Test Probability Theory for TCS-BIT, June 12, 2020.

You may consult the reader and your own notes. A conventional calculator is allowed, but a programmable calculator is not. Always provide your arguments. For your convenience, a formula sheet and the normal table is provided with this test.

Read the following carefully: By testing you remotely in this fashion, we express our trust that you will adhere to the ethical standard of behaviour expected of you. This means that we trust you to answer the questions and perform the assignments in this test to the best of your own ability, without seeking or accepting the help of any source that is not explicitly allowed by the conditions of this test.

We ask you to copy the following statement to your first answering sheet and sign it with your name and student number:"*I will make this test to the best of my own ability, without seeking or accepting the help of any source not explicitly allowed by the conditions of the test.*"

1. Given the following joint probability distribution, i.e. the probabilities $P(X = x \text{ and } Y = y)$ of two random variables X and Y:

$x \setminus y$	0	1	2
0	1/9	1/9	1/9
1	0	1/9	2/9
2	2/9	1/9	0

   a) Determine the probability distribution of X, and further find $P(X \geq 1)$, $E(X)$, and $\text{Var}(X)$.
   b) Determine the correlation between X and Y.
   c) Are X and Y independent? Motivate your answer.
   d) Determine the conditional distribution of Y given $X = 1$ and calculate $E[Y|X = 1]$.

2. If X is $N(\mu, \sigma^2)$-distributed, then we say that $Z = e^X$ is log-normal(μ, σ^2)-distributed. These kind of random variables appear in financial and economic models.

a) Find the density function of Z (i.e. f_Z).

b) For the special case $\mu = 3$ and $\sigma^2 = 9$, find $P(1/2 \leq Z \leq 2)$.

c) Let X_1 and X_2 be independent and respectively $N(\mu_1, \sigma_1^2)$-distributed and $N(\mu_2, \sigma_2^2)$-distributed. Defining $Z_i = e^{X_i}$, for $i \in \{1, 2\}$, find the density of the product random variable $Z_1 \cdot Z_2$.

3. An empty train, consisting of three wagons, arrives at a station where eight people are waiting. These passengers choose at random, and independently of each other, which wagon to board.

   a) What is the probability that nobody boards in the first wagon?
   b) What is the probability that at most two people (i.e. 0, 1 or 2) end up in the second wagon?

c) Let X_1, X_2, X_3 denote respectively de number of people in the first, second, and third wagons. Are X_1, X_2, X_3 independent? Compute the probability of the event $A = \{X_1 = 2 \text{ and } X_2 = 4 \text{ and } X_3 = 2\}$.

4. A group of ten friends meets at the beach to play volleyball. Six of the ten friends are considered good players. In a usual beach volleyball match, two players play against two players. If these four players are chosen completely at random from the group of friends, what is the probability that the match will be fair, i.e. that both teams have the same number of good players?

5. Assume that the time (in years) between two consecutive iPhone releases is a random variable which is exponentially distributed with parameter $\lambda = 1$. These random variables are assumed to be independent of each other.

   a) Let S_n denote the total time until the n-th release of an iPhone (i.e. the iPhone n). Justify that the event "$S_n \leq 75$" is equal to the event "there are at least n iPhone releases in 75 years".
   b) Compute or approximate the probability of the event: "there are at least 60 iPhone releases in 75 years".

6. For the following questions, answer "True" or "False". If you answer "True", provide a proof, and if you answer "False", provide a simple counterexample.

   a) If two events A and B are independent, then they are mutually exclusive (i.e. disjoint).
   b) If $P(A|B) = P(A)$, then $P(B|A) = P(B)$.
   c) If X and Y are two random variables such that $E[XY] = 0$, then they are independent random variables.
   d) If X and Y are independent and positive random variables, then $E[X^Y] = E[X]^{E[Y]}$.
   e) If X is a random variable with $P[X \geq 2] > 0$, then $P[X \geq 3 | X \geq 2] \geq P[X \geq 3]$.

Conversion table for the exercise numbering:

This test is Version 6	1	2	3	4	5	6
Numbering in Version 3	4	6	5	2	1	3
Numbering in Version 7	4	5	6	3	2	1

Points:

Question	1a	1b	1c	1d	2a	2b	2c	3a	3b	3c	4	5a	5b	6a	6b	6c	6d	6e	Total
Points	3	2	2	2	2	4	2	2	2	4	4	2	4	1	1	1	1	1	40

Grade: $\frac{\text{Your Points}}{40} * 9 + 1$ (rounded to one decimal).

Solutions:

1.

a) First $P(X = 1) = 3 * 1/9 = 1/3$, $P(X = 2) = 1/9 + 2/9 = 1/3$ and $P(X = 2) = 2/9 + 1/9 = 1/3$. Thus $P(X \geq 1) = 2/3$.

