Test Probability Theory, Friday June 14, 2016, 8.45-10.45 h.

This test consists out of 6 exercises. The formula sheet and table of the standard normal distribution are added. A normal, non-programmable calculator is allowed, a programmable calculator is not allowed.

1. In a population, the following events are defined: $A =$ "person has an *Apple* computer" and $H =$ "person is Highly educated."
If we choose at random a person out of the population, we have the following probabilities:
$P(A) = 0.3$ and $P(H) = 0.4$.

a. Calculate $P(A \cup H)$, if (for his part) we can assume that A and H are independent.
b. Calculate the probability that an owner of an *Apple* is highly educated, if $P(A|H) = 0.5$.

2. After an extensive selection and a screening session, a group of 25 international students is left.
All students are qualified for the 6 available scholarships for a master CS in Twente.
Of those 25 students, 10 students live in Asia, 5 in Africa and 10 in South-America.
There has been decided that the scholarships are awarded by drawing lots.
X is the number of Asian students that receive a scholarship.

a. Calculate $P(X = 2)$ and $E(X)$.
b. If $Y =$ "the number of African students that receive a scholarship", calculate the probability $P(X = 2 \text{ and } Y = 2)$.

3. In the table on the right, the probability function of the joint distribution $P(X = x \text{ and } Y = y)$ is given.

x \ y	0	2	4
0	0.09	0.12	0.09
2	0.12	0.16	0.12
4	0.09	0.12	0.09

a. Calculate $P(X > Y)$.
b. Determine $E(X)$ and $var(X)$.
c. Are X and Y independent? Motivate your answer briefly.

4. The density function of a variable X is given by: $f(x) = \begin{cases} \frac{1}{8}x, & \text{if } 0 < x < 4 \\ 0, & \text{otherwise} \end{cases}$

a. Calculate $P(X \leq 2)$.
b. Calculate $E(X)$ and $var(X)$.
c. Determine the density function of $Y = \sqrt{X}$ (First, give the range of the values that Y can attain.)
d. Suppose, we have 72 independent variables $X_1, \ldots, X_{72}$, which have the same distribution as X.
Then, calculate or approximate the probability $P(\sum_{i=1}^{72} X_i \leq 200)$.
(If you were not able to solve b., choose $E(X) = 2.5$ and $var(X) = 1$)

5. The initial salary X of starting IT-specialists, in thousands of Euros gross per year, is on average k€ 35 in a country, with a standard deviation of k€ 3.
Research showed that the normal distribution applies to X, approximately.
The country has a tax-free rate of 5 thousand Euro and a "*flat tax*" of 40%, so for a salary of X thousand Euro, the tax $Y = (X - 5) \cdot 0.4$, so $\boldsymbol{Y = 0.4X - 2}$.
(Remark: the probability of having an initial salary below 5 thousand euro is negligibly small.)

a. Calculate $P(X > 40)$.
b. Calculate $P(X > 40 | X > 35)$.
c. What is the value of $\rho(X, Y)$? (**Argue** your answer.)
d. Verify the given value for $\rho(X, 0.4X - 2)$ in c. by determining it using the definition of the correlation coefficient and the properties of the covariance and variance.

6. At a ticket window, a customer sees two customer in front of him: those customers must be served before he can be served. We consider the operating times (in minutes) of the customers in front of him to be independent variables X and Y, both exponentially distributed with parameter $\lambda = \frac{1}{3}$. $W = X + Y$ is the waiting time for the customer.

a. Calculate $E(X)$ and $P(X > E(X))$
b. Determine the median, which is the value M such that $P(X \geq M) = P(X \leq M) = \frac{1}{2}$
c. Deduce the density function of W (using the convolution-integral) and determine $E(W)$.

