MAP531 : Statistics
PC 3, 4 – Estimation

1 To warm up

Exercise 1. Estimators for the Gaussian mean
We consider a $n-$sample $(X_1, \ldots, X_n)$ distributed from a normal distribution $\mathcal{N}(\theta, \sigma^2)$ where $\theta \in \mathbb{R}$ and $\sigma^2 > 0$ are unknown.

1. Which of the following quantities are estimators of θ:

$$\frac{1}{n}\sum_{i=1}^n X_i, \quad \frac{1}{3}(X_1+X_2+X_3), \quad 1, \quad \theta, \quad X_1, \quad \mathsf{E}[X_1], \quad \frac{1}{n}\sum_{i=1}^n \left(X_i - \frac{1}{n}\sum_{j=1}^n X_j\right)^2, \quad \sigma^2, \quad X_n, \quad \frac{2}{n}\sum_{i=1}^n X_i - \theta.$$

2. Which estimator would you choose? Why?
3. Compute the expectation and the variance of all estimators.

We now assume that the variance is known and equals to 1.

4 What is the m.l.e. $\hat{\theta}_{\mathrm{MLE}}$ of θ?

5 Give an estimator $\hat{\theta}_{\mathrm{mom}}$ of θ using the method of moments.

6 What happens to the variance of $\hat{\theta}_{\mathrm{MLE}}$ when n increases?

We now consider a prior distribution Π on θ which is a normal distribution $\mathcal{N}(m, \tau)$.

7 How would you interpret the choice of a large τ?

8 What is the posterior distribution $\Pi(\cdot|X_1, \ldots, X_n)$ of θ? What happens when τ is large?

9 What is the posterior mean $\hat{\theta}_{\mathrm{pm}}$?

10 Compute the MAP estimator $\hat{\theta}_{\mathrm{MAP}}$, i.e. the parameter θ which maximises the p.d.f. of the posterior distribution.

11 What happens to $\hat{\theta}_{\mathrm{MLE}}$, $\hat{\theta}_{\mathrm{mom}}$, $\Pi(\cdot|X_1, \ldots, X_n)$, $\hat{\theta}_{\mathrm{pm}}$ and $\hat{\theta}_{\mathrm{MAP}}$ when n increases?

Solution

1. By definition of an estimator (measurable function defined on the observation space with value in the parameter space which does not depend on θ), the estimators of θ are:

$$\frac{1}{n}\sum_{i=1}^n X_i, \quad \frac{1}{3}(X_1 + X_2 + X_3), \quad 1, \quad X_1, \quad \frac{1}{n}\sum_{i=1}^n \left(X_i - \frac{1}{n}\sum_{j=1}^n X_j\right)^2, \quad X_n.$$

2. Choice of the estimator with the smallest mean square error among the candidates: $(1/n)\sum_{i=1}^n X_i$. Recall that the mean squared error (MSE) of an estimator $T(X)$ of $g(\theta)$ is

$$MSE_\theta(T(X)) = \mathsf{E}_\theta[(T(X)-g(\theta))^2] = \underbrace{\mathsf{E}_\theta[(T(X)-\mathsf{E}_\theta(T(X))^2]}_{\mathsf{Var}_\theta(T(X))} + (\underbrace{\mathsf{E}_\theta(T(X))-g(\theta)}_{Bias_\theta(T(X),g(\theta))})^2.$$

Or by using the law of large number and CLT. Note that 1 is a good estimator only in the case where $\theta = 1$, yet we don't know θ otherwise we wouldn't estimate it!

3. For all $\theta \in \mathbb{R}$ and $\sigma > 0$,

$$\mathsf{E}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n}\sum_{i=1}^n \mathsf{E}(X_i) = \theta,$$

by linearity of the expectation;

$$Var\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n^2}Var\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2}\sum_{i=1}^n Var(X_i) = \frac{\sigma^2}{n},$$

using the independence of the observations;

$$\mathsf{E}\left(\frac{1}{3}(X_1+X_2+X_3)\right) = \theta, \quad Var\left(\frac{1}{3}(X_1+X_2+X_3)\right) = \frac{\sigma^2}{3},$$

$$\mathsf{E}(1) = 1, \quad Var(1) = 0,$$

$$\mathsf{E}(X_1) = \theta, \quad Var(X_1) = \sigma^2,$$

$$\mathsf{E}\left[\frac{1}{n}\sum_{i=1}^n\left(X_i - \frac{1}{n}\sum_{j=1}^n X_j\right)^2\right] = \frac{1}{n}\mathsf{E}\left[\sum_{i=1}^n X_i^2 - \frac{1}{n}\Big(\sum_{j=1}^n X_j\Big)^2\right] = \mu^2+\sigma^2 - \frac{1}{n^2}(n^2\mu^2+n\sigma^2) = \frac{n-1}{n}\sigma^2$$

using that $n\sigma^2 = Var(\sum_{i=1}^n X_i) = \mathsf{E}\left[\left(\sum_{j=1}^n X_j\right)^2\right] - n^2\mu^2$. We could also use the fact that $\frac{1}{\sigma^2}\sum_{i=1}^n\left(X_i - \frac{1}{n}\sum_{j=1}^n X_j\right)^2$ is distributed from a $\chi^2(n-1)$ distribution which has a variance of $2(n-1)$ so that

$$Var\left(\frac{1}{n}\sum_{i=1}^n\left(X_i - \frac{1}{n}\sum_{j=1}^n X_j\right)^2\right) = \frac{2(n-1)\sigma^4}{n^2},$$

we can deduce from these computations that $\frac{1}{n}\sum_{i=1}^n\left(X_i - \frac{1}{n}\sum_{j=1}^n X_j\right)^2$ 'converge' to σ^2 'in distribution'. It is a usual estimator of σ^2.

$$\mathsf{E}(X_n) = \theta, \quad Var(X_n) = \sigma^2.$$

4. Done in the lecture, or if a repetition is needed, the log-likelihood associated is

$$\ell_n(\theta) = -\log((\sqrt{2\pi})^n) + \sum_{i=1}^n\left(-\frac{(X_i-\theta)^2}{2}\right).$$

$$\ell_n'(\theta) = \sum_{i=1}^n (X_i - \mu) = n\left(\frac{1}{n}(\sum_{i=1}^n X_i) - \theta\right)$$

$$\ell''(\theta) = -n$$

then the log-likelihood is a concave function of θ so admits a unique maximum which is $\hat{\theta}_{MLE} = \frac{1}{n}\sum_{i=1}^n X_i$.

