MAP531 : Statistics
Mid-term exam

Stéphanie Allassonnière, Wei Jiang, Geneviève Robin, Zoltan Szabo and Elodie Vernet

General instructions :
— You have **two hours** to solve these exercises.
— The two exercises are independent from each other.
— You are allowed to use your handwritten notes from the class and PC sessions as well as the material that was provided to you on moodle.
— Usage of **cell/smart phones and laptops** is **forbidden**.
— If you have to go to **bathroom**, please do it **now**.
— Generally, it is a good idea to start working *after* reading through all the exercises (your unconscious mind is solving them in the background ⇒ you are more efficient). Good luck !

EXERCISE 1 -Geometric distribution

Let $p \in]0,1[$. We say that the random variable X follows a geometric law with parameter p (we write it as $X \sim Geom(p)$) if X takes values in $\{1, 2, \ldots\}$ and

$$P(X = k) = (1-p)^{k-1}p \quad (k = 1, 2, \ldots). \tag{1}$$

This distribution satisfies :

$$\begin{cases} \mathbb{E}[X] & = & 1/p \\ \mathbb{V}ar[X] & = & \frac{1-p}{p^2}\,. \end{cases} \tag{2}$$

1. Let $X_1, ..., X_n$ be an i.i.d. sample of geometric random variable with unknown parameter p. Using the method of moment, compute two estimators $\hat{p}_n^{MOM_1}$ and $\hat{p}_n^{MOM_2}$ of p.
2. Compute the maximum likelihood estimate, $\hat{p}_n^{MLE}$ of p ; check the 1st (derivative) and 2nd-order (Hessian) optimality conditions.
3. Is it easy to compute the MSE of these three estimators ? Why ? Propose a pseudo-code to approximate this MSE.

Let now $\alpha \in]1, +\infty[$ and Y a discrete random variable such that :

$$P(Y = k) = \left(1 - \frac{1}{\alpha}\right)^{k-1} \frac{1}{\alpha} \quad (k = 1, 2, \ldots). \tag{3}$$

4. Give the expression of $\mathbb{E}[Y]$ and $\mathbb{V}ar[Y]$.
5. Compute the maximum likelihood estimate, $\hat{\alpha}_n^{MLE}$ of α ; check the 1st and 2nd-order optimality conditions.
6. Compute the MSE of $\hat{\alpha}_n^{MLE}$.
7. What can you say about this MSE when $n \to \infty$?

We are now considering the Bayesian framework going back to $Geom(p)$. We recall that the $Beta(a, b)$ distribution for $a > 0$ and $b > 0$ is a probability distribution on the real line with p.d.f. given by :

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}\mathbb{1}_{(0,1)}(x)\,. \tag{4}$$

Remember that the expectation of a $Beta(a, b)$ distribution is

$$\frac{a}{a+b}. \tag{5}$$

8. Assuming that the prior on p is $Beta(a, b)$, what is the posterior probability distribution of p given X ?
9. Compute the posterior mean. What happens to the posterior mean as $n \to \infty$?
10. Compute the MAP estimator : $\hat{p}_n^{MAP} = \arg\max_p P(p|X_1, \dots, X_n)$; check the 1st and 2nd-order optimality conditions. What happens to the MAP estimator as $n \to \infty$?

SOLUTION :

1. Method of moment estimator based on
   (a) $\mathbb{E}[X] = 1/p$: By expressing parameter p as a function of the expectation, and replacing the expectation with empirical average one gets

   $$\mathbb{E}[X] = 1/p \quad \Rightarrow \quad p = \frac{1}{\mathbb{E}[X]} \quad \Rightarrow \quad \hat{p}_n^{MOM_1} = \frac{1}{\frac{1}{n}\sum_{i=1}^n X_i} = \frac{1}{\bar{X}_n},$$

   using the $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ shorthand.

