Exercise sheet 4

Exercise 1 Consider independent Gaussian random variables $X_1, \dots, X_n, Y_1, \dots, Y_n$ with $X_j, Y_j \sim N(\mu_j, \sigma^2)$ with $\mu_1, \dots, \mu_j, \sigma^2$ unknown. Find the maximum likelihood estimator. Is it consistent?

Solution 1 The log likelihood is

$$\ell_n(\mu_1, \dots, \mu_n, \sigma^2) = -n \log 2\pi - n \log \sigma^2 - \tfrac{1}{2\sigma^2} \sum (X_j - \mu_j)^2 + (Y_j - \mu_j)^2.$$

For any σ^2, the derivative with respect to μ_j is $2(\mu_j - X_j + \mu_j - Y_j)$, increasing in μ_j, so the unique minimiser is $\mu_j = (X_j + Y_j)/2$. After plugging this into ℓ_n, we maximise a concave function over σ^2, and the maximiser is

$$(\widehat{\sigma}_n^2)^{MLE} = \tfrac{1}{4n} \sum (X_j - Y_j)^2.$$

But $X_j - Y_j \sim N(0, 2\sigma^2)$, so this has expectation $\sigma^2/2$ and thus converges almost surely to this value by the law of large numbers. Thus $(\widehat{\sigma}_n^2)^{MLE}$ is not consistent!

Exercise 2 Consider the geometric distribution with parameter $p \in (0,1]$ ($P(X=x) = (1-p)^x p$, $x = 0, 1, \dots$ and $\mathbb{E}_p X = (1-p)/p$).

Find the maximum likelihood estimator $\widehat{p}_n^{MLE}$ and show that it is biased, that is $\mathbb{E}_p \widehat{p}_n^{MLE} \neq p$. *Hint:* Subtract and add $\mathbb{E}_p \overline{X}_n = (1-p)/p$ in the denominator and use an inequality based on two terms of a geometric series.

Solution 2 The log likelihood is concave and $\widehat{p}_n^{MLE} = 1/(1+\overline{X}_n)$.

$$\mathrm{E}_p \tfrac{1}{1+\overline{X}_n} = \mathrm{E}_p \tfrac{p}{1+p(\overline{X}_n - \frac{1-p}{p})} > \mathrm{E}_p p[1 - p(\overline{X}_n - \tfrac{1-p}{p})] = p,$$

where we used the inequality $1/(1+u) \geq 1-u$ which holds for $u > -1$ (which is satisfied for $u = p(\overline{X}_n - \frac{1-p}{p})$) and which is strict for $u \neq 0$ (which is satisfied for $u = p(\overline{X}_n - \frac{1-p}{p})$ with positive probability).

Alternatively, one can use Jensen's inequality for the function $(1+x)^{-1}$.

Exercise 3 Let $X_1, \dots, X_n$ be a sample from the distribution with density

$$f(x;\theta,p) = (1-p)1_{(-1,0)}(x) + p\theta^{-1}1_{(0,\theta)}(x)$$

(a mixture of $U(-1,0)$ and $U(0,\theta)$), where $p \in [0,1]$ and $\theta > 0$. Find the maximum likelihood estimators of the parameters p and θ.

Solution 3 The likelihood equals

$$(1-p)^{\sum_{i=1}^n 1_{(-1,0)}(X_i)}(p\theta^{-1})^{n-\sum_{i=1}^n 1_{(-1,0)}(X_i)}1_{[X_{(n)}^+<\theta]}1_{[X_{(1)}>-1]}.$$

Consider two cases: (1) If $X_{(n)} := \max(X_1,\dots,X_n) > 0$, then the value $\widehat{\theta} = X_{(n)}$ maximises the likelihood independently of p. Then the estimator $\widehat{p}_n^{MLE}$ is obtained straightforwardly (binomial likelihood), $\widehat{p}_n^{MLE} = 1 - \sum_{i=1}^n 1_{(-1,0)}(X_i)/n$.

(2) If $X_{(n)}^+ \le 0$, then all the observations come only from one component of the mixture. The likelihood then equals $(1-p)^n$ which is maximized for $\widehat{p} = 0$. All the observations follow the uniform distribution on $(-1,0)$, which does not depend on θ, hence the sample contains no information about the parameter θ and nothing can be said about θ. All values of θ are equally likely, as the likelihood does not depend on θ.

Exercise 4 Find the asymptotic covariance of the maximum likelihood estimator from question 5 of last week.

Solution 4 The Fisher information matrix (in ites alternative form under regularity) is

$$J_1(\lambda,\gamma) = -\mathrm{E}_{\lambda,\gamma}\begin{pmatrix} -D/\lambda^2 & 0 \\ 0 & -(1-D)/\gamma^2 \end{pmatrix} = \begin{pmatrix} \frac{1}{\lambda(\lambda+\gamma)} & 0 \\ 0 & \frac{1}{\gamma(\lambda+\gamma)} \end{pmatrix}$$

(where we used the fact that D is Bernoulli distributed with parameter $\lambda/(\lambda+\gamma)$). Hence $\sqrt{n}(\widehat{\lambda}-\lambda,\widehat{\gamma}-\gamma)^\top$ is asymptotically bivariate normal with mean zero and covariance matrix

$$J_1(\lambda,\gamma)^{-1} = \begin{pmatrix} \lambda(\lambda+\gamma) & 0 \\ 0 & \gamma(\lambda+\gamma) \end{pmatrix}.$$

