Final #1

Mark all correct answers in each of the following questions.

1. A coin is tossed n times. Let X be the number of heads and Y the sum of the round numbers at which heads are tossed. For example, if $n = 7$, and the results are H, T, H, T, T, H, T, then $X = 3$ and $Y = 1 + 3 + 6 = 10$.

   (a) Y is binomially distributed.

   (b) $P(Y = 6) = 4/2^n$ for all sufficiently large n.

   (c) $P(X = 3|Y = 7) = 1/4$.

   (d) $E(Y) = n(n + 1)/2$.

   (e) $V(Y) = n(n + 1)(2n + 1)/6$.

   (f) $\rho(X, Y) \underset{n\to\infty}{\longrightarrow} \frac{\sqrt{2}}{2}$.

2. We select n random numbers as follows. The first number is 1. The second number is either 1 or 2, each being selected at a probability of 1/2. In general, for $1 \leq k \leq n$ we select at the k-th stage one of the numbers $1, 2, \ldots, k$, each with probability $1/k$. Let X be the sum of all selected numbers, Y – their product, Z – the number of those equal to 1, and W the number of numbers between 1 and n chosen at least once. For example, if $n = 7$ and the numbers 1, 1, 3, 2, 1, 6, 3 are selected, then $X = 17$, $Y = 108$, $Z = 3$, $W = 4$.

   (a) For $n \geq 2$, the probability that the number 1 is selected more times than the number 2 is 1/2.

(b) Markov's inequality implies:

$$P(X \geq 0.4n(n+3)) \underset{n\to\infty}{\longrightarrow} \frac{5}{8}\,.$$

(c) $E(Y^2) = \frac{(2n+3)!}{2\cdot 12^n}\,.$

(d) Chebyshev's inequality implies that for every $\varepsilon > 0$:

$$P(|Z - \log n| > \varepsilon \log n) \underset{n\to\infty}{\longrightarrow} 0.$$

(e) $\frac{E(W)}{n} \underset{n\to\infty}{\longrightarrow} 1 - \frac{1}{e}\,.$

(f) For sufficiently large n, there exist eactly two quadruples (a, b, c, d) for which:

$$P(Y = b|X = a) = P(Z = c|X = a) = P(W = d|X = a) = 1.$$

3. The variable (X, Y) is uniformly distributed in the planar region

$$S = \{(x, y) : 0 \leq x \leq \pi,\ 0 \leq y \leq \sin x\}.$$

(That is, since the area of S is 2, the probability of (X, Y) to assume a value in some set $S' \subseteq S$ is half the area of S'.)

(a) The distribution function of X is given by:

$$F_X(x) = \begin{cases} 0, & x < 0, \\ \frac{1}{2} - \frac{\cos x}{2}, & 0 \leq x \leq \pi, \\ 1, & \pi < x. \end{cases}$$

(b) The density function of Y is given by:

$$f_Y(y) = \begin{cases} 3(1 - y^{1/2}), & 0 \leq y \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

(c) $E(Y) = \pi/4$.

(d) $E(XY) = \pi^2/16$.

(e) The random variables $X_1, X_2, \ldots, X_{100}$ are independent and have the same distribution as X. Then:

$$P\left(\sum_{i=1}^{100} X_i \geq 170\right) \leq 0.05.$$

4. Let X, Y, Z be random variables.

(a) If X, Y are independent and $V(X) = V(Y)$, then $X+Y, X-Y$ are uncorrelated but not necessarily independent.

(b) If $X+Y$ has a finite variance, then at least one of X and Y has a finite variance as well.

(c) Suppose $E(X) = E(Y) = E(Z) = 0$ and all three correlation coefficients $\rho(X,Y), \rho(X,Z), \rho(Y,Z)$ are strictly positive. If $E(XYZ)$ exists, then $E(XYZ) \geq 0$.

Solutions

1. If $Y \sim B(m,p)$ for some m and p, then, since $P(Y=0) = P(Y=1) = P(Y = n(n+1)/2) = 1/2^n$, we obtain $(1-p)^m = mp(1-p)^{m-1} = p^m$. The equality of the first and third expressions yields $p = 1/2$, and then the equality of the first and second expressions yields $m = 1$. It follows that Y is not binomially distributed unless $n = 1$.

The equality $Y = 6$ means that the outcomes of all tosses were T, except for either (i) toss 6, or (ii) tosses 1 and 5, or (iii) tosses 2 and 4, or (iv) tosses $1, 2$ and 3. The probability of each exception (for $n \geq 6$) is $1/2^n$, and thus $P(Y = 6) = 4/2^n$. Similarly, we easily verify that $Y = 7$ is obtained for 5 sequences only, out of which only the sequence H, H, T, H, T, T, T, …, T yields $X = 3$. Consequently, $P(X = 3 | Y = 7) = 1/5$.

