Solutions to the final exam for Probability 2018/19

Ignacio Cascos, Department of Statistics, Universidad Carlos III de Madrid

2018/19

1. A gambler has a fair coin and a two-headed coin in his pocket. He selects one of the coins at random. When he flips it, it shows heads.
   a. What is the probability that it is the fair coin?
   b. Suppose that he flips the same coin a second time and, again, it shows heads. Now what is the probability that it is the fair coin?
   c. Suppose that he flips the same coin a third time and it shows tails. Now what is the probability that it is the fair coin?

$$F \equiv \text{'fair coin'}, \quad P(F) = 1/2$$
$$H_i \equiv \text{'heads in } i\text{-th toss'}, \quad P(H_i|F) = 1/2, \quad P(H_i|\overline{F}) = 1.$$

The events $H_1, H_2, H_3, \ldots$ are conditionally independent on F.

$$P(F|H_1) = \frac{P(H_1|F)P(F)}{P(H_1|F)P(F) + P(H_1|\overline{F})P(\overline{F})} = \frac{1/2 \times 1/2}{1/2 \times 1/2 + 1 \times 1/2} = \frac{1}{3}.$$
$$P(F|H_1 \cap H_2) = \frac{P(H_1 \cap H_2|F)P(F)}{P(H_1 \cap H_2|F)P(F) + P(H_1 \cap H_2|\overline{F})P(\overline{F})} = \frac{1/4 \times 1/2}{1/4 \times 1/2 + 1 \times 1/2} = \frac{1}{5}.$$
$$P(F|H_1 \cap H_2 \cap \overline{H}_3) = \frac{P(H_1 \cap H_2 \cap \overline{H}_3|F)P(F)}{P(H_1 \cap H_2 \cap \overline{H}_3|F)P(F) + P(H_1 \cap H_2 \cap \overline{H}_3|\overline{F})P(\overline{F})} = \frac{1/8 \times 1/2}{1/8 \times 1/2 + 0 \times 1/2} = 1.$$

2. The Zero-Truncated Poisson (ZTP) distribution is the conditional probability distribution of a Poisson random variable, given that the value of the random variable is not zero (discrete distribution whose support is formed by the positive integers).
   a. Determine the probability mass function of a ZTP random variable with parameter λ.
   b. Determine the mean and variance of a ZTP random variable with parameter λ.
   c. The number of items that a client buys in a supermarket follows a Poisson distribution with parameter $\lambda = 10$. Only when the client has bought some item will he will go through the checkout line. What is the probability that the number of items in a shopper's basket at the supermarket checkout line is greater than 5?
   d. Write a piece of code to simulate 1000 observations of a ZTP random variable with parameter $\lambda = 10$. Use your simulations to approximate the answer to part c).

Denote $X \sim \text{ZTP}(\lambda)$ and $Y \sim \mathcal{P}(\lambda)$.

a. For $k \in \{1, 2, 3, \ldots\}$,

$$P(X = k) = P((X = k) \cap (X > 0))/P(X > 0) = P(Y = k)/P(Y > 0) = e^{-\lambda}\frac{\lambda^k}{k!}(1 - e^{-\lambda})^{-1}.$$

b. $\mathbb{E}[X] = \sum_{k\geq 1} kP(X = k) = \sum_{k\geq 1} ke^{-\lambda}\frac{\lambda^k}{k!}(1 - e^{-\lambda})^{-1} = \lambda(1 - e^{-\lambda})^{-1}$
$\mathbb{E}[X^2] = \sum_{k\geq 1} k^2P(X = k) = \sum_{k\geq 1} k^2e^{-\lambda}\frac{\lambda^k}{k!}(1 - e^{-\lambda})^{-1} = \mathbb{E}[Y^2](1 - e^{-\lambda})^{-1} = (\lambda + \lambda^2)(1 - e^{-\lambda})^{-1}$
$\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = (\lambda - \lambda e^{-\lambda} - \lambda^2 e^{-\lambda})(1 - e^{-\lambda})^{-2}$.

