**SAMPLE PROBLEMS #1**

**Problem 1**

1. Let $X$ denotes the # of black balls
   \[ X \sim \text{hypergeometric distribution} \]
   \[ P(X = m) = \frac{\binom{m}{w} \binom{n-m}{x-w}}{\binom{n}{x}} \]

2. Let $M$ denotes amount of $\frac{x}{y}$ you get at a certain trial
   \[ EM = 0 \times P(x=0) + 2 \times P(x=1) + 4 \times P(x=2) + 6 \times P(x=3) + 28 \times P(x=4) \]
   \[ = 2x \times \left( \frac{\binom{4}{3}}{\binom{8}{3}} \right) + 4x \times \left( \frac{\binom{5}{2}}{\binom{8}{3}} \right) + 6x \times \left( \frac{\binom{6}{1}}{\binom{8}{3}} \right) + 28 \times \left( \frac{\binom{7}{0}}{\binom{8}{3}} \right) \]
   \[ = \frac{30}{7} \]

**Problem 2**

1. \( P(D) = \frac{1}{100} \) \( P(\neg D) = \frac{1}{100} \) \( P(H) = \frac{3}{100} \)
   \[ P(D+) = \frac{P(D) \cdot P(H)}{P(H) + P(\neg H)P(D)} \]
   \[ = \frac{\frac{1}{100} \cdot \frac{3}{100}}{\frac{3}{100} + \frac{99}{100} \cdot \frac{1}{100}} \]
   \[ = \frac{1}{3} \]

2. \( P(D|\text{twin}) = \frac{P(\text{twin} \mid D)P(D)}{P(D)} \)
   \[ = \frac{P(\text{twin} \mid D) \cdot P(D)}{(1 - P(\neg D))^2 \cdot P(D)} \]
   \[ = \frac{P(\neg D)^2 \cdot P(H) + P(\text{twin} \mid D)P(D)}{(1 - P(\neg D))^2 \cdot P(D)} \]
   \[ = \frac{(1 - \frac{1}{100})^2 \cdot \frac{3}{100} + (1 - P(\neg D))^2 \cdot \frac{1}{100}}{(1 - \frac{1}{100})^2 \cdot \frac{3}{100} + (1 - \frac{1}{100})^2 \cdot \frac{1}{100}} \]
   \[ = \frac{99}{103} \]
Problem 3

1) \[ \theta = \sin^{-1} x \quad \theta \in \left[ 0, \frac{\pi}{2} \right] \]

\[ P(0 \leq \theta \leq \frac{\pi}{2}) = P_x(\sin^{-1} x \leq \frac{\pi}{2}) = P_x(x \leq \frac{\pi}{2}) = \frac{1}{2} \]

2) \[ F_{\theta}(t) = P_{\theta}(\theta \leq t) = P_x(\sin^{-1} x \leq t) = P_x(x \leq \sin t) = \sin t \]

3) \[ f_{\theta}(t) = F_{\theta}'(t) = (\sin t)' = \cos t \]
7. Suppose \((X, Y)\) is uniform in the region \(\Delta\) bounded by the positive \(x\)-axis, the positive \(y\)-axis, and the line \(x + y = 2\).

1. Compute the marginal density of \(X\).

\[
f_X(x) = \int_0^{2-x} f_{X,Y}(x, y) dy = \int_0^{2-x} \frac{1}{2} dy = \frac{2-x}{2} \text{ for } x \in [0, 2].
\]

2. Compute \(F_{X+Y}(t)\), the c.d.f. of \(X + Y\). Deduce the p.d.f. of \(X + Y\).

Computation of CDF:

\[
F_{X+Y}(t) = \mathbb{P}(x + y \leq t) = \mathbb{P}(x \leq t, y \leq t - x) \\
= \int_0^t \int_0^{t-x} f_{X,Y}(x, y) dy dx = \int_0^t \int_0^{t-x} \frac{1}{2} dy dx \\
= \frac{1}{2} \int_0^t (t - x) dx = \frac{t^2}{4} \text{ for } t \in [0, 2].
\]

Computation of PDF:

Thus \(f_{X+Y}(t) = F'_{X+Y}(t) = t/2\) for \(t \in [0, 2]\).

