Probability and Statistics

Exercise sheet 12

Exercise 12.1 The goal of this exercise is to show that if $X = (X_1, ..., X_n)^T \sim \mathcal{N}(\mu, \Sigma)$ with Σ invertible, then X admits a density with respect to the Lebesgue measure on $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$, given by

$$f(x) = f_X(x) = \frac{1}{(\sqrt{2\pi})^n} \frac{1}{\sqrt{\det(\Sigma)}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)} \tag{1}$$

for any $x = (x_1, ..., x_n)^T \in \mathbb{R}^n$.

Before showing this, we first settle some questions around the covariance matrix Σ (this is done in the first two parts). In (a) and (b), the random vector X can have any distribution (not necessarily normal).

(a) Recall that the covariance matrix of X, Σ, has entries $\Sigma_{ij} = \mathrm{cov}(X_i, X_j)$ for $1 \leq i, j \leq n$. Show that

$$\Sigma = E[(X - \mu)(X - \mu)^T].$$

Remark: Expectations are evaluated componentwise, i.e. if M is a random matrix,

$$E\left[\begin{pmatrix} M_{11} & \dots & M_{1n} \\ \vdots & \ddots & \vdots \\ M_{n1} & \dots & M_{nn} \end{pmatrix}\right] = \begin{pmatrix} E(M_{11}) & \dots & E(M_{1n}) \\ \vdots & \ddots & \vdots \\ E(M_{n1}) & \dots & E(M_{nn}) \end{pmatrix}.$$

(b) Let $A \in \mathbb{R}^{p \times n}$ be a fixed (deterministic) matrix. Show that the covariance matrix of AX is $A\Sigma A^T$.

If $A = a^T \in \mathbb{R}^{1\times n}$, what is the covariance of $a^T X$? Conclude that Σ is semi-positive definite.

(c) Now take $X \sim \mathcal{N}(\mu, \Sigma)$. By definition, $X \overset{\mathrm{d}}{=} \mu + AZ$ with $AA^T = \Sigma$ (i.e., A is a square root of Σ), and Z is standard normal, i.e. $Z = (Z_1, ..., Z_n)^T$ for $Z_1, ..., Z_n \overset{\mathrm{iid}}{\sim} \mathcal{N}(0, 1)$.

- Check that Σ is indeed the covariance matrix of X.
- Assuming that Σ is invertible, show that A is also invertible. Using the Jacobian formula, show that X has density given by (1) almost everywhere.

(d) Suppose you are given a density in the form (1). Can you find the marginal density of X_i ($i \in \{1, ..., n\}$) without additional calculations?

(e) (optional).

For $d = 2$, if $\sigma_1^2 = \mathrm{var}(X_1) > 0$, $\sigma_2^2 = \mathrm{var}(X_2) > 0$ and $\mathrm{cov}(X_1, X_2) = \sigma_1\sigma_2\rho$ with ρ the correlation between X_1 and X_2.What is the condition on ρ for Σ to be invertible? What is the expression of the density in this case?

Exercise 12.2 (some training) Let $X_1, ..., X_n$ be i.i.d with density $f(\cdot \mid \theta_0)$, where the true value of θ_0 is unknown.

(a) For the following models, find the moment estimator and MLE for $\theta_0 \in \Theta$ as well as the Fisher information $I(\theta_0)$ (you may assume that all regularity conditions are fulfilled).

1. (Geometric)

$$f(x \mid \theta) = (1-\theta)^{x-1}\theta$$

for $x \in \mathbb{N}_{\geq 1}$,where $\theta \in \Theta = (0, 1)$.

2. (Bernoulli)

$$f(x \mid \theta) = \theta^x(1-\theta)^{1-x}$$

for $x \in \{0, 1\}$, where $\theta \in \Theta = (0, 1)$.

3. (Beta$(1, \theta)$)

$$f(x \mid \theta) = \theta(1-x)^{\theta-1}\mathbb{1}_{x\in(0,1)},$$

where $\theta \in \Theta = (0, +\infty)$.