c. The number of items in the client's basket is $Y \sim \mathcal{P}(\lambda = 10)$,

$$P(Y > 5|Y > 0) = \frac{P(Y > 5)}{P(Y > 0)} = \frac{\texttt{1-ppois(5,lambda=10)}}{\texttt{1-ppois(0,lambda=10)}} = 0.9329564$$

d.

```
n=1000
x=rep(0,n)
set.seed(1)
for(i in 1:n){
  while(x[i]==0) {x[i]=rpois(1,lambda=10)}
}
sum(x>5)/n
```

```
## [1] 0.913
```

3. Each of the members of a 7-judge panel independently makes a correct decision with probability 0.7. Assume the panel's decision is made by majority rule and it is mandatory for all judges to take a decision (guilty or not guilty).
   a. What is the probability that the panel makes the correct decision? Denote by X the number of judges in the panel that make a correct decision, then $X \sim \mathrm{B}(n = 7, p = 0.7)$, $P(X \geq 4) = 1 - P(X \leq 3) =$ `1-pbinom(3,size=7,prob=.7)` $= 0.873964$.
   b. Given that 4 of the judges agreed (the other 3 did also agree in the other decision), what is the probability that the panel made the correct decision? We have to compute the conditional probability

$$P(X \geq 4 | X \in \{3, 4\}) = \frac{P(X = 4)}{P(X = 3) + P(X = 4)} = \frac{\texttt{dbinom(4,size=7,prob=.7)}}{\texttt{dbinom(3,size=7,prob=.7)+dbinom(4,size=7,prob=.7)}} = 0.7.$$

   Assume now that each member of the panel can make the decision that the individual is guilty or not, but she can also reject to make a decision. As before, with probability 0.7 she will make the correct decision, but with probability 0.1 she will reject to make any decision.
   c. What is the probability that 4 judges make the correct decision, one rejects to make any decision, and two make the wrong decision? Denote by Y the Multinomial random vector whose first componen is the number of correct decisions, the second component the number of no decisions, and third component number of wrong decisions, $Y \sim \mathrm{M}(n = 7, \mathbf{p} = (0.7, 0.1, 0.2))$. Finally compute $P(Y_1 = 4, Y_2 = 1, Y_3 = 2) =$ `dmultinom(c(4,1,2),prob=c(.7,.1,.2))` $= 0.100842$. Assume now that the panel is selected at random from a group of 50 judges. Out of them, 35 will take the correct decision, while 15 will take the wrong one.
   d. What is the probability that the panel makes the correct decision? Denote by V the Hypergeometric random variable with the number of correct decisions out of $k = 7$ judges selected from a group of 50 out of which 35 will make a correct decision and 15 a wrong one, $V \sim \mathrm{H}(N_1 = 35, N_2 = 15, k = 7)$. Finally, $P(V \geq 4) = 1 - P(V \leq 3) =$ `1-phyper(3,m=35,n=15,k=7)` $= 0.8908506$.
4. IQ scores are commonly assumed to be normally distributed. An extensive Scot- tish survey1 suggests that the mean IQ score for 11-year-old girls is 100.64 and its standard deviation is 14.1, while the mean IQ score for 11-year-old boys is 100.48 and its standard deviation is 14.9. The test was ran on 39343 girls (49.6%) and 40033 boys. Assume that the joint distribution of the IQs of a couple of twins is bivariate normal with correlation 0.8 regardless of their gender.
   a. What percentage of (11-year-old Scottish) girls have an IQ greater than 115? Denote the IQ of a girl by $X \sim \mathrm{N}(\mu = 100.64, \sigma = 14.1)$. $P(X > 115) =$ `1-pnorm(115,mean=100.64,sd=14.1)` $= 0.1542345$.
   b. If 10 girls are selected at random, what is the probability that at least 6 of them have an IQ greater than 115? Denote by V the number of girst (out of 10) with an IQ greater than 115, $V \sim \mathrm{B}(n = 10, p = 0.1542345)$, $P(V \geq 6) = 1 - P(V \leq 5) =$ `1-pbinom(5,size=10,prob=0.1542345)` $= 0.0016079$.
   c. What IQ is exceeded by 75% of girls? `qnorm(.25,mean=100.64,sd=14.1)` $= 91.1296945$.