5. We know that $E(X_1) = \theta$ so an estimator of θ by the method of moment is $\hat{\theta}_{mom} = \frac{1}{n}\sum_{i=1}^n X_i$ which is the same as the MLE.

6. Given the answer of Question 1, $\hat{\theta}$ is an unbiased estimator of θ and its variance is decreasing when n is increasing.

7. The parameter τ represents the variance of the prior distribution, a choice of a large τ means we are not confident about our knowledge of θ.

8. The posterior density is proportional to

$$\exp\left(-\sum_{i=1}^{n}\frac{(X_i-\theta)^2}{2}-\frac{(\theta-m)^2}{2\tau}\right)\propto\exp\left(-\frac{\theta^2}{2}\Big(n+\frac{1}{\tau}\Big)+\theta\Big(\sum_{i=1}^{n}X_i+\frac{m}{\tau}\Big)\right)$$
$$\propto\exp\left(-\frac{1}{2\tau/(1+\tau n)}\Big(\theta-\frac{\frac{1}{n}\sum_{i=1}^{n}X_i+\frac{m}{n\tau}}{1+1/(\tau n)}\Big)^2\right).$$

So that the posterior distribution of θ is $\mathcal{N}\left(\frac{\frac{1}{n}\sum_{i=1}^{n}X_i+\frac{m}{n\tau}}{1+1/(\tau n)},\frac{\tau}{1+\tau n}\right)$. We can notice that the mean of the posterior distribution is a barycenter of $\frac{1}{n}\sum_{i=1}^{n}X_i$ and m with respective weight $1/(1+1/(\tau n))$ and $1/(\tau n(1+1/(\tau n)))$. When τ is large the weight of the observations $\frac{1}{n}\sum_{i=1}^{n}X_i$ becomes larger and the weight of the prior mean smaller.

9. The posterior mean is $\hat{\theta}_{pm}=\frac{\frac{1}{n}\sum_{i=1}^{n}X_i+\frac{m}{n\tau}}{1+1/(\tau n)}$.

10. The MAP estimator is the same as the posterior mean.

11. By the strong law of large number $\hat{\theta}_{MLE}=\hat{\theta}_{mom}$ converge to θ. It is always centered at θ (unbiased) and its variance decreases to zero.

$$\mathsf{E}(\hat{\theta}_{pm})=\frac{\theta+\frac{m}{n\tau}}{1+1/(\tau n)}$$

so that $\hat{\theta}_{pm}$ is typically biased since $m\neq 0$. But it is asymptotically unbiased, since its expectation tends to θ when $n\to\infty$.

$$Var(\hat{\theta}_{pm})=\frac{\sigma^2}{n(1+1/\tau n)^2}$$

which tends to 0 when n tends to ∞, then $\hat{\theta}_{pm}$ also converge to θ when $n\to\infty$.

As to the posterior distribution, its expectation gets closer to $\frac{1}{n}\sum_{i=1}^{n}X_i$ which tends to θ and its variance gets closer to 0. Thus it concentrates its mass at θ when $n\to\infty$.

Exercise 2. Exponential distribution and method of moments
Let $(X_1,...,X_n)$ be a $n-$sample from an exponential distribution with parameter λ.

1. Give a simple estimator of λ using the method of moments.
2. Compute $\mathsf{E}(X_1^2)$.
3. Compute the following function $t_0\mapsto\bar{F}(t_0)=\mathbb{P}(X_1>t_0)$ as an expectation.
4. Deduce from the two previous questions two other estimators of λ, using the method of moments.

Solution

1. $\mathsf{E}(X_1)=1/\lambda$ so that $\lambda=f(\mathsf{E}(X_1))$ with $f:x\mapsto 1/x$. Then an estimator of λ using the method of moments is
$$\widehat{\lambda}_1=\frac{n}{\sum_{i=1}^{n}X_i}.$$

2. We know that $\mathsf{Var}(X_1)=1/\lambda^2$. So that,
$$\mathsf{E}(X_1^2)=\mathsf{Var}(X_1)+\mathsf{E}(X_1)^2=\frac{2}{\lambda^2}$$

and $\lambda = \sqrt{\frac{2}{\mathsf{E}(X_1^2)}}$. Then

$$\widehat{\lambda}_2 = \sqrt{\frac{2}{\frac{1}{n}\sum_{i=1}^n X_i^2}}$$

is another estimator using the method of moments.

3.
$$\mathbb{P}(X_1 > t_0) = \mathsf{E}(\mathbb{1}_{(t_0,+\infty)}(X_1)) = \exp(-\lambda t_0)$$
Then $\lambda = -\log(\mathsf{E}(\mathbb{1}_{(t_0,+\infty)}(X_1)))/t_0$ and

$$\widehat{\lambda}_3 = -\frac{\log\left(\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{(t_0,+\infty)}(X_i)\right)}{t_0}$$

is another estimator using the method of moments.

2 To train

Exercise 3. Distribution of genotypes in a population
When the gene frequencies are reaching their equilibrium, the genotypes AA, Aa and aa appear in the population with probability $(1-\theta)^2$, $2\theta(1-\theta)$ and θ^2 respectively, where θ in an unknown parameter. Plato et al. (1964) have published the following data on the haptoglobine type in a sample composed of 190 subjects:

Haptoglobine type:

Hp-AA	*Hp-Aa*	*Hp-aa*
10	68	112

1. How would you interpret the parameter θ? Propose a statistical model for this problem.
2. Propose an estimator using the method of moments.
3. Compute the MLE $\hat{\theta}_n$ of θ.

Solution

1. θ represents the proportion of alleles "a" in the population.
   Statistical model: $\mathcal{X} = \{AA, Aa, aa\}$, $\mathcal{A} = \mathcal{P}(\mathcal{X})$, $\mathcal{P} = \{\mathbb{P}_\theta, \theta \in (0,1)\}$ where $\mathbb{P}_\theta$ is a distribution on $\mathcal{X}$ such that
   $$\mathbb{P}_\theta(AA) = (1-\theta)^2, \quad \mathbb{P}_\theta(Aa) = 2\theta(1-\theta), \quad \mathbb{P}_\theta(aa) = \theta^2.$$
   We assume that the observations X_i are i.i.d. from some $\mathbb{P}_\theta$ and we want to estimate θ. Let us denote
   $$N_{AA} = \sum_{i=1}^n \mathbb{1}_{X_i=AA}, \quad N_{Aa} = \sum_{i=1}^n \mathbb{1}_{X_i=Aa}, \quad N_{aa} = \sum_{i=1}^n \mathbb{1}_{X_i=aa}.$$
   We have observed a realization $N_{AA}(\omega) = 10$, $N_{Aa}(\omega) = 68$ and $N_{aa}(\omega) = 112$.

