Test Probability Theory, 14 June 2018, 13:45-15:45h, Therm (201300180)

Lecturers: Aleida Braaksma, Judith Timmer - Module coordinator: Klaas Sikkel

This test has 6 exercises, a formula sheet and tables of the binomial, Poisson and normal distribution. You should motivate all answers. A regular scientific calculator is allowed, a programmable calculator ('GR') is not allowed.

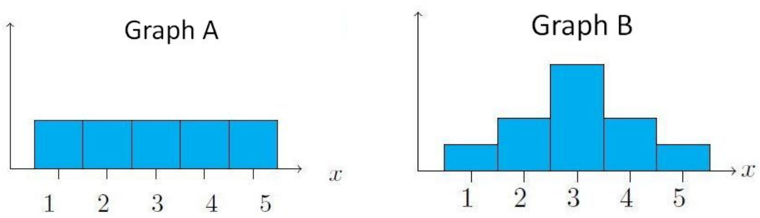
1. Answer the following questions and give an adequate motivation.

   (a) True or false: the following table represents a valid probability table.

outcomes	1	2	3	4	5
probability	0.2	0.3	0.4	-0.1	0.2

   (b) In three tosses of a coin with outcomes T (tails) and H (heads), which of the following equals the event 'exactly two heads'?
   $A = \{THH, HTH, HHT, HHH\}$
   $B = \{THH, HTH, HHT\}$
   $C = \{THH, HTH\}$

   (c) The graphs below show the probability functions of two random variables. Which has the smallest standard deviation?



2. The time X (measured in years) required to complete a software project has the following density function.

$$f_X(x) = \begin{cases} cx(1-x), & 0 \le x \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

   (a) Show that $c = 6$.

   (b) What is the probability that a project is completed in less than four months?

   (c) Calculate $E(X)$ and var(X).

3. In the elevators of a university building one can read: "At most 1200 kg or 16 people". So, if 16 people weigh, on average, 75 kg or less then the elevator is not overloaded (16×75 = 1200). Assume that each of the users of the elevator has a weight that is normally distributed with expectation 72 kg and a standard deviation of 6 kg. $X_1, X_2, \ldots, X_{16}$ are the independent weights of 16 persons who'd like to use the elevator together.

   (a) Calculate $P(X_1 > 75)$, the probability that a single person's weight is over 75 kg.

   (b) Calculate the probability that all 16 people have a weight below 75 kg.

   (c) Calculate the probability that the elevator is overloaded if the 16 people use it together.

4. A student has developed a prototype of a new app to support the choice for bars, discos and dining rooms in the region when going out. Before he invests in further development of the app he would like to have a proper indication of the number of interested students for this app. For this, he uses a sample of n students. The students in the sample are asked if they are seriously interested in buying the app for 1 Euro. Let p be the probability that a random student is seriously interested and let X be the number of seriously interested students in the sample.

   (a) What is the probability distribution of X? Motivate your choice by checking the assumptions of the distribution.
   (b) Calculate or approximate as good as possible: $P(X > 0)$ if $n = 10$ and $p = 0.2$.
   (c) Calculate or approximate as good as possible: $P(X > 25)$ if $n = 100$ and $p = 0.2$.
   (d) Calculate or approximate as good as possible: $P(X > 5)$ if $n = 240$ and $p = 0.01$.

5. In the Netherlands 1 out of 100 people has rheumatoid arthritis. There is a test, the 'rheuma test', available that usually gives a positive result for rheumatology patients and a negative result for non-rheumatology patients. However, the test is not 100% precise. It has a so-called specificity (that is, the probability of a negative result if the person does not have the disease) of 80% and a sensitivity (probability of a positive result if the person has the disease) of 70%.

   (a) What is the probability that a random person has a positive test result?
   (b) What is the probability that a person has the disease if the test result is positive?

6. An insurance company has a call centre that automatically pre-sorts customers. Via a menu, where questions have to be answered and data has to be provided, the customer is lead to an employee being an expert on the matter. The time that is needed until the first contact with the employee is set, X, is modelled as an exponentially distributed variable with $\lambda_1 = 1/2$. The operating time Y (in minutes) of the employee is supposed to be exponentially distributed with parameter $\lambda_2 = 1$. We suppose that the time spent in the menu, X, and the operating time, Y, are independent. Let $V = X + Y$ be the sojourn time of a customer in the system.

   (a) Calculate $P(X > 2\mu_X | X > \mu_X)$.
   (b) Determine the density function of V using a convolution integral.
   (c) Two customers arrive at the same time in the telephone system; their times spent in the menu are called X_1 and X_2. Determine the *distribution function* of the maximum of the two times spent in the menu (the maximum of the times at which both customers are finished).

Points:

1			2			3			4				5		6			total
a	b	c	a	b	c	a	b	c	a	b	c	d	a	b	a	b	c	
1	1	1	2	2	3	1	2	2	2	1	2	2	2	2	2	3	3	34

Grade: $\dfrac{\text{number of points}}{34} \times 9 + 1$ (rounded to one decimal)

Formula sheet Probability Theory for BIT and TCS in module 4

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

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

Probability Theory, 14 June 2018

1

(a) False : $p(4) < 0$ is not allowed.

(b) Exactly two heads in three tosses, so, there should be one tails in first, second or third toss. Event B is correct.

(c) Both probability functions have expectation 3 (symmetry). The outcomes in graph B have a larger probability to be near the expectation, so, a smaller variance. Hence, the standard deviation of graph B is smaller.

