Exam Probability & Statistics I – 24th of May 2018, 15.00-17.00 hrs

This test consists of 5 exercises, a formula sheet and the standard normal table.
IT-students may choose to solve exercise 6 instead of exercise 5.

1. The following information about the events A and B is available: $P(A) = 0.2$ and $P(B) = 0.5$. Determine the probability $P(A \cup B)$ for each of the following additional information:

a. The events A and B are mutually exclusive.
b. The events A and B are independent.
c. $P(A|B) = 0.3$.

2. The police is checking cars for speeding: if the speed limit is exceeded by more than 30 km/h, the driver is arrested and his car is seized.
Let X be the number of seized cars during the speed check, where the speed of n cars is measured and p is the proportion of drivers whose speed is more than 30 km/h above the limit.
Compute or approximate the following probabilities:

a. $P(X > 0)$ if $n = 10$ and $p = 0.2$.
b. $P(X \geq 2)$ if $n = 200$ and $p = 0.01$.
c. $P(X < 30)$ if $n = 400$ and $p = 0.1$.

3. A small factory has a morning and evening shift. Absenteeism by the employees was monitored over the years and these statistics enable us to give the joint probability function of the variables X and Y, where X = "the number of absent employees in the morning shift" and Y = "the number of absent employees in the evening shift".
The probabilities $P(X = x \text{ and } Y = y)$ are given in the table:

$x \setminus y$	0	1	2
0	0.20	0.10	0
1	0.10	0.15	0.15
2	0	0.15	0.15

a. Determine $E(X)$ and $var(X)$.
b. Compute the correlation coefficient and explain the meaning of the computed value with respect to the relation of X and Y.
c. Compute $E(X + Y)$ and $var(X + Y)$.
d. Determine the distribution of Y, given $X = 0$, and $E(Y|X = 0)$.

4. The return (in % per year) on investment in stocks is often modelled as a normally distributed variable with an expected return μ and a variance σ^2. Suppose that for specific stocks these parameters are given by the expected return $\mu = 8$ and standard deviation $\sigma = 10$.

a. Compute the probability of loss (negative return).
b. Determine the 99th percentile c of the returns, i.e. the value c such that $P(X \geq c) = 0.01$

Suppose that a sample of 10 independent returns $X_1, X_2, \ldots, X_{10}$ of 10 years is available: the mean return is given by $\overline{X} = \frac{1}{10}\sum_{i=1}^{10} X_i$.

c. Compute the probability of a negative mean return: $P(\overline{X} \leq 0)$.

*Exercise **5** is **not for IT students**: they can solve the extra exercise 6 below instead.*

5. At a counter a queue of customers is lined up, waiting to be served. Assume that the service times of the customers, $X_1, X_2, \ldots$, are independent and all exponentially distributed with an expected service time of 2 minutes. (Of course, X_1 is the service time of the first customer to be served, etc.)

   a. Compute $P(X_1 > 3)$ and $P(X_1 > 5 \mid X_1 > 2)$.
   b. Compute $P(X_1 > 3 \text{ and } X_2 > 3)$.
   c. Compute or approximate $P(X_1 + X_2 + \cdots + X_{36} > 90)$.
   d. If we want to simulate the exponential distribution of the service times, we can use the relation between X and a random (uniformly distributed) number U between 0 and 1, by applying the function $X = -2\ln(U)$.
      Show that X has the desired exponential distribution, if U is $U(0,1)$-distributed.

Extra exercise, ***only for IT-students,*** *instead of exercise 5 (grading a-d:* 3, 3, 2 and 2)

6. 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 $P(A|\overline{B}) = P(A)$?

Grading:

1			**2**			**3**				**4**			**5**				**Total**
a	b	c	a	b	c	a	b	c	d	a	b	c	a	b	c	d	
1	1	2	2	2	3	2	3	3	3	2	2	3	2	2	3	3	39

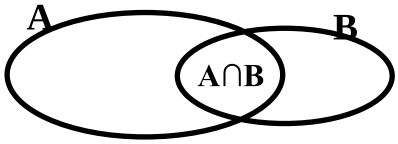
Grade = 1 + 9×(# points)/39

Formula sheet Probability & Statistics I

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	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

Solutions

Exercise 1



a. If A and B are mutually exclusive, we have: $P(A \cap B) = 0$,
so $P(A \cup B) = P(A) + P(B) = 0.2 + 0.5 = \mathbf{0.7}$

b. If A and B are independent, then

$$P(A \cap B) \overset{\text{ind.}}{=} P(A) \times P(B) = 0.2 \times 0.5 = 0.10$$

Hence: $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.2 + 0.5 - 0.1 = \mathbf{0.60}$

c. $P(A|B) = \frac{P(A \cap B)}{P(B)} = 0.3$,
or: $P(A \cap B) = 0.3 \times P(B) = 0.3 \times 0.5 = 0.15$.
Then: $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.2 + 0.5 - 0.15 = \mathbf{0.55}$

Exercise 2

a. X has a binomial distribution with $n = 10$ and $p = 0.2$:
$P(X > 0) = 1 - P(X = 0) = 1 - 0.8^{10} = 89.3\%$

b. If $n = 200$ and $p = 0.01$ we can apply the Poisson approximation of the binomial distribution of X, where $\mu = np = 2 < 10$. So $P(X \geq 2) = 1 - P(X < 2) = 1 - P(X = 0) - P(X = 1)$

$$= 1 - e^{-2} - 2e^{-2} = 1 - 3e^{-2} \approx 59.4\%$$

c. If $n = 400$ and $p = 0.1$ X is approximately $N(np, np(1-p))$- so $N(40, 36)$-distributed, according to the CLT.
$P(X < 30) = P(X \leq 29.5)$ (*continuity correction*)

$$= P\left(\frac{X-40}{\sqrt{36}} \leq \frac{29.5-40}{6}\right) \approx P(Z \leq -1.75) = \frac{1}{2} - P(0 \leq Z \leq 1.75) = \frac{1}{2} - 0.4599 = 4.01\%$$

