This exercise sheet contains problems coming from Part 0 of the course.

Problem 1

Mathematical statistics basics I:

(a) For any constants $a, b \in \mathbb{R}$, prove each of the following properties:

(1) $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$

(2) $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$

(3) $Var(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$

(4) If X and Y are independent, then: $Var(X + Y) = Var(X) + Var(Y)$

(5) $Var(aX + b) = a^2 Var(X)$

(b) For a random variable Z, its mean and variance are defined as $\mathbb{E}[Z]$ and $\mathbb{E}[(Z - \mathbb{E}[Z])^2]$, respectively.

(1) A random variable $X \sim \mathcal{N}(\mu, \sigma^2)$ is Gaussian distributed with mean μ and variance σ^2. Given that for any $a, b \in \mathbb{R}$, we have that $Y = aX + b$ is also Gaussian, find a, b such that $Y \sim \mathcal{N}(0, 1)$.

(2) Let $X_1, \ldots, X_n$ be independent and identically distributed random variables, each with mean μ and variance σ^2. If we define $\widehat{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, what is the mean and variance of $\sqrt{n}(\widehat{X}_n - \mu)$?

(c) Suppose that $X_n \geq 0$ and $\mathbb{E}(X_n) = O(r_n)$. Prove that $X_n = O_{\mathrm{P}}(r_n)$.

(d) Suppose that $X_n \geq 0$ and $X_n = O_{\mathrm{P}}(r_n)$. Give an example to show that in general, this does not imply that $\mathbb{E}(X_n) = O(r_n)$.

(e) Suppose that $X_n \geq 0$ and $X_n = O_{\mathrm{P}}(r_n)$, the latter bound holding "exponentially" fast, meaning that there are constants $\gamma_0, n_0 > 0$ such that for all $\gamma \geq \gamma_0$ and $n \geq n_0$, we have $X_n \leq \gamma r_n$ with probability at least $1 - \exp(-\gamma)$. Prove that $\mathbb{E}(X_n) = O(r_n)$. (Hint: use the formulation for $\mathbb{E}(X_n)$ from Problem 2 (a) below.)

(f) Let $X_n \sim Exp(n)$, show that $X_n \xrightarrow{P} 0$.

(g) Let X be a random variable, and $X_n = X + Y_n$, where

$$\mathbb{E}(Y_n) = \frac{1}{n}, \qquad \mathrm{Var}(Y_n) = \frac{\sigma^2}{n},$$

where $\sigma > 0$ is a constant. Show that $X_n \xrightarrow{P} X$.

Solution.

(a) (1) **Linearity of Expectation with a Constant** $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$ **Proof:** By the definition of expectation: $\mathbb{E}(aX + b) = \int_{-\infty}^{\infty} (aX + b) f_X(x)\, dx$ Using the linearity of integration: $= a \int_{-\infty}^{\infty} X f_X(x)\, dx + b \int_{-\infty}^{\infty} f_X(x)\, dx$ The first term is $a\mathbb{E}(X)$ and the second term is b because $\int_{-\infty}^{\infty} f_X(x)\, dx = 1$ (total probability): $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$

(2) **Linearity of Expectation for Two Variables** $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$. **Proof:** By the definition of expectation: $\mathbb{E}(X + Y) = \int_{-\infty}^{\infty} (X + Y) f_{X,Y}(x, y)\, dx\, dy$. Using the linearity of integration: $= \int_{-\infty}^{\infty} X f_{X,Y}(x, y)\, dx\, dy + \int_{-\infty}^{\infty} Y f_{X,Y}(x, y)\, dx\, dy$. These are $\mathbb{E}(X)$ and $\mathbb{E}(Y)$ respectively, so: $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$

