Pre-Final Review Questions

Mark all correct answers in each of the following questions.

1. Let $V = \{v_1, v_2, \ldots, v_n\}$, where $n \geq 3$. Construct a random graph $G = (V, E)$ by taking E as a random subset of size n of the set of all $\binom{n}{2}$ unordered pairs (v_i, v_j) with $i \neq j$. Let $i + j$ be the *weight* of the edge $(v_i, v_j) \in E$. Denote by X_i the number of neighbors of the vertex v_i, by W_i the sum of weights of all edges one of whose endpoints is v_i, and by W the total sum of weights of all edges of G. For example, if $n = 4$ and $E = \{(v_1, v_2), (v_1, v_4), (v_2, v_3), (v_3, v_4)\}$, then $X_2 = 2$, $W_2 = 8$ and $W = 20$.

   (a) $V(X_1) = \frac{2(n-3)}{n+1}$.

   (b) $E(W_i) = \frac{n(n+1)+2(n-2)i}{n-1}$.

   (c) Calculating the variance of $\sum_{i=1}^{n} X_i$ in two ways, one can show that $\text{Cov}(X_1, X_2) = -V(X_1)/n$.

   (d) $E(W) = n(n + 1)$.

   (e) A random edge, one of whose is endpoints is v_i, weighs on the average 1 more than a random edge one of whose endpoints is v_{i+1}. Therefore, $E(W_{i+1} - W_i) = E(X_1)$.

   (f) For sufficiently large n we have $\rho(X_i, W) \neq 0$ for every $i = 1, 2, \ldots, n$. (Hint: Do not calculate the correlation coefficient. Use intuition.)

   (g) None of the above.

2. One of the ideas raised in order to increase the number of visitors in DisneyRandomWorld, Beer Sheva, was to contribute to EMACS (Enhancing Mice And Cats Solidarity) organization a certain amount of

money, related to the number of visitors, as follows. Each day, a random minute will be chosen. All visitors arriving to DisneyRandomWorld during that minute will participate at an experiment, in which the success probability is p (where $0 < p < 1$). If at least one of them succeeds, DisneyRandomWorld will contribute $\$M$ to EMACS. Suppose the number of visitors arriving per minute is distributed $P(\lambda)$. Let X denote the number of people arriving during the selected minute and Y the amount (0 or M) of dollars contributed that day.

(a) $E(Y) = M\left(1 - e^{-\lambda p}\right)$.

(b) For every $\lambda > 0$ and $0 < p < 1$ we have $\sigma(Y) > M\left(1 - e^{-\lambda p}\right)$.

(c) The number of people arriving during the selected minute and succeeding in the experiment is distributed $P(\lambda p)$.

(d) Suppose after some statistical data has been gathered it turned out that the number of visitors arriving per minute is distributed $P(\lambda/2)$ (and not $P(\lambda)$ as supposed earlier), where the numbers of visitors in distinct minutes are independent. The park administration decided to select 2 random minutes each day and let all visitors arriving during those minutes to participate at the experiment (where again, $\$M$ will be donated if at least one of all those visitors succeeds). The combination of these changes makes no difference whatsoever for EMACS.

(e) The changes in (d) do not change the expected contribution to EMACS, but they do change the distribution of this contribution.

(f) Suppose the park administration decides to make the following change (with respect to the initial situation). Denoting $q = 1 - p$, if k people arrive during the selected minute, the first participates at an experiment with failure probability of $\frac{q}{2}$, the second – at an experiment with failure probability of $\frac{q}{3}$, ..., the kth – at an experiment with failure probability of $\frac{q}{k+1}$. If $\lambda = 2\ln 2,\ q = 1/2$ and $M = 10000$, then the expected increase in the contribution is less than \$2000 per day.

(g) None of the above.

3. The numbers $1,\ 2, \ldots,\ n$ are ordered randomly. Let X_i be 1 if i is to the left of 1 and 0 otherwise, $2 \le i \le n$. Denote by X the number

of numbers to the left of 1, and by S the sum of these numbers. For example, if $n = 4$ and the ordering is $2, 4, 1, 3$, then $X_2 = X_4 = 1$, $X_3 = 0$, $X = 2$ and $S = 6$.

(a) $X \sim B(n-1, 1/2)$.

(b) $V(X) = \frac{n^2-1}{12}$.

(c) $E(S) = \frac{(n+2)(n-1)}{2}$.

(d) X_i, X_j are independent for $i \neq j$.

(e) $\rho(X_i, S) < \rho(X_j, S)$ for $2 \leq i < j \leq n$.

(f) $\rho(X, S) = \sum_{i=2}^{n} \rho(X_i, S)$.

(g) None of the above.

