Test Probability Theory (M4-201300180), July 3, 2017, 13.45-15.45 h.

Lecturer: Dick Meijer - Module coordinator: Klaas Sikkel

This test consists of 5 exercises, a formula sheet and the Poisson and standard normal tables.
A regular scientific calculator is allowed, a programmable calculator ("GR") is not allowed.

1. Answer the following questions and give an adequate motivation:

a. The traffic authority is checking the safety after maintenance of cars via random samples.
Suppose that a garage repaired and serviced 15 cars: 4 of them are not safe enough to drive the public roads (according to the rules of the authority), but the other 11 are.
If the traffic authority chooses (at random) 3 of the 15 serviced cars in its sample, what is the probability that at least one of the chosen cars is unsafe?

b. True or false: the probability that you need more than 12 rolls of a dice to have 6 as a result is less than 10%?

c. True or false: if the random variables X and Y are not correlated, then they are independent?

d. True or false: if $P(A|B) = P(A)$, then $P\left(A\middle|\overline{B}\right) = P(A)$?

2. In the supply conditions of "mass products" (produced in large quantities such as eggs, screws, bricks, etc.), it is often stated which proportion of the products is allowed not to meet all agreed specifications, for instance "at most 1% substandard (bad) products". We will use this proportion of 1% in this exercise.
Suppose the proportion of substandard products in a (very) large order is checked with a random sample of n of these products: X, the number of substandard products among the n products in the sample, should be **at least** k: if $X \geq k$, then the order is refused by the buyer.
Compute or approximate $P(X \geq k)$, the probability of refusal though the order just met the condition (1% substandard) for the following values of sample size n and critical value k:

a. $n = 10$ and $k = 1$.
b. $n = 200$ and $k = 3$.
c. $n = 2000$ and $k = 25$.

3. The department for claims of a car insurance company gave the following information:

- The **monthly** number of claims has a Poisson distribution with mean 5.
- The amount X (in thousands of Euro's) is on average $\mu = 10$ (thousand Euro) with a standard deviation $\sigma = 10$ (thousand Euro).

We define $N =$ "the total number of claims **in a year**" and $S =$ "Total amount of claims in a year".
If $N = 50$, then S is the summation of the independent amounts $X_1, X_2, \ldots, X_{50}$: $S = \sum_{i=1}^{50} X_i$

a. Compute or approximate the probability that the number of claims in a month is greater than 10.
b. Compute or approximate the probability $P(X > 20)$, if possible.
If computation is not possible, explain why not.
c. Compute or approximate $P\left(\sum_{i=1}^{50} X_i > 600\right)$, if possible.
If computation is not possible, explain why not.
d. Compute or approximate, if possible: $P(N > 70)$
If computation is not possible, explain why not.
e. Express $E(S|N = n)$ in n and use this expression to find the value of $E(S)$.

4. X and Y are independent and $N(45, 144)$- and $N(60, 81)$-distributed, respectively.

- **a.** Compute $P(X + Y > 120)$.
- **b.** Compute $\rho(X, X + Y)$.

5. How long does it take before a perishable product is used by consumers?
For fresh milk researchers found the distribution of the time X of finishing a pack of milk, given by its density function f, where the moment of production is set to $x = 0$ and the expire date to $x = 1$.

$$f(x) = \begin{cases} 6x(1-x) & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

- **a.** Compute $P\left(X \le \frac{3}{4}\right)$.
- **b.** Determine $E(X)$ and $var(X)$.
- **c.** Derive the **density function** of $= \frac{1}{X}$.