(3) **Variance Definition** $Var(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$ **Proof:** By definition, variance is: $Var(X) = \mathbb{E}[(X - \mathbb{E}(X))^2]$ Expanding the square: $Var(X) = \mathbb{E}[X^2 - 2X\mathbb{E}(X) + (\mathbb{E}(X))^2]$ Using the linearity of expectation: $Var(X) = \mathbb{E}[X^2] - 2\mathbb{E}(X)\mathbb{E}(X) + (\mathbb{E}(X))^2$ Simplifying: $Var(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$

(4) **Variance of the Sum of Independent Variables** $Var(X+Y) = Var(X) + Var(Y)$ **Proof:** By the definition of variance: $Var(X+Y) = \mathbb{E}[(X+Y)^2] - (\mathbb{E}[X+Y])^2$ Expanding $(X+Y)^2$: $Var(X+Y) = \mathbb{E}[X^2 + 2XY + Y^2] - (\mathbb{E}(X) + \mathbb{E}(Y))^2$ Using linearity of expectation: $= \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2] - (\mathbb{E}(X))^2 - 2\mathbb{E}(X)\mathbb{E}(Y) - (\mathbb{E}(Y))^2$ Since X and Y are independent, $\mathbb{E}[XY] = \mathbb{E}(X)\mathbb{E}(Y)$, so the middle terms cancel out: $Var(X+Y) = Var(X) + Var(Y)$

(5) **Variance of a Linear Transformation** $Var(aX+b) = a^2 Var(X)$ **Proof:** By the definition of variance: $Var(aX+b) = \mathbb{E}[(aX+b-\mathbb{E}(aX+b))^2]$ Since $\mathbb{E}(aX+b) = a\mathbb{E}(X)+b$, the expression simplifies: $= \mathbb{E}[(aX - a\mathbb{E}(X))^2]$ Factor out a^2: $= a^2\mathbb{E}[(X-\mathbb{E}(X))^2] = a^2 Var(X)$

(b) (1) Given that $X \sim \mathcal{N}(\mu, \sigma^2)$, and $Y = aX + b$, we need to find a and b such that $Y \sim \mathcal{N}(0,1)$. $\mathbb{E}[Y] = \mathbb{E}[aX+b] = a\mathbb{E}[X] + b = a\mu + b$. For $Y \sim \mathcal{N}(0,1)$, we require $\mathbb{E}[Y] = 0$, so: $a\mu + b = 0 \quad \Rightarrow$ $b = -a\mu$. $\text{Var}(Y) = \text{Var}(aX+b) = a^2\text{Var}(X) = a^2\sigma^2$. For $Y \sim \mathcal{N}(0,1)$, we require $\text{Var}(Y) = 1$, so: $a^2\sigma^2 = 1 \quad \Rightarrow \quad a = \frac{1}{\sigma}$.

Thus, the transformation is: $a = \frac{1}{\sigma}, \quad b = -\frac{\mu}{\sigma}$. So, $Y = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$.

(2) (i). Mean of $\sqrt{n}(\widehat{X}_n - \mu)$: Since the expected value of the sample mean $\widehat{X}_n$ is the population mean μ, we have: $\mathbb{E}[\widehat{X}_n] = \mu$. Thus, for the term $\sqrt{n}(\widehat{X}_n - \mu)$, we get: $\mathbb{E}[\sqrt{n}(\widehat{X}_n - \mu)] = \sqrt{n}(\mathbb{E}[\widehat{X}_n] - \mu) = \sqrt{n}(\mu - \mu) = 0$. So, the mean of $\sqrt{n}(\widehat{X}_n - \mu)$ is 0.

(ii). Variance of $\sqrt{n}(\widehat{X}_n - \mu)$: First, we recall that the variance of the sample mean $\widehat{X}_n$ is given by: $\text{Var}(\widehat{X}_n) = \frac{\sigma^2}{n}$. Now, the variance of $\sqrt{n}(\widehat{X}_n - \mu)$ is: $\text{Var}(\sqrt{n}(\widehat{X}_n - \mu)) = n \cdot \text{Var}(\widehat{X}_n) = n \cdot \frac{\sigma^2}{n} = \sigma^2$. So, the variance of $\sqrt{n}(\widehat{X}_n - \mu)$ is σ^2.

