FIRST COURSE IN PROBABILITY AND STATISTICS (MS-A0501/2/3/4)
Department of Mathematics and Systems Analysis, Aalto University
Pekka Pere, Natalia Vesselinova, Jukka Kohonen

Suggested solutions for the questions of the 16.04.2024 exam

1.

a)

- $\mathsf{P}(A \text{ or } B \text{ or both}) = \mathsf{P}(A \text{ or } B) - \mathsf{P}(A \text{ and } B) = \mathsf{P}(A \cup B) - \mathsf{P}(A \cap B) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(A \cap B) = 0.5 + 0.3 - 0.2 = 0.6.$
- $\mathsf{P}(A \text{ only}) = \mathsf{P}(A \text{ but not } B) = \mathsf{P}(A \cap B^C) = 0.5 - 0.2 = 0.3.$
- $\mathsf{P}(\text{at least } A) = \mathsf{P}(A) = 0.5.$
- $\mathsf{P}(\text{at most } A) = \mathsf{P}(B^C) = 1 - \mathsf{P}(B) = 1 - 0.3 = 0.7.$

b) The first option “John is married with two children” is more probable because the description “likes to spend his evenings tackling mathematical puzzles and playing computer games” restricts the set of men fitting the description. This answer suffices.

Venn-diagram of Figure 1 illustrates ($A =$ “married men with two children”, $B =$ “men tackling mathematical puzzles and playing computer games” ja $S =$ “men”). In principle it is possible (though unlikely) that the two events are of equal probability (Figure 4 displaying case 3).

Three cases:

1. Married men with two children and men who like to spend evenings tackling mathematical puzzles and playing computer games are disjoint sets ($A \cap B = \emptyset$; Figure 2). Probability of John having two children and tackling mathematical puzzles and playing computer games is especially small or nill under this circumstance.

2. Men who like to spend evenings tackling mathematical puzzles and playing computer games are a (proper) subset of married men with two children ($B \subset A$; Figure 3). Also under this circumstanc it is more probable that John is married with two children than that he is married with two children and likes to spend evenings tackling mathematical puzzles and playing computer games.

3. All married men with two children like to spend evenings tackling mathematical puzzles and playing computer games ($A \subset B$; Figure 3 switching the labels and meanings of sets A and B) or married men with two children and men who like to spend evenings tackling mathematical puzzles and playing computer games are the same set ($A = B$, Figure 4). In these two cases the probabilities of the events “John is married with two children” and “John is married with two children, and likes to spend his evenings tackling mathematical puzzles and playing computer games” are equal.

Daniel Kahneman and Amos Tversky say that a *conjunction fallacy* is committed when it is reasoned that a conjunction of two events is more probable than one of the events. Kahneman is a 2002 Nobel laureate of economics.

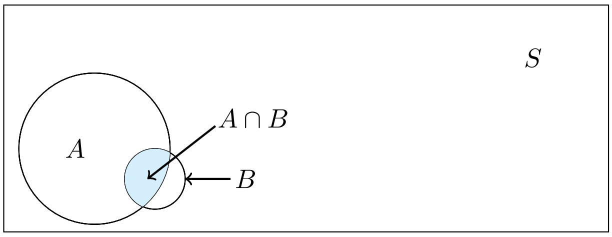


Figure 1: Some married men with two children like to spend evenings tackling mathematical puzzles and playing computer games but not all.

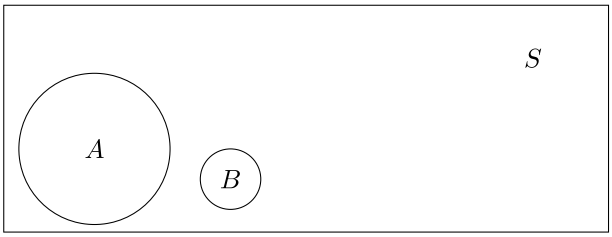


Figure 2: None of the married men with two children like to spend evenings tackling mathematical puzzles and playing computer games.

