Final #2

Mark all correct answers in each of the following questions.

1. Let $a_1,\ a_2, \dots,\ a_n$, where $n \geq 3$, be a finite sequence consisting of distinct numbers, ordered randomly (with all $n!$ permutations equiprobable). Let M be the number of pairs (i, j) such that $1 \leq i < j \leq n$ and $a_i > a_j$. Let S be the number of indices $1 \leq i \leq n-2$ for which the three terms $a_i,\ a_{i+1},\ a_{i+2}$ do not form a monotonic subsequence (namely, they satisfy either $a_i < a_{i+1} > a_{i+2}$ or $a_i > a_{i+1} < a_{i+2}$). For example, if $n = 10$ and the ordering is 1, 2, 3, 6, 7, 4, 5, 8, 9, 10, then $M = 4$ and $S = 2$. For convenience denote $m = \binom{n}{2}$.

   (a) $M \sim B\left(m, \frac{1}{2}\right)$.

   (b) $S \sim B\left(n-2, \frac{1}{2}\right)$.

   (c) $E(M) = \frac{n(n-1)}{4}$.

   (d) $E(S) > \frac{n-2}{2}$ for all sufficiently large n.

   (e) $P(M = m) = P(S = n-2)$.

   (f) The more the terms of the sequence are shuffled, the more disorder we should expect to have. Hence, if in addition to ordering the elements randomly we pick randomly a pair of distinct terms and swap them, and denote by M' and S' the values corresponding to M and S, respectively, after this change, then $E(M') > E(M)$ and $E(S') > E(S)$.

   (g) None of the above.

2. A player draws with replacement cards from a full deck of 52 cards, until he draws for the first time the king of spades. Let M be the number

of distinct cards drawn (between 1 and 52), N – the total number of drawings (including the last one) and X the number of times a queen is drawn.

(a) $E(M) = 26$.

(b) $E(N) = 104$.

(c) $V(X) = 20$.

(d) The probability for equal numbers of drawings of the kings of hearts and the king of diamonds is strictly between $\frac{1}{3}$ and $\frac{3}{5}$.

(e) Every two out of the three random variables M, N and X are dependent.

(f) The correlation coefficient of any two of the three random variables M, N and X is positive. (Hint: Use your intuition. Do not try to calculate the correlation coefficients.)

(g) None of the above.

3. In some programming languages, when a program is run the garbage collector is activated each time when the available memory gets below some threshold. More precisely, the system checks periodically, say every unit time, whether the available memory is below the threshold, in which case the garbage collector is activated. Suppose the time T it takes the available memory to get below the threshold, after the garbage collector has been last activated, is distributed $\mathrm{Exp}(\theta)$. Thus the time between consecutive activations of the garbage collector is $S = \lceil T \rceil$ (where $\lceil t \rceil$ denotes the least integer not smaller than the real number t). Let $Y = S - T$ the amount of time between the moment the memory level reached the threshold and the moment the garbage collector started taking care of the problem.

(a) The distribution of S is geometric.

(b) $Y \sim U(0, 1)$.

(c) The random variables S, Y are independent.

(d) The random variables T, Y are dependent, yet uncorrelated.

(e) $E(Y) = \frac{1}{1-e^{-\theta}} - \frac{1}{\theta}$.

(f) $S = T + Y$, which in our case implies that $E(S) = E(T) + E(Y)$ and $V(S) = V(T) + V(Y)$.

(g) None of the above.

4. Let $X_1,\ X_2, \dots,\ X_m \sim U[1, n]$ be independent. Put $Y = \min_{1 \le i \le m} X_i$ and $Z = \max_{1 \le i \le m} X_i$.

(a) Z is one of the variables X_i, and in particular it is uniformly distributed.

(b) The limiting distribution of Y as $m = n \longrightarrow \infty$ is geometric.

(c) The limiting distribution of Y as $m = n \longrightarrow \infty$ is Poissonian.

