Final #2 – Questions 4-6

Mark all correct answers in each of the following questions.

4. A carpenter has a batch of n wooden sticks of unit length each. He breaks each stick at a random point, so that the distance X of the breaking point from the left endpoint is distributed $U(0,1)$. Then he takes out of each pair of pieces the long one, and combines all these pieces into a single stick of length L. Similarly, he combines all small pieces into one stick of length S. (For example, if $n = 3$, and the breaking points are at distances of $0.2, 0.4$ and 0.7 from the left endpoints of the three sticks, then $L = 0.8+0.6+0.7 = 2.1$ and $S = 0.2+0.4+0.3 = 0.9$.)

   (a) When breaking each of the initial sticks, the length of the long piece is distributed according to the following distribution function:

   $$F(x) = \begin{cases} 0, & x < 1/2, \\ 4(x-1/2)^2, & 1/2 \le x \le 1, \\ 1, & x > 1 \end{cases}$$

   (b) $E(S) = \frac{n}{4}$.

   (c) $E\left(\frac{S}{L}\right) = \frac{1}{3}$.

   (d) Let X be the number of sticks, out of n, such that, when broken, form a large piece of length at least 0.9 and a small piece of length at most 0.1. Let Y be the total length of the large pieces generated out of these X sticks. Then X, Y are uncorrelated but dependent.

   (e) The normal approximation gives, for $n = 19200$,

   $$P(L \ge 14360) \approx 0.84.$$

5. The random variable (X, Y) is uniformly distributed in the region:

$$S = \{(x, y) : 0 \leq x \leq \pi/4, \sin x \leq y \leq \operatorname{tg} x\}.$$

Namely, denoting by s the area of S, the probability for (X, Y) to assume values in a sub-region $S' \subseteq S$ is area$(S')/s$. You may verify that $s = \frac{1}{2}\ln 2 + \frac{\sqrt{2}}{2} - 1$.

(a) The distribution function of X is:

$$F_X(x) = \begin{cases} 0, & x < 0, \\ \frac{-\ln\cos x + \cos x - 1}{s}, & 0 \leq x \leq \frac{\pi}{4}, \\ 1, & x > \frac{\pi}{4}. \end{cases}$$

(b) The density function of Y is:

$$f_Y(y) = \begin{cases} \frac{\arcsin y - \operatorname{arctg} y}{s}, & 0 \leq y \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

(c) $P(Y \geq X) = 1/2$.

(d) $\rho(X, Y) > 0$. (Hint: Do not calculate it exactly.)

(e) Let $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ be independent random variables, all distributed as (X, Y). For $0 \leq k \leq n$, denote by I_k the number of indices j in the range from 1 to k for which $Y_j \geq X_j$. Suppose n is even. Then:

$$P\left(\min_{0\leq k\leq n}\left(I_k - \frac{k}{2}\right) = 0 \middle| I_n = \frac{n}{2}\right) = \frac{1}{n}.$$

6. (a) The random variable X assumes all values $\pm 1/2^n, n = 0, 1, 2, \ldots$, with probabilities:

$$P(X = 1/2^n) = P(X = -1/2^n) = 1/2^{n+2}, \qquad n = 0, 1, 2, \ldots.$$

Then F_X is continuous at the point 0.

(b) A gambler tosses a coin until the upface shows T. Denote by X the number of tosses. If X is even, the player wins 2^X shekels, while if it is odd, then he needs to pay 2^X shekels. Then the expected value of his winnings is 0.

(c) X is a random variable with finite expectation and variance. S is a random variable, assuming the values 1 and -1, with probability $1/2$ each. It is known that X, S are independent. Then X and SX may be dependent, but in any case are uncorrelated.

(d) X is a random variable with finite variance. X_1, X_2 are independent random variables, each distributed as X. Then:

$$E\left((X_1 - X_2)^2\right) = V(X).$$

(e) $(X_n)_{n=1}^{\infty}$ is a sequence of independent random variables, with the same expectation μ and the same variance σ^2 to all of them. Then:

$$V\left(\frac{X_1 + 2X_2 + 3X_3 + \ldots + nX_n}{n^{3/2}}\right) \underset{n\to\infty}{\longrightarrow} 2\sigma^2.$$

