Final #2

Mark the correct answer in each part of the following questions.

1. Computer Science students need to decide which of two electives – Approximation Algorithms (AA) and Logic Programming (LP) – they take. Each student decides to take or not to take each course independently of other students and of his own choice concerning the other course. The probability that a random student takes AA is 0.3, while the probability that he takes LP is 0.4. (Thus, there are four possibilities for each student: a) choose none of the two courses, b) choose only AA, c) choose only LP, and d) choose both courses.)

   The students of the department are labeled by numbers from 1 on. For each $k \geq 1$, denote by R_k the number of students, out of first k students, taking AA, and by S_k the number of those taking only AA. (For example, if students $3, 5, 8$ take none of the two courses, students $1, 2, 6$ – only AA, students $7, 9$ – only LP, and students $4, 10$ – both, then $R_5 = 3, S_5 = 2, R_{10} = 5, S_{10} = 3$.)

   (a) The random variables R_1 and S_1 are:

      (i) independent.

      (ii) dependent, but uncorrelated.

      (iii) correlated with $|\rho(R_1, S_1)| < 1$.

      (iv) linearly dependent, but not equal.

(v) equal.

(b) $\rho(R_{20}, S_{30}) \approx$

(i) 0.

(ii) 0.24.

(iii) 0.33.

(iv) 0.58.

(v) None of the above.

(c) The value of k, for which the probability $P(S_{20} = k)$ is maximal, is:

(i) 3.

(ii) 4.

(iii) 5.

(iv) 6.

(v) None of the above.

(d) $P(R_5 < 2 | S_5 < 2) \approx$

(i) 0.5.

(ii) 0.679.

(iii) 0.7.

(iv) 0.932.

(v) None of the above.

(e) A direct application of Markov's inequality shows that the inequality $P(S_{20} \geq a) \leq 0.05$ holds for $a \geq$

(i) 63.

(ii) 72.

(iii) 81.

(iv) 90.

(v) None of the above.

Remark: We mean here the best bound that can be obtained by Markov's inequality. For example, if Markov's inequality implies that inequality $P(S_{20} \geq a) \leq 0.05$ is true for $a \geq 63$, then it is evidently true for $a \geq 72$, $a \geq 81$ and $a \geq 90$, but only (i) should be marked as the correct answer.

2. A gambler participates in a 2-stage game. At the first stage, he tosses a die till the result 4 is obtained for the first time. Let X be the number of his tosses at this stage. If $X = k$, then at the second stage the gambler tosses a coin k times. Let Y be the number of heads obtained during these k coin tosses.

(a) $E(Y) =$

(i) 2.1.

(ii) 3.

(iii) 4.25.

(iv) 5.

(v) None of the above.

(b) $V(Y) =$

(i) 6.

(ii) 7.

(iii) 8.

(iv) 9.

(v) None of the above.

(c) $\text{Cov}(X, Y) =$

(i) 14.

(ii) 15.

(iii) 16.

(iv) 17.

(v) None of the above.

(d) $P(X = 5|Y = 3) =$

(i) 0.2.

(ii) 0.3.

(iii) 0.45.

(iv) 0.6.

(v) None of the above.

(e) The value of the moment generating function $\psi_Y(t)$ at the point $t = \ln(6/5)$ is

(i) 1.3.

(ii) 2.2.

(iii) 3.5.

(iv) 4.1.

(v) None of the above.

3. In the Knesset there are 900 phones. Let X_i be the waiting time until the next call to be received at the i-th phone, $i = 1, 2, ..., 900$. The X_i's are independent and $X_i \sim \text{Exp}(1)$ for each i. Denote $S_n = \sum_{i=1}^{n} X_i, \quad n = 1, 2, ..., 900$.

(a) $\rho(X_1 X_2, S_2) =$

(i) $-\sqrt{\frac{3}{5}}$.

(ii) $-\sqrt{\frac{1}{3}}$.

(iii) $\sqrt{\frac{1}{5}}$.

(iv) $\sqrt{\frac{2}{3}}$.

(v) None of the above.

(b) For $s > 0$ we have $f_{S_2}(s) =$

(i) $2e^{-2s}$.

(ii) se^{-2s}.

(iii) $2e^{-s}$.

(iv) se^{-s}.

(v) None of the above.

(c) $P(X_1 \geq 1 | X_2 \leq X_1) =$

(i) $\frac{1}{e} - \frac{1}{e^2}$.