(d) $P_{Y,Z}(k, l) = \left(\frac{l-k+1}{n}\right)^m, \qquad 1 \le k \le l \le n.$

(e) $P_{Y,Z}(k, l) < \left(\frac{l-k+1}{n}\right)^m, \qquad 1 \le k \le l \le n.$

(f) The minimum of numbers tends to be small, while the maximum tends to be large. In our case, this means that Y and Z have opposite tendencies. In particular, for every fixed $n \ge 2$, if m is sufficiently large then $\rho(Y, Z) < 0$. However, if either $m = 1$ or $n = 1$, then $\rho(Y, Z) = 1$.

(g) None of the above.

5. A point (X, Y) is selected uniformly in the annular region consisting of all points whose distance from the origin is between 1 and 2. Put $R = \sqrt{X^2 + Y^2}$, $S = X^2+Y^2$ and $Q = Y/X$. (Note that Q is undefined at the points where $X = 0$, but as this set is of probability 0 we may ignore it.)

(a) $R \sim U(1, 2)$.

(b) $S \sim U(1, 4)$.

(c) The distribution of Q is symmetric about the point 0, and in particular $E(Q) = 0$.

(d) $F_R(r) = \frac{r^2-1}{3}$ for $1 \le r \le 2$.

(e) $V(R) = \frac{13}{162}$.

(f) The random variables X, R are independent.

(g) None of the above.

Solutions

1. We have $M = 0$ if and only if the sequence is sorted, which occurs with probability $\frac{1}{n!}$. If we had $M \sim B\left(m, \frac{1}{2}\right)$, then we would have $P(M = 0) = 1/2^m$. Similarly, the event $\{M = m\}$ consists of a single point, whereas the event $\{S = n - 2\}$ occurs for numerous orderings of the sequence (for example, 1, n, 2, $n - 1, \ldots$, as well as the inverse ordering).

   The event $\{S = 0\}$ occurs if and only if the sequence is either increasing or decreasing. Hence $P(S = 0) = \frac{2}{n!} \neq \frac{1}{2^{n-2}}$, and in particular S is not $B\left(n - 2, \frac{1}{2}\right)$-distributed.

   If M assumes the value l for some ordering, then it assumes the value $m-l$ for the inverse ordering. Hence the distribution of M is symmetric about the point $\frac{m}{2}$, and in particular $E(M) = \frac{m}{2} = \frac{n(n-1)}{4}$. (A more "conventional" way to obtain this result is by expressing M in the form $\sum_{1 \leq i < j \leq n} M_{ij}$, where $M_{ij} = 1$ if $a_i > a_j$ and $M_{ij} = 0$ otherwise.)

   For $1 \leq i \leq n - 2$, define the random variable S_i by $S_i = 1$ if the three terms $a_i,\ a_{i+1},\ a_{i+2}$ do not form a monotonic subsequence and $S_i = 0$ otherwise. Clearly, $S = \sum_{i=1}^{n-2} S_i$. By symmetry, all $3! = 6$ orderings of $a_i,\ a_{i+1},\ a_{i+2}$ are equi-probable. Since for 4 of these orderings the terms form a monotonic subsequence, we have $P(S_i = 1) = \frac{4}{6} = \frac{2}{3}$, and therefore $E(S) = \frac{2(n-2)}{3} > \frac{n-2}{2}$.

   Ordering the sequence randomly, and then letting an additional random transposition act on the sequence, amounts to still having a random permutation, with each of the $n!$ *a priori* possible orderings equiprobable.

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

2. We may assume that the experiment is held indefinitely, even though we care only about the part until the first king of spades has been drawn. To understand the distribution of M, we may ignore all drawings of cards which were drawn already. Thus, we just look at the 52! possible

orders of drawing the cards. By symmetry, all these orders are equiprobable, and in particular the king of spades has equal probability of being any of the 52. Hence $M \sim U[1,52]$, and therefore $E(M) = 53/2$.

Obviously, $N \sim G(1/52)$, so that $E(N) = 52$.