d. Compute the mean and variance of the IQ in the previous population. Denote by Z the IQ of an individual selected at random

$$\mathbb{E}[Z] = 0.496\mathbb{E}[X] + 0.505\mathbb{E}[Y] = 100.55936$$
$$\mathbb{E}[Z^2] = 0.496\mathbb{E}[X^2] + 0.505\mathbb{E}[Y^2] = 1.0322694 \times 10^4$$
$$\text{Var}[Z] = \mathbb{E}[Z^2] - \mathbb{E}[Z]^2 = 210.5091996.$$

e. What is the probability that an individual selected at random from that population has an IQ greater than 115? Denote the IQ of a boy by $Y \sim \text{N}(\mu = 100.48, \sigma = 14.9)$. If an individual is selected at random, with probability 0.496 it will be a girl, and with probability 0.504 a boy, then (mixture of probabilities,or total probability rule)

`0.496(1-pnorm(115,mean=100.64,sd=14.1))+0.504(1-pnorm(115,mean=100.48,sd=14.9))` $= 0.1596124.$

f. If an individual selected at random has an IQ greater than 115. What is the probability that it is a girl? Apply Bayes Theorem to obtain

$$\frac{0.496P(X > 115)}{0.496P(X > 115) + 0.504P(Y > 115)} = 0.479288.$$

g. Consider now a couple of twins (one sister and one brother). What is the distribution of their average IQ? Denote by W the average IQ, then $W \sim \text{N}(\mu = (100.64 + 100.48)/2 = 100.56, \sigma = \sqrt{14.1^2/4 + 14.9^2/4 + 2 \times 0.8 \times 14.1 \times 14.9/4} = 13.7565)$.

h. What is the probability that the average IQ from part g) is greater than 115? $P(W > 115) =$ `1-pnorm(115,mean=100.56,sd=13.7565)` $= 0.1469313$.

5. From past experience, a professor knows that the test score of a student taking her final examination is a random variable with mean 75.

a. Give an upper bound for the probability that a student's test score will exceed 85. Denote the student's test score as X. We know $\mathbb{E}[X] = 75$ and then applying Markov's inequality

$$P(X \geq 85) \leq \frac{\mathbb{E}[X]}{85} = \frac{75}{85} = 0.8823529.$$

Suppose, in addition, that the professor knows that the variance of a student's test score is equal to 25.

b. What can be said about the probability that a student will score between 65 and 85? We know further that $\text{Var}[X] = 25$, and now we can apply Chebishev's inequaliy to obtain

$$P(65 \leq X \leq 85) = P(\mathbb{E}[X] - 10 \leq X \leq \mathbb{E}[X] + 10) \geq 1 - \frac{\text{Var}[X]}{10^2} = 0.75.$$

c. How many students would have to take the examination to ensure, with probability at least 0.9, that the class average would be within 5 of 75 (between 70 and 80)? Do not use the central limit theorem. Denote the class average by $\overline{X}_n$, where n is the number of students in the class. Since we know that $\mathbb{E}[\overline{X}_n] = 75$ and $\text{Var}[\overline{X}_n] = 25/n$, we can apply Chebishev's inequality to obtain

$$P(75 - 5 \leq \overline{X}_n \leq 75 + 5) \geq 1 - \frac{\text{Var}[\overline{X}_n]}{5^2} = 1 - \frac{25/n}{25} = 0.9.$$

In conclusion $n = 10$.

d. How many students would have to take the examination to ensure, with probability at least 0.95, that the class average would be within 1 of 75 (between 74 and 76)? Use the central limit theorem. We apply now the CLT, and (as long there is a large number of students in the class) the approximate distribution of the sample mean (class average) is $\overline{X}_n \approx \text{N}(\mu = 75, \sigma = 5/\sqrt{n})$. We must determine n such that $P(74 \leq \overline{X}_n \leq 76) = 0.95$, so $(76 - 75)/(5/\sqrt{n}) =$ `qnorm(.975)` $= 1.96$, so $n = 96.04$. The number of students that should take the exam is 96 (97 in case we need to be really sure that we can guarantee the statement).