(ii) $\frac{1}{e} - \frac{1}{2e^2}$.

(iii) $\frac{2}{e} - \frac{1}{2e^2}$.

(iv) $\frac{2}{e} - \frac{1}{e^2}$.

(v) None of the above.

(d) A direct application of Chebyshev's inequality yeilds
$P(810 \leq S_{900} \leq 990) \geq$

(i) 23/27.

(ii) 24/27.

(iii) 25/27.

(iv) 26/27.

(v) None of the above.

Remark: We mean here the best bound that can be obtained by Chebyshev's inequality. For example, if Chebyshev's inequality implies that the above probability is at most 26/27, hence it is also at most 25/27, and (24/27, and 23/27), but only (iv) should be marked as the correct answer.

(e) A direct application of the Central Limit Theorem implies $P(810 \leq S_{900} \leq 990) \approx$

(i) 0.8889.

(ii) 0.9234.

(iii) 0.9561.

(iv) 0.9974.

(v) None of the above.

Solutions

1. (a)-(b) Define the random variables

$$X_i = \begin{cases} 1, & \text{the } i\text{-th student takes AA}, \\ 0, & \text{otherwise}, \end{cases}$$

and

$$Y_i = \begin{cases} 1, & \text{the } i\text{-th student takes only AA}, \\ 0, & \text{otherwise}. \end{cases}$$

Clearly,

$$X_i \sim B(1, 0.3), \quad Y_i \sim B(1, 0.3 \cdot (1 - 0.4)),$$

and

$$R_k = \sum_{i=1}^{k} X_i \sim B(k, 0.3), \quad S_k = \sum_{i=1}^{k} Y_i \sim B(k, 0.18).$$

Therefore:

$$\begin{aligned} \mathrm{Cov}(X_i, Y_i) &= E(X_i \cdot Y_i) - E(X_i) \cdot E(Y_i) \\ &= E(Y_i) - E(X_i) \cdot E(Y_i) \\ &= 0.18 - 0.3 \cdot 0.18 = 0.126. \end{aligned}$$

For $i \neq j$, the variables X_i, Y_j are independent. Therefore

$$\begin{aligned} \mathrm{Cov}(R_k, S_m) &= \sum_{i=1}^{k} \sum_{j=1}^{m} \mathrm{Cov}(X_i, Y_j) \\ &= \sum_{i=1}^{\min(k,m)} \mathrm{Cov}(X_i, Y_i) \\ &= 0.126 \cdot \min(k, m), \end{aligned}$$

so that

$$\rho(R_k, S_m) = \frac{0.126 \cdot \min(k,m)}{\sqrt{V(R_k) \cdot V(S_m)}} = \frac{0.126 \cdot \min(k,m)}{\sqrt{0.030996 \cdot k \cdot m}}. \qquad (1)$$

By (1), with $k = m = 1$, and with $k = 20,\ m = 30$, we obtain

$$\rho(R_1, S_1) = \frac{0.126}{\sqrt{0.030996}} \approx 0.716,$$

and

$$\rho(R_{20}, S_{30}) = \frac{0.126 \cdot 20}{\sqrt{0.030996 \cdot 20 \cdot 30}} \approx 0.584,$$

respectively.
Thus, (a.iii) and (b.iv) are true.

(c) We have:

$$P(S_{20} = k) = \binom{20}{k} \cdot 0.18^k \cdot 0.82^{20-k}, \qquad 0 \le k \le 20.$$

The condition $P(S_{20} = k) > P(S_{20} = k+1)$ is equivalent to

$$(k+1)!(19-k)! \cdot 0.82 > k!(20-k)! \cdot 0.18,$$

namely

$$(k+1) \cdot 0.82 > (20-k) \cdot 0.18,$$

which is valid if and only if $k \ge 3$.
Thus, (i) is true.

(d) Evidently, $\{R_5 < 2\} \subset \{S_5 < 2\}$, so

$$\begin{aligned} P(R_5 < 2 | S_5 < 2) &= \frac{P(R_5 < 2, S_5 < 2)}{P(S_5 < 2)} \\ &= \frac{P(R_5 < 2)}{P(S_5 < 2)} \\ &= \frac{\binom{5}{0} \cdot 0.7^5 + \binom{5}{1} \cdot 0.3 \cdot 0.7^4}{\binom{5}{0} \cdot 0.82^5 + \binom{5}{1} \cdot 0.18 \cdot 0.82^4} \\ &\approx 0.679. \end{aligned}$$

Thus, (ii) is true.

