Solution to the Exam

November 2020

Exercise 1

1. The c.d.f of a single Pareto distribution is given by :

$$P(X \leqslant x) = \begin{cases} 0 & \text{if } x \leqslant a \\ 1 - \left(\frac{a}{x}\right)^\theta & \text{if } x > a \end{cases}$$

Therefore, the p.d.f is given by its derivative :

$$p_\theta(x) = \frac{d\ P(X \leqslant x)}{dx} = \begin{cases} 0 & \text{if } x \leqslant a \\ \frac{\theta a^\theta}{x^{\theta+1}} & \text{if } x > a \end{cases}$$

2. The density of the vector $X = (X_1,\ ...,\ X_n)$ of n i.i.d. observations is given by the product of each density :

$$p_\theta(x) = \prod_{i=1}^n p_\theta(x_i) = \theta^n \frac{a^{n\theta}}{x_i^{(\theta+1)}} 1_{\{x>a\}}$$

3. The maximum likelihood estimator $\widehat{\theta}_{ML}(x)$ of θ is :

$$\widehat{\theta}_{ML}(x) = \arg\max_{\theta \in \Theta} \log p_\theta(x)$$

Now,

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

$$\forall x > a,\ \frac{\partial \log p_\theta}{\partial \theta}(x) = \frac{n}{\theta} + n \log a - \sum_{i=1}^n \log x_i$$

As such, the ML estimator verifies :

$$\frac{n}{\widehat{\theta}_{ML}(x)} - \sum_{i=1}^n \log \left(\frac{x_i}{a}\right) = 0$$

Thus :

$$\widehat{\theta}_{ML}(x) = \frac{n}{\sum_{i=1}^n \log \left(\frac{x_i}{a}\right)}$$

4. According to the strong law of large numbers, with $X = (X_1,\ ...,\ X_n)$ being i.i.d. samples :

$$\widehat{\theta}_{ML}(X) = \frac{n}{\sum_{i=1}^n \log \left(\frac{X_i}{a}\right)} \xrightarrow[n\to\infty]{\text{a.s.}} \frac{1}{\mathbb{E}\left[\log \left(\frac{X_i}{a}\right)\right]}$$

And :

$$\mathbb{E}\left[\log \left(\frac{X_i}{a}\right)\right] = \int_a^\infty \log \frac{x}{a}\ p_\theta(x) dx = \int_0^\infty y\theta e^{-\theta y} dy = \frac{1}{\theta}$$

(Expected value of a random variable following the exponential distribution of parameter θ).

$$\widehat{\theta}_{ML}(X) \xrightarrow[n\to\infty]{\text{a.s.}} \theta$$

Exercise 2

1.

$$X - \theta \sim \mathcal{E}(1)$$

$$\mathbb{E}[X-\theta] = 1 = \mathbb{E}[X] - \theta$$

Using the method of moments :

$$\mathbb{E}[X] = 1 + \theta \approx \frac{1}{n}\sum_{i=1}^{n} x_i$$

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

2. Let us calculate the quadratic risk of this estimator.

$$\mathbb{E}[\theta(X)] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] - 1 = \mathbb{E}[X_1] - 1 = \theta$$

Thus the estimator is unbiased.

$$R(\theta,\ \widehat{\theta}) = \mathrm{Var}(\widehat{\theta}(X)) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{\mathrm{Var}(X_1)}{n} = \frac{\mathrm{Var}(X_1-\theta)}{n} = \frac{1}{n}$$

$$R(\theta,\ \widehat{\theta}) = \frac{1}{n}$$

3. Let us define

$$\forall x \in \mathbb{R}^n, \tilde{\theta}(x) = \min_{1 \leqslant i \leqslant n} x_i$$

$$Y = \tilde{\theta}(X)$$

$$\forall y \geqslant 0,\ P(Y > y) = \prod_{i=1}^{n} P(X_i - \theta > y - \theta) = \prod_{i=1}^{n}\int_{y-\theta}^{\infty} e^{-t}dt = \prod_{i=1}^{n} e^{-y+\theta} = e^{-n(y-\theta)}$$

$$P(Y > y + \theta) = P(Y - \theta > y) = e^{-ny}$$

Thus

$$Y - \theta \sim \mathcal{E}(n)$$

4.

$$\mathbb{E}\left[\tilde{\theta}(X) - \theta\right] = \frac{1}{n}$$

as it follows an exponential distribution of parameter n. Thus, the estimator is biased.

5.

$$\mathrm{Var}(Y) = \mathrm{Var}(Y-\theta) = \frac{1}{n^2}$$

$$R\left(\theta, \tilde{\theta}\right) = b\left(\theta, \tilde{\theta}\right)^2 + \mathrm{Var}\left(\tilde{\theta}(X)\right) = \frac{2}{n^2}$$

6. $\tilde{\theta}$ is better than $\widehat{\theta}$ in terms fo quadratic risk for all $n \geq 2$.

