Final #1

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

1. The title of the well-known book "Men Are from Mars, Women Are from Venus" derives from the fact that in Mars three-fourths of the births are of boys and only one-fourth of girls, while on Venus two-thirds are of girls and only one-third of boys.

   (a) A random family with five children is chosen on Venus. It turns out that most children in the family are boys. The probability that exactly three of them are boys is

   (i) 4/243.

   (ii) 40/243.

   (iii) 40/51.

   (iv) 16/17.

   (v) none of the above.

   (b) One of the planets Mars and Venus is chosen randomly. Next a random family with five children on this planet is chosen. The probability that exactly three of the children are boys is:

   (i) $\dfrac{20}{3^5}$.

   (ii) $\dfrac{3^3 \cdot 5}{2^{10}}$.

   (iii) $\dfrac{5^2}{2^8 \cdot 3^2}$.

   (iv) $\dfrac{5 \cdot (2^{12} + 3^8)}{12^5}$.

   (v) none of the above.

(c) As in the preceding part, one of the planets Mars and Venus has been chosen randomly and a random family with five children on this planet is chosen. It turned out that exactly three of the children are boys. The probability that the planet chosen is Venus is

(i) $\dfrac{2^{12}}{2^{12}+3^8}$.

(ii) $\dfrac{2^{12}+3^8}{2^{12}+2\cdot 3^8}$.

(iii) $\dfrac{2^{12}+2\cdot 3^8}{2^{12}+3\cdot 3^8}$.

(iv) $\dfrac{2^{12}+3\cdot 3^8}{2^{12}+4\cdot 3^8}$.

(v) none of the above.

2. Each person in a group of $k \geq 2$ people is asked to rank n of the sonatas of Mozart according to their quality. (Assume that n is large.) As in fact all these sonatas are equally good, each person selects one of the $n!$ possible rankings, with equal probabilities to all rankings. All people are assumed to be independent.

(a) The probability that all k people give the same ranking is

(i) $1/k^{n!-1}$.

(ii) $1/k^{n!}$.

(iii) $1/n!^{k-1}$.

(iv) $1/n!^{k}$.

(v) none of the above.

(b) We choose randomly two of the k people. The probability that none of the n sonatas is at the same place in the rankings of the two people lies in the interval

(i) $[0, 0.2)$.

(ii) $[0.2, 0.4)$.

(iii) $[0.4, 0.6)$.

(iv) $[0.6, 0.8)$.

(v) $[0.8, 1]$.

(c) Suppose $k = n!$. Let X be the number of people whose ranking coincides with the ordering of the sonatas by the time they were composed. Then $P(X = 5) \approx$

(i) $e^{-5}/120$.

(ii) $e^{-5}/8$.

(iii) $e^{-5}/2$.

(iv) e^{-5}.

(v) none of the above.

(d) Let U be the number of people who rank the last composed sonata as the best. Let W be the number of people who rank the last composed sonata as the best and the second last as the second best. Then $\mathrm{Cov}(U, W) =$

(i) $-1/n$.

(ii) $-k/n^2$.

(iii) k/n^2.

(iv) $1/n$.

(v) none of the above.

(e) Suppose that $k = n^3$, and let U be as in the preceding part. Then $P(n^2 - n + 1 \leq U \leq n^2 + n - 1)$ lies in the interval

(i) $[0, 0.2)$.

(ii) $[0.2, 0.4)$.

(iii) $[0.4, 0.6)$.

(iv) $[0.6, 0.8)$.

(v) $[0.8, 1]$.

(**Hint:** Normal approximation.)

3. Let $k \geq 2$ be an integer. We choose randomly a k-digit integer, with the same probability for each. Let X be the chosen number, $D_1, D_2, \ldots, D_k$ its digits from left to right, and Y the number obtained from X by reversing the order of the digits. (For example, if $k = 4$ then X is

selected in the interval $[1000, 9999]$. If $X = 7383$ then $D_1 = 7, D_2 = 3, D_3 = 8, D_4 = 3, Y = 3837$. Note that if X ends with a 0, then Y has less than k digits; for example, if $X = 8500$ then $Y = 58$.)