2. For instance, $\mathsf{E}(\mathbb{1}_{X_1=aa}) = \theta^2$, where $\hat{\theta}_m = \sqrt{N_{aa}/n}$.

3. The likelihood is
   $$p_n(X_1, \ldots, X_n, \theta) = (1-\theta)^{2N_{AA}}(2\theta(1-\theta))^{N_{Aa}}\theta^{2N_{aa}}$$
   and the log-likelihood:
   $$\mathcal{L}_n(X_1, \ldots, X_n, \theta) = 2N_{AA}\log(1-\theta) + N_{Aa}\log(2\theta(1-\theta)) + 2N_{aa}\log(\theta).$$
   The M.L.E is
   $$\hat{\theta}_{MLE} = \frac{N_{Aa} + 2N_{aa}}{2n}.$$
   With our experiment, our estimate is $\hat{\theta}_{MLE}(\omega) = 146/190 \simeq 0.77$.

Exercise 4. Translation and scaling model

An experimenter realizes n independent observations of a quantity. We assume that:

$$X_i = \mu + \sigma\zeta_i \,,$$

where $\{\zeta_i\}_{i=1}^n$ are i.i.d random variables and their pdf w.r.t the Lebesgue measure on $\mathbb{R}$ g is known. μ is the *translation* parameter and σ a *scale*.

1. Write the statistical model.
2. When g is the pdf of a standard normal distribution, propose two estimators of the two parameters using the method of moments and the maximum likelihood estimation.
3. Assume now that g is a Laplace pdf. Propose two estimators of the two parameters.

Solution

1. Statistical model: the observed values $x_1, \dots, x_n$ are realizations of the random vector $(X_1, \dots, X_n)$ with values in $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$.
   Let $\theta = (\mu, \sigma^2)$ in $\Theta = \mathbb{R} \times \mathbb{R}_+$. We consider the family of distributions $\{\mathbb{P}_\theta, \theta \in \Theta\}$. $\mathbb{P}_\theta$ is defined thanks to its density p^θ:
   $$p^\theta : (y_1, \dots, y_n) \in \mathbb{R}^n \mapsto \frac{1}{\sigma^n} \prod_{i=1}^n g\left(\frac{y_i - \mu}{\sigma}\right).$$
   Indeed, for any bounded continuous function f,
   $$\mathsf{E}(f(X_1)) = \int f(\mu + \sigma t) g(t) dt = \int f(y) \frac{1}{\sigma} g\left(\frac{y - \mu}{\sigma}\right) dy.$$
2. **Method of moments**: $\mathsf{E}(X_1) = \mu$ and $\mathsf{E}(X_1^2) - \mathsf{E}(X_1)^2 = \mathsf{Var}(X_1) = \sigma^2$. So that
   $$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^n X_i \quad \widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \left(\frac{1}{n}\sum_{i=1}^n X_i\right)^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \widehat{\mu})^2$$
   are estimators of μ and σ^2 using the method of moments.
   MLE The likelihood is
   $$\frac{1}{(\sqrt{2\pi\sigma^2})^n} \exp\left(-\sum_{i=1}^n \frac{(X_i - \mu)^2}{2\sigma^2}\right)$$
   The log likelihood is
   $$\ell_n(\theta) = -\frac{n}{2}\log(2\pi\sigma^2) - \sum_{i=1}^n \frac{(X_i - \mu)^2}{2\sigma^2}.$$
   $$\frac{\partial \ell_n}{\partial \mu}(\theta) = \sum_{i=1}^n \frac{X_i - \mu}{\sigma^2} = \frac{n}{\sigma^2}\left(\frac{1}{n}\sum_{i=1}^n X_i - \mu\right)$$
   $$\frac{\partial \ell_n}{\partial \sigma^2}(\theta) = -\frac{n}{2\sigma^2} + \sum_{i=1}^n \frac{(X_i - \mu)^2}{2\sigma^4} = \frac{n}{2\sigma^2}\left(-1 + \frac{1}{\sigma^2}\frac{1}{n}\sum_{i=1}^n (X_i - \mu)^2\right)$$
   Here, the log-likelihood admits a unique maximum which is given by
   $$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^n X_i \quad \widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \widehat{\mu})^2.$$
3. $g(x) = \frac{1}{\sqrt{2}}\exp(-\sqrt{2}|x|)$ **Method of moments**: same as before.
   MLE The likelihood is
   $$\frac{1}{\sqrt{2\sigma^2}^n} \exp\left(-\sqrt{2}\sum_{i=1}^n \frac{|X_i - \mu|}{\sqrt{\sigma^2}}\right)$$

The log likelihood is

$$\ell_n(\theta) = constant - n\log(\sigma) - \frac{\sqrt{2}}{\sigma}\sum_{i=1}^{n}|X_i - \mu|.$$

Maximizing it with respect to σ, gives $\hat{\sigma} = \frac{\sqrt{2}}{n}\sum_{i=1}^{n}|X_i - \mu|$. Maximizing it with respect to μ is equivalent to minimizing $\mu \mapsto \sum_{i=1}^{n}|X_i - \mu|$. We recognize the definition of the median. When n in uneven, $\hat{\mu} = X_{((n+1)/2)}$. When n is odd, any value in $(X_{(n/2)}, X_{(n/2+1)})$ maximizes the quantity.

Exercise 5. Binomial models

Suppose that X is distributed from a Binomial distribution $\mathrm{Bin}(n,\theta)$ and we want to estimate θ.

1. Compute a m.l.e. $\hat{\theta}_{\mathrm{MLE}}$, its bias and its variance.
2. Consider an alternate estimator $\hat{\theta}_{\mathrm{alt.}} = (X+1)/(n+2)$. Compute its bias and variance.
3. Compare the mean square error of $\hat{\theta}_{\mathrm{MLE}}$ and $\hat{\theta}_{\mathrm{alt.}}$.

Notice that the Beta distribution $\mathrm{Beta}(a,b)$ with parameters $a>0$ and $b>0$ has the following p.d.f.:

$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}\mathbb{1}_{[0,1]}(x)$$

with respect to the Lebesgue measure.

4 Compute the expectation of a Beta distribution $\mathrm{Beta}(a,b)$ with parameters $a>0$ and $b>0$. You can use that $\Gamma(a+1) = a\Gamma(a)$.

We now take a Bayesian point of view and consider a prior distribution Π for θ in $[0,1]$.