   (b) $\mathbb{V}ar[X] = \frac{1-p}{p^2}$: Similarly to the previous bullet

   $$\mathbb{V}ar[X] = \frac{1-p}{p^2} \quad \Rightarrow \mathbb{V}ar[X]p^2 + p - 1 = 0 \quad \Rightarrow \quad p_{1,2} = \frac{-1 \pm \sqrt{1 + 4\mathbb{V}ar[X]}}{2\mathbb{V}ar[X]} \Rightarrow$$
   $$p = \frac{\sqrt{1 + 4\mathbb{V}ar[X]} - 1}{2\mathbb{V}ar[X]} \quad \Rightarrow \quad \hat{p}_n^{MOM_2} = \frac{\sqrt{1 + \frac{4}{n}\sum_{i=1}^n \left(X_i - \bar{X}_n\right)^2} - 1}{\frac{2}{n}\sum_{i=1}^n \left(X_i - \bar{X}_n\right)^2}$$

   by discarding the solution on the '$-$' thread (since p has to be positive).
2. $\hat{p}_n^{MLE}$ maximizes the likelihood (LL) function or equivalently its logarithm (since LL> 0)

   $$\hat{p}_n^{MLE} = \arg\max_p L(X_1, \dots, X_n; p)$$

   where

   $$L(X_1, \dots, X_n; p) := \prod_{i=1}^n \left[(1-p)^{X_i - 1} p\right] = (1-p)^{\left(\sum_{i=1}^n X_i\right) - n} p^n \tag{6}$$

   So that

   $$\ln\left[L(X_1, \dots, X_n; p)\right] = c\ln(1-p) + n\ln(p), \text{ with } c := \left(\sum_{i=1}^n X_i\right) - n,$$
   $$0 = \frac{\partial \ln\left[L(X_1, \dots, X_n; p)\right]}{\partial p} = \frac{-c}{1-p} + \frac{n}{p} \quad \Leftrightarrow \quad \frac{c}{1-p} = \frac{n}{p} \quad \Leftrightarrow \quad cp = n - np$$
   $$\Leftrightarrow (c+n)p = n \quad \Rightarrow \quad \hat{p}_n^{MLE} = \frac{n}{c+n} = \frac{n}{\left(\sum_{i=1}^n X_i\right) - n + n} = \frac{n}{\sum_{i=1}^n X_i} = \frac{1}{\bar{X}_n}.$$

   Since $X_i \geq 1$ so that $c \geq 0$, for all $p \in (0, 1)$

   $$\frac{\partial^2 \ln\left[L(X_1, \dots, X_n; p)\right]}{\partial^2 p} = -\frac{c}{(1-p)^2} - \frac{n}{p^2} \quad \leq 0,$$

   the obtained p candidate is the unique maximum because the log-likelihood is concave.

3. MSE : $\hat{p}_n^{MLE} = \hat{p}_n^{MOM_1}$ looks somewhat 'nicer' (no square root) than $\hat{p}_n^{MOM_2}$, but still one should be able to compute

$$\text{MSE}\left(\hat{p}_n^{MLE}\right) = \mathbb{E}\left(\frac{1}{\frac{1}{n}\sum_{i=1}^{n} X_i} - p\right)^2.$$

However, the average shows up in the nominator, so it is not easy to check the unbiasedness of the estimator or to compute its MSE. To approximate the MSE MSE$(\hat{p}, p)$ of an estimator $\hat{p}(X_1, \ldots, X_n)$ at p we can use the Monte Carlo method. Then the MSE MSE$(\hat{p}, p) = \mathbb{E}_p\left((\hat{p}(X_1, \ldots, X_n) - p)^2\right)$ can be approximated by

$$\frac{1}{K}\sum_{j=1}^{K}\left(\hat{p}(X_1^j, \ldots, X_n^j) - p\right)^2,$$

where $(X_1^j, \ldots, X_n^j)$, $j \le K$ are K samples from $\mathbb{P}_p$. In a pseudo-code :

Algorithm 1 Aproximation of MSE

Input: n, K, p

Output: mse, approximation of the MSE of $\hat{p}(X_1, \ldots, X_n)$ at p

```
mse ← 0
for j ← 1 to K do
    (x_1, ..., x_n) ← n-sample of a Geometric G(p)
    p̂ ← p̂(x_1, ..., x_n)
    error^2 ← (p̂ − p)^2
    mse ← mse + error^2
end for
mse ← mse/K
return mse
```