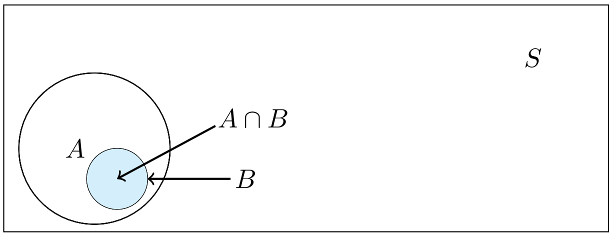
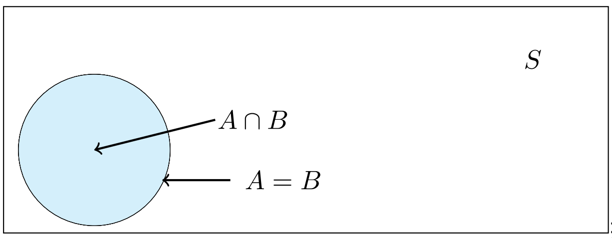


Figure 3: Men who like to spend evenings tackling mathematical puzzles and playing computer games are a (proper) subset of married men with two children.



;

Figure 4: Married men with two children and men who like to spend evenings tackling mathematical puzzles and playing computer games are the same set.

2.

a) If the distribution were symmetric then the average weight would be the average of the 0.025. and 0.975. quantiles: $(3700 + 6900)/2 = 5300$. The distribution is not symmetric because the average of the quantiles does not equal the average weight 5200. We can not infer from the information given that the distribution is symmetric.

Part B of Figure 2 of the article reveals that the distribution of the ecological body mass is slightly skewed to the right.

b) The average weight equals the average of the 0.025th and 0.975th quantiles: $(5100 + 9100)/2 = 7100$. The quantiles are as distant from the average weight so the distribution may be symmetric. The distribution is not necessarily symmetric: The other (corresponding) quantiles (e.g. the 0.1th and 0.9th) of the distribution may place unsymmetricaly relative to the average. We can not infer from the information given that the distribution is symmetric.

c) Let us denote the weight of a full-grown *Tyrannosaurus rex* by X. We have been told that the weight follows a Normal distribution and that

$$\mathsf{P}(X \leq 5100) = \mathsf{P}(X \geq 9100) = 0.025.$$

The 0.025th and 0.975th quantiles of the Standard Normal distribution are -1.959964 and 1.959964, respectively (`qnorm(0.025)` and `qnorm(0.975)`). Hence it must be the case that

$$\mathsf{P}\left(\frac{5100 - 7100}{\sigma} \leq -1.959964\right) = \mathsf{P}\left(\frac{9100 - 7100}{\sigma} \geq 1.959964\right) = 0.025.$$

A way to solve the standard deviation σ is the following:

$$\begin{aligned} \frac{5100 - 7100}{\sigma} &= -1.959964 && \Leftrightarrow \\ -2000 &= -1.959964\sigma && \Leftrightarrow \\ \sigma &= \frac{-2000}{-1.959964} = 1020.427. \end{aligned}$$

The standard deviation and variance of weight of a full-grown *Tyrannosaurus rex* are 1020 kg and $1020.427^2 = 104\,1271$ kg^2.

3. The lower and upper bounds of the 95 % Wald confidence interval are

$$\frac{y}{n} - 1.960\sqrt{\frac{\frac{y}{n}\left(1 - \frac{y}{n}\right)}{n}}$$

and

$$\frac{y}{n} + 1.960\sqrt{\frac{\frac{y}{n}\left(1 - \frac{y}{n}\right)}{n}}.$$

Above 1.960 is the 0.975th quantile of the Standard normal distribution (`qnorm(0.975)`). In the present application the bounds are

$$\frac{1311}{1700} - 1.960\sqrt{\frac{\frac{1311}{1700}\left(1 - \frac{1311}{1700}\right)}{1700}} = 0.7512074.$$

and

$$\frac{1311}{1700} + 1.960\sqrt{\frac{\frac{1311}{1700}\left(1-\frac{1311}{1700}\right)}{1700}} = 0.7911456.$$

The bounds can be calculated with R as follows:

```
n <- 1700
p <- 1311/n
q <- 1-p
p-1.960*sqrt(p*q/n)
p+1.960*sqrt(p*q/n)

# Alternatively, install the Mosaic package and use binom.test():
install.packages("mosaic")
library(mosaic)
binom.test(x=1311,n=1700,ci.method="Wald")
 - -
## number of successes = 1311, number of trials = 1700, p-value < 2.2e-16
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
## 0.7512077 0.7911452
 - -
```

The confidence interval is $[0.751, 0.791]$. It does not cover the actual voter turnout percentage 70.5.