To understand the distribution of X, we may ignore all drawings of cards other than the king of spades and the 4 queens. The number of cards drawn out of these five cards until the king of spades is drawn for the first time is distributed $G(1/5)$. Since X does not take into account the last drawing, we have $E(X) = 5 - 1 = 4$. When calculating the variance, the last drawing is immaterial, so that

$$V(X) = \frac{4/5}{(1/5)^2} = 20.$$

Denote by A the event in (d). First note that this event strictly contains the event B whereby the king of hearts and the king of diamonds do not appear at all before the first time the king of spades does. Now the event B occurs exactly if the king of spades is the first to appear among the three kings in question. By symmetry we have $P(B) = 1/3$, and therefore $P(A) > 1/3$. On the other hand, A is strictly contained in the event C, whereby the king of hearts and the king of diamonds appear together an even number of times before the first time the king of spades does. In other words, C is the event that the $G(1/3)$-distributed variable, counting the number of drawings of these three kings until the first occurrence of the king of spades, assumes an odd value. This latter probability is

$$\frac{1}{3} + \left(\frac{2}{3}\right)^2 \cdot \frac{1}{3} + \left(\frac{2}{3}\right)^4 \cdot \frac{1}{3} + \ldots = \frac{1}{1-(2/3)^2} \cdot \frac{1}{3} = \frac{3}{5}.$$

Hence $P(A) < 3/5$. (We mention in passing that it is actually possible to calculate the probability of A precisely. In fact, note that this event happens if, for some n, we have exactly n drawings of each of the king of hearts and the king of diamonds before the first drawing of the king of spades. In other words, denoting the latter event by A_n, we have $A = \bigcup_{n=0}^{\infty} A_n$, where the union is disjoint. When calculating $P(A_n)$ we may ignore all drawings of cards other than the three kings in question. Hence

$$P(A_n) = \frac{\binom{2n}{n}}{3^{2n+1}},$$

so that

$$P(A) = \sum_{n=0}^{\infty} P(A_n) = \sum_{n=0}^{\infty} \frac{\binom{2n}{n}}{3^{2n+1}} = \frac{1}{3}\sum_{n=0}^{\infty} \binom{2n}{n} \left(\frac{1}{9}\right)^n .$$

Now recall that

$$\sum_{n=0}^{\infty} \binom{2n}{n} x^n = \frac{1}{\sqrt{1-4x}}, \qquad -\frac{1}{4} < x < \frac{1}{4},$$

and therefore

$$P(A) = \frac{1}{3}\frac{1}{\sqrt{1-4/9}},$$

so that $P(A) = 1/\sqrt{5}$.)

A relatively large value of any of the three variables M, N and X means that there have been many drawings until the king of spades was drawn, and hence hints that each of the other two is also relatively large. Therefore all correlation coefficients should be positive.

Thus, only (c), (d), (e) and (f) are true.

3. For any positive integer m we have

$$P(S = m) = P(m-1 < T \le m) = \left(1 - e^{-\theta m}\right) - \left(1 - e^{-\theta(m-1)}\right).$$

Therefore

$$P(S = m) = \left(e^{-\theta}\right)^{m-1}\left(1 - e^{-\theta}\right), \qquad m = 1, 2, \ldots,$$

so that $S \sim G\left(1 - e^{-\theta}\right)$. For Y we have

$$\begin{aligned} F_Y(y) &= P(Y \le y) = \sum_{m=1}^{\infty} P(m - y \le T \le m) \\ &= \sum_{m=1}^{\infty} \left(\left(1 - e^{-\theta m}\right) - \left(1 - e^{-\theta(m-y)}\right)\right) \\ &= \frac{e^{\theta y} - 1}{e^{\theta} - 1}, \qquad 0 \le y < 1. \end{aligned}$$

Consequently

$$f_Y(y) = \frac{\theta}{e^\theta - 1} e^{\theta y}, \qquad 0 \le y < 1,$$

and a routine calculation yields

$$E(Y) = \frac{\theta}{e^\theta - 1} \int_0^1 y e^{\theta y}\, dy = \frac{1}{1 - e^{-\theta}} - \frac{1}{\theta}\,.$$

Since

$$\begin{aligned} P(S = m, Y \le y) &= P(m - y \le T \le m) = \left(e^{\theta y} - 1\right) e^{-m\theta} \\ &= P(S = m) P(Y \le y), \end{aligned}$$

the variables S and Y are independent. Now

$$\mathrm{Cov}(T, Y) = \mathrm{Cov}(S - Y, Y) = \mathrm{Cov}(S, Y) - \mathrm{Cov}(Y, Y) = -V(Y),$$

so that T and Y are negatively correlated.