(e) Markov's inequality for S_{20} reads $P(S_{20} \geq a) \leq E(S_{20})/a$. Since $E(S_{20}) = 20 \cdot 0.18 = 3.6$, we need $\dfrac{3.6}{a} \leq 0.05$, namely $a \geq 72$. Thus, (ii) is true.

2. Evidently, $X \sim G(1/6)$ and $Y|_{X=k} \sim B(k, 1/2)$. Therefore

$$P(X = k) = \frac{1}{6}\left(\frac{5}{6}\right)^{k-1}, \qquad k = 1, 2, \dots,$$

and

$$P(X = k, Y = m) = \frac{1}{6}\left(\frac{5}{6}\right)^{k-1} \cdot \binom{k}{m}\left(\frac{1}{2}\right)^{k}, \qquad k \geq 1,\ 0 \leq m \leq k,$$

Hence

$$P(Y = m) = \sum_{k=1}^{\infty} P(X = k, Y = m)$$

can easily be calculated to yield:

$$P(Y = m) = \begin{cases} \dfrac{1}{7}, & m = 0, \\ \dfrac{12}{35}\left(\dfrac{5}{7}\right)^{m}, & m \geq 1. \end{cases}$$

(Note that for $m = 0$ the calculation is different than for $m \geq 1$.) However, it is simpler to start our calculations using $P_{X,Y}$ rather than P_Y even for $E(Y)$.

(a) We have:

$$\begin{aligned} E(Y) &= \sum_{k=1}^{\infty}\sum_{m=0}^{k} m\ P_{X,Y}(k, m) \\ &= \sum_{k=1}^{\infty} \frac{1}{6}\left(\frac{5}{6}\right)^{k-1} \sum_{m=0}^{k} m \cdot \binom{k}{m}\left(\frac{1}{2}\right)^{k}. \end{aligned}$$

For each k, the inner sum is the expectation of a $B\left(k, \frac{1}{2}\right)$ random variable. Consequently:

$$E(Y) = \sum_{k=1}^{\infty} \frac{1}{6}\left(\frac{5}{6}\right)^{k-1} \frac{k}{2} = \frac{1}{2} \cdot E(X) = \frac{1}{2} \cdot 6 = 3.$$

Thus, (ii) is true.

(b) Similarly to the previous part:

$$\begin{aligned} E(Y^2) &= \sum_{k=1}^{\infty} \sum_{m=0}^{k} m^2 \ P_{X,Y}(k,m) \\ &= \sum_{k=1}^{\infty} \frac{1}{6}\left(\frac{5}{6}\right)^{k-1} \left(\frac{k}{4} + \frac{k^2}{4}\right) \\ &= \frac{1}{4}(E(X) + E(X^2)) \\ &= \frac{1}{4}(E(X) + V(X) + E^2(X)) \\ &= \frac{1}{4}(6 + 30 + 36) = 18. \end{aligned}$$

Hence $V(Y) = E(Y^2) - E^2(Y) = 9$.
Thus, (iv) is true.

(c) Similarly to (a) and (b),

$$\begin{aligned} E(XY) &= \sum_{k=1}^{\infty} \sum_{m=0}^{k} k \ m \ P_{XY}(k,m) \\ &= \sum_{k=1}^{\infty} k \cdot \frac{1}{6}\left(\frac{5}{6}\right)^{k-1} \cdot \frac{k}{2} \\ &= \frac{1}{2}E(X^2) = 33. \end{aligned}$$

Therefore:

$$\text{Cov}\ (X,Y) = E(XY) - E(Y) \cdot E(X) = 33 - 3 \cdot 6 = 15.$$

Thus, (ii) is true.

(d) Obviously:

$$\begin{aligned} P(X=5|Y=3) &= \frac{P(X=5,Y=3)}{P(Y=3)} \\ &= \frac{\frac{1}{6}\left(\frac{5}{6}\right)^{5-1}\cdot\binom{5}{3}\left(\frac{1}{2}\right)^5}{\frac{12}{35}\left(\frac{5}{7}\right)^3} \approx 0.2. \end{aligned}$$

Thus, (i) is true.

(e) A routine calculation yields:

$$\psi_Y(t) = E(e^{tY}) = \sum_{k=1}^{\infty}\sum_{m=0}^{k} e^{tm}\ P_{XY}(k,m) = \frac{1}{7} - \frac{12e^t}{35e^t-49}.$$

In particular:

$$\psi_Y(\ln(6/5)) = 2.2.$$

Thus, (ii) is true.