Exercise 3

1. Let $x \in \mathbb{N}^n$.

$$\pi(\theta|x) \propto \pi(\theta)p_\theta(x) \propto 1_{(0,1)}(\theta)\theta^n(1-\theta)^{S-n} \quad S = \sum_{i=1}^{n} x_i$$

Thus

$$\pi(\cdot|x) \sim \mathrm{Beta}\left(n+1, S-n+1\right)$$

2.

$$\mathbb{E}[\theta|X = x] = \frac{n+1}{2+S} = \widehat{\theta}$$

3. With the law of large numbers :

$$\widehat{\theta} \xrightarrow[n\to\infty]{\text{a.s.}} \frac{1}{\mathbb{E}[X]} = \theta$$

Exercice 4

1.

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta = \theta_1$$

The Neyman-Pearson test of level α is :

$$\forall x \in \mathbb{R}^n,\ \delta(x) = 1_{\left\{\frac{p_{\theta_1}(x)}{p_{\theta_0}(x)} > c\right\}}$$

Now :

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

$$\frac{p_{\theta_1}(x)}{p_{\theta_0}(x)} = e^{-\frac{1}{2\sigma^2}\sum_i\left[(x_i-\theta_1)^2-(x_i-\theta_0)^2\right]} \propto e^{\frac{1}{2\sigma^2}\left[2(\theta_1-\theta_0)\sum_i x_i\right]}.$$

This is an increasing function of $\bar{x} = \frac{1}{n}\sum_i x_i$ so the test is of the form :

$$\delta(x) = 1_{\bar{x}>c}.$$

It remains to find the constant c. For this, we use the fact that $\bar{X} \sim \mathcal{N}(\theta, \frac{\sigma^2}{n})$, that is $\bar{X} = \frac{\sigma}{\sqrt{n}}Z + \theta$ with $Z \sim \mathcal{N}(0,1)$. We get :

$$P_{\theta_0}(\delta(X) = 1) = P_{\theta_0}(\bar{X} > c) = P(Z > \frac{\sqrt{n}}{\sigma}(c - \theta_0)).$$

For a test at level α,

$$P_{\theta_0}(\delta(X) = 1) = \alpha$$

so that :

$$\frac{\sqrt{n}}{\sigma}(c - \theta_0) = Q(1-\alpha).$$

We obtain :

$$c = \theta_0 + \frac{\sigma}{\sqrt{n}}Q(1-\alpha).$$

2. Using again the fact that $\bar{X} = \frac{\sigma}{\sqrt{n}}Z + \theta$:

$$P_\theta(\bar{X} > c) = P(Z > \frac{\sqrt{n}}{\sigma}(c - \theta)),$$

so that

$$\sup_{\theta\le 0} P_\theta(\bar{X} > c) = P(Z > \frac{\sqrt{n}}{\sigma}c).$$

For $\alpha = 5\%$, we get $\frac{\sqrt{n}}{\sigma}c = Q(1-\alpha)$, that is $c = \frac{\sigma}{\sqrt{n}}Q(1-\alpha) \approx 0.164$.

3. When $\theta = \theta_1 \approx 0.176$, the type II error rate is

$$P_{\theta_1}(\delta(X) = 0) = P_{\theta_1}(\bar{X} < c) = P(Z < \frac{\sqrt{n}}{\sigma}(c - \theta_1)).$$

With $c \approx 0.164$ and $\theta \approx 0.176$, we get $P(Z < -0.001) \approx 0.5$.

4. The test becomes $\delta(x) = 1_{\bar{x}<c}$, with

$$\sup_{\theta \geq 0} P_\theta(\bar{X} < c) = \alpha.$$

We get :

$$P(Z < \frac{\sqrt{n}}{\sigma} c) = \alpha,$$

that is $\frac{\sqrt{n}}{\sigma} c = Q(\alpha)$ and $c = \frac{\sigma}{\sqrt{n}} Q(\alpha) \approx -0.164$.

Exercise 5

1. Since $\bar{X} = \frac{\sigma}{\sqrt{n}} Z + \theta$, we need to find t such that :

$$P(-t \leq Z \leq t) = 1 - \alpha.$$

By symmetry of the distribution of Z, this means $t = Q(1 - \frac{\alpha}{2})$. Then,

$$P(\theta \in [\bar{X} - c, \bar{X} + c]) = 1 - \alpha,$$

with $c = t \frac{\sigma}{\sqrt{n}} \approx 0.258$.

2. Upper confidence bound :

$$P(Z \geq -t) = 1 - \alpha,$$

for $t = Q(\alpha)$ and

$$P(\theta \leq \bar{X} + c) = 1 - \alpha,$$

with $c = t \frac{\sigma}{\sqrt{n}} \approx 0.233$.

Acknowledgement

Many thanks to Huu Tan Mai for writing the first version of this document !

Appendix

Quantiles of the standard normal distribution

The following table gives some approximate values of quantiles of the standard normal distribution.

x	0.0	0.25	0.52	0.84	1.28	1.64	2.33	2.58
$\mathbb{P}(X \leq x)$	0.5	0.6	0.7	0.8	0.9	0.95	0.99	0.995