Test Probability Theory, Friday the 13th of June, 2014, 13.45-15.45 h.

This test consists of 5 exercises. The formula sheet and the tables of the Poisson- and standard normal distribution are added. A regular, non-programmable calculator is allowed.

1. Determine for each of the following statements if it is (in general) ***true*** or ***false*** and give a **short** argumentation for your answer.

a. If 60% of the unemployed persons in a population are male and 30% of the unemployed have an immigrant origin, then the percentage of immigrant men among the unemployed is 18%.
b. Two mutually exclusive events A and B are independent.
c. If in a double-income household is known that the income of the husband, in thousands of euro's, is (approximately) $N(25, 64)$-distributed and the income an arbitrary married woman, in thousands of euro's, is $N(15, 36)$-distributed, then the total income of a random couple in an double-income household is $N(40, 100)$-distributed.

2. The joint probability function $P(X = x \text{ and } Y = y)$ is given by the following table:

$x \backslash y$	0	1	2
0	0.15	0.05	0.20
1	0.05	0.20	0.05
2	0.20	0.05	0.05

Determine
a. $E(X)$ and $var(X)$.
b. The correlation coefficient $\rho(X, Y)$ and explain what the computed value means for the relation between X and Y.
c. The probability function of $Z = X + Y$.
d. $E(X|Y = 2)$.

3. The help desk of the customer service of an energy company is available during office hours: the help desk is called twice per minute by customers. We assume that X = "the number of telephone calls in a random interval of 1 minute" is distributed according to the Poisson-distribution.
Consider n consecutive minutes and define X_i = "the number of telephone calls in minute i". We will assume that $X_1, \ldots, X_n$ are independent and all X_i are distributed according to the Poisson-distribution, mentioned above.

a. Calculate the probability that there are at least 3 telephone calls in a period of one minute.
b. Calculate the probability that there are at least 7 telephone calls in a period of 4 minutes.
c. Approximate with the normal distribution the probability that there are at least 105 telephone calls within a period of 1 hour (60 minutes).

4. A pollster conducts a research, commissioned by a panel discussion show, to the opinions of Dutchmen about political matters. The pollster conducts a sample of n = 400 Dutchmen, to investigate whether voters think that the (Dutch populist party) PVV was banned from the government of municipalities deservedly: according to the PVV itself this was "averting".
We call p the proportion of Dutchmen that thinks that the PVV was banned deservedly and X is the number of people that share this point of view.

a. Express $E(X)$ and $var(X)$ in p.
b. Which **approximate** distribution is the best if we want to calculate probabilities with respect to X, if n = 400 and p = 0.01? State the type of distribution and its parameter(s).
c. Calculate or approximate $P(X \geq 50)$, in case that $\boldsymbol{p}$ **= 0.10** (and again n = 400).

5. An Insurance company has a call centre that automatically pre-sorts customers: via a menu, where questions have to be answered and data have to be provided, the customer is lead to an, on the matter, expert employee. The time that is needed until the first contact with an employee is set, called X, is modelled as an exponentially distributed variable with $\lambda_1 = ½$.
The operating time Y (in minutes) of the employee, is supposed to be exponentially distributed parameter $\lambda_2 = 1$.
We suppose that the time spent in the menu, X, and the operating time, Y, are independent.
$V = X + Y$ is the sojourn time of a customer in the system.

   a. Calculate $P(X > 2\mu_X | X > \mu_X)$
   b. Determine the density function of V, using a convolution integral.
   c. Determine also $E(V)$ and $var(V)$.
   d. Two customers arrive at the same time in the telephone system: their times spent in the menu are called X_1 and X_2. Determine the **distribution function** of the maximum of the 2 times spent in the menu (the maximum of the times at which both customers are finished).

Grade calculation: grade = $1 + \frac{number\ of\ points}{31} \times 9$

1			2				3			4			5				Tot
a	b	c	a	b	c	d	a	b	c	a	b	c	a	b	c	d	
1	1	1	3	3	1	2	1	1	3	1	2	3	2	2	1	3	31

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 $P(X = x) = \frac{e^{-\mu}\mu^x}{x!}, x = 0, 1, 2, \ldots$	μ	μ
Uniform on (a, b)	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponential	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Erlang $f_X(x) = \frac{\lambda(\lambda x)^{n-1}e^{-\lambda x}}{(n-1)!}, x \geq 0$	$\frac{n}{\lambda}$	$\frac{n}{\lambda^2}$

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

Solutions:

Exercise 1

a. False: the (joint) probability of both events can only be calculated as a product of 0.60×0.30 if the events "male" and "immigrant origin" are independent. In general this is not true.

b. False: "mutually exclusive" means $P(AB) = 0$. Then, A and B can only be independent when $P(AB) = P(A)P(B) = 0$. If $P(A) > 0$ and $P(B) > 0$, A and B are dependent.

c. False: the incomes X and Y of a husband and his wife are probably dependent ($cov(X, Y) \neq 0$), so the equality $var(X + Y) = var(X) + var(Y)$ does **not** hold.
(in practice, the dependency is caused by the fact that, for example, both partners have the same level of education and therefore have the same income level. Another example is that the fractions of part time work of partners depend.)