Since $n(n+1)/2$ is the maximal possible value of Y, and there are other possible values for Y, we must have $E(Y) < n(n+1)/2$.

Now define random variables X_i, $1 \le i \le n$, by $X_i = 1$ if the outcome of the i-th toss is H and $X_i = 0$ otherwise. Clearly, $X = \sum_{i=1}^n X_i$ and $Y = \sum_{i=1}^n iX_i$. Since $E(X_i) = 1/2$ and $V(X_i) = 1/4$ for each i, we have

$$E(Y) = \sum_{i=1}^n i \cdot \frac{1}{2} = \frac{n(n+1)}{4}$$

and

$$V(Y) = \sum_{i=1}^n i^2 \cdot \frac{1}{4} = \frac{n(n+1)(2n+1)}{24}.$$

Also

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \sum_{i,j=1}^n i\,\mathrm{Cov}(X_i,X_j)\\ &= \sum_{i=1}^n i\,\mathrm{Cov}(X_i,X_i) = \sum_{i=1}^n iV(X_i) = \tfrac{n(n+1)}{8},\end{aligned}$$

and therefore

$$\rho(X,Y) = \frac{n(n+1)/8}{\sqrt{n/4 \cdot n(n+1)(2n+1)/24}} \underset{n\to\infty}{\longrightarrow} \frac{\sqrt{3}}{2}.$$

Thus, only (b) is true.

2. Let $p_<, p_=$ and $p_>$ be the probabilities that the number of 1's in the last $n-1$ tosses is less than, equal or larger, respectively, than the number of 2's in the same tosses. By symmetry, $p_< = p_>$, and therefore the probability required in part (a) is (for $n \ge 3$)

$$p_= + p_> = p_=/2 + (p_< + p_= + p_>)/2 = p_=/2 + 1/2 > 1/2.$$

Denote by X_i the number selected at the i-th round, $1 \le i \le n$. Clearly, $X_i \sim U[1,i]$, so that $E(X_i) = (1+i)/2$ and $V(X_i) = (i^2-1)/2$ for each i. Thus

$$E(X) = \sum_{i=1}^n \frac{i+1}{2} = \frac{n(n+1)}{4} + \frac{n}{2} = \frac{n(n+3)}{4}$$

and

$$V(X) = \sum_{i=1}^{n} \frac{i^2 - 1}{2} = \frac{n(n+1)(2n+1)}{12} - \frac{n}{2} = \frac{n(2n^2+3n-5)}{12}.$$

Markov's inequality implies therefore

$$P(X \geq 0.4n(n+3)) \leq \frac{n(n+3)/4}{0.4n(n+3)} \underset{n\to\infty}{\longrightarrow} \frac{5}{8},$$

so that:

$$\limsup_{n\to\infty} P(X \geq 0.4n(n+3)) \leq \frac{5}{8}.$$

Obviously, $Y = \prod_{i=1}^{n} X_i$, which implies

$$\begin{aligned} E(Y^2) &= \prod_{i=1}^{n} E(X_i^2) = \prod_{i=1}^{n} \tfrac{(i+1)(2i+1)}{6} \\ &= \prod_{i=1}^{n} \tfrac{(2i+2)(2i+1)}{12} = \tfrac{(2n+4)!}{2\cdot 12^n}. \end{aligned}$$

We may write $Z = \sum_{i=1}^{n} Z_i$, where $Z_i = 1$ if the number selected at the i-th stage is 1 and $Z_i = 0$ otherwise. Consequently

$$E(Z) = \sum_{i=1}^{n} \frac{1}{i} = \log n + O(1)$$

and

$$V(Z) = \sum_{i=1}^{n} \frac{1}{i}\left(1 - \frac{1}{i}\right) = \log n + O(1).$$

Chebyshev's inequality then implies for $\varepsilon > 0$:

$$\begin{aligned} P(|Z - \log n| > \varepsilon \log n) &\leq P(|Z - \sum_{i=1}^{n} \tfrac{1}{i}| > \varepsilon \log n + O(1)) \\ &\leq \frac{\log n + O(1)}{(\varepsilon \log n + O(1))^2} \underset{n\to\infty}{\longrightarrow} 0. \end{aligned}$$

