Test Probability Theory, Monday June 29, 2015, 13.45-15.45 h.

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

1. An automated routine end test is conducted at a production line for laptops to check if the laptops meet the quality requirements. From statistics of the factory we know that 97% of the produced laptop is approved. In an evaluation, it turned out that 99% of the approved laptops actually meets the quality requirements and that 15% of the disapproved laptops still meets the quality requirements. Calculate the following probabilities by defining relevant events, express the given and requested probabilities in those events and apply the needed formulas.

a. The probability that a laptop meets the requirements and
b. the probability that a laptop, that meets the requirements, is approved

2. From a bin with 10 balls, having numbers 1 to 10, 3 balls are drawn randomly and without replacement. Let X be the lowest number and Y be the highest number of the 3 drawn balls. Calculate:

a. $P(Y = 5)$
b. $P(X = 2|Y = 5)$
c. $E(X|Y = 5)$

3. A (pseudo) random generator gives us a random number X between 0 and 1.

a. Show that the k-th moment $E(X^k)$ equals $\frac{1}{k+1}$ (for) $k = 1,2,3, ...$).
b. Use part a. to derive the variance of X.
c. Show that $Y = -3\ ln(X)$ is exponentially distributed and determine $E(Y)$.

4. Each of the operating times of 25 customers at a ticket window is exponentially distributed with an expectation of 2 minutes. Consider the operating times as independent stochastic variables $X_1, X_2, \ldots, X_{25}$.

a. Determine the expectation and variance of the total operating time $\sum_{i=1}^{25} X_i$.
b. Calculate, using the Central Limit Theorem, an approximation for the probability that the total operating time is more than 1 hour. (Give the answer in tenths of percentages accurate).
c. Calculate the correlation coefficient of the operating time of customer 1 (X_1) and the total operating time $\sum_{i=1}^{25} X_i$.

5. From a population of 20-years old men, whose weights are $N(80, 100)$-distributed, two men are chosen at random: their weights are X and Y, respectively.

a. Calculate $P(X > 90 \text{ and } Y > 90)$
b. Calculate $P(X + Y > 180)$

6. An internet service provider wants to investigate the effect of a possible change in their tariff system. To do that, a group of 54 people is randomly chosen out of the customer database of the biggest competitor. For that group is determined how many of them would become customer of the internet service provider, after introduction of the new tariff system. Let X be the number of future customers, out of the group of 54 randomly chosen people, after introduction of the new tariff system. For a positive decision regarding the new tariff system, at least half of the people must transfer from the competitor. The question is whether this is really the case, if at least half of the people in the sample expresses to transfer.
To give an answer to this question, we assume that for the following questions **in reality only 40%** of the customers from the competitor would transfer.

a. Write down the probability distribution of X (type + parameters) and determine $E(X)$ and $var(X)$.

b. Calculate (or approximate) the probability $P(X \geq 27)$, the probability that at least half of the people in the sample would transfer.

Grading:

Grade = $1 + 9 \times \frac{\text{number of points}}{36}$

1		**2**			**3**			**4**			**5**		**6**		**Total**
a	b	a	b	c	a	b	c	a	b	c	a	b	a	b	
3	2	2	2	2	3	2	3	2	2	2	2	2	3	4	**36**

Formula sheet Probability Theory for BIT and TCS in module 4

Distribution	$E(X)$	$var(X)$
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 $P(X = x) = \frac{e^{-\mu}\mu^x}{x!}, x = 0, 1, 2, ...$	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. Events: M occurs if the laptop "Meets the requirements" and A if the laptop "Approved"

$P(A) = 0.97$, $P(M|A) = 0.99$ en $P(M|\overline{A}) = 0.15$

$P(M) = P(MA) + P(M\overline{A}) = P(M|A)P(A) + P(M|\overline{A})P(\overline{A})$ = 0.99 · 0.97 + 0.15 · 0.03 = 0.9648

b. $P(A|M) = \frac{P(AM)}{P(M)} = \frac{0.9603}{0.9648} \approx 0.9953$ (or first give the Bayes' rule)

Exercise 2

a. $P(Y=5) = \frac{\binom{4}{2}}{\binom{10}{3}} = \frac{6}{120} = 0.05$ Schematically:

no. 1-4	no. 5	no. 6-10	Total
4	1	5	10
↓	↓	↓	↓
2	1	0	3

(or: $P(Y=5) = \frac{1}{10} \times \frac{4}{9} \times \frac{3}{8} \times 3$)

b. $P(X=2|Y=5) = \frac{P(X=2 \text{ and } Y=5)}{P(Y=5)} = \frac{\frac{2}{120}}{\frac{6}{120}} = \frac{1}{3}$

(if $Y = 5$ then there are only $\binom{4}{2} = 6$ combinations: (1,2,5), (1,3,5), (1,4,5), (2,3,5), (2,4,5) and (3,4,5). Two of those 6 combinations have 2 as the lowest number: then $X = 2$ and $Y = 5$).

c. Similar to part b. we find: $P(X=1|Y=5) = \frac{3}{6}$ and $P(X=3|Y=5) = \frac{1}{6}$

