Point estimation

Exercise 1 (Statistical model)**:**

1. $\mathcal{P} = \{X \sim P \text{ with } \mathrm{E}(X) = \theta + \mu, \mathrm{var}(X) = \sigma^2\}$.
2. Not parametric because the distribution is unknown. Parametric for gaussian distribution, $\mathcal{P} = \{P_\theta \sim \mathcal{N}(\theta + \mu, \sigma^2)\}$
3. The parameter θ is identifiable since $\theta = \mathrm{E}(X) - \mu$ (so a function of the distribution P). Note that if the parameter μ is unknown, θ is no longer identifiable. For instance, (θ, μ) and $(\theta + 1, \mu - 1)$ are valid, with $\theta + \mu = \mathrm{E}(X)$.

Exercise 2 (Identifiability)**:**
The model is $\mathcal{P} = \{X \sim P = \mathcal{B}(pq)\}$. Not identifiable (all parameters $\theta = (p, q)$ with the same value pq give the same distribution).

Exercise 3 (Bernoulli model)**:**
$\mathcal{P} = \{P_\theta \sim \mathcal{B}(\theta)\}$. Density with respect to the counting measure μ^n with $\mu = \delta_0 + \delta_1$:

$$p_\theta(x) = \prod_{i=1}^n \theta^{x_i}(1-\theta)^{1-x_i}$$

Log-likelihood:

$$L(\theta) = \log p_\theta(x) = \sum_{i=1}^n (x_i \log\theta + (1 - x_i)\log(1-\theta)) = S_n \log\theta + (n - S_n)\log(1-\theta),$$

with

$$S_n = \sum_{i=1}^n x_i.$$

Derivative:

$$L'(\theta) = \frac{S_n}{\theta} - \frac{n - S_n}{1-\theta}.$$

We have $L'(\theta) = 0$ for:

$$\widehat{\theta} = \frac{S_n}{n}.$$

This is a maximum because the function L is concave:

$$L''(\theta) = -\frac{S_n}{\theta^2} - \frac{n - S_n}{(1-\theta)^2} \leq 0$$

Exercise 4 (Geometric model)**:**
$\mathcal{P} = \{P_\theta \sim \mathcal{G}(\theta)\}$. Density with respect to the counting measure μ^n with $\mu = \sum_{k\in\mathbb{N}} \delta_k$:

$$p_\theta(x) = \prod_{i=1}^n \theta(1-\theta)^{x_i}.$$

Log-likelihood:

$$L(\theta) = \log p_\theta(x) = \sum_{i=1}^n (\log\theta + x_i \log(1-\theta)) = n\log\theta + S_n \log(1-\theta),$$

with

$$S_n = \sum_{i=1}^n x_i.$$

Derivative:

$$L'(\theta) = \frac{n}{\theta} - \frac{S_n}{1-\theta}.$$

We have $L'(\theta) = 0$ for:

$$\widehat{\theta} = \frac{n}{S_n + n}.$$

This is a maximum because the function L is concave:

$$L''(\theta) = -\frac{n}{\theta^2} - \frac{S_n}{(1-\theta)^2} \leq 0$$

Exercise 5 (Gaussian model – mean)**:**
Density with respect to the Lebesgue measure:

$$p_\theta(x) \propto \prod_{i=1}^n e^{-\frac{1}{2\sigma^2}(x_i-\theta)^2}.$$

Log-likelihood:

$$L(\theta) = \log p_\theta(x) = c - \frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \theta)^2.$$

Derivative:

$$L'(\theta) = \frac{1}{\sigma^2}\sum_{i=1}^n (x_i - \theta).$$

We have $L'(\theta) = 0$ for:

$$\sum_{i=1}^n (x_i - \theta) = 0,$$

that is

$$\widehat{\theta} = \frac{S_n}{n}, \quad S_n = \sum_{i=1}^n x_i.$$

This is a maximum because the function L is concave:

$$L''(\theta) = -\frac{n}{\sigma^2} \leq 0.$$

Exercise 6 (Gaussian model – variance)**:**
Density with respect to the Lebesgue measure:

$$p_\theta(x) \propto \prod_{i=1}^n \frac{1}{\sqrt{\theta}} e^{-\frac{1}{2\theta}(x_i-\mu)^2}.$$

Log-likelihood:

$$L(\theta) = \log p_\theta(x) = c - \frac{n}{2}\log\theta - \frac{1}{2\theta}\sum_{i=1}^n (x_i-\mu)^2.$$

Derivative:

$$L'(\theta) = -\frac{n}{2}\frac{1}{\theta} + \frac{1}{2\theta^2}\sum_{i=1}^n (x_i-\mu)^2.$$

We have $L'(\theta)=0$ for:

$$\frac{n}{\theta} = \frac{1}{\theta^2}\sum_{i=1}^n (x_i-\mu)^2$$

that is

$$\widehat{\theta} = \frac{1}{n}\sum_{i=1}^n (x_i-\mu)^2.$$

This is a maximum because L is C^1 over $]0,+\infty[$, with

$$\lim_{\theta\to 0} L(\theta) = \lim_{\theta\to+\infty} L(\theta) = -\infty.$$

Note that L is not concave.