3. (a) Since $X_i \sim \text{Exp}(1)$, we have $E(X_i) = 1$ and $E(X_i^2) = V(X_i) + E^2(X_i) = 1 + 1 = 2$. We have:

$$\begin{aligned} \text{Cov}(X_1X_2, S_2) &= \text{Cov}(X_1X_2, X_1 + X_2) \\ &= \text{Cov}(X_1X_2, X_1) + \text{Cov}(X_1X_2, X_2) \\ &= 2\cdot\text{Cov}(X_1X_2, X_1) \\ &= 2\cdot\big(E(X_1^2)\cdot E(X_2) - E^2(X_1)\cdot E(X_2)\big) \\ &= 2\cdot(2\cdot 1 - 1^2\cdot 1) = 2. \end{aligned}$$

Also,

$$\begin{aligned} V(X_1X_2) &= E(X_1^2X_2^2) - E^2(X_1X_2) \\ &= E(X_1^2)\cdot E(X_2^2) - E^2(X_1)\cdot E^2(X_2) \\ &= 2\cdot 2 - 1^2\cdot 1^2 = 3, \end{aligned}$$

and $V(S_2) = V(X_1 + X_2) = V(X_1) + V(X_2) = 2$. Hence:

$$\rho(X_1X_2, S_2) = \frac{\text{Cov}(X_1X_2, S_2)}{\sqrt{V(X_1X_2) \cdot V(S_2)}} = \frac{2}{\sqrt{2 \cdot 3}} = \sqrt{\frac{2}{3}}.$$

Thus, (iv) is true.

(b) Obviously, $F_{S_2}(s)$ for $s < 0$. For $s \geq 0$:

$$\begin{aligned} F_{S_2}(s) &= P(X_1 + X_2 \leq s) \\ &= \int_0^s dx_1 \int_0^{s-x_1} e^{-x_1-x_2} dx_2 \\ &= \int_0^s e^{-x_1} \left(1 - e^{-s+x_1}\right) dx_1 \\ &= 1 - e^{-s} - se^{-s}. \end{aligned}$$

It follows that:

$$f_{S_2}(s) = F'_{S_2}(s) = \begin{cases} se^{-s}, & s \geq 0, \\ 0, & s < 0. \end{cases}$$

Thus, (iv) is true.

(c) We have:

$$\begin{aligned} P(X_1 \geq 1, X_2 \leq X_1) &= \int_1^\infty e^{-x_1} dx_1 \int_0^{x_1} e^{-x_2} dx_2 \\ &= \int_1^\infty e^{-x_1} \left(1 - e^{-x_1}\right) dx_1 \\ &= \frac{1}{e} - \frac{1}{2e^2}. \end{aligned}$$

By symmetry, $P(X_2 \leq X_1) = \frac{1}{2}$. Therefore:

$$P(X_1 \geq 1 | X_2 \leq X_1) = \frac{P(X_1 \geq 1, X_2 \leq X_1)}{P(X_2 \leq X_1)} = \frac{2}{e} - \frac{1}{e^2}.$$

Thus, (iv) is true.

(d) Clearly, $E(S_{900}) = 900E(X_1) = 900$ and $V(S_{900}) = 900V(X_1) = 900$. Since S_{900} is continuous,

$$\begin{aligned} P(810 \leq S_{900} \leq 990) &= P(810 < S_{900} < 990) \\ &= 1 - P(|S_{900} - E(S_{900})| \geq 90) \\ &\geq 1 - \frac{V(S_{900})}{90^2} = \frac{8}{9}. \end{aligned}$$

Thus, (ii) is true.

(e) We have:

$$P(810 \leq S_{900} \leq 990) = P(810 \leq \sum_{i=1}^{900} X_i \leq 990).$$

Therefore, the required probability is

$$\begin{aligned} P\left(\frac{810 - 900}{\sqrt{900}} \leq \frac{\sum_{i=1}^{900} X_i - E(S_{900})}{\sqrt{V(S_{900})}} \leq \frac{990 - 900}{\sqrt{900}} \right) &\approx \Phi(3) - \Phi(-3) \\ &= 2\Phi(3) - 1 \\ &\approx 0.9974. \end{aligned}$$

Thus, (iv) is true.