Solutions

4. (a) Denote by L_i the length of the long piece of the i-th stick. Clearly, L_i assumes values between $1/2$ and 1. For $1/2 \le x \le 1$, we have $L_i \le x$ if and only if the breaking point is at a distance of at least $1 - x$ and at most x from the left endpoint of the stick. Hence:

$$F(x) = \begin{cases} 0, & x < 1/2, \\ 2x - 1, & 1/2 \le x \le 1, \\ 1, & x > 1. \end{cases}$$

(b) The formula for the distribution function of L_i shows that $L_i \sim U(1/2, 1)$. Hence $E(L_i) = 3/4$. Since $L = \sum_{i=1}^{n} L_i$, we obtain $E(L) = 3n/4$. Now $S = n - L$, so that $E(S) = n/4$.

(c) The claim is false already for $n = 1$. Indeed, in this case we have $S/L \le t$ (for $0 \le t \le 1$) if $\frac{1-L}{L} \le t$, which is equivalent to $L \ge \frac{1}{1+t}$. It follows that

$$F_{S/L}(t) = \begin{cases} 0, & t < 0, \\ 1 - F_L\left(\frac{1}{1+t}\right), & 0 \le t \le 1, \\ 1, & t > 1 \end{cases}$$

Consequently:

$$E(S/L) = \int_0^\infty (1 - F_{S/L}(t))dt = \int_0^1 \left(2 \cdot \frac{1}{1+t} - 1\right) dt$$
$$= [2\ln(1+t) - 1]_{t=0}^1 = 2\ln 2 - 1.$$

(d) We clearly have $X = \sum_{i=1}^n X_i$ and $Y = \sum_{i=1}^n Y_i$, where:

$$X_i = \begin{cases} 1, & 0.9 \le L_i < 1, \\ 0, & \text{otherwise}, \end{cases} \qquad Y_i = \begin{cases} L_i, & 0.9 \le L_i < 1, \\ 0, & \text{otherwise}, \end{cases}$$

for $1 \le i \le n$. It follows easily that $E(X_i) = 0.2$ and $E(Y_i) = 0.19$. Therefore:

$$E(X)E(Y) = 0.2n \cdot 0.19n = 0.038n^2.$$

Now:

$$E(XY) = \sum_{i=1}^n \sum_{j=1}^n E(X_i Y_j).$$

Split the sum into two sub-sums, one formed by all pairs of indices (i, j) with $i \ne j$ and the other by those with $i = j$. Since $X_i Y_i = Y_i$ for each i, and X_i, Y_j are independent for $i \ne j$, we have:

$$E(XY) = n(n-1) \cdot 0.2 \cdot 0.19 + n \cdot 0.19 = 0.038n^2 + 0.152n.$$

Since $E(XY) > E(X)E(Y)$, the variables X, Y are positively correlated. (Note that the result is very intuitive; the larger X is, the more larger pieces there are, and therefore their total length should be expected to be larger.)

(e) We have seen earlier that $E(L_i) = 3/4$, and we similarly have $V(L_i) = \frac{(1-1/2)^2}{12} = \frac{1}{48}$. Hence:

$$P(L \ge 14360) = P\left(\sum_{i=1}^{19200} L_i \ge 14360\right)$$
$$= P\left(\frac{\sum_{i=1}^{19200} L_i - 19200 \cdot \frac{3}{4}}{\sqrt{19200 \cdot \frac{1}{48}}} \ge \frac{14360 - 19200 \cdot \frac{3}{4}}{\sqrt{19200 \cdot \frac{1}{48}}}\right).$$

The normal approximation gives:

$$P(L \geq 14360) \approx P(Z \geq -2),$$

where Z is a standard normal variable. Thus, the required probability is approximately 0.977.

Thus, only (b) is true.

5. (a) The area between the curves $y = \sin t$ and $y = \operatorname{tg} t$, from $t = 0$ up to $t = x$, is:

$$\int_0^x (\operatorname{tg} t - \sin t)dt = [-\ln\cos t + \cos t]_{t=0}^{x} = -\ln\cos x + \cos x - 1.$$

Hence the distribution function of X is:

$$F_X(x) = \begin{cases} 0, & x < 0, \\ \dfrac{-\ln\cos x + \cos x - 1}{s}, & 0 \leq x \leq \frac{\pi}{4}, \\ 1, & x > \frac{\pi}{4}. \end{cases}$$

