Trial exam Probability& Statistics I

This exam consists of 5 exercises, a formula sheet and a table of the normal distribution.
***IT**-students should only solve exercises 1-4*

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

x \ 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

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. The distribution of X given $Y = 2$, and $E(X|Y = 2)$.

2.

a. Suppose, a recruiter has to hire one person for a job that is available this month. To determine if a person should be hired, the recruiter has a job interview with that person and then determines if that person gets hired. After he has made that decision, a next person may get a job interview. This means that recruiter stops doing job interviews after he has made the decision to hire a person. From data we know that 1 out of 7 persons that have a job interview gets hired.
Let X be the number of consecutive interviews that are held by the recruiter. Determine the probability distribution of $X, E(X)$ and $P(X > 7)$.
b. Determine the probability that 6 is the highest number if one randomly takes three balls out of a box with ten balls, being numbered 1 till 10.
c. Determine the probability that, in a group of 200 students, there are (exactly) 8 dyslectic students, if, on average, 2% of all students suffer from dyslexia. Use an appropriate approximation.

3. The cylinder of water pumps in developing countries is the most critical part of the water pump: if the pump breaks down, the cylinder (almost) always has to be replaced. From historical data we can conclude that the expected lifespan of a cylinder is **5 years on average**, but the variation in lifespan is large: as an approximating model, the normal distribution with a standard deviation of 2 years is used. An NGO therefore places pumps and for maintenance on the spot a spare cylinder is supplied as well. Let X be the lifespan of the first cylinder that is used and Y the lifespan of the spare cylinder.

a. Determine $P(X > 6)$.
b. Calculate $P(X + Y > 12)$.
First state on which (reasonable) assumption(s) you will base your calculation.

4. The University of Twente wants to conduct a market research on the interest in the new Twente Educational Model (TOM) among high school students that have a science profile. The university hopes to attract more students than was the case with the “old” technical studies. Earlier research showed that old studies altogether could

count on serious interest among 25% of the high school students with a science profile in the North-Eastern region of The Netherlands. After an information campaign, a survey is conducted to measure the interest in the TOM-model among these students.
In a sample of 192 students, 60 students are seriously interested in the TOM-model. Show that the probability, that you encounter **60 or more students with serious interest** in such a sample, equals 3% (rounded), assuming that still 25% of all of these students in the region have serious interest in the TOM-model and applying continuity correction.

Exercise is not for IT-students

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

a. Determine $P(X > 0.8)$ and $E(X)$.

If we choose the largest number out of three randomly chosen numbers between 0 and 1, then this maximum Y has the following density function:

$$f_Y(y) = \begin{cases} 3y^2 & \text{if } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

b. Determine $P(Y > 0.8)$, $E(Y)$ and $var(Y)$

c. Show that Y has the aforementioned density function if $Y = \max(X_1, X_2, X_3)$ and X_1, X_2 and X_3 are independent random numbers between 0 and 1.

-- END --------------------------------

Formula sheet Probability & Statistics I

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

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

Solution trial exam P&S I

Exercise 1

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

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

$E(X) = \sum_x xP(X = x) = 0 + 1\times0.3 + 2\times0.3 = 0.9$
$E(X^2) = \sum x^2P(X = x) = 0 + 1\times0.3 + 4\times0.3 = 1.5$
$var(X) = E(X^2) - (EX)^2 = 1.5 - 0.81 = 0.69$

b. Y has the same distribution as X, so $E(X) = E(Y) = 0.9$ and $var(X) = 0.69$
$E(XY) = \sum\sum xyP(X = x \text{ en } Y = y)$
$= 1\times1\times0.2 + 1\times2\times0.05 + 2\times1\times0.05 + 2\times2\times0.05 = 0.6$

$$\rho(X, Y) = \frac{cov(X,Y)}{\sigma_X\sigma_Y} = \frac{E(XY) - EX\cdot EY}{\sigma_X\sigma_Y} = \frac{(0.6 - 0.9\times0.9)}{\sqrt{0.69}\sqrt{0.69}} = -\frac{0.21}{0.69} \approx -0.304$$

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 2

a. X is distributed according to the Geometric distribution,
with parameter $p = 1/7$.
This distribution applies, because the experiment stops after one success (decision to hire a person) occurs.|

$E(X) = \frac{1}{p} = \frac{1}{1/7} = 7$ and $P(X > 7) = (1 - p)^7 = \left(\frac{6}{7}\right)^7 \approx 0.340$ (7 times no success)