(c) $X_n \in O_p(r_n)$ is defined as follows:

$$X_n \in O_p(r_n) \Leftrightarrow \forall \varepsilon > 0 \exists M, N \in \mathbb{N}, \text{ such that } P(|\frac{X_n}{r_n}| > M) < \varepsilon \forall n > N.$$

$\mathbb{E}[X_n] \in O(r_n)$ is defined as follows:

$$\mathbb{E}[X_n] \in O(r_n) \Leftrightarrow \exists \tilde{M}, N \in \mathbb{N} \text{ such that } |\mathbb{E}[X_n]| \leq \tilde{M} r_n \forall n > N.$$

Hence, we can derive the following result:

$$P\Big[\frac{|X_n|}{r_n} > M\Big] \leq \frac{\mathbb{E}[|X_n|]}{r_n M} \leq \frac{\tilde{M} r_n}{r_n M} = \frac{\tilde{M}}{M} < \varepsilon,$$

in case that M is chosen to be $\frac{\tilde{M}}{\varepsilon}$ (we can do that!).

(d) We show that by constructing a counterexample. We consider the sequence of random variables, that is defined as follows:

$$X_n := \begin{cases} 1 \text{ with probability } \frac{1}{\sqrt{n}} \\ 0 \text{ with probability } 1 - \frac{1}{\sqrt{n}} \end{cases} \quad \text{for all } n \in \mathbb{N}.$$

We can show, that $X_n \in O_p(r_n)$ for $r_n = \frac{1}{n}$:

$$P\Big[\frac{|X_n|}{r_n} > M\Big] = P[|X_n| > \frac{M}{n}] \overset{\text{always}}{\leq} P[X_n = 1] = \frac{1}{\sqrt{n}} \overset{\text{choose } N > \frac{1}{\varepsilon^2}}{<} \varepsilon \text{ for all } n > N,$$

but we have that $\mathbb{E}[X_n] = \frac{1}{\sqrt{n}} \notin O(r_n)$, obviously.

(e) We use the identity of the exercise below to show that

$$\frac{\mathbb{E}[X_n]}{r_n} = \mathbb{E}[\frac{X_n}{r_n}] = \int_0^\infty P\Big[\frac{X_n}{r_n} > t\Big] dt \leq \int_0^\infty e^{-t} dt = 1.$$

(f) We can calculate that

$$\lim_{n\to\infty} P[|X_n| - 0 > \varepsilon] = \lim_{n\to\infty} P[X_n > \varepsilon] = \lim_{n\to\infty} e^{-n\varepsilon} = 0.$$

(g) We can calculate that

$$|Y_n| = |Y_n - \mathbb{E}[Y_n] + \mathbb{E}[Y_n]| \leq |Y_n - \mathbb{E}[Y_n]| + |\mathbb{E}[Y_n]| = |Y_n - \mathbb{E}[Y_n]| + \frac{1}{n},$$

and we can derive that

$$P[|X_n - X| > \varepsilon] = P[|Y_n| > \varepsilon] \leq P[|Y_n - \mathbb{E}[Y_n]| + \frac{1}{n} > \varepsilon] \overset{\text{Tscheb.}}{\leq} \frac{Var[Y_n]}{(\varepsilon - \frac{1}{n})^2} = \frac{\sigma^2}{n}\frac{1}{(\varepsilon - \frac{1}{n})^2} \to 0.$$

Problem 2

Mathematical statistics basics II:

(a) Prove that for $X \geq 0$, it holds that $\mathbb{E}(X) = \int_0^\infty \mathrm{P}(X > t)dt$. You may assume that X is continuously distributed and hence has a probability density function.