5 Compute the posterior distribution, its mean $\hat{\theta}_{m1}$ and the MAP estimator $\hat{\theta}_{MAP1}$ when the prior distribution is a Uniform distribution on $[0,1]$.

6 Compute the posterior distribution, its mean $\hat{\theta}_{m2}$ when the prior distribution is a Beta distribution $\mathrm{Beta}(a,b)$. Compute the MAP estimator $\hat{\theta}_{MAP2}$ in the case where $a>1$ and $b>1$.

Solution

1. The log-likelihood is
$$\ell_n(\theta) = \text{constant} + X\log(\theta) + (n-X)\log(1-\theta),$$
$$\ell_n'(\theta) = \frac{X}{\theta} - \frac{n-X}{1-\theta}, \quad \ell_n''(\theta) = -\frac{X}{\theta^2} - \frac{n-X}{(1-\theta)^2} < 0.$$
So the unique MLE is $\hat{\theta}_{MLE} = X/n$. It is unbiased:
$$\mathsf{E}(\hat{\theta}_{MLE}) = \theta, \quad Var(\hat{\theta}_{MLE}) = \frac{\theta(1-\theta)}{n}, \quad mse(\hat{\theta}_{MLE}) = \frac{\theta(1-\theta)}{n}.$$
2. $\hat{\theta}_{alt}$ is biased but asymptotically unbiased, it has smaller variance than $\hat{\theta}_{MLE}$:
$$\mathsf{E}(\hat{\theta}_{alt}) = \frac{n\theta}{n+2} + \frac{1}{n+2}, \quad Var(\hat{\theta}_{alt}) = \frac{n\theta(1-\theta)}{(n+2)^2}, \quad mse(\hat{\theta}_{alt}) = \frac{(n-4)\theta(1-\theta)+1}{(n+2)^2}.$$
3. See Figure **??**, the mse of $\hat{\theta}_{MLE}$ is smaller than the one of $\hat{\theta}_{alt}$ when θ is close to 0 and 1 otherwise it is the opposite.
4. The expectation of a $\mathrm{Beta}(a,b)$ distribution is
$$\int_0^1 x\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}dx = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)} = \frac{a}{a+b}.$$

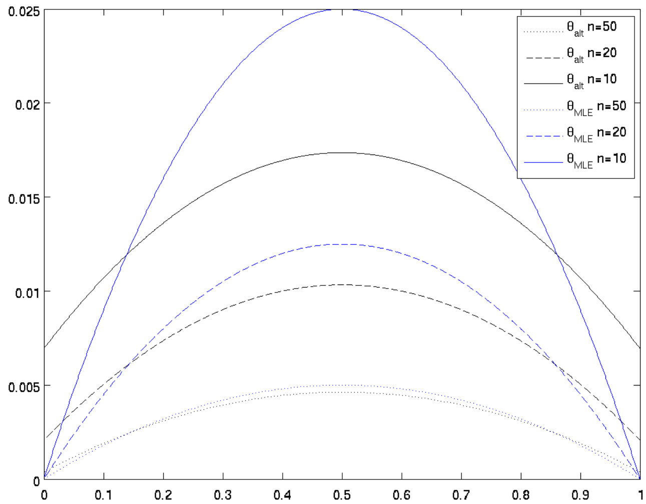

Figure 1: Mean square error of different estimators with respect to θ

5. When the prior distribution is uniform:

$$\Pi(\theta|X) \propto \theta^X (1-\theta)^{n-X} * \mathbb{1}_{[0,1]}(\theta)$$

the posterior distribution is a Beta$(X+1, n-X+1)$. Using Question 4, the posterior mean is $\hat{\theta}_{m1} = \frac{X+1}{n+2} = \hat{\theta}_{alt}$. And the MAP estimator is equal to the MLE.

6. When the prior distribution is Beta(a, b):

$$\Pi(\theta|X) \propto \theta^X (1-\theta)^{n-X} \theta^{a-1} (1-\theta)^{b-1} \propto \theta^{X+a-1} (1-\theta)^{n-X+b-1}$$

the posterior distribution is a Beta$(X+a, n-X+b)$. Note that it is coherent with what we found in the previous question since the uniform distribution on $[0,1]$ is a Beta$(1,1)$ distribution. Using Question 4, the posterior mean is $\hat{\theta}_{m2} = \frac{X+a}{n+a+b}$. Note that $\hat{\theta}_{m2} = \hat{\theta}_{alt}$ when $a = b = 1$. To find the MAP estimator we maximize $f_{a,b}(\theta) = (X+a-1)\log(\theta) + (n-X+b-1)\log(1-\theta)$,

$$f'_{a,b}(\theta) = \frac{X+a-1}{\theta} - \frac{n-X+b-1}{1-\theta}, \quad f''_{a,b}(\theta) = -\frac{X+a-1}{\theta^2} - \frac{n-X+b-1}{(1-\theta)^2} < 0.$$

Then the MAP estimator is $\hat{\theta}_{MAP} = (X+a-1)/(n+a+b-2)$.

Exercise 6. Waiting time at a red light

We consider the random variable which models the waiting time at a red light. Let $\theta > 0$ be the maximum waiting duration at this light. We observe a $n-$sample $(t_1, ..., t_n)$ of the waiting duration of n people. We assume that the associated random variables $(T_1, ..., T_n)$ are i.i.d. from a Uniform distribution on $[0, \theta]$.

1. Compute the expectation and the variance of this distribution.

2. We want to estimate the parameter θ. Give the bias and the variance of the statistic $\bar{T}_n = \frac{1}{n}\sum_{i=1}^{n} T_i$.

3. Show that $\hat{\theta}_1 = 2\bar{T}_n$ is an unbiased estimator of θ.

4. Show that the MLE of θ is $\hat{\theta}_n = \sup_{i\in\{1,..,n\}} T_i$.

5. Compute the cumulative distribution function of $\hat{\theta}_n$. Deduce its p.d.f. and compute its expectation and variance.