An explanation for the outcome is that the interviewed people have hesitated to admit that they had not voted so they have told the interviewer that they had.

4.

a) The likelihood function for the observed data and stochastic model (Poisson distribution) is:

$$\begin{aligned} \mathsf{L}(\lambda) &= \prod_{i=1}^{3} p_X(x_i) \\ &= \prod_{i=1}^{3} e^{-\lambda}\frac{\lambda^{x_i}}{x_i!} \\ &= e^{-3\lambda}\frac{\lambda^{x_1+x_2+x_3}}{x_1!x_2!x_3!} \\ &= e^{-3\lambda}\frac{\lambda^{22}}{10!5!7!}. \end{aligned}$$

The log-likelihood function can simplify our calculations of the MLE and is obtained by taking the natural logarithm of the likelihood function:

$$\begin{aligned} l(\lambda) &= \ln(\mathsf{L}(\lambda)) \\ &= \ln\left(e^{-3\lambda}\frac{\lambda^{22}}{10!5!7!}\right) \\ &= \ln e^{-3\lambda} + \ln\left(\frac{\lambda^{22}}{10!5!7!}\right) \\ &= -3\lambda + 22\ln\lambda - \ln(10!5!7!) \end{aligned}$$

To find the turning point(s) we take the first derivative of the log-likelihood function:

$$
\begin{aligned}
l'(\lambda) &= \frac{\partial l(\lambda)}{\partial \lambda} \\
&= -3 + 22\frac{1}{\lambda}
\end{aligned}
$$

and set it to zero:

$$
\begin{aligned}
l'(\lambda) &= \frac{\partial l(\lambda)}{\partial \lambda} = 0 \\
-3 + 22\frac{1}{\lambda} &= 0 \\
\Rightarrow \lambda &= \frac{22}{3}.
\end{aligned}
$$

We take the second derivative to determine if this turning point is a minimum or a maximum value:

$$
\begin{aligned}
l''(\lambda) &= \frac{\partial^2 l(\lambda)}{\partial \lambda^2} \\
&= -22\frac{1}{\lambda^2} < 0.
\end{aligned}
$$

We must also check the boundaries: $\lambda \to 0^+$ and $\lambda \to \infty$. For both of these boundaries, $\ln(\mathsf{L}(\lambda)) \to -\infty$. Therefore, (and since the second derivative is negative), $\hat{\lambda} = 22/3 \approx 7.33$ is the maximum. Note that the MLE we found is the sample mean of the observations.

b) The unnormalised posterior is the product of the prior and the likelihood:

$$
p(\lambda) \cdot p(x|\lambda) = p(\lambda) \cdot \mathsf{L}(\lambda).
$$

The posterior is:

$$
\begin{aligned}
&\propto e^{\lambda} \cdot e^{-3\lambda}\lambda^{22} \\
&\propto e^{-4\lambda}\lambda^{22}.
\end{aligned}
$$

This is the core of the Gamma distribution, which we have mentioned in the lectures. We can follow the same steps as in a) to find out that the MAP estimate is $\hat{\lambda} = 22/4 = 11/2 = 5.5$.

c) Now, our prior is the (unnormalised) posterior from the preceding step, part b). And the likelihood considering the new data is:

$$
\begin{aligned}
\mathsf{L}(\lambda) &= p_X(x_4) \\
&= e^{-\lambda}\frac{\lambda^{20}}{20!}.
\end{aligned}
$$

Hence, our updated unnormalised posterior is

$$
e^{-4\lambda}\lambda^{22} \cdot e^{-\lambda}\lambda^{20} = e^{-5\lambda}\lambda^{44},
$$

and the new MAP estimate is $\hat{\lambda} = 44/5 = 8.8$.