So $E(X|Y=5) = \sum_x xP(X=x|Y=5) = 1 \times \frac{3}{6} + 2 \times \frac{2}{6} + 3 \times \frac{1}{6} = \frac{5}{3}$

Exercise 3

a. $E(X^k) = \int_{-\infty}^{\infty} x^k f(x)dx = \int_0^1 x^k \cdot 1dx = \left[\frac{1}{k+1}x^{k+1}\right]_{x=0}^{x=1} = \frac{1}{k+1}$, for = 1,2, ... ,

(so $E(X) = \frac{1}{2}$ and $E(X^2) = \frac{1}{3}$.)

b. $var(X) = E(X^2) - (EX)^2 = \frac{1}{2+1} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}$

c. $F_Y(y) = P(Y \le y) = P(-3\ln(X) \le y) = P\left(X \ge e^{-\frac{1}{3}y}\right) = 1 - P\left(X \le e^{-\frac{1}{3}y}\right) = 1 - F_X\left(e^{-\frac{1}{3}y}\right)$

$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{d}{dy}\left(1 - F_X\left(e^{-\frac{1}{3}y}\right)\right) = +\frac{1}{3}e^{-\frac{1}{3}y} f_X\left(e^{-\frac{1}{3}y}\right)$

Because $f_X(x) = 1$ if $0 \le x \le 1$, $f_X\left(e^{-\frac{1}{3}y}\right) = 1$ if $y \ge 0$, so:

$f_Y(y) = \frac{1}{3}e^{-\frac{1}{3}y}$, for $y \ge 0$, so Y is exponentially distributed with $\lambda = \frac{1}{3}$

So $E(Y) = \frac{1}{\lambda} = 3$

Exercise 4

a. Because $E(X_i) = 2 = \frac{1}{\lambda}$ and $var(X_i) = \frac{1}{\lambda^2} = 4$, $E\left(\sum_{i=1}^{25} X_i\right) = \sum_{i=1}^{25} E(X_i) = 25 \times 2 = 50$ and $var\left(\sum_{i=1}^{25} X_i\right) = \sum_{i=1}^{25} var(X_i) = 25 \times 4 = 100$.

b. According to the Central Limit Theorem, $\sum_{i=1}^{25} X_i$ is, when approximating, $N(25 \times 2, 25 \times 4)$-distributed: $P\left(\sum_{i=1}^{25} X_i > 60\right) = P\left(\frac{\sum_{i=1}^{25} X_i - 50}{\sqrt{100}} > \frac{60-50}{\sqrt{100}}\right) \approx 1 - \Phi(1) = 1 - 0.8413 = 15.87\%$

c. $\varrho\left(X_1, \sum_{i=1}^{25} X_i\right) = \frac{cov\left(X_1,\ \sum_{i=1}^{25} X_i\right)}{\sqrt{varX_1}\sqrt{\sum_{i=1}^{25} X_i}}$

$$cov\left(X_1, \sum_{i=1}^{25} X_i\right) = cov(X_1, X_1) + cov(X_1, X_2) + \cdots + cov(X_1, X_{25})$$
$$= var(X_1) + 0 + \cdots + 0 = 4$$

and (see a.) $var(X_1) = 4$ respectively. $var\left(\sum_{i=1}^{25} X_i\right) = \sum_{i=1}^{25} var(X_i) = 100$.
So $\varrho\left(X_1, \sum_{i=1}^{25} X_i\right) = \frac{4}{\sqrt{4}\sqrt{100}} = \frac{1}{5}$

Exercise 5

a. $P(X > 90 \text{ and } Y > 90) \overset{\text{i.i.}}{=} P(X > 90) \cdot P(Y > 90) = \left(1 - \Phi\left(\frac{90-80}{10}\right)\right)^2 \approx 2.52\%$

b. $X + Y$ is $N(80 + 80, 100 + 100)$-distributed, so
$P(X + Y > 180) = P\left(Z > \frac{180-160}{\sqrt{200}}\right) \approx 1 - \Phi(1.41) = 7.93\%$

Remark: don't use continuity correction at exercise 5a and 4b: the variables are already continuous!

Exercise 6

a. Let X be the number of future customers, after introduction of the new tariff system, out of the group of n = 54 randomly chosen customers of the competitor: $X \sim B(54, 0.40)$.
$E(X) = np = 54 \cdot 0.4 = 21.6$ and $var(X) = np(1-p) = 54 \cdot 0.4 \cdot 0.6 = 12.96$

b. Because $n > 25$, $np = 21.6 > 5$ and $n(1-p) = 32.4 > 5$, we could approach X with the normal distribution with $\mu = 21.6$ and $\sigma = \sqrt{var(X)} = \sqrt{12.96} = 3.6$
$P(X \geq 27) \overset{\text{cont. corr.}}{=} P(X \geq 26.5) \approx P\left(Z \geq \frac{26.5-21.6}{3.6}\right) \approx 1 - \Phi(1.36) = 1 - 0.9131 = 8.69\%$