4. A coin is tossed consecutively until it falls on a head. If the number of tosses has been odd – the experiment is finished; otherwise – the coin is tossed again until the next head shows up. If the number of tosses at the second stage is odd – the experiment is finished; otherwise – we move on to the third stage. We continue similarly until, for the first time, the number of tosses at that stage is odd. Denote by X the total number of times the coin is tossed, by Y – the number of times it is tossed at the last stage, and by S – the number of stages. For example, if the sequence of results is T,T,T,H,T,H,T,T,H, then the experiment is finished after the ninth toss, with $X = 9$, $Y = 3$ and $S = 3$.

(a) The three random variables X, Y, S are dependent. Out of the three pairs of variables we can generate from them, exactly one pair consists of independent variables.

(b) $V(S) = \frac{3}{4}$.

(c) $E(X) = 3$.

(d) $E(Y^2) = \frac{41}{9}$.

(e) $0 < \text{Cov}(X, Y) < \frac{8\sqrt{2}}{3}$.

(f) If the coin is not fair, but has a probability p (where $0 < p < 1$) for a head, then the distribution of S remains the same.

(g) None of the above.

5. The 2-dimensional random variable (X, Y) is uniformly distributed in the intersection $\{(x, y) : x^2+y^2 \leq 1\} \bigcap \{(x, y) : (1-x)^2+(1-y)^2 \leq 1\}$.

   (a) The probability for (X, Y) to obtain a value within the circle of radius ε around the point $(1/2, 1/2)$ is $\pi\varepsilon^2$ for sufficiently small $\varepsilon > 0$.

   (b) $E(X + Y) = 1$.

   (c) $P(X < 1/2) = 1/2$, yet X is not distributed $U(0, 1)$.

   (d) The distribution of $Y - X$ is symmetric around 0.

   (e) The random variables X and Y are dependent, yet uncorrelated.

   (f) $\rho(X, Y) < 0$.

   (g) None of the above.

Solutions

1. There are $\binom{n}{2}$ potential edges, out of which $n-1$ have v_1 as one of their endpoints and $\binom{n-1}{2}$ do not. As we choose n edges for E, we have $X_1 \sim H\left(n, n-1, \binom{n-1}{2}\right)$. In particular

$$E(X_1) = \frac{n(n-1)}{\binom{n}{2}} = 2$$

and

$$V(X_1) = \frac{n(n-1)\binom{n-1}{2}}{\binom{n}{2}^2}\left(1 - \frac{n-1}{\binom{n}{2}-1}\right) = \frac{2(n-3)}{n+1} \ .$$

Put, for $i \neq j$:

$$X_{ij} = \begin{cases} 1, & (v_i, v_j) \in E, \\ 0, & \text{otherwise.} \end{cases}$$

Then:

$$W_i = (1+i)X_{i1} + \ldots + (2i-1)X_{i,i-1} + (2i+1)X_{i,i+1} + \ldots + (n+i)X_{in} \ .$$

Since $E(X_{ij}) = P(X_{ij} = 1) = n/\binom{n}{2} = \frac{2}{n-1}$, this implies:

$$\begin{aligned} E(W_i) &= \tfrac{2}{n-1}\left((1+i) + \ldots + (2i-1) + (2i+1) + \ldots + (n+i)\right) \\ &= \tfrac{n(n+1)+2(n-2)i}{n-1} . \end{aligned}$$

Hence:

$$E(W_{i+1} - W_i) = E(W_{i+1}) - E(W_i) = \frac{2(n-2)}{n-1} < 2 = E(X_1).$$

(The reason that the argument in (d) is false is that one of the possible contributions to both W_i and W_{i+1} is through the edge (v_i, v_{i+1}), which contributes an equal amount to both. Hence $E(W_{i+1})$ exceeds $E(W_i)$ by slightly less than 2.)

As $W = \frac{1}{2}\sum_{i=1}^{n} W_i$, we have:

$$\begin{aligned} E(W) &= \frac{1}{2}\sum_{i=1}^{n} E(W_i) = \frac{n^2(n+1)}{2(n-1)} + \frac{2(n-2)}{2(n-1)}\sum_{i=1}^{n} i \\ &= \frac{n^3 + n^2 + (n-2)n(n+1)}{2(n-1)} = n(n+1). \end{aligned}$$

(Try to calculate $E(W)$ without using $E(W_i)$!)

