This first homework is meant to refresh the mathematical background needed for this class. In particular some probability and linear algebra.

Solution Sheet.

Problem 1

Mathematical statistics warm-up I:

(a) For a random variable Z, its mean and variance are defined as $\mathrm{E}[Z]$ and $\mathrm{E}[(Z-\mathrm{E}[Z])^2]$, 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 $X_1, \dots, X_n$ 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 $\sqrt{n}(\widehat{X}_n - \mu)$?

(b) Suppose $X \sim Pois(\lambda)$. Show that $\mathrm{Var}(X) = \lambda$.

(c) Suppose $X \sim Exponential(\lambda)$. Show that $\mathrm{Var}(X) = \frac{1}{\lambda^2}$.

(d) Suppose X is a random variable distributed according to F_X with $\mathrm{E}(X) = 0$ and density f_x. Show that $\mathrm{Var}(aX + b) = a^2\,\mathrm{Var}(X)$ for constants $a, b \in \mathbb{R}$.

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

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

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

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

Solution.

(a) (1)

$$\begin{aligned}
&\mathrm{E}[Y] = \mathrm{E}[aX+b] = a\,\mathrm{E}[x] + b \quad \implies \mathrm{E}[Y] = a\mu + b\\
&\mathrm{Var}[Y] = \mathrm{Var}[aX+b] = a^2\,\mathrm{Var}[x] \implies \mathrm{Var}[Y] = a^2\sigma^2\\
&\text{Now, as } Y \sim \mathcal{N}(0,1) \text{ we have:}\\
&0 = a\mu + b \ \wedge\ \sigma^2 = a^2\\
&\implies a = \pm\frac{1}{\sigma} \ \wedge\ b = \mp\frac{\mu}{\sigma}.
\end{aligned}$$

(2)
$$\mathrm{E}\left[\sqrt{n}(\hat{X}-\mu)\right] = \sqrt{n}\left(\mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}X_i\right]-\mu\right) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left[X_i\right]-\mu\right) = 0,$$
$$\mathrm{Var}\left[\sqrt{n}(\hat{X}_n-\mu)\right] = n\,\mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)\right] = \frac{n}{n^2}\sum_{i=1}^{n}\mathrm{Var}\left[(X_i-\mu)\right] = \sigma^2.$$

(b) As $X \sim Pois(\lambda)$, we know for the lecture that $\mathrm{E}[X] = \lambda$. Now, to calculate the variance we use the formula:
$$\mathrm{Var}[X] = \mathrm{E}\left[X^2\right] - \mathrm{E}[X]^2.$$

$$\mathrm{E}\left[X^2\right] = \mathrm{E}\left[X^2 - X + X\right] = \mathrm{E}[X(X-1)+X] = \mathrm{E}[X(X-1)] + \mathrm{E}[X] = \mathrm{E}[X(X-1)] + \lambda.$$
$$\mathrm{E}[X(X-1)] = \sum_{k\geq 0} k(k-1)\frac{1}{k!}\lambda^k e^{-\lambda} = \sum_{k\geq 2} k(k-1)\frac{1}{k!}\lambda^k e^{-\lambda} = \sum_{k\geq 2}(k-1)\frac{1}{(k-1)!}\lambda^k e^{-\lambda} =$$
$$= \lambda^2\sum_{k\geq 2}(k-2)\frac{1}{(k-2)!}\lambda^{k-2}e^{-\lambda} \overset{j=k-2}{\equiv} \lambda^2\sum_{j\geq 0} j\frac{1}{j!}\lambda^j e^{-\lambda} = \lambda^2.$$
$$\implies \mathrm{E}\left[X^2\right] = \lambda^2 + \lambda.$$
$$\implies \mathrm{Var}[X] = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

(c) As $X \sim Exponential(\lambda)$, we know for the lecture that $\mathrm{E}[X] = \frac{1}{\lambda}$. Now, to calculate the variance we use the formula:
$$\mathrm{Var}[X] = \mathrm{E}\left[X^2\right] - \mathrm{E}[X]^2.$$

$$\mathrm{E}\left[X^2\right] = \int_0^\infty x^2\lambda e^{-\lambda x}dx \overset{\text{Integration by Parts}}{=} \left[-x^2\lambda e^{-\lambda x}\right]_0^\infty + \int_0^\infty 2xe^{-\lambda x}dx$$
$$\overset{\text{Taking the respective limits}}{\implies} 0 + \frac{2}{\lambda}\int_0^\infty x\lambda e^{-\lambda x}dx = \frac{2}{\lambda}\mathrm{E}[X] = \frac{2}{\lambda^2}.$$
$$\implies \mathrm{E}\left[X^2\right] = \frac{2}{\lambda^2}.$$
$$\implies \mathrm{Var}[X] = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