6. Show that the statistic $\hat{\theta}_2 = \frac{n+1}{n}\hat{\theta}_n$ is an unbiased estimator of θ.

7. Which estimator would you choose to estimate θ: $\hat{\theta}_1$ or $\hat{\theta}_2$?

Solution

1. $\mathsf{E}_\theta(X_1) = \theta/2$ and $\mathsf{Var}_\theta(X_1) = \theta^2/12$ because

$$\mathsf{Var}(X_1) + \mathsf{E}(X_1)^2 = \mathsf{E}(X_1^2) = \frac{1}{\theta}\int_0^\theta x^2 dx = \left[\frac{x^3}{3\theta}\right]_0^\theta = \frac{\theta^2}{3}.$$

2. $\mathsf{E}_\theta(\bar{T}_n) = \mathsf{E}_\theta(T_1) = \theta/2$ so that the bias of $\bar{T}_n$ is $\theta/2$. $\bar{T}_n$ is a biased estimator of θ which is not asymptotically unbiased.
Since the T_i are i.i.d., $\mathsf{Var}_\theta(\bar{T}_n) = \frac{1}{n^2}\sum_{i=1}^n \mathsf{Var}_\theta(T_i) = \frac{\theta^2}{12n}$. The mean squared error of $\bar{T}_n$ is $bias_\theta^2(\bar{T}_n) + \mathsf{Var}_\theta(\bar{T}_n) = \theta^2\frac{3n+1}{12n}$, which is not converging to 0 when n increases!

3. $\hat{\theta}_1$ is an estimator of θ using the method of moments. By linearity of the expectation, $\mathsf{E}_\theta(\hat{\theta}_1) = 2\mathsf{E}_\theta(\bar{T}_n) = \theta$ for all $\theta > 0$. So that $\hat{\theta}_1$ is an unbiased estimator of θ.

4. The likelihood is

$$\mathcal{L}(T_1, \ldots, T_n, \theta) = \frac{1}{\theta^n}\prod_{i=1}^n \mathbb{1}_{[0,\theta]}(T_i) = \frac{1}{\theta^n}\prod_{i=1}^n \mathbb{1}_{0\leqslant \min_{i=1,\ldots,n}(T_i)\leqslant \max_{i=1,\ldots,n}(T_i)\leqslant\theta}.$$

$\theta \mapsto \mathcal{L}(T_1, \ldots, T_n, \theta)$ is null when $\theta < \max_{i=1,\ldots,n}(T_i)$ and is positive and non-increasing on $[\max_{i=1,\ldots,n}(T_i), +\infty)$. So that the likelihood is maximized at $\max_{i=1,\ldots,n}(T_i) = \hat{\theta}_n$, the unique m.l.e.

5. Let $t \in \mathbb{R}$, by independence,

$$P_\theta(\hat{\theta}_n \leqslant t) = P_\theta(T_i \leqslant t, \text{ for all } i \leqslant n) = \prod_{i=1}^n P_\theta(T_i \leqslant t) = \begin{cases} 0 & \text{if } t < 0 \\ \left(\frac{t}{\theta}\right)^n & \text{if } 0 \leqslant t \leqslant \theta \\ 1 & \text{otherwise} \end{cases} = \int_{-\infty}^t \frac{n}{\theta}\left(\frac{x}{\theta}\right)^{n-1}\mathbb{1}_{[0,\theta]}(x)dx.$$

So that the p.d.f. of $\hat{\theta}_n$ is $\theta \mapsto \frac{n}{\theta}\left(\frac{x}{\theta}\right)^{n-1}\mathbb{1}_{(0,\theta)}(x)$.

$$\mathsf{E}_\theta(\hat{\theta}_n) = \int_\mathbb{R} x\frac{n}{\theta}\left(\frac{x}{\theta}\right)^{n-1}\mathbb{1}_{[0,\theta]}(x)dx = \frac{n}{n+1}\theta$$

then $\hat{\theta}_n$ is biased but asymptotically unbiased.

$$\mathsf{E}_\theta(\hat{\theta}_n^2) = \int_\mathbb{R} x^2\frac{n}{\theta}\left(\frac{x}{\theta}\right)^{n-1}\mathbb{1}_{[0,\theta]}(x)dx = \frac{n}{n+2}\theta^2,$$

so that

$$\mathsf{Var}_\theta(\hat{\theta}_n) = \mathsf{E}_\theta(\hat{\theta}_n^2) - \mathsf{E}_\theta(\hat{\theta}_n)^2 = \frac{n}{(n+1)^2(n+2)}\theta^2.$$

6. By linearity of the expectation,

$$\mathsf{E}_\theta(\hat{\theta}_2) = \frac{n+1}{n}\mathsf{E}_\theta(\hat{\theta}_n) = \theta$$

for all θ. So that $\hat{\theta}_2$ is unbiased.

7. One would choose the estimator with the smallest mean square error. As these two estimators are unbiased, it's the same as choosing the estimator with the smallest variance.

$$\mathsf{Var}_\theta(\hat{\theta}_2) = \left(\frac{n+1}{n}\right)^2\mathsf{Var}_\theta(\hat{\theta}_n) = \frac{1}{n(n+2)}\theta^2$$

which decreases quadratically with n instead of linearly for $\mathsf{Var}_\theta(\hat{\theta}_1)$. I then would choose $\hat{\theta}_2$.
Note that the rate of convergence of $\hat{\theta}_2$ is n and not the usual $\sqrt{n}$ (usually obtained with the CLT) because the model is not regular (the support of the distribution depends on the parameter).

Exercise 7. Suppose that $X_1, \dots, X_n$ are i.i.d. distributed from a distribution with p.d.f.

$$\theta x^{\theta-1} \mathbb{1}_{(0,1)}(x)$$

with respect to the Lebesgue measure, for some $\theta > 0$.

1. Give a m.l.e. of θ.
2. Give an estimator of θ using the method of moments.

Assume that the prior distribution is a Gamma distribution Gamma(a, b) for some $a > 0$, $b > 0$.

4. Find the posterior distribution of θ.
5. Compute the posterior mean.
6. Compute a MAP estimator.

Solution

1. The log-likelihood is

$$\ell_n(\theta) = n \log(\theta) + (\theta - 1) \sum_{i=1}^{n} \log(X_i)$$

$$\ell_n'(\theta) = \frac{n}{\theta} + \sum_{i=1}^{n} \log(X_i)$$

$$\ell_n''(\theta) = -\frac{n}{\theta^2} < 0.$$

So that the unique maximum of ℓ_n is at $\hat{\theta}_{MLE} = -\frac{n}{\sum_{i=1}^n \log(X_i)}$.

2.

$$\mathsf{E}(X_1) = \int_0^1 x\theta x^{\theta-1} dx = \frac{\theta}{\theta+1}$$

so that $\theta = \frac{\mathsf{E}(X_1)}{1-\mathsf{E}(X_1)}$ so an estimate of θ by the method of moment is $\hat{\theta}_{mom} = \frac{1/n \sum_{i=1}^n X_i}{1-1/n \sum_{i=1}^n X_i}$.