(a) Consider the random variables $D_1, D_2, \ldots, D_k$.

- (i) They are independent.
- (ii) They are dependent but uncorrelated.
- (iii) $\rho(D_i, D_j) < 0$ for every $i \neq j$.
- (iv) $\rho(D_i, D_j) > 0$ for every $i \neq j$.
- (v) None of the above.

(b) Consider the relation between of the variables D_i and X.

- (i) D_i and X are independent for each i.
- (ii) D_i and X are dependent but uncorrelated for each i.
- (iii) $\rho(D_1, X) = \rho(D_2, X) = \ldots = \rho(D_k, X) > 0$.
- (iv) $\rho(D_1, X) \underset{k\to\infty}{\longrightarrow} \sqrt{80}/9$.
- (v) None of the above.

(c) $F_Y(6000) =$

- (i) $3/5$.
- (ii) $6001/10000$.
- (iii) $2/3$.
- (iv) $6001/9000$.
- (v) none of the above.

(d) $E(Y) =$

- (i) $4.5 \cdot 10^{k-1}$.
- (ii) $5 \cdot 10^{k-1}$.
- (iii) $5.5 \cdot 10^{k-1} - 1$.
- (iv) $5.5 \cdot 10^{k-1}$.
- (v) none of the above.

4. We hold a 5200-stage experiment. At the first stage, a full deck of 52 cards is shuffled, and a single card is drawn randomly. At the second

stage, two full decks are shuffled together and a single card is drawn. We continue similarly, until at the 5200-th stage we shuffle 5200 full decks and draw a single card. Let X be the number of the stage at which a spade jack has been drawn for the third time, where we set $X = 5201$ if at most two spade jacks are drawn throughout the experiment. Let Y be the total number of spade jacks drawn. Finally, let W be the number of i's between 1 and 5199 such that a spade jack is drawn at both the i-th and the $(i+1)$-st stage. (For example, if jack spades are drawn at stages $37, 38, 39, 40, 512, 513, 4700, 5200$, then $X = 39, Y = 8, W = 4$.)

(a) X is distributed approximately:

(i) $B(5200, 3/52)$.
(ii) $H(3, 5200, 5200^2 - 5200)$.
(iii) $\bar{B}(3, 1/52)$.
(iv) $P(100)$.
(v) None of the above.

(b) A direct application of Markov's inequality shows that $P(Y \geq 800) \leq$

(i) 1/800.
(ii) 52/800.
(iii) 1/8.
(iv) 800/5200.
(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 $P(Y \geq 800) \leq 1/800$, then the inequalities implied by all parts (i)-(iv) hold, but only (i) should be marked as correct.

(c) $V(W) =$

(i) $\frac{5199}{52^2} + \frac{2 \cdot 5198}{52^3} - \frac{3 \cdot 5198+1}{52^4}$.
(ii) $\frac{5199}{52^2} + \frac{2 \cdot 5198}{52^3} + \frac{3 \cdot 5198+1}{52^4}$.
(iii) $\frac{5199}{52^2} + \frac{3 \cdot 5198}{52^3} - \frac{2 \cdot 5198+1}{52^4}$.
(iv) $\frac{5199}{52^2} + \frac{3 \cdot 5198}{52^3} + \frac{2 \cdot 5198+1}{52^4}$.
(v) none of the above.

5. The two-dimensional density function of a continuous random variable (X, Y) is defined by:

$$f_{XY}(x, y) = \begin{cases} C(2 - x^2 + y^2), & x^2 + y^2 \leq 1 \\ 0, & \text{otherwise,} \end{cases}$$

where C is a constant.