Since S and Y are independent

$$V(T) = V(S - Y) = V(S) + V(-Y) = V(S) + V(Y).$$

Thus, only (a), (c) and (e) are true.

4. We have $Y = k$ if and only if all X_i's are at least k, but not all of them are at least $k + 1$. Hence:

$$\begin{aligned} P(Y = k) &= \left(\frac{n - k + 1}{n}\right)^m - \left(\frac{n - k}{n}\right)^m \\ &= \left(1 - \frac{k - 1}{n}\right)^m - \left(1 - \frac{k}{n}\right)^m. \end{aligned}$$

As $m = n \longrightarrow \infty$ we have:

$$P(Y = k) \longrightarrow e^{-(k-1)} - e^{-k} = \left(e^{-1}\right)^{k-1} \left(1 - e^{-1}\right), \qquad k = 1, 2, \ldots, n.$$

The number on the right hand side is exactly the probability of a $G\left(1 - e^{-1}\right)$-distributed random variable to assume the value k. Hence the limiting distribution of Y as $m = n \longrightarrow \infty$ is geometric.

Similarly:

$$P(Z=l)=\left(\frac{l}{n}\right)^m-\left(\frac{l-1}{n}\right)^m, \qquad l=1,2,\ldots,n.$$

Obviously

$$P_{Y,Z}(k,l)\le P(k\le X_1,X_2,\ldots,X_m\le l)=\left(\frac{l-k+1}{n}\right)^m, \quad k\le l.$$

The inequality is usually strict. However, if $k=l$, for example, it is actually an equality.

For $n\ge 2$, a large value of Y means that all X_i's are large, and hence Z should be expected to be relatively large. Hence $\rho(Y,Z)$ should be expected to be positive. In fact, one can show, for example, that for $n=2$ we have $\mathrm{Cov}(Y,Z)=1/2^{2m}>0$, so that $\rho(Y,Z)>0$.

Thus, only (b) is true.

5. The random variable R clearly assumes values in the interval $[1,2]$. For r in this range, the value of $F_R(r)$ is the ratio between the area of the annular region consisting of all points whose distance from the origin is between 1 and r and the area of the region consisting of all points whose distance is between 1 and 2. Therefore

$$F_R(r)=\frac{\pi\left(r^2-1^2\right)}{\pi\left(2^2-1^2\right)}=\frac{r^2-1}{3}, \qquad 1\le r\le 2,$$

and hence:

$$f_R(r)=\frac{2r}{3}, \qquad 1\le r\le 2.$$

Thus

$$E(R)=\int_1^2 r\cdot\frac{2r}{3}\,dr=\frac{14}{9},$$

and

$$E\left(R^2\right)=\int_1^2 r^2\cdot\frac{2r}{3}\,dr=\frac{5}{2},$$

so that:

$$V(R)=E\left(R^2\right)-E(R)^2=\frac{13}{162}.$$

The variable S assumes values in $[1,4]$, and for s in this interval we have

$$F_S(s) = P(S \leq s) = P(R \leq \sqrt{s}) = \frac{s-1}{3},$$

which implies that $S \sim U(1,4)$.

The angle between the positive x-axis and the line leading from a randomly chosen (X,Y) is clearly distributed $U(0,2\pi)$. Hence Q is Cauchy distributed. Its distribution is symmetric with respect to the point 0, but $E(Q)$ does not exist.

The variables X, R are dependent as, for example, if X is close to 2, then R must be close to 2 as well.

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