3. The posterior density function is proportional to

$$\theta^n \exp\left((\theta - 1) \sum_{i=1}^{n} \log(X_i)\right) \theta^{a-1} \exp(-b\theta) \propto \theta^{n+a-1} \exp\left(-\Big(b - \sum_{i=1}^{n} \log(X_i)\Big)\theta\right).$$

So that the posterior distribution is a Gamma $(n + a, b - \sum_{i=1}^n \log(X_i))$ distribution.

4. We know that the expectation of a Gamma(a, b) distribution is a/b so that the posterior mean is $(n + a)/(b - \sum_{i=1}^n \log(X_i))$. Note that it is equal to the mle when $a = b = 0$.

5. Let $c, d > 0$, we want to find the maximum, if it exists of the function $\theta^{c-1} \exp(-d\theta)$, let us define $f_{c,d}(\theta) = \log(\theta^{c-1} \exp(-d\theta)) = (c - 1) \log(\theta) - d\theta$.

$$f_{c,d}'(\theta) = \frac{c-1}{\theta} - d, \quad f_{c,d}''(\theta) = -\frac{c-1}{\theta^2}$$

So that the maximum of f is reached at

$$\begin{cases} \frac{c-1}{d} & \text{if } c > 1 \\ 0 & \text{otherwise.} \end{cases}$$

Then the MAP estimator is

$$\hat{\theta}_{MAP} = \begin{cases} \frac{n+a-1}{b-\sum_{i=1}^n \log(X_i)} & \text{if } n + a > 1 \\ 0 & \text{otherwise.} \end{cases}$$

Exercise 8. Connection duration
We can model the connection duration on a website by a Gamma(a, b), $(a > 0, b > 0)$ distribution which p.d.f. is given by

$$p_{a,b}(x) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx) \mathbf{1}[0, +\infty)(x).$$

Let $\theta = (a, b)$.
To fix the advertisement prices, one aims at estimating the parameter θ from an observed sample $X_1, ..., X_n$ of n connection durations.

1. Knowing that $\Gamma(a+1) = a\Gamma(a)$, propose an estimator of the parameter θ with the method of moments.
2. Write the likelihood equations. Let a be fixed, determine the maximum $\hat{b}_n(a))$ of the likelihood function.
3. Prove that the MLE is $\hat{\theta}_n = (\hat{a}_n, \hat{b}(\hat{a}_n))$ where $\hat{a}_n$ is the maximum of

$$L_n(a) = na \ln(a) - na \ln(\bar{X}_n) - n \ln(\Gamma(a)) + n(a-1)\overline{\ln(X)}_n - na$$

   where $\overline{\ln(X)}_n = n^{-1} \sum_{i=1}^{n} \ln(X_i)$.
4. Propose a numerical scheme to determine this MLE.

Exercise 9. Soccer score We recall that the *Poisson distribution* with parameter $\lambda > 0$ has a pdf given by $(p(\lambda, k), k \in \mathbb{N})$ w.r.t the counting measure on $\mathbb{N}$:

$$p(\lambda, k) = \exp(-\lambda) \frac{\lambda^k}{k!} .$$

1. Compute the probability generating function of the Poisson distribution given by $G_X(s) = \mathsf{E}[\exp(sX)]$ for $s \in \mathbb{R}$.
2. Deduce the mean and the variance of this distribution as a function of λ.

   Note that if X_1 and X_2 are two independent random variables following a Poisson distribution with respective parameters $\lambda_1 > 0$ and $\lambda_2 > 0$, then $X_1 + X_2$ has a Poisson distribution of parameter $\lambda_1 + \lambda_2$.

We are provided with n independent observations of a Poisson random variable of parameter $\theta \in \Theta = \mathbb{R}_+^*$.

3. Write the corresponding statistical model.
4. Propose a method to estimate θ.

We want to model the number of goals during a soccer game by a Poisson distribution. We first consider that the number of goals of the local team and of the visiting team are independent Poisson, resp. with parameter $\lambda > 0$ and $\mu > 0$.

5. Write the statistical model associated to the observation of n match results.
6. Propose a method to estimate λ and μ.
7. We are provided with the following results from of the *premier league* for seasons from 2004-2005 to 2008-2009. The number of matches is $n = 1,900$:

   The total number of goals is in average equal to $2,523$ with variance $2,640$,

   The number of goals of the local team is in average equal to $1,468$ with variance $1,617$.

   The number of goals of the visiting team is in average equal to $1,055$ with variance $1,158$.

   Estimate λ and μ. Does the Poisson assumption look correct?

	Observed	Poisson ($\lambda = 1.468$)		Observed	Poisson ($\mu = 1.055$)
0	469	437.7	0	692	661.6
1	621	642.6	1	680	697.9
2	456	471.7	2	335	368.2
3	217	230.8	3	131	129.5
4	100	84.7	4	51	34.1
$\geqslant 5$	37	32.5	$\geqslant 5$	11	8.7
Total	1900		Total	1900	

Figure 2: Results of the local team (left) and visiting one (right)

Instead of fitting a Poisson distribution, it looks cleverer in this case to consider another family of distribution such that the mean and the variance can be different (called "over-dispersion")
Let us first analyze the numbers of goals of the local team and of the visiting one (see figure 1). This table shows that the null result is underestimated and one or two goals are overestimated. We therefore introduce another distribution with an atom at O.

$$p_{\pi,\lambda}(k) = (1-\pi)\mathbf{1}_{\{0\}}(k) + \pi \exp^{-\lambda} \frac{\lambda^k}{k!} \,,$$

where $\pi \in (0,1)$ is the mixture proportion.

8. Compute the mean and the order 2 moment.

9. Propose an estimation method of π and λ.

3 To go further

Exercise 10. Autoregressive model
Let us consider the autoregressive model with $X_1, ..., X_n$, n observations such that the X_i's are given by: $X_i = \theta X_{i-1} + \xi_i, \quad i = 1, ..., n, \quad X_0 = 0$, with ξ_i i.i.d. following $\mathcal{N}(0, \sigma^2)$ and $\theta \in \mathbb{R}$. Compute the MLE $(\hat{\theta}_n, \hat{\sigma}_n^2)$ of (θ, σ^2).

