**Test Probability Theory for TCS-BIT, June 30, 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 $Y$:

| $x \setminus y$ | -2 | 0 | 2 |
|---|---|---|---|
| -1 | 0.1 | 0.1 | 0.05 |
| 0 | 0.1 | 0.2 | 0.1 |
| 1 | 0.05 | 0.1 | 0.2 |

   (a) Determine the probability distribution of $Y$, and further find $E(Y)$ and $\mathrm{Var}(Y)$.

   (b) Determine the covariance between $X$ and $Y$, namely $Cov(X, Y)$.

   (c) Determine the standard deviation of $Z = X + Y$, namely $\sigma_Z$.

   (d) Calculate $E[X|Y = 0]$.

2. The annual income of an adult in an affluent country is modelled with a Normal distribution with mean $\mu = 60000$ and standard-deviation $\sigma = 20000$. By definition, a person is called "rich" if his/her income is in the top 1% of the population.

   (a) Determine (or approximate) the level $C$ so that a person being "rich" is the same as his/her income being greater than or equal to $C$.

   (b) If 100 adults are chosen randomly from the population, what is the probability that 4 or more of them are "rich"?

3. After much work, the parliament of a given country drafts a 350-pages-long law. It is known that mistakes are randomly distributed over the draft, and on average one in every seven pages contains mistakes. Parliamentarians are supposed to find these. A lazy parliamentarian decides he will only check the first 100 pages of the draft.

   (a) Determine the distribution of the number of mistakes identifies by the lazy parliamentarian. Spell out your assumptions.

   (b) Determine or approximate the probability that the number of mistakes determined by the lazy parliamentarian is greater or equal than 15.

4. For the coming flu season, experts determined that a certain virus will be particularly active. However, the virus can come in one of three mutations ($m_1$, $m_2$ and $m_3$), each of them equally likely. It is assumed that in the case of mutation $m_1$, the probability of a serious outbreak is 10%. In the case of $m_2$ this number is 40%, and for $m_3$ it is 70%.

   (a) Determine the conditional distribution of the mutation of the virus, given that the outbreak is serious. Do so defining all relevant events an their (conditional) probabilities.

   (b) A big insurance company is supposed to pay the government an amount of $k$-billion-euros if mutation $m_k$ is detected ($k = 1, 2, 3$). Determine the average amount paid by the company if a serious outbreak happened.

5. The waiting time $X$ of a costumer at a call-center is exponentially distributed with parameter $\lambda = 2$. After this waiting period, the costumer spends an amount of time $Y$ discussing with the operator at the call-center. Suppose that $Y$ is uniformly distributed over the interval $(0, 1)$, and that $X$ and $Y$ are independent. Define $Z = X + Y$, the total length of the call.

   (a) Identify $S_Z$, the range of $Z$.

   (b) Determine $f_Z(z)$, the density function of $Z$. (Hint: distinguish the cases $0 \le z \le 1$ and $z > 1$.)

Order of questions:

| This Version (Version 2) | 1a | 1b | 1c | 1d | 2a | 2b | 3a | 3b | 4a | 4b | 5a | 5b |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Version 5 | 4a | 4b | 4c | 4d | 3a | 3b | 2a | 2b | 1a | 1b | 5a | 5b |
| Version 6 | 2a | 2b | 2c | 2d | 1a | 1b | 4a | 4b | 3a | 3b | 5a | 5b |

Points:

| Question | 1a | 1b | 1c | 1d | 2a | 2b | 3a | 3b | 4a | 4b | 5a | 5b | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Points | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 3 | 2 | 2 | 1 | 4 | 23 |

Grade: $\frac{\text{Your Points}}{23} * 9 + 1$ (rounded to one decimal).

**Solutions:**

**1.**

(a) We have $P(Y = -2) = 0.25$, $P(Y = 0) = 0.4$ and $P(Y = 2) = 0.35$.

So $E[Y] = -2 * 0.25 + 2 * 0.35 = 0.2$ and $E[Y^2] = 4 * 0.25 + 4 * 0.35 = 2.4$. Finally $Var(Y) = 2.4 - (0.2)^2 = 2.36$.

(b) $E[XY] = 2*(0.1+0.2) - 2*(0.05+0.05) = 0.4$. Observe that $P(X = a) = P(Y/2 = a)$, so $E[X] = E[Y/2] = 0.1$ by Part (a). Hence $Cov(X, Y) = 0.4 - 0.2 * 0.1 = 0.38$.