Grade calculation:
Grade = 1 + 9 × $\frac{\text{number of points}}{36}$

1		**2**		**3**			**4**				**5**				**6**			**Tot**
a	b	a	b	a	b	c	a	b	c	c	a	b	c	d	a	b	c	
2	2	2	2	1	2	2	2	2	2	3	2	2	1	2	2	2	3	**36**

Formula sheet Probability Theory for BIT and TCS in module 4

Distribution	$\boldsymbol{E(X)}$	$\boldsymbol{var(X)}$
Geometric	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Hypergeometric	$n \cdot \frac{R}{N}$	$n \cdot \frac{R}{N} \cdot \frac{N-R}{N} \cdot \frac{N-n}{N-1}$
Poisson $\quad P(X = x) = \frac{e^{-\mu}\mu^x}{x!}, x = 0, 1, 2, \ldots$	μ	μ
Uniform on (a, b)	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponential	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Erlang $\quad f_X(x) = \frac{\lambda(\lambda x)^{n-1}e^{-\lambda x}}{(n-1)!}, x \geq 0$	$\frac{n}{\lambda}$	$\frac{n}{\lambda^2}$

$$var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} var(X_i) + \sum_{i \neq j}\sum cov(X_i, X_j)$$

Solutions

Exercise 1

a. $P(A \cup H) = P(A) + P(H) - P(AH) = P(A) + P(H) - P(A) \cdot P(H)$, due to independence of A and H. So $P(A \cup H) = 0.3 + 0.4 - 0.3 \cdot 0.4 = 0.58$.

b. Required: $P(H|A) = \frac{P(A \cap H)}{P(A)}$, in which $P(A) = 0.3$ and $P(A \cap H) = P(A|H)P(H) = 0.5 \cdot 0.4 = 0.2$
So $P(H|A) = \frac{0.2}{0.3} = \frac{2}{3}$

Exercise 2

a. For the distribution of X, it is only important to distinguish Asians and non-Asians. Therefore X is hypergeometrically distributed, because the drawing are without replacement (2 scholarships per student should be excluded)

$P(X = 2) = \frac{\binom{10}{2}\binom{15}{4}}{\binom{25}{6}} \approx 34.7\%.$

$E(X) = n \cdot \frac{R}{N} = 6 \cdot \frac{10}{25} = 2.4$

b. Schematically:

Asia	Africa	S.America	Total
10	5	10	25
↓	↓	↓	↓
2	2	2	6

$P(X = 2 \text{ and } Y = 5) = \frac{\binom{10}{2}\binom{5}{2}\binom{10}{2}}{\binom{25}{6}} \approx 11.4\%$

Exercise 3

a. $P(X > Y) = P(X = 2 \text{ and } Y = 0) + P(X = 4 \text{ and } Y + 0) + P(X = 4 \text{ and } Y = 2)$
$= 0.12 + 0.09 + 0.12 = 33\%$

x	0	2	4	Total
$P(X = x)$	0.3	0.4	0.3	1
$x^2P(X = x)$	0	1.6	4.8	6.4=$E(X^2)$

b. Adding the columns gives the distribution of X, see table:
$E(X) = 2$ due to symmetry
$var(X) = E(X^2) - (EX)^2 = 6.4 - 2^2 = 2.4$
(the value of $E(X^2)$ is determined in the last row of the table)

c. X and Y are independent:
Because $P(X = x \text{ and } Y = y) = P(X = x) \cdot P(Y = y)$, in this case always holds. For example, if $x = 0$ and $y = 0$: $0.09 = P(X = 0 \text{ and } Y = 0) = P(X = 0) \cdot P(Y = 0) = 0.3 \cdot 0.3$

Exercise 4

a. $P(X \leq 2) = \int_0^2 \frac{1}{8} x dx = \left[\frac{1}{16}x^2\right]_{x=0}^{x=2} = \frac{1}{4}$
(or geometrically from the graph of f: area of a triangle = $\frac{1}{2} \cdot 2 \cdot \frac{1}{4} = \frac{1}{4}$)

b. $E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_0^4 x \cdot \frac{1}{8} x dx = \left[\frac{1}{8} \cdot \frac{1}{3} x^3\right]_{x=0}^{x=4} = \frac{8}{3}$

$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx = \int_0^4 x^2 \cdot \frac{1}{8} x dx = \left[\frac{1}{8} \cdot \frac{1}{4} x^4\right]_{x=0}^{x=4} = 8$

$var(X) = E(X^2) - (EX)^2 = 8 - \left(\frac{8}{3}\right)^2 = \frac{8}{9}$

c. If $Y = \sqrt{X}$ and$0 \leq X \leq 4$, then $0 \leq Y \leq \sqrt{4}$. So for $0 \leq y \leq 2$ we have:
$F_Y(y) = P(Y \leq y) = P(\sqrt{X} \leq y) = P(X \leq y^2) = F_X(y^2)$.