4. Since (1) and (3) only differ in parameterization ($p = \frac{1}{\alpha}$) the asked expectation and variance are

$$\mathbb{E}[X] = \alpha, \qquad \mathbb{V}ar[X] = \frac{1-\frac{1}{\alpha}}{\frac{1}{\alpha^2}} = \frac{\alpha-1}{\alpha}\alpha^2 = (\alpha-1)\alpha.$$

5. Following the derivation at the p-parameterization, we have

$$\begin{aligned}
\ln\left[L(X_1, \ldots, X_n; \alpha)\right] &= c\ln\Big(\underbrace{1-\frac{1}{\alpha}}_{\frac{\alpha-1}{\alpha}}\Big) + n\ln\left(\frac{1}{\alpha}\right), \text{ with } c := \left(\sum_{i=1}^{n} X_i\right) - n, \\
\frac{\partial \ln\left[L(X_1, \ldots, X_n; \alpha)\right]}{\partial \alpha} &= c\frac{\alpha}{\alpha-1}\frac{1}{\alpha^2} - \frac{n}{\alpha} = \frac{c}{(\alpha-1)\alpha} - \frac{n}{\alpha} = 0 \quad \Leftrightarrow \quad c = n(\alpha-1) \Rightarrow \\
\hat{\alpha}_n^{MLE} &= \frac{c}{n} + 1 = \frac{1}{n}\sum_{i=1}^{n} X_i, \\
\frac{\partial^2 \ln\left[L(X_1, \ldots, X_n; \alpha)\right]}{\partial^2 \alpha} &= -c\frac{2\alpha-1}{[\alpha(\alpha-1)]^2} + \frac{n}{\alpha^2} < 0 \quad \Leftrightarrow \quad n(\alpha-1)^2 < c(2\alpha-1) \quad \Leftrightarrow \quad \frac{(\alpha-1)^2}{2\alpha-1} < \frac{c}{n} \\
&\Leftrightarrow \frac{(\alpha-1)^2}{2\alpha-1} + 1 < \bar{X}_n, \text{ with } \alpha = \bar{X}_n, \\
&\Leftrightarrow \frac{(\alpha-1)^2}{2\alpha-1} + 1 < \alpha \quad \Leftrightarrow \quad (\alpha-1)^2 < (\alpha-1)(2\alpha-1) \quad \Leftrightarrow \quad \alpha - 1 < 2\alpha - 1 \\
&\Leftrightarrow 0 < \alpha = \bar{X}_n \quad [\Leftarrow X_i > 0, \forall i].
\end{aligned}$$

Hence, $\hat{\alpha}_n^{MLE} = \bar{X}_n$ is a local maximum of the log-likelihood. As the log-likelihood tends to $-\infty$ when α goes to 1 and $+\infty$ it is the global maximum and $\bar{X}_n$ is the unique ML estimator.

6. Since $\mathbb{E}[\bar{X}_n] = \mathbb{E}[X_1] = \alpha$, the ML estimator is unbiased. As a result

$$\mathrm{MSE}\left(\bar{X}_n\right) = \mathrm{var}\left(\bar{X}_n\right) = \frac{1}{n}\mathrm{var}(X_1) = \frac{(\alpha-1)\alpha}{n}.$$

7. $\mathrm{MSE}\left(\bar{X}_n\right) = \frac{(\alpha-1)\alpha}{n} \xrightarrow{n\to\infty} 0.$
8. The posterior distribution of p is proportional to the product of the likelihood and the prior. Using the likelihood expression already computed in (6)

$$\begin{aligned}\pi(p|X_1,\ldots,X_n) &\propto L(X_1,\ldots,X_n|p)\pi(p) \propto (1-p)^{\left(\sum_{i=1}^n X_i\right)-n} p^n p^{a-1}(1-p)^{b-1}\mathbb{1}_{(0,1)}(p) \\ &= p^{n+a-1}(1-p)^{\left(\sum_{i=1}^n X_i\right)-n+b-1}\mathbb{1}_{(0,1)}(p),\end{aligned}$$

hence the posterior is $Beta\left(n+a, \sum_{i=1}^n X_i - n + b\right)$.
9. Thanks to the posterior obtained in the previous point and (5), the posterior mean is

$$\frac{n+a}{(n+a)+\left(\sum_{i=1}^n X_i - n + b\right)} = \frac{n+a}{a+b+\sum_{i=1}^n X_i} = \frac{1+\frac{a}{n}}{\frac{a+b}{n}+\frac{1}{n}\sum_{i=1}^n X_i}.$$