Grading: grade = $1 + \frac{\#\, points}{38} \times 9$

1				2			3					4		5			Tot
a	b	c	d	a	b	c	a	b	c	d	e	a	b	a	b	c	
2	2	2	2	2	2	3	2	2	3	2	2	2	3	2	3	2	38

--

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 $f_X(x) = \frac{\lambda(\lambda x)^{n-1}e^{-\lambda x}}{(n-1)!}, x \ge 0$	$\frac{n}{\lambda}$	$\frac{n}{\lambda^2}$

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

Solutions

Exercise 1

Unsafe	Safe	Total
4	11	15
↓	↓	↓
0	3	3

a. $X =$ "# Unsafe cars in the sample of 3 out of 15"

The draws are without replacement, so

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{\binom{11}{3}}{\binom{15}{3}} = 1 - \frac{11}{15} \cdot \frac{10}{14} \cdot \frac{9}{13} \approx 63.7\%$$

b. False: X, the required number of rolls to get a 6, is geometric with success probability $p = \frac{1}{6}$.

So $P(X > 12) = (1-p)^{12} = \left(\frac{5}{6}\right)^{12} \approx 11.2\% > 10\%$

c. False, no correlation (no linear dependence) does not exclude dependence.
(We only know that from independence it follows that $\rho = 0$, not reversely.)

d. True: From $P(A|B) = P(A)$ we know that then $\frac{P(AB)}{P(B)} = P(A)$ or $P(AB) = P(A)\,P(B)$

Since $P(A) = P(AB) + P(A\overline{B})$ we have

$$P(A\overline{B}) = P(A) - P(A)P(B) = P(A)[1 - P(B)] = P(A)P(\overline{B})$$

So $P(A|\overline{B}) = \frac{P(A\overline{B})}{P(\overline{B})} = \frac{P(A)P(\overline{B})}{P(\overline{B})} = P(A)$

Exercise 2

a. $n = 10$ and $k = 1$.
$X \sim B(10, 0.01)$, so $P(X \geq 1) = 1 - P(X = 0) = 1 - 0.99^{10} \approx 9.56\%$

b. $n = 200$ and $k = 3$. $X \sim B(200, 0.01)$.
"Large $n > 25$ and small $p = 0.01$": $\mu = np = 2 < 10$, so X is approximately Poisson ($\mu = 2$).
Using the Poisson-table: $P(X \geq 3) = 1 - P(X \leq 2) = 1 - 0.677 = 32.3\%$
<u>Note:</u> *in this case you can also use the exact $B(200{,}0.01)$-distribution of X to compute $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$, but if $P(X > 10)$ is requested, the Poisson approximation would be preferred for simple calculation, using the Poisson-tables.*

c. $n = 2000$ and $k = 25$: large $n > 25$ and $\mu = np = 20 > 5$, so using the CLT:
X is approximately $N(np, np(1-p)) = N(20, 19.8)$

$P(X \geq 25) \overset{\text{c.c.}}{=} P(X \geq 24.5) \overset{\text{CLT}}{\approx} P\left(Z \geq \frac{24.5 - 20}{\sqrt{19.8}}\right) \approx 1 - \Phi(1.01) = 15.62\%$

Exercise 3

a. The number of claims in a month: $Y \sim$ Poisson(5), so $P(Y > 10) = 1 - P(Y \leq 10) = 1.4\%$ (using the Poisson-table with $\mu = 5$).
(If you interpreted the question as the annual number of claims (N), your answer is considered to be the answer of this question)

b. No, this is impossible, since we do not have any information about the distribution of X.
Note: e.g. a normal distribution cannot apply: then we should have $X \sim N(10, 10^2)$, but the probability of a "negative claim" would be $P(X < 0) = 1 - \Phi(1) \approx 15\% = P(X > 20)$.
If $\mu = \sigma = 10$, then , theoretically, an $Exp\left(\lambda = \frac{1}{10}\right)$-distribution could be possible (accepted as a possibly correct answer).