$P(\text{"highest number is 6"}) = \frac{\binom{5}{2}}{\binom{10}{3}} = \frac{10}{120} \approx 8.3\%$ (*If the highest number is 6, there are* $\binom{5}{2}$ *combinations with two numbers lower than* 6)

b. "The number of students with dyslexia" = $X \sim B(200, 0.02)$,
so by approximation Poisson with $\mu = np = 4$, because $n > 25$ and $np < 10$.
Hence $P(X = 8) = \frac{4^8}{8!}e^{-8} \approx 2.98\%$

Exercise 3

a. $P(X > 6) = P\left(\frac{X-5}{2} > \frac{6-5}{2}\right) = P(Z > 0.5) = 1 - \Phi(0.5) = 1 - 0.6915 =$ 30.85%

b. If we assume independence for the lifespans,
$X + Y$ is $N(5+5, 2^2 + 2^2)$ distributed.
So $P(X + Y > 12) = P\left(Z > \frac{12-10}{\sqrt{8}}\right) \approx 1 - P(Z \leq 0.71) = 1 - 0.7611 \approx$ 23.9%

Exercise 4

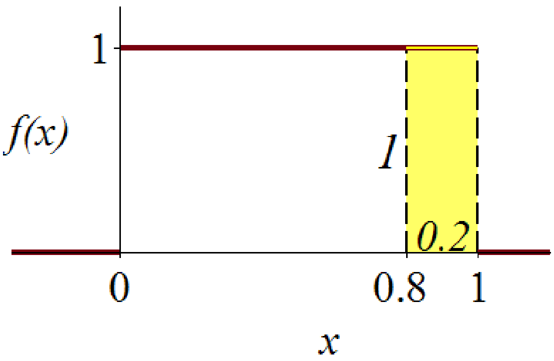
We assume that X is $B(192, 0.25)$-distributed: $E(X) = np = 48 > 5$
($np = 48$ is large enough for using an approaching normal distribution)
X is approximately N(192× ¼ , 192× ¼ × ¾) = N(48, 36)-distributed.

$$P(X \geq 60) \overset{\text{c.c.}}{=} P(X \geq 59.5) \overset{\text{CLT}}{=} P\left(\frac{X-48}{\sqrt{36}} \geq \frac{59.5-48}{\sqrt{36}}\right)$$
$$= 1 - P(Z \leq 1.92) = 1 - 0.9726$$
$$= 2.74\%$$

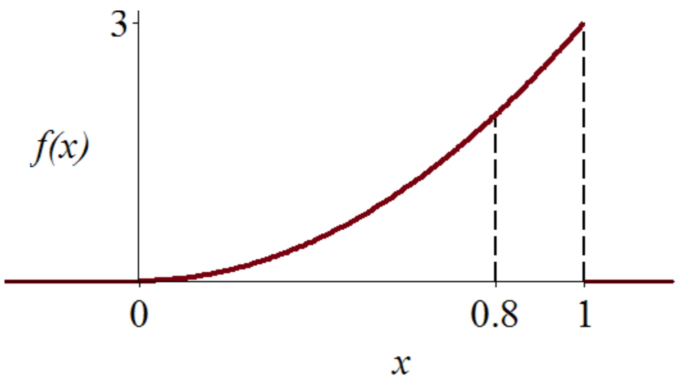
Exercise 5

a. $P(X > 0.8) = 0.2$ and (formula sheet or due to symmetry) $E(X) = \frac{1}{2}$

Uniform density on [0, 1]



Density function of the maximum of 3 random numbers



b. $P(Y > 0.8) = \int_{0.8}^{1} 3y^2 dy$

$= [y^3]_{y=0.8}^{y=1} = 1 - 0.8^3 = 48.8\%$ and $E(Y) = \int_0^1 y \cdot 3y^2 dy = \left[\frac{3}{4}y^4\right]_{y=0}^{y=1} = \frac{3}{4}$

c. $F_Y(y) = P(\max(X_1, X_2, X_3) \leq y) = P(X_1 \leq y \text{ and } X_2 \leq y \text{ and } X_3 \leq y)$
$\overset{\text{ind.}}{=} P(X_1 \leq y) \cdot P(X_2 \leq y) \cdot P(X_3 \leq y) = [F_X(y)]^3$
$F_X(y)$ is the distribution function of X: $F_X(y) = y$ if $0 \leq y \leq 1$,
(*In the graph of* f_X *this is the area above the interval* $[0, y]$:
$length \times height = y \times 1$)
so $F_Y(y) = y^3$ and $f_Y(y) = \frac{d}{dy}F_Y(y) = 3y^2$ for $0 \leq y \leq 1$
(and $f_Y(y) = 0$ elsewhere)