Then $E[X] = 1 * P(X = 1) + 2P(X = 2) = 1$. Similarly $E[X^2] = 1 * P(X = 1) + 4 * P(X = 2) = 5/3$. Finally $\text{Var}(X) = 5/3 - 1^2 = 2/3$.

b) Let ρ be this correlation, so $\rho = \frac{E[XY]-E[X]E[Y]}{\sigma_X \sigma_Y}$. Clearly the distribution of Y and X are the same, so by Part (a) we have $E[X]E[Y] = 1^2 = 1$ and $\sigma_X = \sigma_Y = \sqrt{2/3} \approx 0.81$. On the other hand, $E[XY] = 1*1*1/9+1*2*2/9+2*1*1/9 = 7/9$. Hence $\rho = \frac{7/9-1}{2/3} = -1/3 \approx -0.333$.

c) By Part (b) the correlation is non-zero, so X and Y cannot be independent. Alternatively, we may note that $P(X = 1 \text{ and } Y = 0) = 0$ while $P(X = 1)P(Y = 0) = 1/3 * 1/3 \neq 0$.

d) $P(Y = 0|X = 1) = 0$, $P(Y = 1|X = 1) = \frac{P(Y=1 \text{ and} X=1)}{P(X=1)} = (1/9)/(1/3) = 1/3$, and $P(Y = 2|X = 1) = \frac{P(Y=2 \text{ and} X=1)}{P(X=1)} = (2/9)/(1/3) = 2/3$. Hence $E[Y|X = 1] = 1 * 1/3 + 2 * 2/3 = 5/3$.

2.

a) The range of Z is $(0, \infty)$ so $f_Z(z) = 0$ for $z \leq 0$. Otherwise, we compute $F_Z(z) = P(Z \leq z) = P(e^X \leq z) = P(X \leq \log(z))$. Differentiation of this last expression, and the chain rule, give $f_Z(z) = f_X(\log(z)) * 1/z = \frac{1}{z\sqrt{2\pi\sigma^2}} e^{-\frac{(\log(z)-\mu)^2}{2\sigma^2}}$.

b) By definition

$P(1/2 \leq Z \leq 2) = P(\log(1/2) \leq X \leq \log(2)) = P(-\log(2) \leq X \leq \log(2)) \approx P(-0.3 \leq X \leq 0.3)$. We now standardize X, by using that $\frac{X-3}{3} \sim N(0, 1)$, so that

$$P(1/2 \leq Z \leq 2) \approx P\left(-1.1 \leq \frac{X-3}{3} \leq -0.9\right) = P\left(0.9 \leq \frac{X-3}{3} \leq 1.1\right) = \Phi(1.1) - \Phi(0.9),$$

where we used the symmetry of the standard normal distribution. Finally $P(1/2 \leq Z \leq 2) \approx 0.8643 - 0.8159 = 0.0484$.

c) $Z_1 \cdot Z_2 = e^{X_1+X_2}$. Since $X_1+X_2 \sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$, we deduce that $Z_1 \cdot Z_2$ is lognormal$(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$-distributed. By Part (a) its density is

$$\frac{1}{z\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}} e^{-\frac{(\log(z)-\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}}.$$

3.

a) There is a 2/3 probability for a passenger not choosing the first wagon. By independence, the probability that the 8 passengers do not choose the first wagon, is $(2/3)^8 \approx 0.039$.

b) The distribution of the number of passengers in the second wagon is Binomial$(n = 8, p = 1/3)$. So the desired probability can be obtained from the table (approx.), or directly computed, as

$$\binom{8}{0}(2/3)^8 + \binom{8}{1} 1/3 \cdot (2/3)^7 + \binom{8}{2}(1/3)^2 \cdot (2/3)^6 \approx 0.468.$$

c) X_1, X_2, X_3 are not independent. For instance $P(X_1 = 0, X_2 = 0, X_3 = 8) = P(X_3 = 8) = (1/3)^8$, since "$X_3 = 8$" already implies "$X_1 = 0, X_2 = 0$', while $P(X_1 = 0)P(X_2 = 0)P(X_3 = 8) = (2/3)^8(2/3)^8(1/3)^8 \neq (1/3)^8$.

For $P(A)$ we can immediately use the formula $\frac{8!}{2!4!2!}(1/3)^2(1/3)^4(1/3)^2$. Or we can reason as follows: since "$X_1 = 2, X_3 = 2$" already implies "$X_2 = 4$", we have

$$P(A) = P(X_1 = 2, X_3 = 2) = P(X_1 = 2|X_3 = 2)P(X_3 = 2),$$

and as X_1 is Binomial$(6, 1/2)$ given $X_3 = 2$, we have

$$P(A) = \binom{6}{2}(1/2)^2(1/2)^4 \cdot \binom{8}{2}(1/3)^2(2/3)^6 = \frac{8!}{2!4!2!}(1/3)^2(1/3)^4(1/3)^2.$$

In any case, $P(A) = \frac{8!}{2!4!2!}(1/3)^8 \approx 0.064$.