3. Suppose \((X_1, Y_1), \ldots, (X_n, Y_n)\) are points chosen uniformly on \(\Delta\). Compute \(\mathbb{P}(\max_{i \leq n} \{X_i + Y_i\} \geq 1)\).

\[
\mathbb{P}(\max_{i \leq n} \{X_i + Y_i\} \geq 1) = 1 - \mathbb{P}(\max_{i \leq n} \{X_i + Y_i\} < 1)
\] (1)

\[
= 1 - \mathbb{P}(X_1 + Y_1 < 1, X_2 + Y_2 < 1, \ldots, X_n + Y_n < 1)
\] (2)

\[
= 1 - \prod_{i=1}^n \mathbb{P}(X_i + Y_i < 1)
\] (3)

\[
= 1 - \mathbb{P}(X_1 + Y_1 < 1)^n
\] (4)

\[
= 1 - \frac{1}{4^n}
\] (5)

Justifications:

(2) \(\rightarrow\) (3) by independence

(3) \(\rightarrow\) (4) by identical distribution

(4) \(\rightarrow\) (5) since \(F_{X+Y}(t)|_{t=1} = \frac{1}{4} |_{t=1} = \frac{1}{4}\)
(1) \( P(\text{All couples opposite sex}) = \frac{m}{2m-1} \cdot \frac{m-1}{2(m-1)-1} \cdot \frac{m-2}{2(m-2)-1} \cdots \frac{m-(m-1)}{2(m-(m-1))-1} \)

\[ = \frac{m!}{\prod_{k=1}^{m} [2(m-k)+1]} \]

(2) \( \mathbb{E}[X] = \sum_{i=1}^{m} \mathbb{E}[X_i] = \sum_{i=1}^{m} \frac{m}{2m-1} \]

\[ \mathbb{E}[X] = \frac{m^2}{2m-1} \]

(3) \( \text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 \)

\[ \mathbb{E}[X^2] = \mathbb{E}\left[ \sum_{i=1}^{m} X_i^2 \right] = \mathbb{E}\left( \sum_{i=1}^{m} X_i + \sum_{i \neq j} X_i X_j \right) \]

\[ = \sum_{i=1}^{m} \frac{m}{2m-1} + \sum_{i \neq j} \frac{m}{2m-1} \cdot \frac{m-1}{2m-3} \cdot m-1 \]

\[ = \frac{m^2}{2m-1} + \frac{(m^2-m)^2}{(2m-1)(2m-3)} \]

\( \text{Var}(X) = \frac{m^2}{2m-1} + \frac{(m^2-m)^2}{(2m-1)(2m-3)} - \frac{m^4}{(2m-1)^2} \)
(1) \[ f_{Y \mid X=x} = \frac{f_{X,Y}}{f_X(x)} \quad X \sim \text{uniform (0,30)} \]
\[ f_X(x) = \begin{cases} \frac{1}{10} & \text{if } x \in [0,30] \\ 0 & \text{otherwise} \end{cases} \]
\[ f_{Y \mid X=x} = \frac{1}{X} \quad (y \in [20, 20+x]) \]
\[ \text{or} \quad \frac{1}{20+x-20} = \frac{1}{X} \]

(2) \[ E[Y \mid X=x] = \int y \cdot f_{Y \mid X=x}(y) \, dy \]
\[ = \frac{1}{X} \int_{20}^{20+x} y \, dy \]
\[ = \frac{1}{X} \left[ \frac{y^2}{2} \right]_{20}^{20+x} = \frac{20+x}{2} \]

\[ E[Y \mid X=x] = \frac{40 + x}{2} = 20 + \frac{x}{2} \]

(3) \[ f_Y(y) = \int f_{X,Y}(x,y) \, dx \]
\[ P(Y > b \mid X=x) = \int_b^{20+x} f_{Y \mid X=x}(y) \, dy \]
\[ P(Y > 25 \mid X=x) = \int_{25}^{20+x} \frac{1}{X} \, dy \]
\[ = \int_{25}^{20+x} \frac{1}{X} \, dy \left[ X \geq 5 \right] + \int_{25}^{20+x} \frac{1}{X} \, dy \left[ X < 5 \right] \]
\[ = \left( \frac{20+x-25}{X} \right) \left[ X \geq 5 \right] + 0 \]

\[ P(Y > 25 \mid X=x) = \frac{x-5}{X} \left[ X \geq 5 \right] \]
(4) \[ P(Y \geq 25) = E[P(Y \geq 25 | X)] \]
= \[ E \left[ \frac{x-5}{x} \mathbb{1}_{[X \geq 5]} \right] \]
= \[ \int \frac{x-5}{x} \cdot \frac{1}{30} \, dx \]
= \[ \frac{1}{30} \int_{5}^{20} \left(1 - \frac{5}{x}\right) \, dx \]
= \[ \frac{1}{30} \left( \int_{5}^{30} \, dx - \int_{5}^{30} \frac{1}{x} \, dx \right) \]
= \[ \frac{1}{30} \left( 30 - 5 \right) - \frac{1}{6} \left( \ln 6 \right) \]

\[ P(Y \geq 25) = \frac{5 - \ln 6}{6} \]
Problem 11: You flip a fair coin 100 times. Denote $X$ the number of Heads that you get.