Each edge contributes 2 to $\sum_{i=1}^{n} X_i$, so that

$$\sum_{i=1}^{n} X_i = 2n, \qquad V\left(\sum_{i=1}^{n} X_i\right) = 0.$$

On the other hand:

$$V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} V(X_i) + \sum_{1 \le i < j \le n} 2\mathrm{Cov}(X_i, X_j).$$

By symmetry, all X_i's have the same variance and all pairs X_i, X_j with $i \neq j$ have the same covariance. Hence

$$nV(X_1) + n(n-1)\mathrm{Cov}(X_1, X_2) = 0,$$

which yields

$$\mathrm{Cov}(X_1, X_2) = \frac{-V(X_1)}{n-1} .$$

A large value of X_i means that a (relatively) large part of W is due to edges, one of whose endpoints is v_i. For small i, such edges tend to have a small weight, while for large i – a large weight. In other words, $\rho(X_i, W)$ should be negative for small i and positive for large i. For i exactly in the middle of the range, namely $i = \frac{n+1}{2}$ (which is possible only for odd n), we should have $\rho(X_i, W) = 0$.

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

2. Denote by S the number of people arriving during the selected minute and succeeding in the experiment. Clearly $P_{X,S}(k,l) = \frac{\lambda^k}{k!}e^{-\lambda}\cdot\binom{k}{l}p^l q^{k-l}$, and therefore

$$\begin{aligned} P(S=l) &= \sum_{k=l}^{\infty} P_{X,S}(k,l) = \left(\frac{p}{q}\right)^l e^{-\lambda} \sum_{k=l}^{\infty} \frac{(\lambda q)^k}{k!}\binom{k}{l} \\ &= \left(\frac{p}{q}\right)^l e^{-\lambda} \cdot \frac{1}{l!} \sum_{k=l}^{\infty} \frac{(\lambda q)^k}{(k-l)!} = \left(\frac{p}{q}\right)^l e^{-\lambda} \cdot \frac{1}{l!}(\lambda q)^l \sum_{j=0}^{\infty} \frac{(\lambda q)^j}{j!} \\ &= \frac{(\lambda p)^l}{l!} e^{-\lambda} e^{\lambda q} = \frac{(\lambda p)^l}{l!} e^{-\lambda p} . \end{aligned}$$

Consequently, $S \sim P(\lambda p)$, and therefore

$$P(Y = M) = P(S > 0) = 1 - e^{-\lambda p} .$$

Hence

$$E(Y) = M\left(1 - e^{-\lambda p}\right)$$

and

$$E(Y^2) = M^2\left(1 - e^{-\lambda p}\right) .$$

Thus

$$V(Y) = M^2 e^{-\lambda p}\left(1 - e^{-\lambda p}\right) ,$$

which implies that $\sigma(Y) > M\left(1 - e^{-\lambda p}\right)$ if and only if $e^{-\lambda p} > 1 - e^{-\lambda p}$, namely if and only if $\lambda p < \ln 2$.

Under the changes in (d) we have $X = X_1 + X_2$, where $X_1, X_2 \sim P(\lambda/2)$ are independent. Since the sum of independent Poisson distributed random variables is again a Poisson distributed random variable, whose parameter is the sum of the parameters of the addends, we have $X \sim P(\lambda)$, so nothing changes statistically with respect to the original assumptions.

Let Y' be the contribution under the assumptions in (f). Then:

$$P(Y' = 0 \,|\, X = k) = \frac{q}{2} \cdot \frac{q}{3} \cdot \ldots \cdot \frac{q}{k+1} = \frac{q^k}{(k+1)!},$$

and consequently

$$P(Y' = M) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} \left(1 - \frac{q^k}{(k+1)!}\right) = 1 - e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda q)^k}{k!(k+1)!}.$$

With the values $\lambda = 2\ln 2$, $q = 1/2$ and $M = 10000$, the expected original contribution is

$$E(Y) = M\left(1 - e^{-\lambda p}\right) = 10000\left(1 - e^{-\ln 2}\right) = 5000,$$

while the expected contribution after the change is

$$\begin{aligned} E(Y') &= 10000\left(1 - \frac{1}{4} \sum_{k=0}^{\infty} \frac{(\ln 2)^k}{k!(k+1)!}\right) \\ &\leq 10000\left(1 - \frac{1}{4} - \frac{1}{4} \cdot \frac{\ln 2}{2}\right) = 6634. \end{aligned}$$

Hence the expected increase in the contribution is less than \$1634.

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

3. The number has equal probabilities of being placed at each of the n possible places, whence $X \sim U[0, n-1]$. In particular, $V(X) = \frac{n^2-1}{12}$.

   Each $i \geq 2$ has the same probability of being placed before 1 as of being placed after it. Hence $X_i \sim B(1, 1/2)$. Since

$$S = \sum_{i=2}^{n} iX_i,$$

we have

$$E(S)=\sum_{i=2}^{n} iE(X_i)=\frac{1}{2}\sum_{i=2}^{n} i=\frac{(n+2)(n-1)}{4}\,.$$

By symmetry, each of the $3! = 6$ orderings of the numbers 1, i and j in the sequence is equi-probable, so that

$$P(X_i=X_j=1)=\frac{2}{6}=\frac{1}{3}\neq\frac{1}{4}=P(X_i=1)P(X_j=1),$$

and thus X_i, X_j are dependent.