Thus the difference of the posterior mean and $\hat{p}_n^{MLE} = \frac{1}{\bar{X}_n}$ becomes arbitrarily small as $n \to \infty$.
10. We got that the posterior is $Beta\left(n+a, \sum_{i=1}^n X_i - n + b\right)$. By (4), the MAP estimator is the solution of

$$\max_p \left[c p^{n+a-1}(1-p)^{\sum_{j=1}^n X_j - n + b - 1}\right], \tag{7}$$

where $c = \frac{\Gamma\left(a+b+\sum_{i=1}^n X_i\right)}{\Gamma(n+a)\Gamma\left(\sum_{i=1}^n X_i - n + b\right)} > 0$. (7) is equivalent to maximize F with respect to p where

$$J(p) = \underbrace{(n+a-1)}_{=:e}\ln(p) + \underbrace{\left(\sum_{j=1}^n X_j - n + b - 1\right)}_{=:f}\ln(1-p).$$

$$\begin{aligned}\frac{\partial J}{\partial p} &= \frac{e}{p} - \frac{f}{1-p} = 0 \quad \Leftrightarrow \quad e(1-p) = fp \quad \Leftrightarrow \quad e = (e+f)p \quad \Rightarrow \\ \hat{p}_n^{MAP} &= \frac{e}{e+f} = \frac{n+a-1}{(n+a-1)+\left(\sum_{j=1}^n X_j - n + b - 1\right)} = \frac{n+a-1}{(a-1)+(b-1)+\sum_{j=1}^n X_j} \\ &= \frac{1+\frac{a-1}{n}}{\frac{(a-1)+(b-1)}{n}+\frac{1}{n}\sum_{j=1}^n X_j}, \\ \frac{\partial^2 J}{\partial^2 p} &= -\left[\frac{e}{p^2} + \frac{f}{(1-p)^2}\right] < 0.\end{aligned}$$

Hence $\hat{p}_n^{MAP}$ is the unique MAP estimator and the difference between $\hat{p}_n^{MAP}$ and $\hat{p}_n^{MLE} = \frac{1}{\bar{X}_n}$ becomes arbitrarily small as $n \to \infty$.

EXERCISE 2 -

In this exercise, we are interested in modeling the size of oil fields of a hydrocarbon basin in operation. It is assumed that the discovered field sizes (denoted by $R_1, ..., R_n$) are independent and identically distributed. Oil fields that are too small are not visible to the exploration campaigns and they are therefore not found. Hence we model the size of discovered oil fields by a lower-truncated Pareto law having density :

$$f_{\alpha,\eta}(x) = Kx^{-\alpha-1}\mathbb{1}_{x\geq\eta}. \tag{8}$$

1. Compute K when $\alpha > 0$.
2. Discuss the existence of expectation and variance of this distribution.

3. Write down the likelihood of this model.
4. Compute the MLE of both parameters $\theta = (\alpha, \eta)$.

A random variable X follows a lognormal distribution of parameters μ and σ^2 if $\log(X)$ follows the normal distribution $\mathcal{N}(\mu, \sigma^2)$. Its p.d.f. is

$$\frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{[\ln(x)-\mu]^2}{2\sigma^2}} \mathbb{1}_{(0,+\infty)}(x).$$

Instead of the Pareto distribution, assume that R_i-s are i.i.d. from a lognormal distribution with parameters μ and σ^2.

5. Is this equality correct : $\mathbb{E}[R] = \exp(\mu)$? Why ? (Hint : You can reduce the computation to the moment-generating function $M_U(t) = \mathbb{E}\left[e^{tU}\right]$, where $U \sim N(0,1)$).
6. Compute the ML estimator of $\theta = \left(\mu, \sigma^2\right)$; checking 1st-order optimality condition is enough.