Alternatively, we can compute the covariance matrix of $S_1(\lambda,\gamma)$ whose components are $D/\lambda - T$ and $(1-D)/\gamma - T$ with T and D independent, so

$$I_1(\lambda,\gamma) = \mathbb{E}_{\lambda,\gamma}\begin{pmatrix} (D/\lambda-T)^2 & D(1-D)/\lambda\gamma - DT/\lambda - T(1-D)/\gamma + T^2 \\ D(1-D)/\lambda\gamma - DT/\lambda - T(1-D)/\gamma + T^2 & ((1-D)/\gamma - T)^2 \end{pmatrix}$$

$$= J_1(\lambda,\gamma).$$

Exercise 5 Consider the rescaled Beta$(1, \alpha + 1)$ distribution with known $\alpha > -1$ and density

$$f(x;\theta) = (\alpha+1)(\theta - x)^\alpha \theta^{-\alpha-1} 1(x \in [0,\theta]),$$

where $\theta > 0$ is unknown. Let $X_1, \dots, X_n \overset{iid}{\sim} f(x;\theta_0)$ for some $\theta_0 > 0$.

(a) Investigate the regularity conditions $\mathbb{E}_\theta S_n(\theta) = 0$ and $I_n(\theta) = J_n(\theta)$ as a function of $\alpha > -1$.

(b) It is not easy to show that $\widehat{\theta}_n^{MLE}$ is consistent. But show that $T_n \le \widehat{\theta}_n^{MLE} \le (\alpha+1)T_n$, where T_n is a consistent estimator.

Solution 5 The likelihood is 0 if $\theta < \max(X_1, \dots, X_n) := M_n$ and otherwise it is

$$\begin{aligned} \ell_n(\theta) &= n\log(\alpha+1) - n(\alpha+1)\log\theta + \alpha \sum \log(\theta - x_i) \\ \ell_n'(\theta) &= \alpha \sum \tfrac{1}{\theta - x_i} - (\alpha+1)\tfrac{n}{\theta} \\ \ell_n''(\theta) &= (\alpha+1)\tfrac{n}{\theta^2} - \alpha \sum \tfrac{1}{(\theta-x_i)^2} \end{aligned}$$

(a) For any $k \in \mathbb{R}$

$$\mathbb{E}_\theta(\theta - X)^k = \int_0^\theta (\alpha+1)(\theta_0 - x)^{\alpha+k}/\theta^{\alpha+1} dx = \begin{cases} \infty & \alpha + k \le -1 \\ \theta^k \frac{\alpha+1}{\alpha+k+1} & \alpha + k > -1 \end{cases}$$

Thus, if $\alpha > 0$ the regularity condition $\mathbb{E}_\theta S_n(\theta) = 0$ is satisfied. If $\alpha > 1$ then we can take $k = -2$ to get

$$J_n(\theta) = -\mathbb{E}_\theta \ell_n''(\theta) = \tfrac{(\alpha+1)n}{\theta^2(\alpha-1)} > 0$$

and

$$I_n(\theta) = \mathbb{E}_\theta S_n^2(\theta) = var_\theta S_n(\theta) = \tfrac{\alpha^2 n(\alpha+1)}{\theta^2(\alpha-1)} - n\alpha^2 \left(\tfrac{\alpha+1}{\alpha\theta}\right)^2 = J_n(\theta).$$

(b) If $\alpha \le 0$ then ℓ_n is decreasing in θ and the maximum likelihood estimator is $M_n = \max(X_1, \dots, X_n)$. If $\alpha > 0$ then $\ell_n \to -\infty$ as $\theta \to M_n$ or as $\theta \to \infty$, so the maximum is attained at a point where the derivative vanishes. Therefore for $\alpha > 0$,

$$\tfrac{1}{\widehat{\theta}_n^{MLE}} = \tfrac{\alpha+1}{\alpha}\tfrac{1}{n} \sum \tfrac{1}{\widehat{\theta}_n^{MLE} - x_i}.$$

This equation has no explicit solution when $n > 1$. However, if $\theta > (\alpha+1)M_n$ then for all i, $\theta - x_i \ge \theta - M_n > (1 - (\alpha+1)^{-1})\theta$ and

$$\ell_n'(\theta) < \alpha n \tfrac{\alpha+1}{\alpha\theta} - (\alpha+1)\tfrac{n}{\theta} = 0,$$

so θ cannot be the maximum likelihood estimator. Therefore $\widehat{\theta}_n^{MLE} \le (\alpha+1)M_n$. (This cannot be improved, since for $n = 1$, $\widehat{\theta}_n^{MLE} = (\alpha+1)x_1$.) Thus, we can choose $T_n = M_n$. It is easy to see that M_n is consistent for θ for all $\alpha > -1$. Indeed, on one hand $M_n \le \theta$ almost surely, and on the other hand for all $t \in [0, 1)$,

$$\mathbb{P}(M_n \le t/\theta) = (1 - (1-t)^\alpha))^n \to 0, \qquad n \to \infty.$$