Test Probability and Statistics I (Ch. 1-5), April 26, 2018

This exam consists of 5 exercises and formulas. A simple scientific calculator is allowed.

1. In some country 40% of all employees work for large companies (> 100 employees): 80% of these employees have a company pension, whereas only 30% of the employees of small companies (≤ 100 employees) have a company pension.

a. What is the probability that an arbitrary employee in this country has a company pension? Answer this question by first defining relevant events, expressing the given probabilities in these events and using the rules of probability to compute the requested probability.

b. What is the probability that an employee with a company pension is working in a small company?

2. The quality control of the mass production of nails is organized by measuring the nails in a relatively small sample of n nails. The manufacturer guarantees that at most 1% of the nails have a size outside prescribed tolerance bounds. For answering the following questions assume that exactly 1% of the nails are substandard (outside tolerance bounds).
X is the number of substandard nails in a random sample of n nails.

a. Compute $P(X \leq 1)$, the probability that at most one of the nails is substandard, for $n = 15$ nails.

b. Compute or approximate $P(X \geq 2)$ for a random sample of $n = 500$ nails.

3. In the most popular lottery in The Netherlands there are many small prizes in each draw. As a result the probability to win a prize is high: 50% at each draw.
X is the number of draws in which grandma has to participate until she has a prize.

a. What is the expected number of participations (draws) in this lottery until grandma has a prize.

b. What is the probability that she has to participate **at least 7 times** until she wins a prize?

4. In the table the joint probability function $P(X = x \text{ and } Y = y)$ of X and Y is given.

a. Give the probability distribution of X and determine $E(X)$ and $var(X)$.

b. Determine the covariance and the correlation coefficient of X and Y and give an interpretation for the values you found.

c. Are X and Y independent? Motivate your answer.

d. Determine the conditional probability distribution of Y given $X = 1$ and calculate $E(Y|X = 1)$.

x \ 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	$\frac{1}{10}$	$\frac{1}{8}$	$\frac{1}{10}$

5. In a village lottery 100 tickets are sold and John bought 5 of them. There are 10 prizes available, which are given to the owners of tickets that are randomly drawn from a bowl with tickets.
Now give the **probability distribution of X** = "the number of prizes of John for his 5 tickets" and compute the **probability of at least one prize** for John, in two cases:

a. With each ticket you can win at most one prize.

b. Each ticket can give you more than one prize (At every draw all tickets are in the bowl).

Grading:

1a	1b	2a	2b	3a	3b	4a	4b	4c	4d	5a	5b	Total
4	2	2	2	2	2	2	3	3	2	2	2	26

Formulas

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 $P(X = x) = \frac{\mu^x e^{-\mu}}{x!}, x = 0, 1, 2, \ldots$	μ	μ

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

Solutions

Exercise 1

a. C = "employee has a Company pension" and L = "employee works at a Large company"

$$P(L) = 0.4, \quad P(C|L) = 0.80 \text{ and } P(C|\overline{L}) = 0.30$$

$P(\overline{L}) = 0.6$ is the probability of a small company (Use a Venn diagram to see that):

$$P(C) = P(L \cap C) + P(\overline{L} \cap C) = P(L)P(C|L) + P(\overline{L})P(C|\overline{L})$$
$$= 0.40 \times 0.80 + 0.6 \times 0.30 = 0.50$$

b. $P(\overline{L}|C) = \frac{P(\overline{L} \cap C)}{P(C)} = \frac{P(\overline{L})P(C|\overline{L})}{P(C)} = \frac{0.6 \times 0.3}{0.50} = \frac{18}{50} = 36\%$

Note: do not use P for "company pension", since the notation $P(P)$ is confusing.

Exercise 2

a. X is $B(15, 0.01)$, so $P(X \leq 1) = P(X=0) + P(X=1) = 0.99^{15} + 15 \cdot 0.01 \cdot 0.99^{14} \approx 99.0\%$

b. X is $B(500, 0.01)$, so $\mu = 500 \times 0.01 = 5 < 10$, so X is approximately Poisson distr. with par. $\mu = 5$.

$$P(X \geq 3) = 1 - P(X=0) - P(X=1) = 1 - e^{-5} - 5e^{-5} = 1 - 6e^{-5} \approx 96.0\%$$

Exercise 3

a. Consecutive participations are Bernoulli trials with success probability $\frac{1}{2}$, so X, the number of participations until a prize is won, has a geometric distribution with parameter $p = \frac{1}{2}$: $E(X) = \frac{1}{p} = 2$.

b. $P(X \geq 7) = P(X > 6) = \left(\frac{1}{2}\right)^6 = 1.5625\%$ ($\approx 1.6\%$).

Exercise 4

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

a. For probability distribution of X, see the table:
$E(X) = 1$ (symmetry) and
$var(X) = E(X^2) - (EX)^2$
$= \left(0^2 \times \frac{13}{40} + 1^2 \times \frac{14}{40} + 2^2 \times \frac{13}{40}\right) - 1^2 = \frac{13}{20} = 0.65$

b. The covariance: $cov(X,Y) = E(XY) - EX \times EY$
Y has the same distribution as X: $E(X) = E(Y)$ and
$$var(X) = var(Y)$$
$E(XY) = \sum\sum x \cdot y \cdot P(X = x \text{ and } Y = y)$
$= 1\times1\times\frac{1}{10} + 1\times2\times\frac{1}{8} + 2\times1\times\frac{1}{8} + 2\times2\times\frac{1}{10} = 1$
$cov(XY) = 1 - 1 \times 1 = 0$, hence $\rho = 0$.
There is no correlation between X and Y (or: no linear relation between X and Y)

c. X and Y are **not** independent, because (for instance):

$$\frac{1}{10} = P(X=0 \text{ and } Y=0) \neq P(X=0)P(Y=0) = (\tfrac{13}{40})^2.$$

(From $\rho(X,Y) = 0$ we can't (directly) conclude X and Y are independent: ρ "measures" only linear dependence).

d. $P(Y=0|X=1) = \frac{P(Y=0 \text{ and } X=1)}{P(X=1)} = \frac{\frac{1}{8}}{\frac{14}{40}} = \frac{5}{14}$

Likewise $P(Y=1|X=1) = \frac{4}{14}$ and $P(Y=2|X=1) = \frac{5}{14}$. Hence $E(Y|X = 1) = 1$ (symmetry)

Exercise 5

a. 10 draws without replacement from 100 tickets of which 5 are John`s: X has a hypergeometric distribution, so $P(X \geq 1) = 1 - P(X=0) = 1 - \frac{\binom{95}{10}}{\binom{100}{10}} = 1 - \frac{95}{100} \times \frac{94}{99} \times \ldots \times \frac{86}{91} \approx 41.6\%$

An alternative approach is to choose 5 out of 100 tickets of which 10 are prize: $1 - \frac{\binom{90}{5}}{\binom{100}{5}} \approx 41.6\%$

b. 10 draws with replacement, so X is $B(10, 0.05)$-distributed.
$P(X \geq 1) = 1 - P(X=0) = 1 - 0.95^{10} \approx 40.1\%$

Note: in b. 10 lots are drawn with replacement, so John can win up to 10 prizes. Choosing 5 lots from 100 (with replacement) from 100 lots of which 10 are "winning" is a false approach since 1 lot can win 2 (or more) prizes.