c. $\sum_{i=1}^{50} X_i$ is, according to the CLT, approximately $N(n\mu, n\sigma^2)$, with sufficiently large $n = 50 > 25$ and $\mu = \sigma = 10$. So: $P\left(\sum_{i=1}^{50} X_i > 600\right) \overset{\text{CLT}}{\approx} P\left(Z \geq \frac{600 - 50 \cdot 10}{\sqrt{50 \cdot 10^2}}\right) \approx 1 - \Phi(1.41) \approx 7.9\%$

d. N has a Poisson distribution with $\mu = 12 \cdot 5 = 60$, so, using the CLT, approximately $N(60, 60)$,
$P(N > 70) \overset{\text{c.c.}}{=} PN > 70.5) \overset{\text{CLT}}{\approx} P\left(Z > \frac{70.5 - 60}{\sqrt{60}}\right) \approx 1 - \Phi(1.36) = 8.69\%$

e. $E(S|N=n) = E(X_1 + \dots + X_n) = n \cdot E(X) = 10n$
Then, from $E(S|N) = 10N$, it follows that $E(S) = E[E(S|N)] = E[10N] = 10 \cdot E(N) = 600$

Exercise 4

a. $P(X+Y > 120)$, where $X+Y$ is normally distributed as well (since X and Y are independent), with $\mu = 60 + 45 = 105$ en $\sigma^2 = 144 + 81 = 225$.
So $P(X+Y > 120) = P\left(Z > \frac{130-105}{\sqrt{225}}\right) = 1 - \Phi(1) = 15.87\%$.

b.

$$\rho(X, X+Y) = \frac{cov(X,X) + cov(X,Y)}{\sigma_X \cdot \sigma_{X+Y}} = \frac{var(X) + 0}{\sqrt{var(X)} \cdot \sqrt{var(X) + var(Y)}} = \frac{144}{12 \cdot 15} = 0.8$$

Exercise 5

a. $P\left(X \le \frac{3}{4}\right) = \int_0^{\frac{3}{4}} 6x(1-x)dx = 6\left[\frac{1}{2}x^2 - \frac{1}{3}x^3\right]_{x=0}^{x=\frac{3}{4}} = 6\left(\frac{9}{32} - \frac{9}{64}\right) - 0 = \frac{27}{32} \approx 84.4\%$

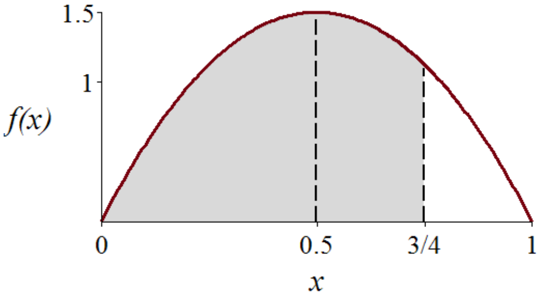
b. $E(X) = \frac{1}{2}$, using symmetry.

$$E(X^2) = \int_{-\infty}^{\infty} x^2 \cdot f(x)dx = \int_0^1 x^2 \cdot 6(x - x^2)dx$$

$$= 6 \cdot \left[\frac{1}{4}x^4 - \frac{1}{5}x^5\right]_{x=0}^{x=1} = 6\left(\frac{1}{4} - \frac{1}{5}\right) = \frac{3}{10} \text{ and}$$

$var(X) = E(X^2) - (EX)^2 = \frac{3}{10} - \left(\frac{1}{2}\right)^2 = \frac{1}{20}$.
(so $\sigma_X = \sqrt{0.05} \approx 0.224$)

The density function f and $P\left(X \le \frac{3}{4}\right)$



c. $F_Y(y) = P\left(\frac{1}{X} \le y\right) = P\left(X \ge \frac{1}{y}\right) = 1 - F_X\left(\frac{1}{y}\right)$
$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{d}{dy}\left[1 - F_X\left(\frac{1}{y}\right)\right] = -\left[-\frac{1}{y^2}f_X\left(\frac{1}{y}\right)\right]$
$f_X(x) = 6x(1-x)$ if $0 \le x \le 1$, so if $0 \le \frac{1}{y} \le 1$, or if $y \ge 1$.
Therefore: $f_Y(y) = \frac{1}{y^2} \cdot 6 \cdot \frac{1}{y}\left(1 - \frac{1}{y}\right) = \frac{6}{y^3} - \frac{6}{y^4} \quad (y \ge 1)$.