SOLUTION :

1. The K constant is such that

$$1 = \int_{\mathbb{R}} f_{\alpha,\eta}(x)\mathrm{d}x = \int_{\eta}^{\infty} Kx^{-\alpha-1}\mathrm{d}x = K\left[\frac{x^{-\alpha}}{-\alpha}\right]_{x=\eta}^{x=+\infty} = \frac{K}{-\alpha}\left(0 - \frac{1}{\eta^\alpha}\right) = \frac{K}{\alpha\eta^\alpha} \quad \Rightarrow \quad K = \alpha\eta^\alpha.$$

2. The variance exists iff the 2nd moment of the random variable exists since $\mathrm{var}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X)$. Thus, the existence of the expectation and variance is equivalent to

$$\infty > \int_{\mathbb{R}} x f_{\alpha,\eta}(x)\mathrm{d}x = \int_{\eta}^{\infty} Kxx^{-\alpha-1}\mathrm{d}x = \int_{\eta}^{\infty} Kx^{-\alpha}\mathrm{d}x \qquad \Leftrightarrow -\alpha < -1 \qquad \Leftrightarrow \alpha > 1,$$
$$\infty > \int_{\mathbb{R}} x^2 f_{\alpha,\eta}(x)\mathrm{d}x = \int_{\eta}^{\infty} Kx^2x^{-\alpha-1}\mathrm{d}x = \int_{\eta}^{\infty} Kx^{-\alpha+1}\mathrm{d}x \qquad \Leftrightarrow -\alpha+1 < -1 \qquad \Leftrightarrow \alpha > 2,$$

respectively.

3. The likelihood is

$$\begin{aligned} L(X_1, \ldots, X_n; \alpha, \eta) &= \prod_{i=1}^n \left[KX_i^{-\alpha-1}\mathbb{1}_{X_i \geq \eta}\right] = K^n \left(\prod_{i=1}^n X_i\right)^{-\alpha-1} \prod_{i=1}^n \mathbb{1}_{X_i\geq\eta} \\ &= K^n \left(\prod_{i=1}^n X_i\right)^{-\alpha-1} \mathbb{1}_{\min(X_1,\ldots,X_n)\geq\eta}. \end{aligned}$$

4. Using the previous point and that $K = \alpha\eta^\alpha$, the likelihood is

$$L(X_1, \ldots, X_n; \alpha, \eta) = \begin{cases} \alpha^n \eta^{n\alpha} \left(\prod_{i=1}^n X_i\right)^{-\alpha-1} (> 0) & \text{if } \min(X_1, \ldots, X_n) \geq \eta, \\ 0 & \text{otherwise.} \end{cases}$$

Since our goal is to maximize the LL we can restrict our attention to $\min(X_1, \ldots, X_n) \geq \eta$ and maximize the following quantity with respect to $\alpha > 0$ and $\eta > 0$

$$\ln\left[L(X_1, \ldots, X_n; \alpha, \eta)\right] = n\ln(\alpha) + n\alpha\ln(\eta) - (\alpha+1)\sum_{i=1}^n \ln(X_i).$$

$\eta \mapsto \ln\left[L(X_1, \ldots, X_n; \alpha, \eta)\right]$ is monotonically increasing, we have to satisfy the $\eta \leq \min(X_1, \ldots, X_n)$ constraint, so we choose $\hat{\eta}_n^{MAP} = \min(X_1, \ldots, X_n)$.

$$\begin{cases} \frac{\partial \ln[L(X_1,\ldots,X_n;\alpha,\eta)]}{\partial\alpha} = \frac{n}{\alpha} + n\ln(\eta) - \sum_{i=1}^n \ln(X_i) = 0 \\ \frac{\partial^2 \ln[L(X_1,\ldots,X_n;\alpha,\eta)]}{\partial\alpha^2} = \frac{-n}{\alpha^2} < 0 \text{ for all } \alpha > 0 \end{cases} \quad \Rightarrow \quad \hat{\alpha}_n^{MAP} = \frac{n}{\left[\sum_{i=1}^n \ln(X_i)\right] - n\ln(\eta)}.$$

