Final #1 – Questions 4-6

Mark all correct answers in each of the following questions.

4. Two players – A and B – play a card game. In the beginning, each of them has a deck of $n \geq 2$ cards, marked by the numbers $1, 2, \ldots, n$. The game consists of n stages. At each stage, A picks a random card from B's pile, and simultaneously B picks a random card from A's pile. (Thus, at the end of each stage, each player still has n cards.) For $1 \leq i, k \leq n$, let X_i be the number on the card which A picks from B at the i-th stage, Y_i be the number on the card which B picks from A at the same stage, and M_k be the number of times a card marked by k is moved from one player to the other. Finally, let N be the number of indices i between 1 and n for which $X_i = Y_i$. (For example, if $n = 3$, at the first stage A picks 2 and B picks 2, at the second stage A picks 2 and B picks 3, and at the third stage A picks 3 and B picks 2, then $X_1 = X_2 = Y_1 = Y_3 = 2, X_3 = Y_2 = 3, M_1 = 0, M_2 = 4, M_3 = 2, N = 1$.)

   (a) $M_k \sim B(2n, 1/n)$ for each $1 \leq k \leq n$.

   (b) If n sufficiently large, then $P(M_k = 0) \approx e^{-2}$ for each $1 \leq k \leq n$.

   (c) If n sufficiently large, then $P(M_k = 1) \approx e^{-2}$ for each $1 \leq k \leq n$.

   (d) $E(N) < 1$.

   (e) The events $E_1 = \{X_i = Y_i, 1 \leq i \leq n\}$ and $E_2 = \{M_1 = 0\}$ are independent.

5. Let $S = \operatorname{tg} T$, where $T \sim U(0, \pi/4)$. Let $S_1, S_2, \ldots, S_n$ be independent random variables, all having the same distribution as S, and set $W = S_1 + S_2 + \ldots + S_n$. Let M and N be the numbers of indices i between 1 and n for which $\frac{\sqrt{3}}{3} \leq S_i \leq 1$ and $0 \leq S_i \leq \frac{1}{1000}$, respectively.

(a) $2S$ has the same distribution as the variable $|V|$, where V is Cauchy distributed. In particular, S does not have a mean.

(b) Markov's inequality implies:

$$P\left(W \geq \frac{n}{2}\right) \leq \frac{4\ln 2}{\pi}\,.$$

(c) Let $n = 900$. Approximation by the central limit theorem implies

$$P(M \in [250, 350]) \geq 0.95.$$

(d) For $n = 1000$:

$$P(N = 3) \approx \frac{1}{6e}\,.$$

(e) Let (X, Y) be a pair of random variables with joint density function f, defined by

$$f(x,y) = \begin{cases} \frac{c}{(1+x^2)(1+y^2)}, & 0 \leq x \leq \frac{1}{3},\ \frac{1}{3} \leq y \leq 1\,, \\ \frac{c}{(1+x^2)(1+y^2)}, & \frac{1}{3} \leq x \leq 1,\ 0 \leq y \leq \frac{1}{3}\,, \\ 0, & \text{otherwise}, \end{cases}$$

where c is the unique constant for which $f(x, y)$ is a 2-dimensional density function. Then the random variable X is Cauchy distributed.

6. (a) Let X, Y be continuous random variables with density functions f_X, f_Y. It is given that $f_X(t) = 2f_Y(t)$ for every $t > 0$, while $f_X(t) = \frac{1}{2}f_Y(t)$ for every $t < 0$. Then $P(X > 0) = 2/3$.

(b) A gambler has to participate in one of two gambles, in the first of which his winnings are governed by the random variable X and in the second by Y. The variables X and Y satisfy the properties assumed in the preceding part. Then he should prefer the gamble where his winnings are X.

(c) The same question as in the preceding part, where X and Y have the distribution functions F_X and F_Y, defined by:

$$F_X(x) = \begin{cases} 0, & x < 0, \\ x^2, & 0 \le x \le 1, \\ 1, & x > 1, \end{cases} \qquad F_Y(y) = \begin{cases} 0, & y < 0, \\ y^3, & 0 \le y \le 1, \\ 1, & y > 1. \end{cases}$$

(d) If X is a non-constant random variable, and there exists a constant c for which $P(|X| < c) = 1$, then $\rho(X^n, X^{n+1}) \ge 0$ for every positive integer n.

(e) Let X, Y be continuous random variables with density functions f_X, f_Y, respectively, which are both known to be positive in the interval $[0, 1]$ and to vanish outside it. If the functions f_X, f_Y are very close, then $\rho(X, Y)$ is very close to 1. More precisely, for every $\varepsilon > 0$ there exists a $\delta > 0$ such that, if $|f_X(t) - f_Y(t)| < \delta$ for every $t \in \mathbf{R}$, then $\rho(X, Y) > 1 - \varepsilon$.