Exercise 3

a. $P(X = 0) = P(X = 2) = 0.3$ and $P(X = 1) = 0.4$
$E(X) = 1$ (using symmetry)
$var(X) = E(X^2) - EX^2 = [0^2 \cdot 0.3 + 1^2 \cdot 0.4 + 2^2 \cdot 0.3] - 1^2 = 0.6$

b. $\rho(X, Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$
Using a.: $EX = EY = 1$ and $\sigma_X = \sigma_Y = \sqrt{0.6}$, since X and Y are identically distributed.
$E(XY) = \sum\sum x \cdot y \cdot P(X = x \text{ en } Y = y) = [1 \cdot 1 + 1 \cdot 2 + 2 \cdot 1 + 2 \cdot 2] \cdot 0.15 = 1.35$
$cov(X, Y) = E(XY) - EX \cdot EY = 1.35 - 1 \cdot 1 = 0.35$
So: $\rho(X, Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{0.35}{0.6} = \frac{7}{12}$ (≈ 0.58)
There is a moderate positive correlation between the numbers of absent employees in the morning and evening shift.

c. $E(X + Y) = E(X) + E(Y) = 2$ and
$var(X + Y) = var(X) + varY) + 2cov(X, Y) = 0.6 + 0.6 + 2 \cdot 0.35 = 1.9$

d. $P(Y = 0|X = 0) = \frac{P(Y=0 \text{ and } X=0)}{P(Y=0)} = \frac{0.2}{0.3} = \frac{2}{3}$ likewise: $P(Y = 1|X = 0) = \frac{1}{3}$
$E(Y|X = 0) = \sum yP(Y = y|X = 0) = 0 \cdot \frac{2}{3} + 1 \cdot \frac{1}{3} = \frac{1}{3}$.

Exercise 4

a. $P(X \leq 0) = P\left(\frac{X-8}{10} \leq \frac{0-8}{10}\right) = P(Z \leq -0.8) = 1 - \Phi(0.8) = 1 - 0.7881 = 21.19\%$

b. $P(X \leq c) = P\left(Z \leq \frac{c-8}{10}\right) = \Phi\left(\frac{c-8}{10}\right) = 0.99$, so $\frac{c-8}{10} = 2.33$, or: $c = 8 + 10 \cdot 2.33 = 31.3$

c. $\overline{X}$ is normally distributed with $\mu_{\overline{X}} = \mu = 8$ and standard deviation $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{10}} = \sqrt{10}$
$P(\overline{X} \leq 0) = P\left(Z \leq \frac{0-8}{\sqrt{10}}\right) = 1 - \Phi(2.53) = 1 - 0.9943 = 0.57\%$

Exercise 5

a. $E(X) = \frac{1}{\lambda} = 2$, so $\lambda = 0.5$

$P(X_1 > 3) = \int_3^\infty 0.5e^{-0.5x}dx = -e^{-0.5x}|_{x=3}^{\infty} = e^{-1.5} \approx 22.3\%$

(or use directly the exponential distribution property: $P(X > x) = e^{-0.5x}$)

$P(X_1 > 5|X_1 > 2) = \frac{P(X_1>5 \text{ and } X_1>2)}{P(X_1>2\,)} = \frac{P(X_1>5)}{P(X_1>2\,)} = \frac{e^{-0.5\cdot 5}}{e^{-0.5\cdot 2}} = e^{-1.5} \approx 22.3\%$

(*Or apply the "lack of memory" property:* $P(X_1 > 5 \mid X_1 > 2) = P(X_1 > 3) = 22.3\%$.)

b. $P(X_1 > 3 \text{ and } X_2 > 3) \overset{\text{ind.}}{=} P(X_1 > 3) \cdot P(X_2 > 3) = 0.2231^2 \approx 5.0\%$

c. $X_1 + X_2 + \cdots + X_{36}$ is according the CLT approximately $N\left(36 \cdot 2,\ 36 \cdot \frac{1}{0.5^2}\right)$-distributed

$P(X_1 + X_2 + \cdots + X_{36} > 90) \approx P\left(Z > \frac{90-72}{\sqrt{144}}\right) = 1 - \Phi(1.5) = 1 - 0.9332 \approx 6.7\%$

d. $F_X(x) = P(-2\ln(U) \le x) = P(U \ge e^{-0.5x}) = 1 - F_U(e^{-0.5x})$

$f_X(x) = \frac{d}{dx}F_X(x) = 0.5e^{-0.5x}f_U(e^{-0.5x})$

$0 < e^{-0.5x} < 1$ if $\boldsymbol{x > 0}$: then we have $f_U(e^{-0.5x}) = 1$ and hence: $f_X(x) = 0.5e^{-0.5x}$

(otherwise $f_X(x) = 0$)

X has an exponential distribution with parameter $\lambda = 0.5$ and expected value $\frac{1}{\lambda} = 2$.

Extra IT-exercise (instead of exercise 5):

Exercise 6

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

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

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 "independence implies $\rho = 0$" , 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 $P(A|\overline{B}) = \frac{P(A\overline{B})}{P(\overline{B})} = \frac{P(A)P(\overline{B})}{P(\overline{B})} = P(A)$