Now:

$$\begin{aligned}\mathrm{Cov}(X_i,S) &= \sum_{j=2}^{n} j\mathrm{Cov}(X_i,X_j)=\sum_{j\neq i} j\cdot\left(\frac{1}{3}-\frac{1}{4}\right)+i\cdot\frac{1}{4}\\ &= \frac{(n+2)(n-1)+4i}{24}\,.\end{aligned}$$

Since the right hand side increases with i, while $V(X_i)$ is the same for each i, this implies that $\rho(X_i, S)$ increases with i. (Note that this is also intuitively clear.)

Since

$$\mathrm{Cov}(X,S)=\sum_{i=2}^{n}\mathrm{Cov}(X_i,S)$$

and

$$V(X)=\frac{n^2-1}{12}>\frac{1}{4}=V(X_i),\qquad n\geq 3,$$

we have

$$\rho(X,S)<\sum_{i=2}^{n}\rho(X_i,S),\qquad n\geq 3.$$

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

4. When a coin is tossed until falling on a head for the first time, the probability for the number of tosses to be odd is

$$\sum_{n=0}^{\infty}\frac{1}{2^{2n+1}}=\frac{2}{3}\,.$$

Hence $S \sim G(2/3)$, and in particular $E(S) = 3/2$ and $V(S) = 3/4$. By the same considerations, if the coin has a probability of p for a head, then $S \sim G\left(\frac{1}{1+q}\right)$.

The total number of tosses X is the number of tosses until the first time a head shows up at an odd-numbered toss. Hence, letting X' denote the number of odd-numbered tosses, we have $X = 2X' - 1$. Obviously, $X' \sim G(1/2)$, so that

$$E(X') = 2, \qquad V(X') = 2$$

and therefore

$$E(X) = 2E(X') - 1 = 3, \qquad V(X) = 2^2 \cdot V(X') = 8.$$

To find the distribution of Y, we note that it is obtained from a $G(1/2)$-distributed variable, conditioned on the event that the number of tosses is odd. Consequently

$$P(Y = 2k - 1) = \frac{1/2^{2k-1}}{2/3} = \frac{3}{2^{2k}} ,$$

which implies that $Y = 2Y' - 1$, with $Y' \sim G(3/4)$. Thus

$$E(Y') = \frac{4}{3} , \qquad V(Y') = \frac{4}{9} ,$$

and

$$E(Y) = 2E(Y') - 1 = \frac{5}{3} , \quad V(Y) = 2^2 \cdot V(Y') = \frac{16}{9} , \quad E(Y^2) = \frac{41}{9} .$$

The number of stages in the experiment does not give any indication regarding the length of the last stage. (Note that it does give information regarding the length of any specific stage. For example, if it is known that $S = 3$, then the lengths of the first two stages must be even and that of the third must be odd. However, it does not give further information, and in particular gives nothing about the length of the last stage.) Hence Y, S are independent. On the other hand, the inequalities $X \geq S$ and $X \geq Y$, for example, show that each of the two pairs X, S and X, Y consists of dependent variables.

The more tosses there are at the last stage, the more there are altogether. Hence, the larger Y is, the larger X should be, so that $\mathrm{Cov}(X, Y) > 0$. On the other hand, they do not completely determine each other, and in particular they are not linear functions of each other, and thus

$$\mathrm{Cov}(X, Y) < \sqrt{V(X)V(Y)} = \frac{8\sqrt{2}}{3} .$$

(It is actually pretty easy to calculate $\mathrm{Cov}(X, Y)$ directly using the observation that the two variables $X - Y, Y$ are independent. Try it!)

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

5. Denote the intersection of the circles in the question by D. Since (X, Y) is uniformly distributed in D, its probability to obtain a value within some region is the area of that region divided by the area of D. This area is easily seen to be $1 - 2(1 - \pi/4) = \pi/2 - 1$, and hence the probability of the event in (a) is $\frac{\pi\varepsilon^2}{\pi/2-1}$.

   The region D is symmetric with respect to the line $x + y = 1$ in the plane. Therefore, the distribution of the random variable $X + Y$ is symmetric with respect to the point 1, and in particular (and since $X + Y$ is bounded) $E(X + Y) = 1$. Similarly, since D is symmetric with respect to the line $y = x$, the distribution of $Y - X$ is symmetric around 0.

   The density function f_X of X at any point x is the height of the fiber of D lying above the point x. Hence f_X is symmetric with respect to the point $1/2$, and in particular $P(X < 1/2) = 1/2$. Since f_X is not constant on $[0, 1]$, X is not distributed $U(0, 1)$.

   Obviously, as X assumes a larger value, Y tends to become smaller. Hence $\rho(X, Y) < 0$.

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