(a) $C =$

(i) $1/2\pi$.
(ii) $1/\pi$.
(iii) $\sqrt{3}/2\pi$.
(iv) $2/\pi$.
(v) none of the above.

(b) $E(X^2) =$

(i) $5C\pi/12$.
(ii) $7C\pi/12$.
(iii) $5C/12$.
(iv) $7C/12$.
(v) none of the above.

(c) $P(0.699 \leq X \leq 0.701, -0.701 \leq Y \leq -0.699) \approx$

(i) $C/10^6$.
(ii) $2C/10^6$.
(iii) $4C/10^6$.
(iv) $8C/10^6$.
(v) none of the above.

Solutions

1. (a) Denote by X the number of boys in the selected family. Clearly, $X \sim B(5, 1/3)$. Hence the required probability is

$$P(X = 3|X \geq 3) = \frac{P(X = 3)}{P(X \geq 3)}.$$

Now

$$\begin{aligned} P(X = 3) &= \binom{5}{3}\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^2 = \frac{40}{3^5}, \\ P(X = 4) &= \binom{5}{4}\left(\frac{1}{3}\right)^4\left(\frac{2}{3}\right)^1 = \frac{10}{3^5}, \\ P(X = 5) &= \binom{5}{5}\left(\frac{1}{3}\right)^5\left(\frac{2}{3}\right)^0 = \frac{1}{3^5}, \end{aligned}$$

and therefore:

$$P(X = 3|X \geq 3) = \frac{40/3^5}{51/3^5} = \frac{40}{51}.$$

Thus, (iii) is true.

(b) Let M and V denote the events whereby the planet Mars and the planet Venus is chosen, respectively. By the law of total probability:

$$P(X = 3) = \frac{1}{2} \cdot P(X = 3|M) + \frac{1}{2} \cdot P(X = 3|V).$$

Now

$$\begin{aligned} P(X = 3|M) &= \binom{5}{3}\left(\frac{3}{4}\right)^3\left(\frac{1}{4}\right)^2 = \frac{10 \cdot 3^3}{4^5}, \\ P(X = 4|V) &= \binom{5}{3}\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^2 = \frac{40}{3^5}, \end{aligned}$$

so that:

$$P(X = 3) = \frac{5 \cdot 3^3}{4^5} + \frac{20}{3^5} = 5 \cdot \frac{2^{12} + 3^8}{12^5}.$$

Thus, (ii) is true.

(c) With the notations in the preceding parts, we have:

$$
\begin{aligned}
P(V|X=3) &= \frac{P(V \cap \{X=3\})}{P(X=3)} \\
&= \frac{20/3^5}{5 \cdot (2^{12}+3^8)/12^5} = \frac{2^{12}}{2^{12}+3^8}.
\end{aligned}
$$

Thus, (i) is true.

2. (a) The probability for a person to give some fixed arbitrary ranking is $1/n!$. Hence the probability for all k people to give some specific ranking is $1/n!^k$. As this ranking may be any of the possible rankings, the required probability is $n!/n!^k = 1/n!^{k-1}$.

   Thus, (iii) is true.

   (b) After the first person has made his choices, the question is about the probability that the permutation selected by the second will differ from that selected by the first in all entries. This is equivalent to the Absent-Minded Secretary Problem. Namely, the required probability is approximately $1/e \approx 0.37$.

   Thus, (ii) is true.

   (c) Consider the event whereby some specific person ranks according to composition dates as a success. Thus, X counts the number of successes in a sequence of $n!$ independent trials, with success probability of $1/n!$ in each. In other words, $X \sim B(n!, 1/n!)$. By the Poissonian approximation of the binomial, X is approximately $P(1)$-distributed. Hence:

$$
P(X=5) \approx e^{-5} \cdot \frac{1^5}{5!} = \frac{e^{-5}}{120}.
$$

   Thus, (i) is true.