(c) Clearly $\sigma_Z = \sqrt{|Var(X) + Var(Y) + 2Cov(X, Y)|}$. Observe that $Var(X) = Var(Y/2) = Var(Y)/4 = 2.36/4 = 0.59$. So $\sigma_Z = \sqrt{|0.59 + 2.36 + 2 * 0.38|} \approx 1.926$.

(d) We have $P(X = -1|Y = 0) = 0.1/0.4 = 0.25$, $P(X = 0|Y = 0) = 0.2/0.4 = 0.5$, and $P(X = 1|Y = 0) = 0.1/0.4 = 0.25$, so $E[X|Y = 0] = 0$.

**2.**

(a) Let $X \sim N(60000, 20000^2)$. We must find $C$ so that $P(X \geq C) = 0.01$. By standardization, we have

$$P(X \geq C) = P\left(\frac{X - 60000}{20000} \geq \frac{C - 60000}{20000}\right) = 1 - \Phi\left(\frac{C - 60000}{20000}\right),$$

so $C$ satisfies $\Phi\left(\frac{C-60000}{20000}\right) = 0.99$. From the table we deduce that $\frac{C-60000}{20000} \approx 2.33$, and so $C \approx 106600$.

(b) The number $Y$ of rich people in the group of 100 adults is binomially distributed with $n = 100$ and $p = 0.01$. We must compute $P(Y \geq 4) = 1 - P(Y \leq 3)$. This cannot be found in the Binomial table, so we approximate it. We can use Poisson approximation ($n > 25$, $np < 10$) with parameter $\mu = np = 1$, so

$$P(Y \leq 3) \approx e^{-1} * [1 + 1 + 1/2 + 1/6] = 0.981,$$

which can also be retrieved from the Poisson table, and so $P(Y \geq 4) \approx 0.019$.

**3.**

(a) Call $X$ this random variable. We assume that whether the pages contain a mistake or not are independent events, so we are in the presence of 100 Bernoulli trials with success probability $p = 0.143$, thanks to part (a). Hence $X$ is binomial distributed with $n = 100$ and $p = 0.143$.

(b) We cannot find this on the binomial table. However $n > 25$, so we may use the CLT. Hence $X$ is approximately normally distributed with mean $np = 14.3$ and variance $np(1 - p) = 12.255$. Using continuity correction we have

$$P(X \geq 15) = 1 - P(X \leq 14.5) \approx 1 - \Phi(0.2/3.5) = 1 - \Phi(0.057) \approx 0.478.$$

**4.**

(a) Denote $M_k$ the event that mutation $m_k$ arises, and $A$ the event that a serious outbreak arises. Then $P(M_k) = 1/3$ and $P(A|M_1) = 0.1$, $P(A|M_2) = 0.4$, $P(A|M_3) = 0.7$. We must compute $P(M_k|A)$, which by Bayes is equal to $P(A|M_k)P(M_K)/P(A)$. First we compute $P(A) = \sum_k P(A|M_k)P(M_k) = 1/3 * 1.2 = 0.4$. Hence

$$P(M_1|A) = 0.1/1.2 \approx 0.083, \quad P(M_2|A) = 0.4/1.2 \approx 0.333, \quad P(M_3|A) = 0.7/1.2 \approx 0.583.$$

(b) We must compute $1*P(M_1|A)+2*P(M_2|A)+3*P(M_3|A) = 0.083+0.666+1.749 = 2.498$ billion euros.

**5.**

(a) Since $S_X = (0, \infty)$ and $S_Y = (0, 1)$ then $S_Z = (0, \infty)$.

(b) If $z \leq 0$ then $f_Z(z) = 0$. By convolution integral $f_Z(z) = \int f_X(z-y)f_Y(y)dy = \int_0^1 f_X(z-y)dy$. Using that $f_X(z-y) = 0$ if $z - y < 0$, we see that for $0 \leq z \leq 1$ we have

$$f_Z(z) = \int_0^z f_X(z-y)dy = \int_0^z 2e^{-2(z-y)} = e^{-2z} \int_0^z 2e^{2y} = e^{-2z} * [e^{2z} - 1] = 1 - e^{-2z}.$$

On the other hand, if $z > 1$ we similarly have

$$f_Z(z) = \int_0^1 2e^{-2(z-y)} = e^{-2z} \int_0^1 2e^{2y} = e^{-2z} * [e^2 - 1].$$