Final #1

Mark all correct answers in each of the following questions.

1. An urn contains n balls, marked with the numbers $1, 2, \ldots, n$.

   (a) Suppose first that the balls are drawn from the urn without replacement until at some stage i a ball different from i is drawn or until all balls have been drawn according to their order. Let X be the number of drawn balls. (For example, if ball #1 is drawn first, then ball #2 and then ball #7, then at this stage the trial is stopped and $X = 3$.) Then:

   $$P(X = 5 | X \geq 4) = \frac{1}{n}\,.$$

   (b) $E(X) \underset{n\to\infty}{\longrightarrow} 1$.

   (c) X is hypergeometrically distributed.

   (d) Now suppose that the trial above is performed again and again, until for the first time all balls are drawn. Let Y be the number of trials. (For example, if $n = 3$ and in the first trial ball #2 is drawn, then balls #1 and #3 are drawn, and then balls #1, #2 and #3 are drawn, then $Y = 3$.) Then Y is geometrically distributed.

   (e) $E(Y) = 2^{n-1}$.

   (f) Now suppose that, unlike in (a), we draw out balls until at some stage i we draw ball i. Let Z be the number of drawn balls. (For example, if first ball #6 is drawn, then ball #1 and then ball #3, then at this stage the trial is stopped and $Z = 3$.) Then:

   $$P(Z = n) \underset{n\to\infty}{\longrightarrow} \frac{1}{e}\,.$$

2. A die is tossed n times. Let X be the sum of all even outcomes and Y the sum of all outcomes not exceeding 3. For example, if $n = 9$ and the outcomes are 6, 2, 2, 1, 1, 5, 4, 3, 1, then $X = 6 + 2 + 2 + 4 = 14$ and $Y = 2 + 2 + 1 + 1 + 3 + 1 = 10$.

   (a) $E(X) = E(2Y)$.

   (b) $V(X) = V(2Y)$.

   (c) $V(X - 2Y) = \frac{32n}{3}$.

   (d) $\mathrm{Cov}(X, Y) < 0$.

   (e) $\rho(X, Y) = -\frac{1}{3}$.

   (f) If n is large, then

$$P(-\sqrt{n} \leq X - 2Y \leq \sqrt{n}) \approx 2\Phi(1/2) - 1,$$

   where Φ is the standard normal distribution function.

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

$$S = \{(x, y) : 0 \leq x \leq 1,\ x^2 - 1 \leq y \leq 1 - x^2\}.$$

   (That is, since the area of S is 4/3, the probability of (X, Y) to assume a value in some set $S' \subseteq S$ is the area of S' divided by 4/3.) Put $T = Y/X^2$.

   (a) $E(X) = \frac{1}{2}$.

   (b) $E(Y) = 0$.

   (c) The density function f_{XY} of the variable XY is even (namely, $f_{XY}(-t) = f_{XY}(t)$ for every $t \in \mathbf{R}$).

   (d) The distribution function of T is given by:

$$F_T(t) = \begin{cases} \frac{1}{2(-t+1)}, & t \leq 0, \\ 1 - \frac{1}{2(t+1)}, & t > 0. \end{cases}$$

   (e) T is symmetric around 0, and in particular $E(T) = 0$.

(f) In view of Chebyshev's inequality, there exists a constant c such that:

$$P(|T| \geq t) \leq \frac{c}{t^2}.$$

(g) $P(0 \leq Y \leq X | X \leq 1/2) = 2/11$.

4. Let X be a random variable.

   (a) If $E\left((X-1)^2\right) = 3$ and $E\left((X+1)^2\right) = 5$, then $V(X) \leq 4$.

   (b) If X assumes only the values $1, 2, \ldots$, and $E(X) = \mu$, then:

   $$P(X \geq 10) \leq \frac{\mu - 1}{9}.$$

   (c) If X is discrete, then for every function $h : \mathbf{R} \to \mathbf{R}$ the random variable $h(X)$ is discrete as well.

   (d) If X is continuous, then for every function $h : \mathbf{R} \to \mathbf{R}$ the random variable $h(X)$ is continuous as well.

   (e) If X is symmetric around 0 and the moment generating function ψ_X exists, then it is even (i.e., $\psi_X(-t) = \psi_X(t)$ for every $t \in \mathbf{R}$).

   (f) If the moment generating function ψ_X exists and is even, then X is symmetric around 0.