1. Give the probability mass function of $X$. Compute $\mathbb{P}(X \geq 75)$. (do not simplify).
2. Compute $\mathbb{E}[X]$ and $\text{Var}(X)$ (as if you didn't know already the answer: use the representation of $X$ as a sum of Bernoulli r.v.).
3. Use Chebyshev's inequality to give a bound on $\mathbb{P}(X \geq 75)$.

Let $X_i$ denote the result of $i^{th}$ flip as a Bernoulli r.v. and let $X_i=1$ if head appears and $X_i=0$ if tail appears. Since the coin is claimed to be fair we have $p=0.5$. The number of heads that we get after 100 trials can be calculated as below:

$$X = \sum_{i=1}^{100} X_i$$

$X$ is a binomial random variable. Therefore its mass probability function is:

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

The above formula is derived as follows: There is $k$ heads with probability of $p^k$ and $n-k$ tails with probability of $(1-p)^{n-k}$. However the $k$ heads can be rearranged in $\binom{n}{k}$ different ways. In our problem $p=1/2$ and $n=100$, so we have:

$$P(X = k) = \binom{100}{k} 0.5^k (1 - 0.5)^{100-k} = \binom{100}{k} 0.5^{100}$$

To calculate the $P(X \geq 75)$ we have:

$$P(X \geq 75) = \sum_{k=75}^{100} \binom{100}{k} 0.5^{100} = 0.5^{100} \sum_{k=75}^{100} \binom{100}{k}$$

Or

$$P(X \geq 75) = 1 - P(X \leq 74) = 1 - \sum_{k=0}^{74} \binom{100}{k} 0.5^{100} = 1 - 0.5^{100} \sum_{k=0}^{74} \binom{100}{k}$$

(2)

The expectation of a Bernoulli variable can be simply calculated as:

$$E(X_i) = \sum_{j=0}^{1} jP(X_i = j) = 1 \times p + 0 \times (1 - p) = p$$
Furthermore

\[ E(X_i^2) = \sum_{j=0}^{1} j^2 P(X_i = j) = 1^2 \times p + 0^2 \times (1 - p) = p \]

Therefore

\[ Var(X_i) = E(X_i^2) - E(X_i)^2 = p - p^2 = p(1 - p) \]

Now we can simply calculate the \( E[X] \) and \( Var(X) \), employing the linearity concept.

\[
E[X] = E \left[ \sum_{i=1}^{100} X_i \right] = \sum_{i=1}^{100} E[X_i] = 100p = 50
\]

\[
Var(X) = Var \left( \sum_{i=1}^{100} X_i \right) = \sum_{i=1}^{100} Var(X_i) = 100p(1 - p) = 25
\]

(3)

The Chebyshev's inequality is

\[ P(|X - E[X]| \geq r) \leq \frac{Var(X)}{r^2} \]

where \( r \) is any positive real number.

We have

\[
P(|X - 50| \geq 25) = P(X - 50 \geq 25 \text{ or } X - 50 \leq -25) = P(X \geq 75 \text{ or } X \leq 25)
\]

\[
= P(X \geq 75) + P(X \leq 25)
\]

Note that \( X \geq 75 \text{ and } X \leq 25 \) cannot happen simultaneously and therefore their intersection is zero. So

\[
P(X \geq 75) < P(|X - 50| \geq 25)
\]

Because \( P(X \leq 25) \) is a positive value. Now we can write the Chebyshev's inequality as below

\[
P(X \geq 75) < P(|X - 50| \geq 25) \leq \frac{Var(X)}{r^2} = \frac{25}{25^2} = \frac{1}{25} = 0.04
\]

Now let's test this result. Using an online binomial calculator I get \( P(X \geq 75) = 2.82 \times 10^{-7} \), which is less than 0.04. A better way to test this result is employing the central limit theorem. As \( X_i \)'s are
Bernouli r.v's, the sum is a binomial and for a large number of n, we can assume the distribution as a normal distribution, therefore:

\[
P\left( \frac{X - 50}{\sqrt{25}} \geq \frac{75 - 50}{\sqrt{25}} \right) \approx P(z \geq 5) = 1 - P(z < 5) = 1 - 0.99999971 = 2.9 \times 10^{-7}
\]

the above results also satisfy our inequality.