(b) (p-moments via tails) Prove that for $X \geq 0$ and $p \in (0, \infty)$, it holds that

$$\mathbb{E}(X^p) = \int_0^\infty pt^{p-1}\,\mathrm{P}(X > t)dt$$

whenever the right hand side is finite. You may assume that X is continuously distributed and hence has a probability density function.

Solution.

(a) Required to prove: $\mathbb{E}(X) = \int_0^\infty \mathbb{P}(X > t)\,dt$.
Since X has a probability density function $f(x)$, the expected value is: $\mathbb{E}(X) = \int_0^\infty x f(x)\,dx$. The survival function is given by: $\mathbb{P}(X > t) = \int_t^\infty f(x)\,dx$.

Consider the integral of the survival function over t: $\int_0^\infty \mathbb{P}(X > t)\,dt = \int_0^\infty \left(\int_t^\infty f(x)\,dx\right) dt$. By Fubini's Theorem (since the integrand is non-negative), we can interchange the order of integration: $\int_0^\infty \left(\int_t^\infty f(x)\,dx\right) dt = \int_0^\infty f(x) \left(\int_0^x dt\right) dx$. We have $\int_0^x dt = x$. Substitute back into the expression: $\int_0^\infty f(x) \left(\int_0^x dt\right) dx = \int_0^\infty f(x) x\,dx = \mathbb{E}(X)$. Hence, $\mathbb{E}(X) = \int_0^\infty \mathbb{P}(X > t)\,dt$.

(b) For $X \geq 0$, we have $\mathbb{E}(X^p) = \int_0^\infty \mathbb{P}(X^p > t)\,dt$.

Since X^p is non-negative and increasing in X, we have: $\mathbb{P}(X^p > t) = \mathbb{P}(X > t^{1/p})$.

Therefore, $\mathbb{E}(X^p) = \int_0^\infty \mathbb{P}(X > t^{1/p})\,dt$.

Make the substitution $s = t^{1/p}$, so $t = s^p$, and $dt = ps^{p-1}\,ds$.

Thus, $\mathbb{E}(X^p) = \int_0^\infty \mathbb{P}(X > s)\,ps^{p-1}\,ds = \int_0^\infty ps^{p-1}\mathbb{P}(X > s)\,ds$.

This gives us the desired result: $\mathbb{E}(X^p) = \int_0^\infty pt^{p-1}\mathbb{P}(X > t)\,dt$.

Problem 3

Vectors and Matrices:

(a) Consider the matrix X and the vectors y and z below:

$$X = \begin{pmatrix} 2 & 4 \\ 1 & 3 \end{pmatrix} \qquad y = \begin{pmatrix} 1 \\ 3 \end{pmatrix} \qquad z = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

   (a) What is the inner product of the vectors y and z?
   (b) What is the product Xy?
   (c) Calculate the determinant, the trace and the Frobenius and operator norms of the matrix X.
   (d) Is X invertible? If so, give the inverse, and if not, explain why not.
   (e) What is the rank of X? Explain your answer.

(b) Let $A = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 0 & 3 \\ 1 & 1 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 1 & 1 & 2 \end{bmatrix}$. For each matrix A and B,

   (a) What is its rank?
   (b) What is a (minimal size) basis for its column span?

(c) Let $A = \begin{bmatrix} 0 & 2 & 4 \\ 2 & 4 & 2 \\ 3 & 3 & 1 \end{bmatrix}$, $b = \begin{bmatrix} -2 & -2 & -4 \end{bmatrix}^T$, and $c = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^T$.

   (a) What is Ac?
   (b) What is the solution to the linear system $Ax = b$?

Solution.

(a) (a) We calculate that $y^T z = 1 \cdot 2 + 3 \cdot 3 = 11$.

   (b) We calculate that $Xy = \begin{pmatrix} 2 \cdot 1 + 4 \cdot 3 \\ 1 \cdot 1 + 3 \cdot 3 \end{pmatrix} = \begin{pmatrix} 14 \\ 10 \end{pmatrix}$.