2

(a) A density function should have $\int_{-\infty}^{\infty} f_X(x)dx = 1$.

$$\Leftrightarrow 1 = \int_0^1 cx(1-x)\,dx = c\int_0^1 (x-x^2)dx = c\left[\tfrac{1}{2}x^2 - \tfrac{1}{3}x^3\right]_0^1 = \tfrac{1}{6}c$$

$\Leftrightarrow c = 6$.

(b) 4 months is $\frac{1}{3}$ year

$$P(X < \tfrac{1}{3}) = \int_0^{\frac{1}{3}} 6x(1-x)\,dx = \left[3x^2 - 2x^3\right]_0^{\frac{1}{3}} = \tfrac{1}{3} - \tfrac{2}{27} = \tfrac{7}{27}$$

(c) $E(X) = \frac{1}{2}$ (by symmetry), or

$$E(X) = \int_0^1 x \cdot 6x(1-x)\,dx = \left[2x^3 - \tfrac{3}{2}x^4\right]_0^1 = \tfrac{1}{2}$$

$$E(X^2) = \int_0^1 x^2 \cdot 6x(1-x)\,dx = \left[\tfrac{3}{2}x^4 - \tfrac{6}{5}x^5\right]_0^1 = \tfrac{3}{10}$$

$$\text{var}(X) = E(X^2) - (E(X))^2 = \tfrac{3}{10} - (\tfrac{1}{2})^2 = \tfrac{1}{20}.$$

3

(a) $P(X_1 > 75) = P\left(\frac{X_1 - 72}{6} > \frac{75-72}{6}\right) = P(Z > \frac{1}{2})$ $(Z \sim N(0,1))$

$= 1 - \Phi(\frac{1}{2}) = 0.3085$

(b) $P(X_1 < 75$ and $X_2 < 75$ and ... and $X_{16} < 75)$

$= P(X_1 < 75) \cdot P(X_2 < 75) \cdot \ldots \cdot P(X_{16} < 75)$ by independence

$= (P(X_1 < 75))^{16}$ due to the identical distribution

From (a): $P(X_1 < 75) = \Phi(\frac{1}{2})$, so

$(\Phi(\frac{1}{2}))^{16} \approx 0.0027$.

(c) Overloaded if the total weight exceeds 1200 kg.

Total weight $\sum_{i=1}^{16} X_i \sim N(16 \cdot 72, 16 \cdot 6^2) = N(1152, 576)$

(by independence).

$P(\sum_{i=1}^{16} X_i > 1200) = P\left(\frac{\sum_{i=1}^{16} X_i - 1152}{\sqrt{576}} > \frac{1200 - 1152}{\sqrt{576}}\right)$

$= P(Z > 2) = 1 - \Phi(2) = 0.0228$.

4

(a) $X \sim B(n,p)$

Assumptions: the students are independent and each student is interested with probability p. According to the text, these assumptions are satisfied.

(b) $P(X > 0) = 1 - P(X = 0) = 1 - (0.8)^{10} \approx 0.893$

(c) $n \geq 25$, $np = 20 > 5$ and $n(1-p) = 80 > 5$

By CLT: approximately $N(np, np(1-p)) = N(20, 16)$.

$P(X > 25) = P(X \geq 26) = P(X \geq 25.5)$ (continuity correction)

$= P\left(\frac{X-20}{4} \geq \frac{25.5-20}{4}\right) = P(Z \geq 1.38) = 1 - \Phi(1.38) = 0.0845$

(d) $n \geq 25$, $np = 2.4 < 5 \Rightarrow X$ is approximately Poisson (2.4).

$P(X > 5) = 1 - P(X \leq 5) = 1 - 0.964 = 0.036$.

5

(a) Define T: test has positive result

R: person has rheumatoid arthritis

Given: $p(R) = 0.01$, $p(T^c | R^c) = 0.80$, $p(T|R) = 0.70$.

law of total probability

$$p(T) = p(T|R) \cdot p(R) + p(T|R^c) \cdot p(R^c)$$
$$= 0.70 \cdot 0.01 + (1 - 0.80) \cdot (1 - 0.01) = 0.205$$

(b) Bayes' rule: $p(R|T) = \dfrac{p(T|R) \cdot p(R)}{p(T)} = \dfrac{0.70 \cdot 0.01}{0.205} \approx 0.034$.

6

(a) $\mu_X = 1/\lambda_1 = 2$

$P(X > 2\mu_X \mid X > \mu_X) = P(X > 4 \mid X > 2)$

lack of memory

$$= P(X > 2) = \int_2^\infty \tfrac{1}{2} e^{-x/2}\, dx = \left[-e^{-x/2}\right]_2^\infty = e^{-1} \approx 0.368$$

(b) $f_V(v)$ via convolution integral:

$$f_V(v) = \int_{-\infty}^{\infty} f_X(x) f_Y(v - x)\, dx = \int_0^v \tfrac{1}{2} e^{-\frac{x}{2}} e^{-(v-x)}\, dx$$

$$= \int_0^v \tfrac{1}{2} e^{\frac{x}{2}} e^{-v}\, dx = e^{-v} \left[e^{\frac{x}{2}}\right]_0^v = e^{-v/2} - e^{-v}, \quad v \geq 0$$

$f_V(v) = 0$, otherwise.

(c) Define $W = \max(X_1, X_2)$

$$F_W(w) = P(W \leq w) = P(\max(X_1, X_2) \leq w)$$

independence

$$= P(X_1 \leq w \text{ and } X_2 \leq w) = P(X_1 \leq w) \cdot P(X_2 \leq w)$$

identically distributed

$$= (1 - e^{-w/2})^2, \quad w \geq 0$$

$F_W(w) = 0$ if $w < 0$.