Test Probability Theory, Tuesday June 16, 2015, 8.45-11.00 h.

This test consists of 6 exercises, the formula sheet and tables of the binomial and standard normal distribution. A regular calculator is allowed, a programmable ("GR") is not.

1. A firm places a bid to bring in a big project. The management of the firm estimates the probability of getting the project at 60%. After placement, the Appraisal body, which assigns the project, can ask for extra information. From the past is known that in 75% of the honoured applications extra information was asked and that in 40% of the non-assigned applications extra information was asked.
During a bid procedure the firm is asked for extra information.
Calculate, using this information, the probability that the application of the firm for the project is assigned to the firm. First define a number of relevant events and write down the given probabilities in terms of (conditional) probabilities of those events.

2. Person C claims to be clairvoyant. To test this 10 boxes are shown to him.
Every box contains a hermetically sealed bottle. The bottles are randomly filled with oil or water (always with equal probability).
Person C has to say for each of the 10 boxes if the bottle contains oil or water.
X is the number of correct answers.
If we assume that C is not clairvoyant at all and answers randomly for each box, calculate $P(X \geq 8)$, $E(X)$ and $var(X)$.

3. The joint probability function $P((X = x \text{ and } Y = y)$ of the stochastic variables X and Y is given in the table on the right.

x \ y	0	1	4
-1	0.04	0.10	0.15
0	0.16	0.05	0.10
1	0.20	0.10	0.10

a. Determine the marginal distributions of X and Y.
b. Calculate $cov(X, Y)$.
c. Calculate $P(X^2 + Y = 1)$
d. Calculate $E(Y|X = 1)$

4. The (continuous) stochastic variable X has a density function f which is given by

$$f(x) = \begin{cases} cx^2, & \text{for } 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

a. Show that $c = 3$.
b. Calculate $E(X)$ and $var(X)$.
c. Determine the density function of $Y = X^2$.

5. The computer facilities service of a big university wants to investigate if there is enough interest among students in the purchase of laptops in the context of a specified laptop offer for students.
Via a sample of 100 of the 60000 students, they want to determine the proportion p of students that are seriously interested in the concrete offering of laptops.
X is the number of interested students in the sample and the estimated proportion is $\frac{X}{100}$.

a. Why can we still use the binomial distribution as a model for X, while conducting a sample with draws without replacement (from students of the population)?

b. If we assume that $p = 0.20$, calculate the probability that the measured proportion in the sample differs at least 5% from this assumed proportion:
calculate or approximate $2 \cdot P(X \geq 25 | p = 0.20)$.

6. Two types of customers arrive at a ticket window to be served. The corresponding operating times X and Y can be modelled as independent and exponentially distributed random variables with parameter $\lambda = 1$ respectively $\lambda = 2$.

a. Determine $E(X+Y)$ and $var(X+Y)$.

b. Determine $\rho(X, X+Y)$.

c. Of each type of customers, 100 arrive at the window on a certain day: 100 operating times are considered to be a random sample $X_1, \ldots, X_{100}$ of X (so 100 independent and $Exp(\lambda = 1)$-distributed operating times) and the other 100 operating times a random sample $Y_1, \ldots, Y_{100}$ of Y. Which probability distribution, with parameters, does the total operating time of the 200 customers have, approximately? Give the properties that you are using explicitly.

Grading: grade = $1 + \frac{number\ of\ points}{36} \times 9$

1	2	3				4			5		6			Tot
		a	b	c	d	a	b	c	a	b	a	b	c	
4	4	2	3	2	2	2	2	3	1	4	2	2	3	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 $P(X = x) = \frac{e^{-\mu}\mu^x}{x!}, x = 0, 1, 2, \ldots$	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

Note A = "application Assigned" and E = "extra information asked";
then $P(A) = 0.60$, $P(E|A) = 0.75$ and $P(E|\overline{A}) = 0.40$.
Asked: $P(A|E)$. The Bayes' rule applies:

$$P(A|E) = \frac{P(A\cap E)}{P(E)} = \frac{P(E|A)P(A)}{P(E|A)P(H)+P(E|\overline{A})P(\overline{A})} = \frac{0.75\cdot 0.60}{0.75\cdot 0.60+0.40\cdot 0.40} = \frac{0.45}{0.61} = 0.74.$$

Exercise 2

$X \sim B\left(10,\frac{1}{2}\right)$, so with the table: $P(X \geq 8) = 1 - P(X \leq 7) = 1 - 0.945 = 5.5\%$

or: $P(X \geq 8) = P(X=8) + P(X=9) + P(X=10) = \left[\binom{10}{8}+\binom{10}{9}+\binom{10}{10}\right]\times\left(\frac{1}{2}\right)^{10} = \frac{56}{1024} \approx 5.47\%$