Solutions

4. (a) It is true that each M_k is a sum of $2n$ random variables, assuming the values 0 and 1. Indeed, by considering each of the $2n$ card drawings, and taking the random variable which is 1 if a card marked by k was drawn that time and 0 otherwise, then M_k is the sum of these variables. Moreover, by symmetry, the *a priori* probability of each of these variables to assume the value 1 is $1/n$. However, the variables are dependent, so that the above does not imply that $M_k \sim B(2n, 1/n)$. Indeed, already for $n = 2$ one can easily verify that $P(M_1 = 1) = 1/8$, whereas this probability would be $1/4$ if M_1 was $B(2n, 1/n)$-distributed.

(b) The event $\{M_k = 0\}$ occurs if, in each of the $2n$ drawings in the game, a card other than k is chosen. This is easily seen to imply:

$$P(M_k = 0) = \left(1 - \frac{1}{n}\right)^{2n} \underset{n\to\infty}{\longrightarrow} e^{-2}.$$

(c) The event $\{M_k = 1\}$ occurs if, for some i between 1 and n, within the first $i-1$ stages none of the players draws one of the two cards marked by k, at the i-th stage exactly one of them draws one of these two cards, and in the last $n-i$ stages again none of these two cards is drawn. Noting that the (conditional) probability for this to happen during the last $n-i$ stages is $(1-2/n)^{n-i}$, we obtain:

$$P(M_k = 1) = \sum_{i=1}^{n} \left(1 - \frac{1}{n}\right)^{2(i-1)} \cdot 2 \cdot \frac{1}{n} \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right)^{n-i} . \tag{1}$$

Now on the one hand

$$\begin{aligned} \left(1 - \frac{1}{n}\right)^{2(i-1)} \left(1 - \frac{2}{n}\right)^{n-i} &\le \left(1 - \frac{1}{n}\right)^{2(i-1)} \left(\left(1 - \frac{1}{n}\right)^{2}\right)^{n-i} \\ &= \left(1 - \frac{1}{n}\right)^{2(n-1)} \xrightarrow[n\to\infty]{} e^{-2} , \end{aligned}$$

and on the other hand

$$\begin{aligned} \left(1 - \frac{1}{n}\right)^{2(i-1)} \left(1 - \frac{2}{n}\right)^{n-i} &\ge \left(1 - \frac{2}{n}\right)^{i-1} \left(1 - \frac{2}{n}\right)^{n-i} \\ &= \left(1 - \frac{2}{n}\right)^{n-1} \xrightarrow[n\to\infty]{} e^{-2} . \end{aligned}$$

Using the last two observations in (1), we get

$$P(M_k = 1) \xrightarrow[n\to\infty]{} 2e^{-2} .$$

(d) Obviously, $N = \sum_{i=1}^{n} N_i$, where

$$N_i = \begin{cases} 1, & X_i = Y_i, \\ 0, & \text{otherwise,} \end{cases}$$

and therefore

$$E(N) = \sum_{i=1}^{n} P(X_i = Y_i).$$

Now let S_i denote the event whereby, at the beginning of the i-th stage, each of A and B has exactly one of the cards marked by 1, exactly one of the cards marked by 2, and so forth. Clearly, $P(X_i = Y_i | S_i) = 1/n$, while $P(X_i = Y_i | \bar{S}_i) < 1/n$. Since $P(S_i) < 1$ for $i \geq 2$, we get

$$P(X_i = Y_i) = P(S_i)P(X_i = Y_i | S_i) + P(\bar{S}_i)P(X_i = Y_i | \bar{S}_i) < 1/n.$$

Thus $E(N) < 1$.

(e) The event E_1 occurs if, at each stage, player B picks from A's cards the counterpart of the card picked by A from B's cards. Hence:

$$P(E_1) = \left(\frac{1}{n}\right)^n .$$

As explained above:

$$P(E_2) = \left(1 - \frac{1}{n}\right)^{2n} .$$

The event $E_1 \cap E_2$ occurs if, at each stage, player B picks from A's cards the counterpart of the card picked by A from B's cards and, moreover, these two cards are marked by one of the numbers between 2 and n. It follows that

$$P(E_1 \cap E_2) = \left(\frac{n-1}{n^2}\right)^n \neq P(E_1)P(E_2),$$

and thus E_1 and E_2 are dependent.

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

5. (a) The variable S assumes values only in the interval $[0, 1]$, whereas $|V|$ may assume any non-negative value. Hence $2S$ and $|V|$ cannot have the same distribution.