   (d) For $1 \le i \le n$, let $U_i = 1$ if the i-th person ranks the last sonata as the best and $U_i = 0$ otherwise. Similarly, let $W_i = 1$ if he ranks the last composed sonata as the best and the second last as the second best, and $W_i = 0$ otherwise. Obviously, $U = \sum_{i=1}^{k} U_i$ and

$W = \sum_{i=1}^k W_i$. Since U_i and W_j are independent for $i \neq j$, by symmetry we have:

$$\begin{aligned} \mathrm{Cov}(U,W) &= \mathrm{Cov}\left(\sum_{i=1}^k U_i, \sum_{i=1}^k W_i\right) \\ &= \sum_{i=1}^k \mathrm{Cov}\left(U_i, W_i\right) = k \cdot \mathrm{Cov}\left(U_1, W_1\right). \end{aligned}$$

Now

$$\begin{aligned} \mathrm{Cov}\left(U_1, W_1\right) &= E(U_1 W_1) - E(U_1)E(W_1) \\ &= \frac{1}{n(n-1)} - \frac{1}{n} \cdot \frac{1}{n(n-1)} = \frac{1}{n^2}, \end{aligned}$$

and therefore

$$\mathrm{Cov}(U,W) = k/n^2.$$

Thus, (iii) is true.

(e) We use the above decomposition $U = \sum_{i=1}^k U_i$. Since $U_i \sim B(1, 1/n)$ for each i, and the U_i's are independent, we have by the Central Limit Theorem:

$$\begin{aligned} &P(n^2 - n + 1 \leq U \leq n^2 + n - 1) \\ &= P\left(\frac{n^2 - n + 1 - k \cdot 1/n}{\sqrt{k \cdot 1/n \cdot (1 - 1/n)}} \leq \frac{\sum_{i=1}^k U_i - k \cdot 1/n}{\sqrt{k \cdot 1/n \cdot (1 - 1/n)}} \leq \frac{n^2 + n - 1 - k \cdot 1/n}{\sqrt{k \cdot 1/n \cdot (1 - 1/n)}}\right) \\ &\approx \Phi\left(\frac{1-n}{\sqrt{n(n-1)}}\right) - \Phi\left(\frac{n-1}{\sqrt{n(n-1)}}\right). \end{aligned}$$

Hence:

$$P(n^2 - n + 1 \leq U \leq n^2 + n - 1) \approx \Phi(1) - \Phi(-1) = 0.68.$$

Thus, (iv) is true.

3. (a) For arbitrary integers $1 \leq d_1 \leq 9$ and $0 \leq d_i \leq 9, \ \ 1 \leq i \leq k$ we

have:

$$\begin{aligned}
P(D_1 = d_1, D_2 = d_2, \dots D_k = d_k) &= P(X = d_1 \cdot 10^{k-1} + d_2 \cdot 10^{k-2} + \dots + d_k) \\
&= \frac{1}{9 \cdot 10^{k-1}} \\
&= \frac{1}{9} \cdot \underbrace{\frac{1}{10} \cdot \dots \cdot \frac{1}{10}}_{k-1} \\
&= P(D_1 = d_1) \cdot P(D_2 = d_2) \dots \cdot P(D_k = d_k).
\end{aligned}$$

Therefore, $D_1 \sim U[1, 9]$ and $D_i \sim U[0, 9], \;\; 2 \leq i \leq k$ are independent random variables.

Thus, (i) is true.