So $f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{d}{dy}[F_X(y^2)] = 2y \cdot f_X(y^2)$.
Because $f_X(x) = \frac{1}{8}x$ for $0 \le x \le 4$, $f_X(y^2) = \frac{1}{8}y^2$ for $0 \le y \le 2$,
so $f_Y(y) = 2y \cdot \frac{1}{8}y^2 = \frac{1}{4}y^3$, for $0 \le y \le 2$ (*check that* $\int_{-\infty}^{\infty} f_y(y)\,dy = \int_0^2 \frac{1}{4}y^3 dy = 1$)

d. When approximating, according to the Central Limit Theorem:

$$\sum_{i=1}^{72} X_i \sim N(n\mu, n\sigma^2) = N\left(72 \cdot \frac{8}{3}, 72 \cdot \frac{8}{9}\right) = N(190{,}64)$$

$$P(\sum_{i=1}^{72} X_i \le 200) \approx P\left(Z \le \frac{200-192}{8}\right) = \Phi(1.00) = 84.13\%.$$

Exercise 5

a. $\sigma = 3$, so $\sigma^2 = 9$: X is $N(35{,}9)$, so $P(X > 40) = P\left(Z > \frac{40-35}{3}\right) \approx 1 - \Phi(1.67) = 4.75\%$.

b. $P(X > 40 | X > 35) = \frac{P(X>40 \text{ and } X>35)}{P(X>35)} = \frac{P(X>40)}{P(X>35)} = \frac{0.0475}{0.5} = 9.5\%$.
*remark: the normal distribution **doesn't have the lack of memory property**, so*
$P(X > 40|X > 35) \ne P(X > 5)$

c. $\rho(X, Y) = +1$, because there is a positive linear relation between X and Y: $Y = 0.4X - 2$

d. $$\rho(X,Y) = \rho(X, 0.4X - 2) = \frac{cov(X,\ 0.4X-2)}{\sigma_X \cdot \sigma_{0.4X-2}} = \frac{0.4cov(X,X)}{\sigma_X \cdot 0.4\sigma_X} = \frac{0.4var(X)}{0.4\sigma_X^2} = 1$$
(*note that* $var(X) = \sigma_X^2$)

Exercise 6

a. $E(X) = \frac{1}{\lambda} = 3$, so $P(X > E(X)) = \int_3^\infty \lambda e^{-\lambda x}dx = \left[-e^{-\lambda x}\right]_{x=3}^{x\to\infty} = 0 - \left(-e^{-\frac{1}{3}\cdot 3}\right) = e^{-1} \approx 36.8\%$

b. $P(X \ge M) = e^{-\frac{1}{3}\cdot M} = \frac{1}{2}$, if $-\frac{1}{3} \cdot M = ln\left(\frac{1}{2}\right)$, so if $M = -3 \cdot \ln\left(\frac{1}{2}\right) \approx 2.08$

c. Applying the convolution-integral:

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)dx = \int_0^z \frac{1}{3}e^{-\frac{1}{3}x} \cdot \frac{1}{3}e^{-\frac{1}{3}(z-x)}dx = \int_0^z \frac{1}{9}e^{-\frac{1}{3}z}dx = \left[\frac{1}{9}e^{-\frac{1}{3}z} \cdot x\right]_{x=0}^{x=z}$$

$= \frac{1}{9}e^{-\frac{1}{3}z} \cdot z$, for $z \ge 0$ (and $f_{X+Y}(z) = 0$, if $z < 0$)

$E(W) = E(X + Y) = E(X) + E(Y) = 3 + 3 = 6$.

($E(W)$ *could also be determined by using the density function* $f_{X+Y}(z)$, *applying integration by parts twice, but this solution, using the properties of expectation, is, of course, much easier*).