Solution

$$X_0 = 0, \quad X_i = \theta X_{i-1} + \xi_i, \quad \xi_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$$

Statistical model

- Probability space: $\mathbb{R}^n$ (endowed with the Borel σ-field,
- The family of distribution $P_{\theta,\sigma}$ is parametrized by $\theta \in \mathbb{R}$ and $\sigma^2 \in \mathbb{R}_+$. Under $P_{\theta,\sigma}$,

$$Z^T = \begin{pmatrix} X_1 \\ X_2 \\ \\ \vdots \\ \\ X_n \end{pmatrix} = \begin{pmatrix} 1 & 0 & & & 0 & 0 \\ \theta & 1 & & & & 0 \\ \theta^2 & \theta & 1 & \ddots & & \\ & \ddots & \ddots & \ddots & & \\ \theta^{n-2} & & & \theta & 1 & 0 \\ \theta^{n-1} & \theta^{n-2} & & \theta^2 & \theta & 1 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \\ \vdots \\ \\ \xi_n \end{pmatrix} =: A_\theta \xi$$

where ξ is a Gaussian vector with distribution $\mathcal{N}(\vec{o}, \sigma^2 Id_n)$. So that Z is a Gaussian vector with distribution $\mathcal{N}(\vec{o}, \sigma^2 A_\theta A_\theta^T)$ which admits the following density (w.r.t. the Lebesgue measure)

$$z \in \mathbb{R}^n \mapsto \frac{1}{\sqrt{2\pi\sigma}^n} \frac{1}{\sqrt{\det(\Gamma_\theta)}} \exp\left(-\frac{1}{2\sigma^2} z^T \Gamma z\right),$$

where

$$\Gamma_\theta = (A_\theta A_\theta^T)^{-1} = \begin{pmatrix} 1+\theta^2 & -\theta & 0 & & 0 & 0 \\ -\theta & 1+\theta^2 & -\theta & 0 & & 0 \\ 0 & -\theta & 1+\theta^2 & -\theta & & \\ & \ddots & \ddots & \ddots & \ddots & \\ 0 & & & -\theta & 1+\theta^2 & -\theta \\ 0 & 0 & & 0 & -\theta & 1+\theta^2 \end{pmatrix}.$$

Exercise 11. Inadmissibility Let $(X_1, \ldots, X_n)$ be a random sample such that $\mathsf{E}(X_1^2) < +\infty$. We want to estimate $\mu = \mathsf{E}(X_1) \in \mathbb{R}$ and we use the mean squared error to measure the performance of the estimators. Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$.

1. Show that any estimator $\hat{\theta}_{a,b} = a\bar{X}_n + b$, with $a > 1$ $b \in \mathbb{R}$, satisfies

$$\mathsf{E}\left((\hat{\theta}_{a,b} - \mu)^2\right) > \mathsf{E}\left((\bar{X}_n - \mu)^2\right), \quad \text{for all } \mu.$$

2. Which estimator, $\hat{\theta}_{a,b}$ or $\bar{X}_n$, do you prefer. Why?
3. Show that any estimator $\hat{\theta}_b = \bar{X}_n + b$, with $b \neq 0$, satisfies

$$\mathsf{E}\left((\hat{\theta}_b - \mu)^2\right) > \mathsf{E}\left((\bar{X}_n - \mu)^2\right) \quad \text{for all } \mu.$$

4. Which estimator, $\hat{\theta}_b$ or $\bar{X}_n$, do you prefer. Why?

Solution

1. For all $\mu \in \mathbb{R}$,

$$\begin{aligned}\mathsf{E}\left((\hat{\theta}_{a,b} - \mu)^2\right) &= E\left((a\bar{X}_n + b - \mu)^2\right) = a^2 Var\left(\bar{X}_n\right) + (a\mu + b - \mu)^2 \\ &\geqslant a^2 Var\left(\bar{X}_n\right) > E\left((\bar{X}_n - \mu)^2\right)\end{aligned}$$

2. Then the estimator $\bar{X}_n$ is always (for all μ) better than $\hat{\theta}_{a,b}$ for any $a > 1$, $b \in \mathbb{R}$. We say that $\hat{\theta}_{a,b}$ is inadmissible for all $a > 1$, $b \in \mathbb{R}$.
3. For all $\mu \in \mathbb{R}$,

$$\begin{aligned}\mathsf{E}\left((\hat{\theta}_b - \mu)^2\right) &= E\left((\bar{X}_n + b - \mu)^2\right) = Var\left(\bar{X}_n\right) + (\mu + b - \mu)^2 \\ &> Var\left(\bar{X}_n\right) = E\left((\bar{X}_n - \mu)^2\right)\end{aligned}$$

4. Then the estimator $\bar{X}_n$ is always (for all μ) better than $\hat{\theta}_b$ for any $b \neq 0$. We say that $\hat{\theta}_b$ is inadmissible for all $b \neq 0$.

Exercise 12. Curse of dimensionality
Let X_i, $i \leqslant n$ be independently distributed as $\mathcal{N}(\mu_i, \sigma^2)$. Independently, let Y_i, $i \leqslant n$ be independently distributed as $\mathcal{N}(\mu_i, \sigma^2)$. We want to estimate $\theta = (\mu_1, \mu_2, \ldots, \mu_n, \sigma^2)$ from the observations $X_1, \ldots, X_n, Y_1, \ldots, Y_n$.

1. Give the statistical model.
2. Find the m.l.e. $\hat{\theta} = (\hat{\mu}_1, \hat{\mu}_2, \ldots, \hat{\mu}_n, \hat{\sigma}^2)$ of θ.

3. Are $\hat{\mu}_i$, $i \leqslant n$, and $\hat{\sigma}$ respectively unbiased estimators of μ_i, $i \leqslant n$ and σ?
4. What is the variance of $\hat{\mu}_i$, $i \leqslant n$, and $\hat{\sigma}^2$? What happens when n increases?
5. Comment the previous results.