(b) Obviously, $X = D_1 \cdot 10^{k-1} + D_2 \cdot 10^{k-2} + \dots D_k$, where $D_1 \sim U[1, 9]$ and $D_i \sim U[0, 9], \;\; 2 \leq i \leq k$, are independent random variables. We have:

$$\begin{aligned}
\rho(D_1, X) &= \frac{\mathrm{Cov}(D_1, X)}{\sqrt{V(D_1) \cdot V(X)}} \\
&= \frac{\mathrm{Cov}\left(D_1, \sum_{i=1}^{k} D_i \cdot 10^{k-i}\right)}{\sqrt{V(D_1) \cdot V(X)}} \\
&= \frac{\sum_{i=1}^{k} 10^{k-i} \cdot \mathrm{Cov}\left(D_1, D_i\right)}{\sqrt{V(D_1) \cdot V(X)}} \\
&= \frac{10^{k-1} \cdot V\left(D_1\right)}{\sqrt{V(D_1) \cdot V(X)}} \\
&= \frac{10^{k-1} \cdot \sqrt{V\left(D_1\right)}}{\sqrt{V(X)}}.
\end{aligned}$$

Since $X \sim U[10^{k-1}, 10^k - 1]$, and $D_1 \sim U[1, 9]$, we have $V(X) = \dfrac{(9 \cdot 10^{k-1})^2 - 1}{12}$ and $V(D_1) = \dfrac{80}{12}$, respectively. Hence:

$$\rho(D_1, X) = \frac{10^{k-1} \cdot \sqrt{80}}{\sqrt{(9 \cdot 10^{k-1})^2 - 1}}.$$

Therefore, $\rho(D_1, X) \underset{k\to\infty}{\longrightarrow} \sqrt{80}/9$, which in particular implies that (i) and (ii) are false. One can check similarly that (iii) is false. Thus, only (iv) is true.

(c) Note that Y cannot assume the value 6000. Therefore:

$$\begin{aligned} F_Y(6000) &= P(1 \le Y \le 5999) \\ &= P(D_4 \le 5, D_3 \le 9, D_2 \le 9, , D_1 \le 9) \\ &= \tfrac{6}{10} \cdot 1 \cdot 1 \cdot 1 = \tfrac{3}{5}. \end{aligned}$$

Thus, (i) is true.

(d) We have $Y = \sum_{i=1}^{k} D_i \cdot 10^{i-1}$, and therefore:

$$\begin{aligned} E(Y) &= E(D_k \cdot 10^{k-1} + D_{k-1} \cdot 10^{k-2} + \ldots + D_1,) \\ &= 4.5 \cdot (10^{k-1} + 10^{k-2} + \ldots + 10^1) + 5 \\ &= 5 \cdot 10^{k-1}. \end{aligned}$$

Thus, (ii) is true.

4. (a) Consider the event whereby a spade jack is chosen as a success. Obviously, the probability of a success in each stage is 1/52 Thus, X counts the number of independent trials till the third success, except that, in the case where we have only up to 2 successes within the first 5200 trials, we stop the trials and take X as 5201. It not for this case, X would be distributed $\bar{B}(3, 1/52)$. Now, the probability for up to 2 successes is $\left(\frac{51}{52}\right)^{5200} + \binom{5200}{1} \cdot \left(\frac{51}{52}\right)^{5199} \cdot \frac{1}{52} + \binom{5200}{2} \cdot \left(\frac{51}{52}\right)^{5198} \cdot \left(\frac{1}{52}\right)^2 \approx 7.44 \cdot 10^{-41}$, which is very small, and therefore X is very close to a $\bar{B}(3, 1/52)$-distributed random variable. Therefore, the distribution of X can be approximated by $\bar{B}(3, 1/52)$.
Thus, (iii) is true.

(b) Clearly, $Y \sim B(5200, \frac{1}{52})$. Therefore $Y \ge 0$ and $E(Y) = 5200 \cdot \frac{1}{52} = 100$. Thus, a Markov's inequality implies

$$P(Y \ge 800) \le \frac{E(Y)}{800} = \frac{1}{8}.$$

Thus, (iii) is true.

