Exercises on Session 2: Probabilities and Random Variables

Vasilis Gkolemis

June 2025

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

Let $X_1, X_2, \ldots, X_n$ 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?

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

Let $X_1, X_2, \ldots, X_n$ 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?

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

In a test you are asked to find the *expected value* of two probability distributions;

- the gamma distribution $\Gamma(x; \alpha, \beta)$, where the shape $\alpha = 2$ and the rate $\beta = 3$
- the beta distribution $B(x; \alpha, \beta)$, where the shape parameters are $\alpha = 2$ and $\beta = 5$.

As you do not remember the closed-form formulas, you rely on a computer with R installed to estimate them.

1. Write an R script that estimates (a) $\mathbb{E}_X[\Gamma(X; \alpha = 2, \beta = 3)]$ and (b) $\mathbb{E}_X[B(X; \alpha = 2, \beta = 5)]$.
2. What is the standard deviation of the approximation? Write an R script that estimates the standard deviation of the approximation.
3. Does the standard deviation makes you confident about the approximation? Are you (almost) sure that the approximation is close enough to the real expected value, within a ± 0.1 interval?

Write an R script that estimates (a) $\mathbb{E}_X[\Gamma(X; \alpha = 2, \beta = 3)]$ and (b) $\mathbb{E}_X[B(X; \alpha = 2, \beta = 5)]$.

Hint: Use ‘rgamma(n, shape, rate)‘ where ‘n‘ is the number of samples, ‘shape‘ is α, and ‘rate‘ is β. Use ‘rbeta(n, shape1, shape2)‘ where ‘n‘ is the number of samples, ‘shape1‘ is α, and ‘shape2‘ is β.

4 Exercise: Sample from a more complicated distribution

Now you are asked to estimate the expected value of the product of a beta and a gamma distribution; $p(X) = \Gamma(X; \alpha = 2, \beta = 3) \cdot B(X; \alpha = 2, \beta = 5)$. Can you write an R script to estimate this expected value? Is it straightforward (like two or three lines of code) or you need to think of something more complicated?

5 Exercise: Illustration of the Central Limit Theorem in R

The Central Limit Theorem states that the distribution of the sum (or average) of a large number of independent random variables, regardless of their original distribution, will approximate a Gaussian distribution:

$$Y = \frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow{d} N(\mu, \sigma^2/n)$$

Write down an R script that illustrates the Central Limit Theorem by sampling from three different non-Gaussian distributions (a) Uniform, (b) Exponential, and (c) Bernoulli distributions. Then, generate n many samples of the average of n_1 independent draws from that distribution (e.g., $n_1 = 30$) and plot the histogram of these n averaged samples (average of n_1 independent draws). Plot the histogram of the averaged samples to illustrate the approximate normality predicted by the Central Limit Theorem.
Seems quinter intutive, right? In figure 1 you can see the histograms of the original distributions. Will their averaged samples look like a Gaussian distribution? To me at least, it is surpising. Let’s see!

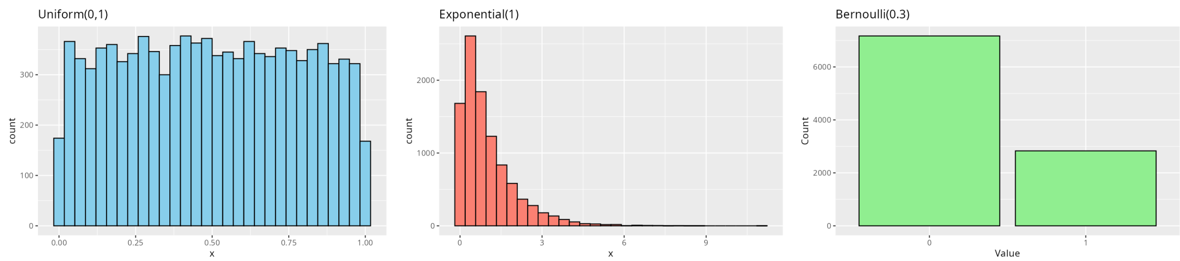

Figure 1: Bernoulli, Uniform, and Exponential distributions (left to right) using $n = 10,000$ samples.