# StatsI — Exercise 4

1. (a) Let $X$ be a random variable with mean $\mu$ and variance $\sigma^2 \in (0, \infty)$. Show that $P(|X - \mu| > 2\sigma) \leq 0.25$. What does this inequality tell us about the distribution of $X$?

   (b) Let $X_1, \cdots, X_n$ be an IID sample from a population with mean $\mu$ and variance $\sigma^2 \in (0, \infty)$. Show that for any $\epsilon > 0$, $P(|\bar{X}_n - \mu| > \epsilon\sigma) \leq \frac{1}{n\epsilon^2}$. Compare this bound with the approximation implied by the CLT when $n$ is large.

   **Note**. The conditions required for these inequalities are minimum.

2. Let $X$ and $Y$ be two r.v.s with positive and finite variances. The correlation coefficient of $X$ and $Y$ is defined as $\rho = \mathrm{Cov}(X,Y)/\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}$. (If $\rho = 0$, $X$ and $Y$ are called <u>uncorrelated</u> or <u>linearly independent</u>.)

   (a) Show that $|\rho| \leq 1$.

   (b) If $Y = aX + b$ for some constants $a(\neq 0)$ and $b$, show that $|\rho| = 1$.

   **Note**. In fact $|\rho| = 1$ if and only if $Y = aX + b$ for some constants $a(\neq 0)$ and $b$.

3. Let $X_1, \cdots, X_n$ be a sample from a distribution with mean $\mu$ and variance $\sigma^2 \in (0, 1)$. Let $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$, where $\bar{X}_n$ is the sample mean.

   (a) Show that $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n X_n^2 - \frac{n}{n-1}\bar{X}_n^2$.

   (b) Using Slutsky's theorem, show that $S_n^2 \xrightarrow{P} \sigma^2$.

4. Construct a Monte Carlo method to estimate the area of the ellipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$. Find the standard error of your estimate. Implement your method in $R$.

5. Compute the following integrals by Monte Carlo simulation. Find the standard error of your estimates.

   (a) $\int_{-2}^{3} \{x^3 + \log(1 + x^2)\}/\{2 + \cos(x)\}dx$.

   (b) $\int_0^\infty \sin\{x^3 + 1/\sqrt{x}\}e^{-x}dx$.

   (c) $\int_{-\infty}^\infty \exp\{\sin(x) - (x - 2)^2\}dx$.

6. Conditionally on a compound index $X$, the exam marks of the ST888 course follow a normal distribution with mean $20+60X$ and variance $10/(1+X^2)$. The index $X$ encapsulates the important factors related to teaching and learning and $X$ has the PDF

$$f_X(x) \propto \begin{cases} \exp\{x + 1/(1 + \sqrt{x})\} & 0 < x < 1 \\ 0 & \text{otherwsie.} \end{cases}$$

   Construct a Monte Carlo method to estimate the PDF of the exam marks, the mean and the standard error of the marks.

7. Here is a method called "Accept-Reject-Sampling" for drawing samples from distribution $f_Y(y) =$
$\int f_{Y|X}(y|x)f_X(x)dx$, when one cannot directly sample from $f_X(\cdot)$.

   Let $p(\cdot)$ be a PDF satisfying:

(i) $f_X(x) \le Mp(x)$ for all $x$, where $M > 0$ is a constant, and
(ii) it is easy to sample from $p(\cdot)$.

Then a sample $Y_1, \cdots, Y_n$ from $f_Y(\cdot)$ is obtained by repeating the following composition $n$ times:

Step 1. Generate $X_i$ as follows:

(i) Draw $X_0 \sim p(\cdot)$ and $U \sim U(0,1)$ independently,
(ii) If $U \le f_X(X_0)/\{Mp(X_0)\}$, take $X_i = X_0$; otherwise return to (i).

Step 2. Draw $Y_i \sim f_{Y|X}(\cdot|X_i)$.

Show that for any set $A$,

$$P\Big\{X_0 \in A \Big| U \le \frac{f_X(X_0)}{Mp(X_0)}\Big\} = \int_A f_X(x)dx.$$