   (c) The determinant of X is

$$\det(X) = 2 \cdot 3 - 1 \cdot 4 = 2.$$

   The trace (sum of diagonal elements) of X is $2 + 3 = 5$. The Frobenius norm of X is $||X||_F = \sqrt{2^2 + 4^2 + 1^2 + 3^2} = \sqrt{30}$. Root of the sum of the squares of entries of X. The operator norm of X equals its largest eigenvalue, which is around 4.5 (the sum of the two eigenvalues is the trace, i.e, 5, and their product is the determinant, 2, which is approximately solved by $\lambda_1 = 4.5$ (a little bit larger) and $\lambda_2 = 0.5$ (a little bit smaller)).

   (d) Yes (determinant is nonzero). The inverse is

$$X^{-1} = \begin{pmatrix} 1.5 & -2 \\ -0.5 & 1 \end{pmatrix}.$$

   (e) The rank of an invertible matrix is full.

(b) (a) We conduct Gaussian elimination to get the matrices in upper triangular form; this delivers no zero row for matrix A (full rank), and one zero row for matrix B (rank 2).

   (b) All columns of A are needed to span the column space, while B only needs the first two columns.

(c) Let $A = \begin{bmatrix} 0 & 2 & 4 \\ 2 & 4 & 2 \\ 3 & 3 & 1 \end{bmatrix}$, $b = \begin{bmatrix} -2 & -2 & -4 \end{bmatrix}^T$, and $c = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^T$.

(a) $Ac = \begin{pmatrix} 6 \\ 8 \\ 7 \end{pmatrix}$.

(b) $A^{-1}b = \begin{pmatrix} -2 \\ 1 \\ -1 \end{pmatrix}$.

Problem 4

Coding problem I: Use a language of your choice (The course material is in R).

Sampling from a distribution.

(a) Draw 100 samples $x = (x_1\ x_2)^T$ from a 2-dimensional Gaussian distribution with mean $(0, 0)^T$ and identity covariance matrix.

(b) Plot them on a scatter plot.

(c) How does the scatter plot change if the mean is $(1, -1)^T$?

(d) (Change the mean back to $(0, 0)^T$.) Change the covariance matrix as follows

$$\Sigma_1 = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} \qquad \Sigma_2 = \begin{pmatrix} 1 & -0.5 \\ -0.5 & 1 \end{pmatrix}$$

and plot the corresponding scatter plots.

Solution.

See python script.

Problem 5

Coding problem II:

(a) Write down the properties a general covariance matrix has to satisfy.

(b) Come up with a 2×2 matrix that satisfies those properties.

(c) Draw 100 samples $x = (x_1\ x_2)^T$ from a 2-dimensional Gaussian distribution with mean $(0, 0)'$ and the covariance matrix you chose in (b). Generate a plot of the support region for the Gaussian random variables. Vary the covariance matrix to demonstrate how the shape of the support region changes depending on the nature of the covariance matrix.

(d) Find a way to generate a covariance matrix of arbitrary dimension.

(e) Recover the histogram plots in the lecture notes. Play around with different dimensions and sample sizes.

Solution.

(a) A covariance matrix is a quadratic, positive semi-definite (non-negative eigenvalues), symmetric matrix.

(b) The unit matrix, e.g.

(c) See code for Problem 4 - this answers it.

(d) Sample n non-negative numbers $\lambda_1, ..., \lambda_n$ (these are the eigenvalues). Initialize a set V as the empty set, and iterate n times: Sample a vector with Euclidean norm 1 randomly from the orthogonal complement of V, and add it to V, calling it v_i in iteration i. Calculate the covariance matrix by $(v_1, ..., v_n)^T \operatorname{diag}(\lambda)(v_1, ..., v_n)$.

(e) Smartly play around with different dimensions and sample sizes.