Solutions

1. The event $\{X \geq 4\}$ occurs if balls #1, #2, #3 are drawn first, second, third, respectively. The event $\{X = 5\}$ occurs if, moreover, ball #4 is drawn at the fourth drawing, but the fifth drawing results in some ball different from #5. Hence:

$$P(X = 5 | X \geq 4) = \frac{1}{n-3} \cdot \frac{n-5}{n-4}.$$

For $1 \leq i \leq n$, let $X_i = 1$ if all $i-1$ first balls are drawn in the correct order and $X_i = 0$ otherwise. Then $X = \sum_{i=1}^{n} X_i$, and therefore

$$\begin{aligned} E(X) &= \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} P(X_i = 1) \\ &= 1 + \frac{1}{n} + \frac{1}{n(n-1)} + \frac{1}{n(n-1)(n-2)} + \dots \\ &\quad + \frac{1}{n(n-1)(n-2) \cdot \ldots \cdot 2} \, . \end{aligned}$$

Hence, on the one hand $E(X) \geq 1$ and on the other hand

$$E(X) \leq 1 + \frac{1}{n} + (n-2) \cdot \frac{1}{n(n-1)} \, .$$

It follows that $E(X) \underset{n \to \infty}{\longrightarrow} 1$.

Intuitively, X has no connection to the trial based on which the hypergeometric distribution is defined, so that X should not be hypergeometrically distributed. (Formally, one may show this as follows. Suppose $X \sim H(m, a, b)$ for some m, a, b. Since an $H(m, a, b)$-distributed variable assumes with positive probability the integer values between $\max(0, m-b)$ and $\min(m, a)$, we must have $m - b = 1$ and $\min(m, a) = n$. Now:

$$E(X) = \frac{am}{a+b} > \frac{am}{a+m} \geq \frac{am}{2\max(a,m)} = \frac{\min(a,m)}{2} = \frac{n}{2} \, .$$

Since $E(X) \underset{n \to \infty}{\longrightarrow} 1$, this is impossible if n is sufficiently large.)

If the trial is performed until all balls are drawn, then it is performed until all balls show up in their natural order. As this happens with probability $1/n!$, we have $Y \sim G(1/n!)$, and in particular $E(Y) = n!$.

The event $\{Z = n\}$ corresponds to the event of no letter being sent to its destination in the lazy secretary problem, but contains in addition the outcomes in which none of the first $n-1$ balls was drawn at the stage corresponding to its number while the n-th ball was drawn at stage n. As we have shown in class, the probability of the first type of outcomes approaches $1/e$ as $n \to \infty$. The probability of the second type is bounded above by $1/n$, and hence does not change the limit.

Thus, only (b), (d) and (f) are true.

2. Let X_i denote the outcome of the i-th toss if it is even and 0 otherwise and Y_i denote the outcome of the i-th toss if it is at most 3 and 0 otherwise, $1 \le i \le n$. Obviously:

$$X = \sum_{i=1}^{n} X_i, \qquad Y = \sum_{i=1}^{n} Y_i.$$

Now

$$E(X_i) = \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 6 + \frac{1}{2} \cdot 0 = 2$$

and

$$E(Y_i) = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{2} \cdot 0 = 1.$$

Consequently

$$E(X) = 2n, \qquad E(Y) = n,$$

and in particular $E(X) = E(2Y)$. To find the variances we calculate

$$E(X_i^2) = \frac{1}{6} \cdot 2^2 + \frac{1}{6} \cdot 4^2 + \frac{1}{6} \cdot 6^2 + \frac{1}{2} \cdot 0^2 = \frac{28}{3}$$

and

$$E(Y_i^2) = \frac{1}{6} \cdot 2^2 + \frac{1}{6} \cdot 4^2 + \frac{1}{6} \cdot 6^2 + \frac{1}{2} \cdot 0^2 = \frac{7}{3}\,.$$

Therefore

$$V(X_i) = \frac{28}{3} - 2^2 = \frac{16}{3}\,, \qquad V(Y_i) = \frac{7}{3} - 1^2 = \frac{4}{3}\,,$$

so that

$$V(X) = \frac{16n}{3}\,, \qquad V(Y) = \frac{4n}{3}\,,$$

and $V(X) = V(2Y)$. Now

$$E(X_i Y_i) = \frac{1}{6} \cdot (0 \cdot 1 + 2 \cdot 2 + 0 \cdot 3 + 4 \cdot 0 + 0 \cdot 0 + 6 \cdot 0) = \frac{2}{3}\,,$$

and (due to independence)

$$E(X_i Y_j) = E(X_i) E(Y_j) = 2, \qquad i \ne j.$$

Hence

$$E(XY) = E(\sum_{i,j=1}^{n} X_i Y_j) = n \cdot \frac{2}{3} + n(n-1)\cdot 2) = 2n^2 - \frac{4n}{3}\,,$$

which gives

$$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) = -\frac{4n}{3}\,.$$

It follows that

$$\begin{aligned} V(X-2Y) &= V(X) + 2\mathrm{Cov}(X,-2Y) + V(-2Y) \\ &= V(X) - 4\mathrm{Cov}(X,Y) + 4V(Y) = 16n \end{aligned}$$

and

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{V(X)V(Y)}} = -\frac{1}{2}\,.$$

Employing the central limit theorem for the variables $X_i - 2Y_i,\ 1 \le i \le n$, we obtain

$$\begin{aligned} P(-\sqrt{n} \le X - 2Y \le \sqrt{n}) &= P\left(-\tfrac{1}{4} \le \tfrac{X-2Y}{\sqrt{16n}} \le \tfrac{1}{4}\right) \\ &\approx \Phi(1/4) - \Phi(-1/4) = 2\Phi(1/4) - 1. \end{aligned}$$

Thus, only (a), (b) and (d) are true.