5. $\log(R) = N\left(\mu,\sigma^2\right) \Leftrightarrow R = e^{N\left(\mu,\sigma^2\right)}$. Unfortunately, the $x \mapsto e^x$ mapping is *not* linear so we expect that the $\mathbb{E}[R] = e^\mu$ relation does not hold. Formally, let $U \sim N(0,1)$ and let $M_U(t) = \mathbb{E}\left[e^{tU}\right]$ be the moment-generating function of U at point t.

$$M_U(t) = \mathbb{E}\left[e^{tU}\right] = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{tu}e^{-\frac{u^2}{2}}\mathrm{d}u \overset{(*)}{=} e^{\frac{t^2}{2}}\int_{\mathbb{R}} \underbrace{\frac{1}{\sqrt{2\pi}}e^{-\frac{(t-u)^2}{2}}}_{\text{pdf of } N(t,1) \text{ at point } u} \mathrm{d}u = e^{\frac{t^2}{2}} \times 1 = e^{\frac{t^2}{2}},$$

$$\mathbb{E}[R] = \mathbb{E}\left[e^{\mu+\sigma U}\right] = e^\mu \mathbb{E}\left[e^{\sigma U}\right] = e^\mu M_U(\sigma) = e^\mu e^{\frac{\sigma^2}{2}} = e^{\mu+\frac{\sigma^2}{2}},$$

where in (*) we used $tu - \frac{u^2}{2} = -\frac{1}{2}\left(u^2 - 2tu\right) = -\frac{1}{2}\left(u^2 - 2tu + t^2 - t^2\right) = -\frac{1}{2}\left[(u-t)^2 - t^2\right]$. Consequently, $\mathbb{E}[R] = e^\mu \Leftrightarrow \frac{\sigma^2}{2} = 0 \Leftrightarrow \sigma = 0$; so $\mathbb{E}[R] = e^\mu$ can not hold since $\sigma > 0$.

6. We compute MLE by maximizing the likelihood :

$$\begin{aligned}
L(X_1,\ldots,X_n;\mu,\sigma) &= \prod_{i=1}^{n}\left[\frac{1}{X_i\sigma\sqrt{2\pi}}e^{-\frac{[\ln(X_i)-\mu]^2}{2\sigma^2}}\mathbb{1}_{(0,+\infty)}(X_i)\right] \\
&= (2\pi)^{-n/2}\sigma^{-n}\prod_{i=1}^{n}\frac{1}{X_i}e^{-\frac{[\ln(X_i)-\mu]^2}{2\sigma^2}}\mathbb{1}_{(0,+\infty)}(X_i), \\
\ln\left[L(X_1,\ldots,X_n;\mu,\sigma)\right] &= \text{constant } - n\ln(\sigma) - \sum_{i=1}^{n}\frac{[\ln(X_i)-\mu]^2}{2\sigma^2}, \\
\frac{\partial \ln\left[L(X_1,\ldots,X_n;\mu,\sigma)\right]}{\partial\mu} &= 2\sum_{i=1}^{n}\frac{[\ln(X_i)-\mu]}{2\sigma^2} = 0 \quad\Leftrightarrow\quad \mu = \frac{1}{n}\sum_{i=1}^{n}\ln(X_i), \\
\frac{\partial \ln\left[L(X_1,\ldots,X_n;\mu,\sigma)\right]}{\partial\sigma} &= -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}\left[\ln(X_i)-\mu\right]^2 = 0 \quad\Leftrightarrow\quad n\sigma^2 = \sum_{i=1}^{n}\left[\ln(X_i)-\mu\right]^2 \\
&\Rightarrow \sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[\ln(X_i)-\mu\right]^2}.
\end{aligned}$$

The resulting estimators are

$$\hat{\mu}_n^{\text{MLE}} = \frac{1}{n}\sum_{i=1}^{n}\ln(X_i), \qquad \hat{\sigma}_n^{\text{MLE}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[\ln(X_i) - \frac{1}{n}\sum_{j=1}^{n}\ln(X_j)\right]^2}.$$