1. Let $a, b, c, d \geq 0$, with $a + b + c + d = 1$. Suppose $X$ and $Y$ are random variables taking values in $\{0, 1\}$ with joint probability function $f_{X,Y}(x, y) = P_{xy}$, where $P = \begin{pmatrix} 0 & 1 \\ a & b \\ c & d \end{pmatrix}$, with rows and columns indexed by $\{0, 1\}$ as shown.

a. [10 points] What is $E[X]$? What is $\text{Var}[X]$?

$E[X] = \Pr(X = 1) = c + d.$

$\text{Var}[X] = E[X^2] - E[X]^2 = c + d - (c + d)^2 = (1 - (c + d))(c + d) = (a + b)(c + d).$

b. [7 points] What is $E[XY]$?

$E[XY] = \Pr(XY = 1) = \Pr(X = 1 \land Y = 1) = d.$

c. [8 points] What is $\text{Corr}[X, Y]$?

$$\text{Corr}[X, Y] = \frac{\text{Cov}[X, Y]}{\text{Var}[X]\text{Var}[Y]} = \frac{E[XY] - E[X]E[Y]}{\sqrt{(a + b)(c + d)(a + c)(b + d)}}$$

$$= \frac{d - (c + d)(b + d)}{\sqrt{(a + b)(c + d)(a + c)(b + d)}}$$

$$= \frac{d(a + b + c + d - (c + d)(b + d))}{\sqrt{(a + b)(c + d)(a + c)(b + d)}}$$

$$= \frac{ad - bc}{\sqrt{(a + b)(c + d)(a + c)(b + d)}}.$$

d. **Extra Credit** [10 points] Show that $-1 \leq \text{Corr}[X, Y] \leq 1$.

This pair of inequalities, together with the definition of correlation as

$$\text{Corr}[X, Y] = \frac{E[(X - \bar{X})(Y - \bar{Y})]}{\sqrt{E[(X - \bar{X})^2]E[(Y - \bar{Y})^2]}},$$

where $\bar{X} = E[X]$ and $\bar{Y} = E[Y]$, should remind you of the dot product of vectors, which satisfies

$$\cos \theta = \frac{x \cdot y}{\sqrt{(x \cdot x)(y \cdot y)}},$$

where $\theta$ is the angle between the vectors $x$ and $y$, and cosine, of course, lies between $-1$ and $1$. So we start by defining two new random variables, $U = X - \bar{X}$ and $V = Y - \bar{Y}$, which take values $0 - (c + d) = -(c + d)$ and $1 - (c + d) = a + b$, and $0 - (b + d) = -(b + d)$ and $1 - (b + d) = a + c$, respectively (using part 1.a), and have joint probability distribution

$$
\begin{pmatrix}
-(c+d) & a+c \\
-(b+d) & a \\
0 & b \\
& c \\
& d \\
\end{pmatrix}.$$
Next we ask what vectors would have dot product equal to

\[ E[UV] = \sum_{u,v} uv \Pr(U = u \land V = v). \]

Since there are four terms in this sum, we should consider vectors in \( \mathbb{R}^4 \), with one component for each outcome. And the components should be the values the random variables take, weighted by the probabilities. Thus we set

\[
x = (-c + d)\sqrt{a}, \quad -(c + d)\sqrt{b}, \quad (a + b)\sqrt{c}, \quad (a + b)\sqrt{d} \\
y = -(b + d)\sqrt{a}, \quad (a + c)\sqrt{b}, \quad -(b + d)\sqrt{c}, \quad (a + c)\sqrt{d}.
\]

Now just compute:

\[
x \cdot x = (c + d)^2(a + b) + (a + b)^2(c + d) = (a + b)(c + d)(a + b + c + d) = (a + b)(c + d) \\
y \cdot y = (b + d)^2(a + c) + (a + c)^2(b + d) = (a + c)(b + d)(a + b + c + d) = (a + c)(b + d), \\
x \cdot y = (c + d)(b + d)a - (c + d)(a + c)b - (a + b)(b + d)c + (a + b)(a + c)d = ad - bc
\]

which gives exactly \( \text{Corr}[X, Y] \) as computed in part 1.c, and satisfies the inequalities because of (*).

There is another, more abstract proof that you may remember from when you learned the Cauchy-Schwartz inequality in linear algebra. For any two real-valued random variables with means 0, \( e.g. \), \( U \) and \( V \), notice that for all \( t \in \mathbb{R} \),

\[
0 \leq E[(U + tV)^2] \\
= E[U^2 + 2tUV + t^2V^2] \\
= \text{Var}[U] + 2t\text{Cov}[U, V] + t^2\text{Var}[V].
\]

The last expression is minimal at \( t = -\frac{\text{Cov}[U, V]}{\text{Var}[V]} \), at which we get

\[
0 \leq \text{Var}[U] - 2\frac{\text{Cov}[U, V]^2}{\text{Var}[V]} + \frac{\text{Cov}[U, V]^2}{\text{Var}[V]},
\]

so \( 0 \leq \text{Var}[U]\text{Var}[V] - \text{Cov}[U, V]^2 \) and thus \( |\text{Cov}[U, V]| \leq \sqrt{\text{Var}[U]\text{Var}[V]} \).