$E(X) = np = 10\cdot\frac{1}{2} = 5$ and
$var(X) = np(1-p) = 2.5$

Exercise 3

x \ y	0	1	4	$P(X=x)$
-1	0.04	0.10	0.15	0.29
0	0.16	0.05	0.10	0.31
1	0.20	0.10	0.10	0.40
$P(Y=y)$	0.40	0.25	0.35	1

a. The marginal distributions of X and Y are added to the table:

b. $cov(X,Y) = E(XY) - EX\cdot EY$, in which
$E(X) = \sum_x xP(X=x)$
$= -1\cdot 0.29 + 1\cdot 0.40 = 0.11$
and $E(Y) = 1\cdot 0.25 + 4\cdot 0.35 = 1.65$
$E(XY) = \sum\sum x\cdot y\cdot P(X = x \text{ en } Y = y)$
$= -1\cdot 1\cdot 0.1 + (-1)\cdot 4\cdot 0.15 + 1\cdot 1\cdot 0.1 + 1\cdot 4\cdot 0.1 = -0.2$
So $cov(X,Y) = -0.2 - 0.11\cdot 1.65 = -0.3815$

c. $P(X^2+Y=1) = P(X=-1 \text{ and } Y=0) + P(X=0 \text{ and } Y=1) + P(X=1 \text{ and } Y=0)$
$= 0.04 + 0.05 + 0.20 = 0.29$

d. $P(Y=0|X=1) = \frac{P(X=1 \text{ and } Y=0)}{P(Y=0)} = \frac{0.20}{0.40} = \frac{1}{2}$, similarly $P(Y=1|X=1) = P(Y=4|X=1) = \frac{1}{4}$,
so $E(Y|X=1) = \sum_y y\cdot P(Y=y|X=1) = 0\cdot\frac{1}{2} + 1\cdot\frac{1}{4} + 4\cdot\frac{1}{4} = 1.25$

Exercise 4

a. $\int_{-\infty}^{\infty} f(x)dx = 1 \quad\Leftrightarrow\quad \int_0^1 cx^2dx = \left[\frac{1}{3}cx^3\right]_{x=0}^{x=1} = \frac{1}{3}c = 1 \quad\Leftrightarrow\quad c = 3$

b. $E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_0^1 x\cdot 3x^2dx = \left[\frac{3}{4}x^4\right]_{x=0}^{x=1} = \frac{3}{4}$

$E(X^2) = \int_{-\infty}^{\infty} x^2f(x)dx = \int_0^1 x^2\cdot 3x^2dx = \left[\frac{3}{5}x^5\right]_{x=0}^{x=1} = \frac{3}{5}$

So $var(X) = E(X^2) - (EX)^2 = \frac{3}{5} - \left(\frac{3}{4}\right)^2 = \frac{3}{80}$ $(= 0.0375)$

c. $F_Y(y) = P(X^2 \leq y) \overset{y>0}{=} P(-\sqrt{y} \leq X \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$

$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{1}{2\sqrt{y}}f_X(\sqrt{y}) + \frac{1}{2\sqrt{y}}f_X(-\sqrt{y})$, still having $y > 0$.

So because $f_X(x) = 3x^2 \text{ for } 0 < x \leq 1$ applies to $0 < \sqrt{y} \leq 1$:

$f_Y(y) = \frac{1}{2\sqrt{y}}\cdot 3(\sqrt{y})^2 + 0 = \frac{3}{2}\sqrt{y}$, if $0 < y \leq 1$

(And $f_Y(y) = 0$ otherwise)

Exercise 5

a. For relative small samples out of large populations, the hypergeometric distribution is approximately equal to the binomial distribution. (Rule of thumb: $60000 = N \geq 5n^2 = 5 \cdot 100^2 = 50000$)

b. For $p = 0.2$, X is, according to the Central Limit Theorem approximately $N\big(np, np(1-p)\big) = N(20{,}16)$-distributed, so:

$$2 \cdot P(X \geq 25 | p = 0.20) \overset{c.c.}{=} 2 \cdot P(X \geq 24.5) \overset{CLT}{\approx} 2 \cdot P\left(Z \geq \frac{24.5-20}{\sqrt{16}}\right) = 2\big(1 - \Phi(1.125)\big)$$

$$= 2(1 - 0.8697) \approx 16.1\%$$

(*the average of* $\Phi(1.12)$ *and* $\Phi(1.13)$ *from the* $N(0{,}1)$-*table is used*)

Exercise 6

a. $E(X+Y) = E(X) + E(Y) = \frac{1}{1} + \frac{1}{2} = 1.5$ and due to independence of X and Y:

$$var(X+Y) \overset{ind.}{=} var(X) + var(Y) = \frac{1}{1^2} + \frac{1}{2^2} = 1.25$$

b. $\rho(X, X+Y) = \frac{cov(X,X+Y)}{\sigma_X \cdot \sigma_{X+Y}} = \frac{cov(X,X) + cov(X,Y)}{\sqrt{var(X)} \cdot \sqrt{var(X+Y)}} = \frac{var(X) + 0}{\sqrt{1} \cdot \sqrt{1.25}} = \frac{1}{\sqrt{1.25}} \approx 0.894$

c. $X_1 + \cdots + X_{100}$ is, according to the CLT ($n > 25$), approximately $N\left(n \cdot \frac{1}{\lambda}, n \cdot \frac{1}{\lambda^2}\right) = N(100{,}100)$-distributed.

Similarly $Y_1 + \cdots + Y_{100} \sim N(50, 25)$.

Because all variables are independent, the sum (of the two sums) is also normally distributed:

The total sum of operating times $(X_1 + \cdots + X_{100}) + (Y_1 + \cdots + Y_{100})$ is thus approximately $N(100 + 50{,}100 + 25)$-distributed.