3. Clearly, for $0 \le x_0 \le 1$:

$$P(X \le x_0) = \frac{3}{4}\int_0^{x_0} ((1-x^2) - (x^2-1))dx = \frac{3}{2}x_0 - \frac{1}{2}x_0^3.$$

Hence

$$f_X(x) = \frac{3}{2} - \frac{3}{2}x^2, \qquad 0 \le x \le 1,$$

and therefore

$$E(X) = \int_0^1 x\left(\frac{3}{2} - \frac{3}{2}x^2\right)dx = \left[\frac{3}{4}x^2 - \frac{3}{8}x^4\right]_0^1 = \frac{3}{8}\,.$$

(Note that, by drawing S, it should have been clear that the distribution of X tends to be more concentrated on the left half of the interval $[0,1]$, so that $E(X) < 1/2$.)

By symmetry we have $E(Y) = 0$. Similarly, the random variable XY is symmetric around 0, and hence its density function is even.

The variable T is obviously symmetric around 0 as well. For fixed $t \geq 0$, the parabola $y = tx^2$ intersects the parabola $y = 1 - x^2$ at the point $\left(\frac{1}{\sqrt{t+1}}, \frac{t}{t+1}\right)$. Hence:

$$P(T \geq t) = P(tX^2 \leq Y) = \frac{3}{4}\int_0^{\frac{1}{\sqrt{t+1}}} (1-x^2-tx^2)dx = \frac{1}{2\sqrt{t+1}}, \quad t \geq 0.$$

It follows that

$$F_T(t) = \begin{cases} \dfrac{1}{2\sqrt{-t+1}}, & t \leq 0, \\ \\ 1 - \dfrac{1}{2\sqrt{t+1}}, & t > 0, \end{cases} \tag{1}$$

and therefore

$$f_T(t) = \begin{cases} \dfrac{1}{4}(-t+1)^{-3/2}, & t \leq 0. \\ \\ \dfrac{1}{4}(t+1)^{-3/2}, & t > 0. \end{cases}$$

Even though T is symmetric around 0, the expectation $E(T)$ does not exist as the integral $\int_{-\infty}^{\infty} t f_T(t)dt$ does not converge absolutely.

By (1) and the symmetry of T, for $t \geq 0$ we have

$$P(|T| \geq t) = \frac{1}{\sqrt{t+1}},$$

so that the left hand side decays to 0 as $t \to \infty$ much slower than $1/t^2$. (Chebyshev's inequality cannot be applied as T does not have finite expectation and variance.)

Finally:

$$\begin{aligned} P(0 \le Y \le X | X \le 1/2) &= \frac{P(0 \le Y \le X \le 1/2)}{P(X \le 1/2)} \\ &= \frac{3/4 \cdot 1/2 \cdot 1/2 \cdot 1/2}{F_X(1/2)} = \frac{3/32}{11/16} = \frac{3}{22}\,. \end{aligned}$$

Thus, only (b) and (c) are true.

4. Under the assumptions of (a) we have

$$E(X) = E\left(\frac{(X+1)^2 - (X-1)^2}{4}\right) = \frac{5-3}{4} = \frac{1}{2}\,,$$

and

$$E(X^2) = E\left(\frac{(X+1)^2 + (X-1)^2 - 2}{2}\right) = \frac{5+3-2}{2} = 3,$$

so that

$$V(X) = E(X^2) - E^2(X) = \frac{11}{4}\,.$$

Under the assumptions of (b), $X-1$ is a non-negative random variable with expectation $\mu - 1$. By Markov's inequality

$$P(X \ge 10) = P(X - 1 \ge 9) \le \frac{\mu - 1}{9}\,.$$

Let X be discrete. If X assumes only the values $x_1, x_2, \ldots$, then $h(X)$ assumes only the values $h(x_1), h(x_2), \ldots$, and hence it is discrete as well. However, if X is any random variable and the function h is constant (or, more generally, its image is finite or countable), then $h(X)$ is discrete.

X is symmetric around 0 if and only if X and $-X$ are identically distributed, which happens (for variables having a moment generating function) if and only if $\psi_X(t) = \psi_{-X}(t)$ for every t (in the domain of ψ_X), which is the case if and only if $\psi_X(t) = \psi_X(-t)$ for every t.

Thus, (a), (b), (c), (e) and (f) are true.