2. Let \( K \) and \( W \) be real-valued random variables such that the probability density function for \( (W|K = k > 1) \) is a “power law distribution”:

\[
f_{W|K}(w|k) = \begin{cases} (k - 1)/w^k & \text{if } w \geq 1, \\ 0 & \text{if } w < 1; \end{cases}
\]

and the exponent \( K \) is uniformly distributed between 2 and 3:

\[
f_K(k) = \begin{cases} 1 & \text{if } 2 < k \leq 3, \\ 0 & \text{otherwise}. \end{cases}
\]
a. [15 points] What is \( f_W(w) \)?

\[
f_W(w) = \int f_{W|K}(w|k)f_K(k)dk = \int e^{-k \log w} \frac{k-1}{w^k} \frac{1}{w^k} \, dk = \int (ke^{-k \log w} - e^{-k \log w}) \, dk
\]

\[
= -\log w \int e^{-k \log w} \frac{k-1}{w^k} \, dk - \int e^{-k \log w} \, dk
\]

\[
= \left( -\frac{k}{w^k \log w} - \frac{1}{w^k (\log w)^2} + \frac{1}{w^k \log w} \right)^2 = \frac{(w-2) \log w + w - 1}{w^3 (\log w)^2}.
\]

b. [10 points] What is the largest term in \( f_W(w) \) as \( w \to \infty \)?

The largest term in the numerator is \( w \log w \), so we write

\[
f_W(w) \sim \frac{w \log w}{w^3 (\log w)^2} = \frac{1}{w^2 \log w} \quad \text{as} \quad w \to \infty.
\]

3. Suppose there is an urn (黑/白), initially containing finite numbers \( R_0 > 0 \) of red balls and \( G_0 > 0 \) of green balls. For \( t = 1, 2, \ldots \) a ball is removed at random, and is replaced with 2 balls of the same color.

a. [10 points] If \( R_0 = 1 = G_0 \), what is the probability that \( R_2 = 3 \)?

The number of balls increases by one at each step, so if \( R_2 = 3 \), a red ball must have been picked twice. Thus \( \Pr(R_2 = 3) = \Pr(R_3 = 3|R_2 = 2) \Pr(R_2 = 2) = \frac{2}{3} \cdot \frac{1}{3} = \frac{2}{9} \).

b. [10 points] Show that \( F_t = \frac{R_t}{R_t + G_t} \) is a Martingale.

First, \( \mathbb{E}[F_t] = \mathbb{E}[F_t] < 1 < \infty \) since \( R_t < R_t + G_t \). Second, if \( R_0 = 1 = G_0 \) then \( R_t + G_t = t + 1 \) since the number of balls increases by 1 at each step. Then

\[
\mathbb{E}[F_{t+1}|F_0, \ldots, F_t] = F_t \frac{F_t(t+1)}{t+2} + (1-F_t) \frac{F_t(t)}{t+2}
\]

\[
= \frac{F_t^2(t+1) + F_t F_t(t+1) - F_t^2(t+1)}{t+2} = F_t.
\]

c. [10 points] Suppose this is a gambling game where you win $1000 if there are ever three times as many red balls as green balls. How much would you be willing to pay to play this game? Why?

If there are three times as many red balls as green balls, then \( F_t = \frac{3}{4} \). Using the maximum Markov inequality, \( \Pr(\max F_t \geq \frac{3}{4}) \leq \frac{F_0/\frac{3}{4} = \frac{1/2}{3}} = \frac{2}{3} \). So if you pay any more than $1000\cdot \frac{2}{3}$ you should expect to lose money.

4. The joint distribution of a mother’s height and her daughter’s height is approximately bivariate normal with the mean of each marginal height distribution being 64 inches and the standard deviation 3 inches. Also, although mother’s and daughter’s heights are not exactly the same, they are positively correlated.
a. [10 points] If Devi is 62 inches tall, is her daughter Freya expected to be shorter than she is, or taller? Why?
Subtract 64 inches from their heights to get a bivariate normal distribution with means 0. Then Freya’s difference from average, conditioned on her mother’s difference from average, is normally distributed with mean $\rho(\sigma_F/\sigma_D)(62 - 64) = -2\rho > -2$, so Freya is expected to be taller than Devi.

b. [10 points] Suppose Devi is much taller than average, say 72 inches tall; is Freya expected to be shorter than she is, or taller? Why?
In this case Freya’s expected difference from average is $\rho(\sigma_F/\sigma_D)(72 - 64) = 8\rho < 8$, so Freya is expected to be shorter than Devi.

The phenomenon illustrated in this problem, namely that the variable being conditioned is closer to the mean than is the condition, is called “regression toward the mean”.