(b) Clearly, F_S vanishes on $(-\infty, 0)$ and is identically 1 on $(1, \infty)$. In between we have

$$F_S(s) = P(S \leq s) = P(T \leq \text{ arctg } s) = \frac{\text{arctg } s}{\pi/4} , \qquad 0 \leq s \leq 1,$$

so that

$$
\begin{aligned}
E(S) =& \int_0^1 \left(1 - \frac{4}{\pi} \operatorname{arctg} s\right) \\
=& 1 - \left[\frac{4}{\pi} s \operatorname{arctg} s\right]_0^1 + \frac{4}{\pi} \int_0^1 \frac{s}{1+s^2} ds \\
=& 1 - \frac{4}{\pi} \operatorname{arctg} 1 + \frac{4}{\pi} \left[\frac{1}{2} \ln(1+s^2)\right]_0^1 = \frac{2\ln 2}{\pi} .
\end{aligned}
$$

It follows that

$$E(W) = nE(S) = \frac{2n \ln 2}{\pi} .$$

Markov's inequality implies therefore:

$$P\left(W \geq \frac{n}{2}\right) \leq \frac{E(W)}{n/2} = \frac{4 \ln 2}{\pi} .$$

(c) Since

$$P\left(\frac{\sqrt{3}}{3} \leq S \leq 1\right) = 1 - \frac{\operatorname{arctg} \sqrt{3}/3}{\pi/4} = \frac{1}{3} ,$$

the random variable M is distributed $B(900, 1/3)$. Approximation by the central limit theorem, for $n = 900$, yields

$$
\begin{aligned}
P(M \in [250, 350]) = P &\left(\frac{250 - 300}{\sqrt{900 \cdot \frac{1}{3} \cdot \frac{2}{3}}} \leq \frac{M - 300}{\sqrt{900 \cdot \frac{1}{3} \cdot \frac{2}{3}}} \leq \frac{350 - 300}{\sqrt{900 \cdot \frac{1}{3} \cdot \frac{2}{3}}}\right) \\
&\approx P(-3.54 \leq Z \leq 3.54),
\end{aligned}
$$

where $Z \sim N(0, 1)$. The right-hand side is very close to 1, and therefore the required probability is certainly at least 0.95.

(d) The probability for S to assume values in the interval $[0, 1/1000]$ is $\frac{\operatorname{arctg} 1/1000}{\pi/4} \approx \frac{4}{1000\pi}$. Hence N is very close to a $B\left(1000, \frac{4}{1000\pi}\right)$ distributed variable. By the Poissonian approximation of the binomial distribution, N is distributed approximately $P\left(\frac{4}{\pi}\right)$. In particular:

$$P(N = 3) \approx \frac{(4/\pi)^3}{3!} e^{-4/\pi} .$$

(e) The random variable X assumes only values between 0 and 1, while a Cauchy distributed variable may assume values on the whole real line. Hence X cannot possibly be Cauchy distributed.

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

6. (a) We have

$$P(X > 0) = \int_0^\infty f_X(t)dt = \int_0^\infty 2f_Y(t)dt = 2P(Y > 0),$$

and

$$P(X < 0) = \int_{-\infty}^0 f_X(t)dt = \int_{-\infty}^0 \frac{1}{2}f_Y(t) = \frac{1}{2}P(Y < 0).$$

Replacing $P(X < 0)$ and $P(Y < 0)$ by $1 - P(X > 0)$ and $1 - P(Y > 0)$, respectively, we obtain two equations in the two unknowns $P(X > 0)$ and $P(Y > 0)$. Solving these equations, we obtain $P(X > 0) = 2/3$ and $P(Y > 0) = 1/3$.

(b) Integrating the two density functions f_X and f_Y, we find that $F_X(t) \leq F_Y(t)$ for every $t \in \mathbf{R}$, which implies that the first gamble is at least as good as the second.

(c) This time we have $F_X(t) \geq F_Y(t)$ for every $t \in \mathbf{R}$, with strict inequality on the interval $(0, 1)$. Hence the first gamble is worse than the second.

(d) Recall that two random variables are positively correlated if large values for one indicate large values of the other, and are negatively correlated if large values for one indicate small values of the other. Hence, to obtain a counter-example, it is natural to take a random variable X which assumes negative values only. To be specific, one may let X assume, say, the two values -1 and -2 with probabilities $1/2$ each. By a routine calculation, we verify that $\rho(X^n, X^{n+1}) < 0$ for every positive integer n.

(e) The condition whereby f_X, f_Y are very close gives no information regarding the issue of dependence/independence of X and Y. Thus, for example, let X and Y be independent identically distributed. The density functions f_X, f_Y are identical in this case, yet X and Y are uncorrelated. (As a more extreme example, one may take X as any continuous random variable with a density function f, which is symmetric around 0, say $X \sim U(-1,1)$ or $X \sim N(0,\sigma^2)$, and put $Y = -X$. Then the density functions f_X, f_Y are identical, but $\rho(X,Y) = -1$.)

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