Test Probability Theory (M4-201300180), June 20, 2017, 8.45-10.45 h. in Therm

Lecturer: Dick Meijer - Module coordinator: Klaas Sikkel

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

1. Given is the following joint probability distribution, i.e. the probabilities $P(X = x \text{ and } Y = y)$ of two random variables X and Y:

$x \backslash y$	0	1	2
0	$\frac{1}{10}$	$\frac{1}{8}$	$\frac{1}{10}$
1	$\frac{1}{8}$	$\frac{1}{10}$	$\frac{1}{8}$
2	$\frac{1}{10}$	$\frac{1}{8}$	$\frac{1}{10}$

a. Determine the probability distribution of X, $E(X)$ and $var(X)$.
b. Determine the correlation coefficient of X and Y.
c. Are X and Y independent? Motivate your answer.
d. Compute $P(Y > X)$.
e. Determine the conditional distribution of Y given $X = 1$ and calculate $E(Y|X = 1)$.

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

a. Determine $P(X > 0.8)$ and $E(X)$.
b. Give the distribution function $F(x)$ of X for $0 \leq x \leq 1$.

If we choose the largest number out of four randomly chosen numbers between 0 and 1, then this maximum Y has the following density function: $f_Y(y) = \begin{cases} 4y^3 & \text{if } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$

c. Determine $P(Y > 0.8)$, $E(Y)$ and $var(Y)$.
d. Show that Y has the aforementioned density function if $Y = \max(X_1, X_2, X_3, X_4)$ and X_1, X_2, X_3 and X_4 are independent random numbers between 0 and 1.

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 lifetime of a cylinder is **5 years on average**, but the variation in lifetime is large: as an approximating model, the normal distribution with a standard deviation of 2 years is used. An NGO places pumps and, to enable prompt maintenance on the spot, a spare cylinder is supplied as well.
Let X be the lifetime of the first cylinder that is used and Y the lifetime of the spare cylinder.

a. Determine $P(X < 3)$.
b. Give the distribution of $X - Y$, the difference in lifetime of the cylinders. State on which (reasonable) assumption(s) you based this distribution.
c. Calculate $P(X + Y > 12)$.

4. The UT 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 UT hopes to attract more students than was the case with the former technical studies. Earlier research showed that former 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 research is conducted to measure the interest in the TOM-model among these students.
In a sample of 192 students, 60 students appeared to have serious interest 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. Apply continuity correction.

5. John has acquired 5 of the 100 lots, sold at a lottery held by a village. Local businesses sponsored the lottery with 10 prizes. Give the type of **the distribution of *X*** = "the number of prizes won by John" and calculate the **probability that John wins at least 1 prize** in the following two situations:

a. Each lot can result in at most one prize.
b. Each lot can result in more than one prize: all 100 lots participate in every draw for a prize.

Grade = $1 + \frac{\text{number of points}}{34} \times 9$, rounded.

1					2				3			4	5		Total
a	b	c	d	e	a	b	c	d	a	b	c		a	b	
3	2	2	2	2	2	1	3	2	2	2	3	4	2	2	34

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

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

Solution:

Exercise 1

x \ y	0	1	2	$P(X=x)$
0	$\frac{1}{10}$	$\frac{1}{8}$	$\frac{1}{10}$	$\frac{13}{40}$
1	$\frac{1}{8}$	$\frac{1}{10}$	$\frac{1}{8}$	$\frac{14}{40}$
2	$\frac{1}{10}$	$\frac{1}{8}$	$\frac{1}{10}$	$\frac{13}{40}$
$P(Y=y)$	$\frac{13}{40}$	$\frac{14}{40}$	$\frac{13}{40}$	1

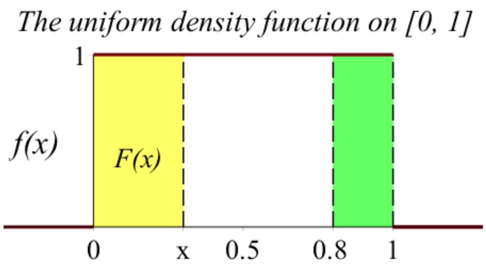
a. For the probability distribution of X, see the table.
$E(X) = 1$ (due to symmetry) and
$var(X) = E(X^2) - (EX)^2$
$= \left(0 \times \frac{13}{40} + 1 \times \frac{14}{40} + 4 \times \frac{13}{40}\right) - 1^2 = \frac{13}{20}$

b. The covariance of X and Y:
$cov(X,Y) = E(XY) - EX \times EY$
$E(XY) = \sum\sum xy\, P(X = x \text{ en } Y = y)$
$= 1\times1\times\frac{1}{10} + 1\times2\times\frac{1}{8} + 2\times1\times\frac{1}{8} + 2\times2\times\frac{1}{10} = 1$
$cov(X,Y) = 1 - 1 \times 1 = 0$, so $\rho = \frac{cov(X,Y)}{\sigma_X\sigma_Y} = 0$

c. X and Y are **dependent**, because (for example):
$\frac{1}{10} = P(X = 0 \text{ and } Y = 0) \neq P(X = 0)P(Y = 0) = (\frac{13}{40})^2$.
(Note that from $cov(XY) = 0$ doesn't follow (automatically) that X and Y are independent, however the reverse holds: if independent, then not correlated.
The statement "if correlated ($\rho \neq 0$), then not independent" is correct).

