ games.

**Problem 5.27**: Two players play a handicapped 3-ahead type match. Compute the player-1 match probability $W$ as a function of $p$, the probability of winning an individual game, if the match is handicapped at 1:5, 2:4, 3:3, 4:2, and 5:1.

*Answer*: There are 7 nodes in the game diagrams for all of these cases, and the Markov transition matrix **M** for all of these cases is given by

$$\mathbf{M} = \begin{pmatrix} 1 & q & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & q & 0 & 0 & 0 & 0 \\ 0 & p & 0 & q & 0 & 0 & 0 \\ 0 & 0 & p & 0 & q & 0 & 0 \\ 0 & 0 & 0 & p & 0 & q & 0 \\ 0 & 0 & 0 & 0 & p & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & p & 1 \end{pmatrix}$$

The different handicaps correspond to different choices for the initial probability. The 1:5 match corresponds to the vector $\mathbf{v}^0=(0,0,0,0,0,1,0)^T$, the 2:4 match corresponds to $\mathbf{v}^0=(0,0,0,0,1,0,0)^T$, the 3:3 match corresponds to $\mathbf{v}^0=(0,0,0,1,0,0,0)^T$, the 4:2 match corresponds to $\mathbf{v}^0=(0,0,1,0,0,0,0)^T$, and the 5:1 match corresponds to $\mathbf{v}^0=(0,1,0,0,0,0,0)^T$. The eigenvectors and eigenvalues of this matrix can be determined in closed form. The stochastic limit for this match situation is determined from

$$\mathbf{M}^\infty = \begin{pmatrix} 1 & \frac{(1-pq)(1-3pq)-p^5}{(1-pq)(1-3pq)} & \frac{(1-pq)(1-3pq)-p^4}{(1-pq)(1-3pq)} & \frac{q^3}{(1-3pq)} & \frac{q^4}{(1-pq)(1-3pq)} & \frac{q^5}{(1-pq)(1-3pq)} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{p^5}{(1-pq)(1-3pq)} & \frac{p^4}{(1-pq)(1-3pq)} & \frac{p^3}{(1-3pq)} & \frac{(1-pq)(1-3pq)-q^4}{(1-pq)(1-3pq)} & \frac{(1-pq)(1-3pq)-q^5}{(1-pq)(1-3pq)} & 1 \end{pmatrix}$$

from which it is seen that the match probability for the various cases are given by relatively simple ratios of polynomials. These probabilities are plotted as a function of $p$ for the various match situations in Fig. 5.11. Note that Fig. 5.11 could have been determined numerically even if closed-form expressions for $\mathbf{M}^\infty$ were not available.

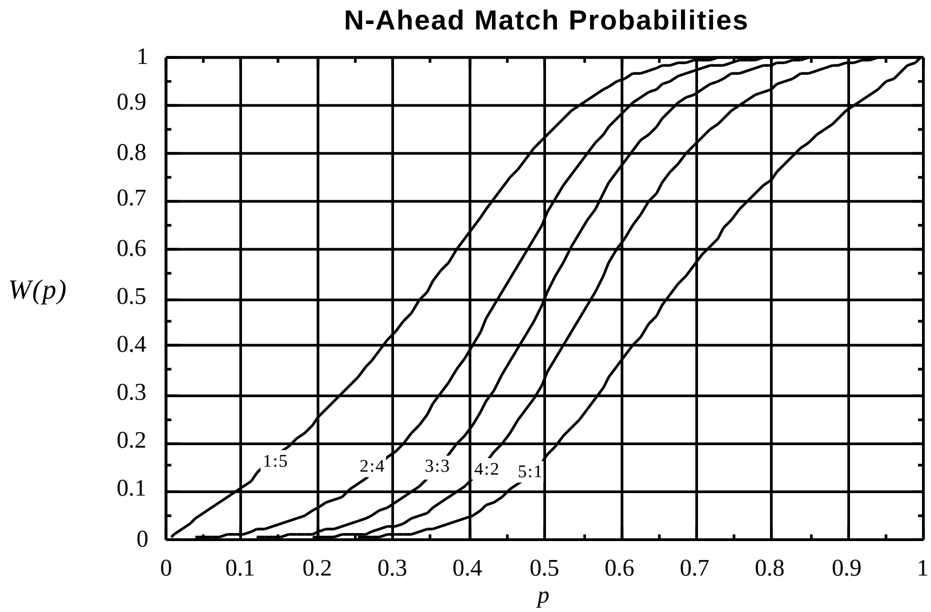



Fig. 5.11. Match probability *W* as a function of the player-1 game probability *p* for *n*-ahead matches handicapped at 1:5, 2:4, 3:3, 4:2, and 5:1.

*Acknowledgments*: Much of this material has been compiled over a long period of time. The author first became interested in the physics of pool during a college physics course (not an uncommon occurrence). Some more recent material has been added as a result of ongoing discussions in the newsgroup `rec.sport.billiard` involving many participants. This newsgroup is highly recommended to anyone interested in discussions involving the various aspects of pool and billiards games.

*Further Reading*: Considering that many important and interesting aspects of pool and billiards may be understood with only simple application of classical physics, and that quite useful results can be obtained even with rather crude approximations, there has been traditionally relatively little physics included in most instructional pool books. Simple physics problems involving pool balls are often included in problem sets in physics text books, but these are not discussed usually in the context of using the results in actual play, but rather as a device to teach a physical principle or in the application of an analytic method. Some of the exceptions to this trend are the regular columns by Bob Jewett in *Billiards Digest*. Another good publication is the book "*The Physics of Pocket Billiards*" by W. C. Marlow. While this present manuscript concentrates mostly on theoretical relations combined with practice exercises, Marlow's book includes descriptions of experimental setups to measure tip-ball contact times, ball-ball contact times, various coefficients of friction, and many other interesting things.