—

4.

We model the situation as follows: first 4 players are chosen at random, and then two of these are assigned at random to Team 1 (the remaining two go to Team 2). Thus, we use the hypergeometric formula. The probability that 0 good players end up in each team is $\frac{\binom{6}{0}\binom{4}{4}}{\binom{10}{4}}$. The probability that 2 good players end up in each team is $\frac{\binom{6}{4}\binom{4}{0}}{\binom{10}{4}}$. The probability that one good player ends up in each team is $\frac{\binom{6}{2}\binom{4}{2}}{\binom{10}{4}} \cdot \frac{\binom{2}{1}\binom{2}{1}}{\binom{4}{2}}$.

Alternatively, we could consider that the players are selected in order, one after the other, where the first two go to Team 1 and the remaining two to Team 2. In this case the probability that 0 good players end up in each team is $\frac{4*3*2*1}{10*9*8*7}$. The probability that two end up in each team is $\frac{6*5*4*3}{10*9*8*7}$. And the probability for one good player per team is $4 * \frac{(6*4)*(5*3)}{10*9*8*7}$.

Yet another alternative is to first form Team 1, and then form Team 2 making sure that the the latter has the same number of good players. Forming Team 1 with 0 good players has probability $\frac{\binom{4}{2}}{\binom{10}{2}}$, and so forming Teams 1 and 2 having each 0 good players, has (by the product rule) probability $\frac{\binom{4}{2}}{\binom{10}{2}} * \frac{\binom{2}{2}}{\binom{8}{2}}$. Similarly, the probability that both teams have one good player each, is $\frac{\binom{4}{1}\binom{6}{1}}{\binom{10}{2}} * \frac{\binom{3}{1}\binom{5}{1}}{\binom{8}{2}}$. Finally, in the case of two good player, we get $\frac{\binom{4}{0}\binom{6}{2}}{\binom{10}{2}} * \frac{\binom{4}{0}\binom{4}{2}}{\binom{8}{2}}$.

In any case, the answer is $\frac{4*3*2*1}{10*9*8*7} + \frac{6*5*4*3}{10*9*8*7} + 4 * \frac{(6*4)*(5*3)}{10*9*8*7} \approx 0.362$.

5.

a) If $S_n \leq 75$, then the time of the n-th release is at most 75 years, and so there are n or more releases within those 75 years. On the other hand, if $S_n > 75$, then the n-th release occurred after year 75, and so there are at most $n-1$ releases within those 75 years. Thus the events are equal.

b) We have to compute $P(S_{60} \leq 75)$, by Part (a). We may write $S_{60} = \sum_{i=1}^{60} X_i$, where the X_i are independent and Exponential($\lambda = 1$)-distributed. We may use the CLT, since S_{60} is approximately Normal$(60, 60)$. Using standardization and the normal table we have

$$P(S_{60} \leq 75) = P\left(\frac{S_{60} - 60}{\sqrt{60}} \leq \frac{15}{\sqrt{60}}\right) \approx \Phi(1.936) \approx 0.9735.$$

Alternatively, the number of releases N within 75 years, is Poisson distributed with $\mu = 75*1$, and this can be approximated by a Normal$(75, 75)$ distribution. Further applying continuity correction, we get

$$P(N \geq 60) \approx P\left(\frac{N - 75}{\sqrt{75}} \geq \frac{60 - 0.5 - 75}{\sqrt{75}}\right) \approx \Phi(1.79) \approx 0.9633.$$

6.

a) False. For instance if $A = B = S$ is the whole sample space (and $S \neq \emptyset$), then $P(A \cap B) = P(S) = 1$ and also $P(A)P(B) = 1 * 1$, although $A \cap B = S$.

b) True. $P(A|B) = P(A)$ implies $P(A \cap B) = P(A)P(B)$ by definition, and dividing the latter by $P(A)$ we get $P(B|A) = P(B)$.

c) False. This is because "uncorrelated" does not imply "independent". For instance if $P(X = 1, Y = 1) = 1/2$, $P(X = 1, Y = -1) = P(X = -1, Y = 1) = 1/4$ and $P(X = -1, Y = -1) = 0$, then $E[XY] = 0$ although X,Y are not independent ($0 = P(X = -1, Y = -1) \neq P(X = -1)P(Y = -1) = (1/4)^2$).

d) False. For instance, if $Y = 2$ is constant, and $X \sim N(0, 1)$ we have $E[X^2] = 1$, whereas $(E[X])^2 = 0$, although X and Y are clearly independendent.

e) True. By definition $P[X \geq 3|X \geq 2] = P(X \geq 3, X \geq 2)/P(X \geq 2) = P(X \geq 3)/P(X \geq 2)$, which is larger than $P[X \geq 3]$ as we are dividing by a number smaller than 1.