4. (Laplace)

$$f(x \mid \theta) = \frac{\theta}{2}e^{-\theta|x|}$$

for $x \in \mathbb{R}$, where $\theta \in \Theta = (0, +\infty)$.

Hint: Note that for $X \sim \text{Laplace}(\theta)$, $E(X) = 0$ and therefore one needs to use the next order moment.

(b) For the first model Geo(θ), construct an asymptotic confidence interval of level $1 - \alpha$ for θ_0, based on the asymptotic normality of the MLE $\hat{\theta}$, and approximating $I(\theta_0)$ by $I(\hat{\theta})$.

(c) In a study of feeding behaviors of birds, the number of hops between flights was counted for $n = 130$ birds. The data are given in the following table.

# Hops	1	2	3	4	5	6	7	8	9	10	11	12
Frequency	48	31	20	9	6	5	4	2	1	1	2	1

For example: in 48 occasions, a bird had just 1 hop before flying again, in 20 occasions they had 3 hops, etc. Assume that the number of hops can be modelled as a geometric random variable with unknown success probability $\theta_0 \in (0, 1)$. Compute the MLE based on the data in the table, and find an asymptotic confidence interval of level 95%.

Exercise 12.3

(a) Find a sufficient statistic for the parameters generating the following models:

1. $$X_1, ..., X_n \overset{\text{iid}}{\sim} \text{U}([0, \theta]), \quad \theta \in (0, +\infty).$$

2. $$X_1, ..., X_n \overset{\text{iid}}{\sim} \text{Exp}(\lambda), \quad \lambda \in (0, +\infty).$$

3. $$X_1, ..., X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2), \quad \theta = (\mu, \sigma)^T \in \mathbb{R} \times (0, +\infty).$$

4. $$X_1, ..., X_n \overset{\text{iid}}{\sim} \text{U}([\theta, \theta + 1]), \quad \theta \in \mathbb{R}.$$

(b) Show that in general, if $T(X_1, ..., X_n)$ is a sufficient statistic for $\theta \in \Theta$ (where $X_1, ..., X_n \overset{\text{iid}}{\sim} f(\cdot \mid \theta)$), then for any $c \in \mathbb{R} \setminus \{0\}$, $cT(X_1, ..., X_n)$ is also sufficient for θ.

Hint: Use the factorisation theorem.

Exercise 12.4 Let $(X, Y)^T$ be a random vector. We want to show that $\mathrm{var}(X \mid Y) = 0$ with probability 1, if and only if there is a measurable function h such that $P(X = h(Y)) = 1$.

We consider only the case where the vector is discrete (takes either finitely many or countably many different values).

(a) State the definition of $\mathrm{var}(X \mid Y = y)$.

(b) Show that $\mathrm{var}(X \mid Y) = 0$ with probability 1 if and only if $P(X = E(X \mid Y)) = 1$.

(c) Conclude.

Exercise 12.5 (optional).

The goal here is to justify why the idea of maximising the likelihood is a good one.

(a) For $X \sim f(\cdot \mid \theta_0)$ and $\theta \in \Theta$, assume that $E[\log f(X \mid \theta)]$ exists.
Show that $E[\log f(X \mid \theta)] \leq E[\log f(X \mid \theta_0)]$.
Hint: Show that $E\left[\log\left(\frac{f(X|\theta_0)}{f(X|\theta)}\right)\right] \geq 0$ by using Jensen's inequality for the convex function $t \mapsto -\log t, t \in (0, +\infty)$.

(b) Recall the weak law of large numbers: if $Y_1, ..., Y_n$ are i.i.d. such that $E(|Y_1|) < \infty$, then

$$\overline{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow{\mathbb{P}} E(Y_1) \quad (n \to \infty).$$

Using the WLLN, explain why the MLE would be a reasonable estimator.