Exercise 7 (Gaussian model – mean and variance)**:**
Density with respect to the Lebesgue measure:

$$p_\theta(x) \propto \prod_{i=1}^n \frac{1}{\sqrt{\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2}.$$

Log-likelihood:

$$L(\theta) = \log p_\theta(x) = c - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2.$$

Derivatives (with $v=\sigma^2$ for convenience):

$$\frac{\partial L}{\partial \mu} = -\frac{1}{v}\sum_{i=1}^n (x_i-\mu), \quad \frac{\partial L}{\partial v} = -\frac{n}{2v} + \frac{1}{2v^2}\sum_{i=1}^n (x_i-\mu)^2.$$

We have $\Delta L(\theta)=0$ for:

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^n x_i, \quad \widehat{v} = \frac{1}{n}\sum_{i=1}^n (x_i-\widehat{\mu})^2.$$

This is a maximum because L is C^1 over $\Theta = \mathbb{R}\times]0,+\infty[$, with

$$\lim_{\mu\to-\infty} L(\theta) = \lim_{\mu\to+\infty} L(\theta) = -\infty,$$

$$\lim_{v\to 0} L(\theta) = \lim_{v\to+\infty} L(\theta) = -\infty.$$

Exercise 8 (Method of moments)**:**
The estimator obtained by the method of moments is the same as the MLE for the Bernoulli model and the gaussian model with unknown mean.

For the geometric model, the expectation of a single observation is:

$$\mathrm{E}(X) = \frac{1-\theta}{\theta}.$$

We deduce the estimator:

$$\frac{S_n}{n} = \frac{1-\widehat{\theta}}{\widehat{\theta}}$$

that is

$$\widehat{\theta} = \frac{n}{S_n + n}.$$

This is the MLE.

For the gaussian model with unknown variance θ, we get:

$$\mathrm{E}(X^2) = \theta + \mu^2,$$

so that:

$$\widehat{\theta} = \frac{1}{n}\sum_{i=1}^n x_i^2 - \mu^2.$$

This differs from the MLE. Another estimator can be obtained with:

$$\mathrm{E}((X-\mu)^2) = \theta.$$

We get:

$$\widehat{\theta} = \frac{1}{n}\sum_{i=1}^n (x_i - \mu)^2.$$

This is the MLE.

Finally, for the gaussian model with unknown mean and variance,

$$\mathrm{E}(X) = \mu, \quad \mathrm{E}(X^2) = \sigma^2 + \mu^2.$$

We get:

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^n x_i, \quad \widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^n x_i^2 - \widehat{\mu}^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \widehat{\mu})^2.$$

This is the MLE.

Exercise 9 (Uniform model)**:**
Density with respect to the Lebesgue measure:

$$p_\theta(x) = \frac{1}{\theta^n}\prod_{i=1}^n 1_{[0,\theta]}(x_i), \quad x_1, \dots, x_n \in [0, \theta].$$

Log-likelihood:

$$\log p_\theta(x) = -n \log \theta, \quad x_1, \dots, x_n \in [0, \theta].$$

Maximum for $\widehat{\theta}_{\text{MLE}} = \max(x_1, \ldots, x_n)$.

Method of moments:

$$\mathrm{E}(X) = \frac{\theta}{2},$$

so that

$$\frac{S_n}{n} = \frac{\widehat{\theta}_{\text{MM}}}{2}$$

and

$$\widehat{\theta}_{\text{MM}} = 2\frac{S_n}{n}.$$

Note that we might have $\widehat{\theta}_{\text{MM}} < \max(x_1, \ldots, x_n)$.

Exercise 10 (Mixture model)**:**
Method of moments:

$$\mathrm{E}(X) = \theta\frac{1}{\mu} + (1-\theta)\frac{1}{\lambda},$$

so that

$$\frac{S_n}{n} = \widehat{\theta}\frac{1}{\mu} + (1-\widehat{\theta})\frac{1}{\lambda}$$

We get:

$$\widehat{\theta} = \frac{\frac{S_n}{n} - \frac{1}{\lambda}}{\frac{1}{\mu} - \frac{1}{\lambda}}$$

Since $\theta \in [0, 1]$, we take:

$$\widehat{\theta} = \max(\min(\frac{\frac{S_n}{n} - \frac{1}{\lambda}}{\frac{1}{\mu} - \frac{1}{\lambda}}, 1), 0)$$

Exercise 11 (Gamma model)**:**
Method of moments:

$$\mathrm{E}(X) = \frac{a}{\lambda}, \quad \mathrm{E}(X^2) = \frac{a(a+1)}{\lambda^2}$$

so that

$$\bar{x} \equiv \frac{1}{n}\sum_{i=1}^n x_i = \frac{\widehat{a}}{\widehat{\lambda}}, \quad \bar{x}^{(2)} \equiv \frac{1}{n}\sum_{i=1}^n x_i^2 = \frac{\widehat{a}(\widehat{a}+1)}{\widehat{\lambda}^2}.$$

We get:

$$\widehat{a} = \frac{\bar{x}^2}{\bar{x}^{(2)} - \bar{x}^2}, \quad \widehat{\lambda} = \frac{\bar{x}}{\bar{x}^{(2)} - \bar{x}^2}.$$

Exercise 12 (Linear regression)**:**
Observations $Y_1, \ldots, Y_n \sim \mathcal{N}(\theta x_1, \sigma^2), \ldots, \mathcal{N}(\theta x_n, \sigma^2)$:

$$p_\theta(y_1, \ldots, y_n) \propto \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - \theta x_i)^2\right)$$

Log-likelihood:

$$\log p_\theta(y_1, \ldots, y_n) = -\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - \theta x_i)^2 + c.$$

Maximum for:

$$\sum_{i=1}^n x_i(y_i - \theta x_i) = 0$$

that is:

$$\widehat{\theta} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}.$$