(d) Consider $\mathrm{Var}[Z] = \mathrm{E}\left[(Z-\mathrm{E}[Z])^2\right]$, and $Z = aX + b$. Now,

$$(Z-\mathrm{E}[Z])^2 \overset{Z=aX+b}{=} (aX+b-\mathrm{E}[aX+b])^2 = (aX+b-(a\,\mathrm{E}[X]+b))^2$$
$$= (aX - a\,\mathrm{E}[X])^2 = a^2(X-\mathrm{E}[X])^2$$
$$\implies \mathrm{E}\left[(Z-\mathrm{E}[Z])^2\right] = a^2\,\mathrm{E}\left[(X-\mathrm{E}[X])^2\right] = a^2\,\mathrm{Var}[X]$$

(e) We can calculate that

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

(f) We can calculate that

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

and we can derive that

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

Problem 2

Mathematical statistics warm-up II:

(a) Prove that for $X \geq 0$, it holds that $\mathrm{E}(X) = \int_0^\infty \mathrm{P}(X > t)dt$. 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

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

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

Solution.

(a) We can calculate that

$$\begin{aligned}
\mathrm{E}[X] &= \int_0^\infty x f(x)dx = \int_0^\infty \int_0^x dt f(x)dx \\
&= \int_0^\infty \int_0^\infty \mathbb{1}_{x\geq t} dt f(x)dx = \int_0^\infty \int_0^\infty \mathbb{1}_{x\geq t} f(x)dxdt = \\
&\int_0^\infty \int_t^\infty f(x)dxdt = \int_0^\infty (1 - F(t))dt = \int_0^\infty P(x > t)dt,
\end{aligned}$$

where $\mathbb{1}$ represents an indicator function, which is 1 if the argument is in the indexed set or satisfies the indexed condition and 0 otherwise. X represents a real-valued continuously distributed random variable. f and F denote its density and distribution functions, respectively. (Wikipedia helps!).

(b) Analogously, we can show that for an arbitrary number $p \in (0, \infty)$,

$$\mathrm{E}[X^p] = \int_0^\infty x^p f(x)dx = \int_0^\infty \int_0^x pt^{p-1} dt f(x) dx$$
$$= \int_0^\infty \int_0^\infty \mathbb{1}_{x \geq t} pt^{p-1} dt f(x) dx = \int_0^\infty \int_0^\infty \mathbb{1}_{x \geq t} pt^{p-1} f(x) dx dt =$$
$$\int_0^\infty \int_t^\infty f(x) dx pt^{p-1} dt = \int_0^\infty (1 - F(t)) pt^{p-1} dt = \int_0^\infty P(x > t) pt^{p-1} dt.$$

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 no, explain why not.
   (e) What is the rank of X? Explain your answer.

(b) Consider $A = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 0 & 3 \\ 1 & 1 & 2 \end{bmatrix}$ 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) Assume $A = \begin{bmatrix} 0 & 2 & 4 \\ 2 & 4 & 2 \\ 3 & 3 & 1 \end{bmatrix}$, $b = \begin{bmatrix} -2 & -2 & -4 \end{bmatrix}^T$, 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 $y^T z = 1 \cdot 2 + 3 \cdot 3 = 10$.

(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 $||X||_F = \sqrt{2^2 + 2^4 + 1^2 + 3^2} = \sqrt{1 + 4 + 9 + 16} = \sqrt{30}$. Square root of the sum of the squares of the 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
$$X^{-1} = \begin{pmatrix} 1.5 & -2 \\ -0.5 & 1 \end{pmatrix}$$
(we get the columns γ_i by solving the equation systems $X\gamma_i = e_i$, with e_i being unit vectors). We can easily check that $X \cdot X^{-1}$ is the unit matrix.

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

(b) (a) We conduct a couple of Gaussian elimination steps 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 two (Any two columns would work, e.g., either the first and the third or the second and the third.)

(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) $Ax = b \overset{\text{if } A^{-1} \text{ exists}}{\Longrightarrow} x = A^{-1}b = \begin{pmatrix} -2 \\ 1 \\ -1 \end{pmatrix}$.

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) Write 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.

1. A covariance matrix is a quadratic, positive semi-definite (non-negative eigenvalues), symmetric matrix.
2. The unit matrix, e.g..
3. See code for Problem 4 - this answers it.

4. 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 $(v_1, ..., v_n)^T \operatorname{diag}(\lambda)(v_1, ..., v_n)$.

5. Be creative.