(b) The region S may be represented in the form:

$$\begin{aligned} S =& \{(x,y) : 0 \leq y \leq \sqrt{2}/2,\ \operatorname{arctg} y \leq x \leq \arcsin y\} \\ & \cup \{(x,y) : \sqrt{2}/2 \leq y \leq 1,\ \operatorname{arctg} y \leq x \leq \pi/4\}. \end{aligned}$$

It follows that:

$$f_Y(y) = \begin{cases} \dfrac{\arcsin y - \operatorname{arctg} y}{s}, & 0 \leq y \leq \frac{\sqrt{2}}{2}, \\ \dfrac{\pi/4 - \operatorname{arctg} y}{s}, & \frac{\sqrt{2}}{2} < y \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

(c) The probability of the event $\{Y \geq X\}$ is given by the ratio of the area of the subset of S, consisting of those points (x,y) satisfying the condition $y \geq x$, and the total area of S. The first area is given by

$$\int_0^{\pi/4} (\operatorname{tg} x - x)dx = \left[-\ln\cos x - \frac{x^2}{2}\right]_{x=0}^{\pi/4} = \frac{1}{2}\ln 2 - \frac{\pi^2}{32}.$$

Hence:

$$P(Y \geq X) = \frac{16 \ln 2 - \pi^2}{16 \ln 2 + 16\sqrt{2} - 32} \neq \frac{1}{2}\,.$$

(d) The curves $y = \operatorname{tg} x$ and $y = \sin x$, bounding the region S from above and from below, respectively, both grow with x. Hence, as the random variable X assumes larger values, the random variable Y tends to assume larger values as well. Thus, $\rho(X, Y) > 0$.

(e) The required probability is the conditional probability that, given that $Y_i \geq X_i$ for exactly half of the indices i, no initial subsequence has the property that more i's satisfy the inverse inequality up to that point. This question is equivalent to the one asked in the ballot problem, and consequently the required probability is $\frac{1}{n/2+1} = \frac{2}{n+2}$.

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

6. (a) Since $P(X = 0) = 0$, the function F_X is continuous at 0. Let us also show it directly in this case. Since X is symmetric around 0, we have $F_X(0) = 1/2$. We have to show that $F_X(x)$ may be made arbitrarily close to $1/2$ by taking x sufficiently close to 0. In fact, take, for example, $x < 0$. If $x > -1/2^m$ for some non-negative integer m, then

$$F_X(x) \geq \sum_{k=0}^{m} \frac{1}{2^{k+2}} = \frac{1}{2} - \frac{1}{2^{m+2}}.$$

The right-hand side converges to $1/2$ as $m \to \infty$, which proves our claim.

(b) Let Y denote the gambler's winnings. The series defining $E(Y)$ may be written in this case in the form

$$\sum_{n=1}^{\infty} (-2)^n \cdot \frac{1}{2^n}.$$

The series does not converge, so that $E(Y)$ does not exist. (In fact, $|Y|$ is the random variable arising is St. Petersburg Paradox. We have shown in class that $E(|Y|)$ is infinite, and therefore $E(Y)$ does not exist as well.)

(c) X and SX are indeed usually dependent, as by knowing the value of X we know that of SX up to sign. However, due to the independence of X and S, and since $E(S) = 0$, we have

$$\begin{aligned}\mathrm{Cov}(X, SX) &= E(SX^2) - E(X)E(SX) \\ &= E(S)E(X^2) - E(S)E^2(X) = 0.\end{aligned}$$

(d) A routine calculation yields:

$$\begin{aligned}E\left((X_1 - X_2)^2\right) &= E\left(X_1^2 - 2X_1X_2 + X_2^2\right) \\ &= E\left(X^2\right) - 2E(X)E(X) + E\left(X^2\right) = 2V(X).\end{aligned}$$

(e) Since the X_i's are independent, so are the iX_i's, and therefore:

$$\begin{aligned}V\left(\frac{X_1 + 2X_2 + \ldots + nX_n}{n^{3/2}}\right) &= \frac{1}{n^3}\left(V(X_1) + \ldots + V(nX_n)\right) \\ &= \frac{\sigma^2 + 2^2\sigma^2 + \ldots + n^2\sigma^2}{n^3} \\ &= \frac{n(n+1)(2n+1)}{6n^3}\sigma^2 \underset{n\to\infty}{\longrightarrow} \frac{\sigma^2}{3}.\end{aligned}$$

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