Exam: Introduction to probability and statistics

M2:Masco 2025-2026

28-11-2025

Exercise 1:

In a game of chance where a coin is flipped, you win when you get heads and you lose when you get tails. A coin is randomly chosen from a bag containing two types of biased coins, A and B. The probability of winning with a type-A coin is 80%, and the probability of winning with a type-B coin is 10%. We also know that the probability of choosing a type-A coin is 30% and that of choosing a type-B coin is 70%.

1. A coin is chosen at random and flipped once. What is the probability of losing?
2. Given that you win, what is the probability that the chosen coin was of type A?

Exercise 2:

We model the number of *spikes* during an experiment by a random variable X following a Poisson distribution with unknown parameter $\Theta = \theta > 0$. We write:

$$P(X = x \mid \Theta = \theta) = e^{-\theta}\,\frac{\theta^x}{x!}, \qquad x = 0, 1, 2, \ldots$$

Based on prior knowledge of the phenomenon, we know that the random variable Θ is discrete and takes the values $\{0.1, 0.5, 1, 2\}$ with associated probabilities $\{0.25, 0.25, 0.4, 0.1\}$.

1. In a first experiment, we observe 2 spikes. Give the posterior distribution of Θ.
2. Determine the posterior mean of Θ.
3. Compute the posterior probability that Θ is smaller than 1.
4. The experiment is repeated and we now observe 1 spike. Assuming the observations are independent, determine the posterior distribution of Θ based on both experiments.

Exercise 3:

Weather modification, or cloud seeding, consists of treating individual clouds or storm systems with various inorganic and organic materials in the hope of increasing precipitation. The dataset `clouds.csv` contains two variables giving the logarithm of precipitation measured in acre-foot (about $1233\,\mathrm{m}^3$) for 26 unseeded clouds (`unseeded`) and 26 seeded clouds (`seeded`).

```r
clouds <- read.csv("clouds.csv")
X <- clouds$unseeded
Y <- clouds$seeded
head(clouds)
```

```
##   unseeded   seeded
## 1 7.092241 7.917755
## 2 6.721546 7.437089
## 3 5.919969 7.412160
```

```
## 4 5.844993 6.885510
## 5 5.772064 6.555926
## 6 5.498397 6.192567
```

1. Explain what each line of the above code does.

You want to examine whether cloud seeding has an impact on precipitation.

2. Propose a procedure to investigate this (state the null hypothesis, the alternative, and the assumptions required, etc.).

You decide to run the following tests

```r
#a. test 1
shapiro.test(X)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  X
## W = 0.97994, p-value = 0.8727
```

```r
#b. test 2
shapiro.test(Y)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Y
## W = 0.96591, p-value = 0.5208
```

```r
#c. test 3
var.test(X,Y)
```

```
##
##  F test to compare two variances
##
## data:  X and Y
## F = 1.0536, num df = 25, denom df = 25, p-value = 0.8971
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.4724176 2.3499219
## sample estimates:
## ratio of variances
##           1.053634
```

```r
#d. test 4
t.test(X,Y, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  X and Y
## t = -2.5444, df = 50, p-value = 0.01408
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.046697 -0.240865
## sample estimates:
## mean of x mean of y
```

```
## 3.990406 5.134187
```

3. Explain what each test is for and why we choose `var.equal = TRUE` in test 4.
4. What are your conclusions?

Exercise 4

1. Here are the observed values of two samples (considered independent):

$$A = (176, 391, 314, 248, 373, 393, 330), \qquad B = (307, 310, 374, 343, 360, 280).$$

   We perform a Wilcoxon rank sum test. What is the value of the test statistic? Show your calculations in detail.

2. If we add 200 to each element of A and B, does the value of the test statistic change?
3. We observe a sample $(X_1, \ldots, X_n)$ of independent random variables such that $X_i \sim \text{Binomial}(10, \theta)$, $\theta \in (0, 1)$. We want to estimate θ. The proposed estimator is

$$\hat{\theta} = \frac{n}{10(n-2)}\bar{X} - \frac{2}{n}, \qquad \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

   Compute the bias and variance of this estimator. Comment on your result.

4. Compare and briefly discuss the Bayesian and frequentist approaches in statistical inference.
5. An experimenter claims: "To reject a null hypothesis of equal means, it is enough to collect sufficiently large samples." Comment on this claim.
6. Briefly explain what Bayesian ANOVA is and what it does.
7. What is the Bayes factor for two models, M_1 and M_2? How do we interpret $BF_{12} = 4$?
8. If $BF_{01} = 3$ and $BF_{12} = 2$, which model is better: M_0 or M_2?