(c) For each stage i, $1 \leq i \leq 5200$, define a random variable X_i by:

$$X_i = \begin{cases} 1, & \text{a spade jack is drawn,} \\ 0, & \text{otherwise.} \end{cases} \qquad 1 \leq i \leq 5200.$$

Clearly $X_i \sim B(1, \frac{1}{52})$, and the X_i's are independent.
In these terms,

$$W = X_1X_2 + X_2X_3 + X_3X_4 + \ldots + X_{5199}X_{5200},$$

and X_iX_{i+1} and X_jX_{j+1} are independent for $j \neq i-1, i, i+1$. Therefore

$$\begin{aligned} V(W) &= 5199V(X_1X_2) + 2 \cdot 5198\,\mathrm{Cov}(X_1X_2, X_2X_3) \\ &= 5199(E(X_1^2X_2^2) - E^2(X_1)E^2(X_2)) \\ &\quad +2 \cdot 5198(E(X_1X_2^2X_3) - E(X_1)E^2(X_2)E(X_3)) \\ &= 5199(E(X_1^2)E(X_2^2) - E^2(X_1)E^2(X_2)) \\ &\quad +2 \cdot 5198(E(X_1)E(X_2^2)E(X_3) - E(X_1)E^2(X_2)E(X_3)). \end{aligned}$$

Since $E(X_i) = E(X_i^2) = \frac{1}{52}$ we obtain:

$$\begin{aligned} V(W) &= 5199\left(\frac{1}{52^2} - \frac{1}{52^4}\right) + 2 \cdot 5198\left(\frac{1}{52^3} - \frac{1}{52^4}\right). \\ &= \frac{5199}{52^2} + \frac{2 \cdot 5198}{52^3} - \frac{3 \cdot 5198 + 1}{52^4}. \end{aligned}$$

Thus, (iii) is true.

5. (a)

$$
\begin{aligned}
1 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)dxdy \\
&= \int_0^1\int_0^{2\pi} f_{XY}(r\cos\theta, r\sin\theta)rd\theta dr \\
&= C\int_0^1\int_0^{2\pi}(2-r^2(\cos^2\theta-\sin^2\theta))rd\theta dr \\
&= C\int_0^1\int_0^{2\pi}(2-r^2\cos 2\theta)rd\theta dr \\
&= C\cdot 2\pi.
\end{aligned}
$$

Hence $C = \frac{1}{2\pi}$. Thus, (i) is true.

(b)

$$
\begin{aligned}
E(X^2) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^2 f_{XY}(x,y)dxdy \\
&= C\int_0^1\int_0^{2\pi} r^3\cos^2\theta\cdot(2-r^2\cos 2\theta)d\theta dr \\
&= C\int_0^1\int_0^{2\pi} r^3\frac{1+\cos 2\theta}{2}\cdot(2-r^2\cos 2\theta)d\theta dr \\
&= \frac{C}{2}\int_0^1\int_0^{2\pi}(2r^3-r^5\cos 2\theta+2r^3\cos 2\theta-r^5\cos^2 2\theta)d\theta dr \\
&= \frac{C}{2}\int_0^1\int_0^{2\pi}(2r^3-r^5\cos 2\theta+2r^3\cos 2\theta-\frac{1}{2}r^5-\frac{1}{2}r^5\cos 4\theta)d\theta dr \\
&= \frac{C}{2}\int_0^1\int_0^{2\pi}(2r^3-\frac{1}{2}r^5)d\theta dr \\
&= \frac{C}{2}\cdot\left(2\pi\cdot 2\cdot\frac{1}{4}-\frac{1}{2}\cdot 2\pi\cdot\frac{1}{6}\right) = \frac{5\pi C}{12}.
\end{aligned}
$$

Thus, (i) is true.

(c)

$$P(0.699 \leq X \leq 0.701, -0.701 \leq Y \leq -0.699)$$

$$\approx f_{XY}(0.7, -0.7) \cdot \left(\frac{2}{1000}\right)^2$$

$$= C(2 - 0.7^2 + (-0.7)^2) \cdot \left(\frac{2}{1000}\right)^2 = \frac{8C}{10^6}.$$

Thus, (iv) is true.