Exercises on Session 2: Probabilities and Random Variables

Vasilis Gkolemis

June 2025

Instructions These exercises are designed to help you practice probabilities and statistics.

1 Exercise: The expected value of the mean (independent case)

Let $X_1, X_2, \dots, X_n$ be independent random variables. Prove that $Y = \frac{1}{n}\sum_{i=1}^{n} X_i$ has expected value:

$$\mathbb{E}[Y] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i].$$

What is the expected value of Y if all X_i are coming from the same distribution with expected value μ?

Solution.

$$\begin{aligned}
\mathbb{E}[Y] &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \\
&= \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n} X_i\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i].
\end{aligned}$$

If all X_i are coming from the same distribution with expected value μ, then:

$$\mathbb{E}[Y] = \frac{1}{n}\sum_{i=1}^{n}\mu = \mu.$$

2 Exercise: The variance of the mean (independent case)

Let $X_1, X_2, \dots, X_n$ be independent random variables with variances $\sigma_i^2 = \mathrm{Var}(X_i)$. Prove that the variance of the mean $Y = \frac{1}{n}\sum_{i=1}^{n} X_i$ is given by:

$$\mathrm{Var}(Y) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2.$$

What is the variance of Y if all X_i are coming from the same distribution with variance σ^2?

Solution. By the properties of variance and independence, we have:

$$\begin{aligned}
\mathrm{Var}(Y) &= \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2.
\end{aligned}$$

If all X_i are coming from the same distribution with variance σ^2, then:

$$\mathrm{Var}(Y) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

3 Exercise: Practical implication of (1) and (2)

You are given a test and you are asked for the expected value of some probability distributions; the gamma distribution $\Gamma(\alpha, \beta)$ and the beta distribution $\mathrm{Beta}(\alpha, \beta)$. Unfortunately, you do not remember the formulas for the expected values of these distributions, but you have access to a computer with R installed. How could you estimate the expected values of these distributions using sampling?

1. Write an R script to estimate the expected value of a Gamma distribution with parameters α and β (shape and rate).
2. Write an R script to estimate the expected value of a Beta distribution with parameters α and β.

Hint: In R, you can use the ‘rgamma‘ and ‘rbeta‘ to sample from the Gamma and Beta distributions, respectively. For ‘rgamma‘, the function signature is ‘rgamma(n, shape, rate)‘ where ‘n‘ is the number of samples, ‘shape‘ is α, and ‘rate‘ is β. For ‘rbeta‘, the function signature is ‘rbeta(n, shape1, shape2)‘ where ‘n‘ is the number of samples, ‘shape1‘ is α, and ‘shape2‘ is β.

After you have run the R script . Your solution will not be exact but it will be close enough to the expected value. Can you answer why it is close enough?

```
# 1. Estimate expected value of a Gamma(alpha, beta) distribution
alpha <- 2
beta <- 3  # Rate parameter
set.seed(123)
samples_gamma <- rgamma(10000, shape = alpha, rate = beta)
expected_gamma <- mean(samples_gamma)
cat("Estimated expected value of Gamma(", alpha, ",", beta, ") = ", expected_gamma, "\n")

# 2. Estimate expected value of a Beta(alpha, beta) distribution
alpha_beta <- 2
beta_beta <- 5
```

```
set.seed(123)
samples_beta <- rbeta(10000, shape1 = alpha_beta, shape2 = beta_beta)
expected_beta <- mean(samples_beta)
cat("Estimated expected value of Beta(", alpha_beta, ",", beta_beta, ") = ", expected_beta,
```

It will be close enough because from (1):

$$\mathbb{E}[Y] = \mu$$

and from (2):

$$\text{Var}(Y) = \frac{\sigma^2}{n}$$

So, for large n, the sample mean Y will converge to the expected value μ due to the Law of Large Numbers. The variance decreases as n increases, making the estimate more precise.

4 Exercise: Sample from a more complicated distribution

You have to sample from $p(x, y) = p(y|x)q(x)$, where $p(y|x)$ is a Gaussian distribution with mean $\mu = x$ and variance σ^2, and $q(x)$ is a uniform distribution on the interval $[a, b]$. Can you do that in R? Is it straightforward (like two or three lines of code) or do you need to write a complicated script?

```
# Sample from p(x,y) = p(y|x)q(x)
set.seed(123)
a <- 0
b <- 10
n_samples <- 10000
# Sample x from uniform distribution
x_samples <- runif(n_samples, min = a, max = b)
# Sample y from Gaussian distribution with mean = x and variance = sigma^2
sigma <- 1
y_samples <- rnorm(n_samples, mean = x_samples, sd = sigma)
# Combine samples into a data frame
samples <- data.frame(x = x_samples, y = y_samples)
# Display first few samples
head(samples)
```

It is straightforward to sample from the joint distribution $p(x, y)$ in R. You first sample x from the uniform distribution and then sample y from the Gaussian distribution with mean equal to the sampled x and a fixed variance σ^2.