Write $W = \sum_{i=1}^{n} W_i$, where $W_i = 1$ if the number i is chosen at least once and $W_i = 0$ otherwise. For $1 \leq i \leq n$:

$$E(W_i) = P(W_i = 1) = 1 - \prod_{j=i}^{n}\left(1 - \frac{1}{j}\right) = 1 - \frac{i-1}{n}.$$

Hence

$$E(W) = \sum_{i=1}^{n} \left(1 - \frac{i-1}{n}\right) = n - \frac{n(n-1)/2}{n} = \frac{n+1}{2}\,,$$

and thus

$$\frac{E(W)}{n} \underset{n\to\infty}{\longrightarrow} \frac{1}{2}\,.$$

In (f), the issue is really which values of X determine, up to order, the outcomes of all tosses. For example, if $X = n$, then all numbers selected in the process are 1, and then $Y = 1, Z = n, W = 1$. Other such values are $X = n+1$, which implies that all numbers selected are 1, except for a single 2, and then $Y = 2, Z = n-1, W = 2$, and $X = n(n+1)/2$, which implies that at each stage i the number i was chosen, and therefore $Y = n!, Z = 1, W = n$.

Thus, only (d) is true.

3. In the "interesting" interval, namely $[0, \pi]$, the density function of X is clearly proportional to the height of the region S. Since the total area of S is 2, this implies:

$$f_X(x) = \begin{cases} 0, & x < 0, \\ \frac{1}{2}\sin x, & 0 \le x \le \pi, \\ 0, & \pi < x, \end{cases}$$

and therefore, by integration, we obtain for $F_X(x)$ the function claimed in (a). Now by symmetry:

$$E(X) = \frac{\pi}{2}\,.$$

Similarly, the density function of Y is proportional to the width of S, that is, for $0 \le y \le 1$:

$$f_Y(y) = \frac{1}{2}\left((\pi - \arcsin y) - \arcsin y\right)\,.$$

Consequently:

$$f_Y(y) = \begin{cases} \frac{\pi}{2} - \arcsin y, & 0 \le y \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

Therefore:

$$E(Y) = \int_0^1 \left(\frac{\pi}{2} - \arcsin y\right) y\,dy.$$

The substitution $y = \sin t$ yields:

$$E(Y) = \int_0^{\pi/2} \left(\frac{\pi}{2} - t\right) \sin t \cos t\,dt = \frac{\pi}{8}\,.$$

By symmetry $E(XY) = E(Y(\pi - X))$, and therefore

$$E(XY) = \frac{1}{2}E(\pi Y) = \frac{\pi}{2}E(Y) = \frac{\pi^2}{16}\,.$$

Next we have

$$E(X^2) = \int_0^{\pi/2} x^2 \cdot \frac{1}{2}\sin x\,dx$$

and routine integration by parts gives

$$E(X^2) = \frac{\pi^2}{2} - 2.$$

Hence:

$$V(X) = E(X^2) - E(X)^2 = \frac{\pi^2}{4} - 2.$$

Estimating the probability in part (e) by the Central Limit Theorem, we obtain:

$$\begin{aligned} P\left(\sum_{i=1}^{100} X_i \geq 170\right) &= P\left(\frac{\sum_{i=1}^{100} X_i - 100\cdot\pi/2}{\sqrt{100(\pi^2/4-2)}} \geq \frac{170-100\cdot\pi/2}{\sqrt{100(\pi^2/4-2)}}\right) \\ &\approx P(Z \geq 1.89). \end{aligned}$$

(where Z is a standard normal random variable), and therefore the required probability is approximately 0.029.

Thus, (a), (d) and (e) are true.

4. In (a), the random variables $X+Y, X-Y$ are indeed uncorrelated since

$$\begin{aligned} E((X+Y)(X-Y)) &= E(X^2 - Y^2) \\ &= V(X) + E(X)^2 - V(Y) - E(Y)^2 \\ &= E(X)^2 - E(Y)^2 \end{aligned}$$

and

$$E(X+Y)E(X-Y) = (E(X)+E(Y))(E(X)-E(Y)) = E(X)^2 - E(Y)^2.$$

For (b), let X be any random variable with infinite variance (say, Cauchy distributed) and $Y = -X$. Then neither X nor Y have finite variances, yet $X + Y = 0$ does.

In (c), let X take the values 1 and -2 with probabilities $2/3$ and $1/3$, respectively, and take $Y = Z = X$. Then $E(X) = E(Y) = E(Z) = 0$, all three correlation coefficients are 1, and $E(XYZ) = E(X^3) = 1^3 \cdot 2/3 + (-2)^3 \cdot 1/3 = -2 < 0$.

Thus, only (a) is true.