Exercise 2

a. The probability function of X is given by: $P(X = 0) = 0.4$ and $P(X = 1) = P(X = 2) = 0.3$
Or in a table:

x	0	1	2	Total
$P(X = x)$	0.4	0.3	0.3	1

$E(X) = \sum_x xP(X = x) = 0 + 1×0.3 + 2×0.3 = 0.9$
$E(X^2) = \sum x^2 P(X = x) = 0 + 1×0.3 + 4×0.3 = 1.5$
$var(X) = E(X^2) - (EX)^2 = 1.5 – 0.81 = 0.69$

b. Y has the same distribution as X, so $E(X) = E(Y) = 0.9$ and $var(X) = 0.69$
$E(XY) = \sum\sum xyP(X = x \text{ en } Y = y) = 1×1×0.2 + 1×2×0.05 + 2×1×0.05 + 2×2×0.05 = 0.6$
$\rho(X, Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E(XY) - EX \cdot EY}{\sigma_X \sigma_Y} = \frac{(0.6 - 0.9×0.9)}{\sqrt{0.69}\sqrt{0.69}} = -\frac{0.21}{0.69} \approx -0.304$
So X and Y have a weak (or moderate) negative correlation.

c. $P(X + Y = 0) = 0.15$, $P(X + Y = 1) = 0.05 + 0.05 = 0.10$, $P(X + Y = 2) = 0.60$, $P(X + Y = 3) = 0.10$ and $P(X + Y = 4) = 0.05$.

(*Or create a table of* $Z = X + Y$:

z	0	1	2	3	4	Total
$P(X = x)$	0.15	0.1	0.6	0.1	0.05	1

d. $P(X = 0|Y = 2) = \frac{P(X=0 \text{ and } Y=2)}{P(Y=2)} = \frac{0.20}{0.30} = \frac{2}{3}$ and $P(X = 1|Y = 2) = P(X = 2|Y = 2) = \frac{0.05}{0.30} = \frac{1}{6}$
So $E(X|Y = 2) = \sum_x xP(X = x|Y = 2) = 0 \times \frac{2}{3} + 1 \times \frac{1}{6} + 2 \times \frac{1}{6} = \frac{1}{2}$

Exercise 3

a. $P(X \geq 3) = 1 - P(X \leq 2) =$ 0.323 (see table).

b. Y = "the number of telephone calls in 4 minutes" is distributed according to the Poisson-distribution with expectation $\mu = 4×2 = 8$
$P(Y \geq 7) = 1 - P(X \leq 6) = 1 - 0.313 = 68.7\%$ (See Poisson table with $\mu = 8$).

c. $S = \sum X_i$ is, when approaching, $N(\mu, \sigma^2)$-distributed with $\mu = 60× 2 = 120$ and $\sigma^2 = 60 × 2 = 120$
$P(S \geq 105) = P(S \geq 105) = P\left(Z \geq \frac{105-120}{\sqrt{120}}\right)$
$\approx P(Z \geq -1.37) = P(Z \leq 1.37) = 91.47\%$
or, with continuity correction: $P(S \geq 105) \overset{\text{c.c.}}{=} P(S \geq 104.5) = \ldots = 92.14\%$
(*Apply c.c. preferably when normally approximating the Poisson distribution: for normal approximation of binomial probabilities c.c. is "mandatory"*)

Exercise 4

a. $E(X) = np = 400p$ and $var(X) = np(1-p) = 400p(1-p)$

b. For $n = 400$ and $p = 0.01$ the Poisson distribution can be approached by the Poisson distribution with μ = 400×0.01 = 4 (< 5). *(np is too small to approach it with the normal distribution!).*

c. See part a. for μ and σ^2, now having p = 0.10: the N(40, 36)-distribution can be used as an approximating distribution of X (check: n > 25, $np = 40 > 5$ and $n(1-p) = 360 > 5$). So:

$$P(X \geq 50) \overset{\text{c.c.}}{=} P(X \geq 49.5) = P\left(\frac{X-40}{6} \geq \frac{49.5-40}{6}\right) \approx 1 - \Phi(1.58) = 5.71\%$$

Exercise 5

a. $E(X) = \frac{1}{\lambda_1} = 2$

$P(X > 2\mu_X | X > \mu_X) = P(X > 4 | X > 2) = P(X > 2)$ (due to the "lack of memory" property)

$$= \int_2^{\infty} \frac{1}{2} e^{-\frac{x}{2}} dx = \left[-e^{-\frac{x}{2}}\right]_2^{\infty} = 0 - (-e^{-1}) = e^{-1} \approx 36.8\%$$

b. $f_V(v) = \int_{-\infty}^{\infty} f_X(x) f_Y(v-x) dx = \int_0^v \frac{1}{2} e^{-\frac{x}{2}} \cdot e^{-(v-x)} dx = \int_0^v \frac{1}{2} e^{-v} e^{\frac{x}{2}} dx = \left[e^{-v} e^{\frac{x}{2}}\right]_0^v$

$= e^{-\frac{v}{2}} - e^{-v}$, if $v \geq 0$.

(and $f_V(v) = 0$ when v < 0).

c. $E(V) = E(X_1) + E(X_2)$ = 2 + 1 = 3

$var(V) = var(X_1) + var(X_2)$ = 4 + 1 = 5, due to independence of X_1 and X_2 (and $var(X) = \frac{1}{\lambda^2}$)

d. $Z = max(X_1, X_2)$

$F(z) = P(max(X_1, X_2) \leq z) = P(X_1 \leq z$ and $X_2 \leq z) = P(X_1 \leq z) \times P(X_2 \leq z) = \left(1 - e^{-\frac{z}{2}}\right)^2$ for $z \geq 0$.