d. $P(Y > X) = P(X = 0 \text{ and } Y = 1) + P(X = 0 \text{ and } Y = 2) + P(X = 1 \text{ and } Y = 2)$
$= \frac{1}{8} + \frac{1}{10} + \frac{1}{8} = \frac{7}{20}$ $(= 0.35)$

e. $P(Y = 0 \mid X = 1) = \frac{P(Y=0 \text{ and } X=1)}{P(X=1)} = \frac{\frac{1}{8}}{\frac{14}{40}} = \frac{5}{14}$
Similarly $P(Y = 1|X = 1) = \frac{4}{14}$ and $(Y = 2|X = 1) = \frac{5}{14}$.
So $E(Y|X = 1) = 1$ (symmetry!)

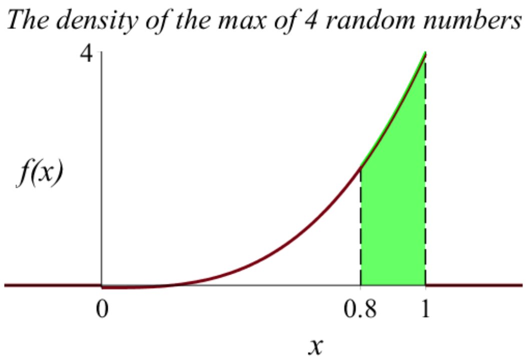


Exercise 2

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

b. $F(x) = x \cdot 1 = x$ for $0 \leq x \leq 1$.

c. $P(Y > 0.8) = \int_{0.8}^{1} 4y^3 dy = [y^4]_{y=0.8}^{y=1} = 1 - 0.8^4 = 59.04\%$
$E(Y) = \int_0^1 y \cdot 4y^3 dy = \left[\frac{4}{5}y^5\right]_{y=0}^{y=1} = \frac{4}{5}$
$E(Y^2) = \int_0^1 y^2 \cdot 4y^3 dy = \left[\frac{4}{6}y^6\right]_{y=0}^{y=1} = \frac{2}{3}$,
so $var(Y) = E(Y^2) - (EY)^2 = \frac{2}{3} - \left(\frac{4}{5}\right)^2 = \frac{2}{75}$



d. $Y = \max(X_1, X_2, X_3, X_4)$:
$F_Y(y) = P(Y \leq y) = P(\max(X_1, X_2, X_3, X_4) \leq y)$
$= P(X_1 \leq y) \times \ldots \times P(X_4 \leq y) = y^4$
since for a $U(0,1)$-distribution we have $P(X_1 \leq y) = y$ $(0 \leq y \leq 1)$, see part b.
$f_Y(y) = \frac{d}{dy}F_Y(y) = 4y^3$, for $0 \leq y \leq 1$.

Exercise 3

a. $P(X < 3) = P\left(\frac{X-5}{2} < \frac{3-5}{2}\right) = P(Z < -1) = 1 - \Phi(1.00) = 1 - 0.8413 = 15.87\%$

b. If we assume independence of the lifetimes, we can use the $N(5{,}4)$-distribution of both X and Y $X - Y$ is $N(\mu_X - \mu_Y, \sigma_X^2 + \sigma_Y^2)$-distributed, so $N(0, 8)$-distribution.

c. $X + Y$ 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

Assuming independence of the trials (serious interest or not) X is $B(192, 0.25)$-distributed, with $E(X) = np = 48$. This expectation is large enough (>5) to apply an approximate normal distribution. X is according the CLT approximately $N\big(np, np(1-p)\big) = N(48, 36)$, so:

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

Exercise 5

a. 10 draws **without** replacement out of 100 lots including the 5 lots of John, so X is hypergeometrically distributed.
$P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{\binom{95}{10}}{\binom{100}{10}} = 1 - \frac{95}{100} \times \frac{94}{99} \times \ldots . \times \frac{86}{91} \approx 41.6\%$
(*Another approach is possible by choosing the 5 lots of John out of the 100 lots, of which 10 lots lead to a prize and 90 don't lead to a prize. Via* $P(X = 0) = \frac{\binom{90}{5}}{\binom{100}{5}}$ *we find the same answer*!)

b. 10 draws **with** replacement: 10 trials with success rate $\frac{5}{100}$, so X has a $B(10, 0.05)$-distribution.
$P(X \geq 1) = 1 - P(X = 0) = 1 - 0.95^{10} \approx 40.1\%$