Solution

1. The statistical model is $\left((\mathbb{R}^{2n}, \mathcal{B}(\mathbb{R})^{2n}), P_{\mu_1,\mu_2,\ldots,\mu_n,\sigma^2}, (\mu_1, \mu_2, \ldots, \mu_n, \sigma^2) \in \mathbb{R}^n \times \mathbb{R}_+^*\right)$, where $P_{\mu_1,\mu_2,\ldots,\mu_n,\sigma^2} = \mathcal{N}(\mu_1,\sigma^2) \otimes \mathcal{N}(\mu_2,\sigma^2) \otimes \ldots \mathcal{N}(\mu_n,\sigma^2) \otimes \mathcal{N}(\mu_1,\sigma^2) \otimes \mathcal{N}(\mu_2,\sigma^2) \otimes \ldots \mathcal{N}(\mu_n,\sigma^2)$.
2. The likelihood function is
$$\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\frac{(X_1-\mu_1)^2}{2\sigma^2}\right) \ldots \exp\left(-\frac{(X_n-\mu_n)^2}{2\sigma^2}\right) \exp\left(-\frac{(Y_2-\mu_2)^2}{2\sigma^2}\right) \ldots \exp\left(-\frac{(Y_n-\mu_n)^2}{2\sigma^2}\right).$$
So the log-likelihood function is
$$\ell_n(\theta) = -n\log(\sqrt{2\pi}) - n\log(\sqrt{\sigma^2}) - \sum_{i=1}^n \frac{(X_i-\mu_i)^2 + (Y_i-\mu_i)^2}{2\sigma^2},$$
$$\frac{\partial \ell_n}{\partial \mu_i}(\theta) = \frac{1}{\sigma^2}(X_i - \mu_i + Y_i - \mu_i), \quad \frac{\partial \ell_n}{\partial \sigma^2}(\theta) = -\frac{n}{2\sigma^2} + \sum_{i=1}^n \frac{(X_i-\mu_i)^2 + (Y_i-\mu_i)^2}{2\sigma^4}.$$
Then $\hat{\mu}_i = (X_i+Y_i)/2$ and $\hat{\sigma}^2 = \frac{1}{2n}(\sum_i (X_i - (X_i+Y_i)/2)^2 + \sum_i (Y_i - (X_i+Y_i)/2))^2 = \frac{1}{4n}\sum_i (X_i - Y_i)^2$.
3. $\hat{\mu}_i$ is unbiased and $\hat{\sigma}^2$ is biased and worse asymptotically biased:
$$\mathsf{E}(\hat{\mu}_i) = \mu_i$$
$$\begin{aligned}\mathsf{E}(\hat{\sigma}^2) &= \frac{1}{4n}\sum_{i=1}^n \mathsf{E}((X_i-Y_i)^2) = \frac{1}{4n}\sum_{i=1}^n (\mathsf{E}(X_i^2) - 2\mathsf{E}(X_iY_i) + \mathsf{E}(Y_i^2)) \\ &= \frac{1}{4n}\sum_{i=1}^n (\mathsf{E}(X_i^2) - 2\mathsf{E}(X_i)\mathsf{E}(Y_i) + \mathsf{E}(Y_i^2)) = \frac{1}{4n}\sum_{i=1}^n (2\sigma^2) = \frac{\sigma^2}{2}\end{aligned}$$
4. The variance of $\hat{\mu}_i$ doesn't decrease when n increases:
$$Var(\hat{\mu}_i) = \frac{1}{4}(Var(X_i) + Var(Y_i)) = \frac{\sigma^2}{2}.$$
Moreover by the law of large number: $\hat{\sigma}^2 \to \frac{\sigma^2}{2}$ so it is not consistent.
5. (Neyman-Scott paradox) the mle fails because the number of parameters grows with the number of observations.

Exercise 13. Canonical exponential models
A large part of the useful models belong to what is called exponential models as for example Gaussian, log-normal, exponential, gamma, Bernoulli or Poisson.
Here, we will study several properties of these models.
Let $(\mathsf{X}, \mathcal{X})$ a measurable space (we will focus on $\mathsf{X} = \mathbb{R}^k$ or $\mathsf{X} = \mathbb{N}^k$) and μ a σ-finite measure on $(\mathsf{X}, \mathcal{X})$. Let $T : \mathsf{X} \to \mathbb{R}$ and $h : \mathsf{X} \to \mathbb{R}^+$ two measurable functions.
A family of distribution which have a p.d.f. w.r.t. μ of the following form
$$x \mapsto q(\eta; x) = h(x)\exp(\eta T(x) - A(\eta)), \quad x \in \mathsf{X}, \tag{1}$$
where $A(\eta)$ is defined by:
$$A(\eta) = \log \int h(x) \exp(\eta T(x))\, \mu(dx) \tag{2}$$
is called a canonical exponential model associated to the couple (T, h). The *natural parameter set* of the canonical family associated to (T, h) is the set
$$\Xi = \{\eta \in \mathbb{R} \mid |A(\eta)| < \infty\}.$$

1. Prove that the exponential family with density

$$x \mapsto p(\eta; x) = \eta \exp(-\eta x) \mathbb{1}_{\mathbb{R}^+}(x)$$

defines a canonical exponential model on $\mathsf{X} = \mathbb{R}$. Give the natural parameter set.

2. Prove that the normal distribution $\mathcal{N}(\eta, 1)$ is a canonical exponential model. Give the natural parameter set.

We now prove several properties of these canonical models.

3 Show that the natural parameter set is a convex subset of $\mathbb{R}$.

Let us assume in the following that Ξ is an open interval.

Let g be a measurable function such that for all $\eta \in \Xi$,

$$\int |g(x)| \exp(\eta T(x)) \mu(dx) < \infty .$$

For $\eta \in \Xi$, let

$$G(\eta) = \int g(x) \exp(\eta T(x)) \mu(dx) .$$

We admit that G is infinitely differentiable on Ξ and, for all $k \in \mathbb{N}^*$ and $\eta \in \Xi$,

$$G^{(k)}(\eta) = \int g(x) T^k(x) \exp(\eta T(x)) \mu(dx) .$$

Let $(X_1, \ldots, X_n)$ be a n-sample of the canonical exponential model associated with (T, h).

4 Give an estimator of η by the method of moments.

5 Give the MLE. What do you notice?

Exercise 14. General exponential models
More generally, let $T : \mathsf{X} \to \mathbb{R}$, $h : \mathsf{X} \to \mathbb{R}^+$, $\varphi : \Theta \to \Xi$ and $B : \Theta \to \mathbb{R}^+$. The family of probability distributions, parametrized by $\theta \in \Theta$, whose p.d.f. are given by

$$x \mapsto p(\theta; x) = h(x) \exp(\varphi(\theta) T(x) - B(\theta))$$

w.r.t. a σ-finite measure on $\mathbb{R}$ or $\mathbb{N}$, is called an exponential family.
We assume that Ξ, the set of natural parameters, is an open subset of $\mathbb{R}$.

1. Show that the Poisson distribution is an exponential model with

$$\varphi(\theta) = \log(\theta), \quad B(\theta) = \theta, \quad T(x) = x, \quad h(x) = 1/x!.$$

2. Show that the binomial distribution is an exponential model with

$$\varphi(\theta) = \log\left(\frac{\theta}{1-\theta}\right), \quad B(\theta) = -n \log(1-\theta), \quad T(x) = x, \quad h(x) = \binom{n}{x}.$$

We assume that $\xi = \Theta$ and the function $\varphi : \Xi \to \Xi$ is a diffeomorphism. We are provided with a $n-$sample $X_1, \ldots, X_n$ with density $(p_\theta, \theta \in \Xi)$

3 Compute the MLE of θ [use the canonical parametrisation].