Statistical Theory MATH-442
Spring 2025 Ho Yun & Yoav Zemel

Exercise sheet 5

In these exercise set, you may use without proof that $\Gamma'(p)/\Gamma(p)$ is strictly increasing in p with limit $-\infty$ as $p \searrow 0$ and ∞ as $p \to \infty$.

Exercise 1

1. Suppose that $X_n \xrightarrow{d} X$ on $\mathbb{R}^k$. Show that (X_n) is bounded in probability.

2. Show that if $X_n \xrightarrow{d} 0$ and Y_n is bounded in probability, then $X_n Y_n \xrightarrow{d} 0$. Here we assume that $X_n Y_n$ makes sense, so that either one of them is scalar, or X_n is a matrix and Y_n and appropriate vector, or X_n and Y_n are vectors of the same dimension and we take an inner product, etc.

Exercise 2 In the one-dimensional case, examine the conditions under which an estimator T attains the Cramér–Rao lower bound.

Solution 1

1. Let $\epsilon > 0$. It is possible to find $a_\epsilon \in \mathbb{R}^k$ such that $P(X \leq a_\epsilon) \leq \epsilon$ and b_ϵ is a continuity point of F_X (since the set of discontinuities is finite or countable). Similarly, it is possible to find b_ϵ such that $P(-X \leq -b_\epsilon) \leq \epsilon$ and b_ϵ is a continuity point of F_{-X}. It follows from $X_n \xrightarrow{d} X$ that
$$\limsup_n P(X_n \leq a_\epsilon \text{ or } X_n \geq b_\epsilon) \leq 2\epsilon,$$
so that for $n > N_\epsilon$ this probability is at most 3ϵ. Increasing (in absolute value) a_ϵ and b_ϵ if necessary, this probability is at most 3ϵ for all n. This proves boundedness in probability.

2. Let $\epsilon > 0$. Let M_ϵ as in the definition of boundedness in probability. Then
$$P(\|X_n Y_n\| > \epsilon) = P(\|X_n Y_n\| > \epsilon, \|X_n\| > M_\epsilon) + P(\|X_n Y_n\| > \epsilon, \|X_n\| \leq M_\epsilon) \leq \epsilon + P(\|Y_n\| \leq \epsilon/M_\epsilon).$$
Since $Y_n \xrightarrow{d} 0$ we obtain
$$\limsup_n P(\|X_n Y_n\| > \epsilon) \leq \epsilon$$
and as $\epsilon > 0$ is arbitrary this proves $X_n Y_n \xrightarrow{d} 0$.

Solution 2 Equality holds if and only if $S_n(\theta, X) = C(\theta)T(X)$ for some $C(\theta) \geq 0$. But $S_n = \nabla_\theta \ell_n$, so the log likelihood is $C(\theta)T(X)$ up to a linear factor depending possibly on X. This is the exponential family situation.

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

$$f(x;\mu) = \frac{\exp\{-(x-\mu)\}}{(1+\exp\{-(x-\mu)\})^2}, \qquad x \in \mathbb{R}, \quad \mu \in \mathbb{R}.$$

(a) Find the maximum likelihood estimator of μ (and verify that the solution actually maximises the likelihood), compute the Fisher information and find the asymptotic distribution of the estimator as $n \to \infty$.

(b) Find an estimator of μ that has an explicit expression and the same asymptotic properties as the maximum likelihood estimator.

Solution 3

(a) The log-likelihood function and its derivative equal

$$\ell_n(\mu) = -\sum_{i=1}^n (X_i - \mu) - 2\sum_{i=1}^n \log(1 + \exp\{-(X_i - \mu)\}),$$

$$\ell_n'(\mu) = n - 2\sum_{i=1}^n \frac{\exp\{-(X_i-\mu)\}}{1+\exp\{-(X_i-\mu)\}} = \sum_{i=1}^n \frac{1-\exp\{-(X_i-\mu)\}}{1+\exp\{-(X_i-\mu)\}}.$$

The estimator $\widehat{\mu}$ is given implicitly as the solution to $\ell_n'(\mu) = 0$. As the second derivative of the log-likelihood

$$\ell_n''(\mu) = -2\sum_{i=1}^n \frac{\exp\{-(X_i-\mu)\}}{[1+\exp\{-(X_i-\mu)\}]^2}$$

is negative, the log-likelihood ℓ_n is concave and the solution is the maximiser. It follows from the corollary to Cramér's theorem (Corollary 2.7, whose conditions are easily verified in this case) that the MLE is consistent. The conditions for asymptotic normality are verified here, since $|\ell'''(\mu)| \le 4n$ and so $|\overline{\ell}_n''(\mu) - \overline{\ell}_n''(\mu + h)| \le 4|h|$ and $\mathbb{E}_\mu(4) < \infty$. Moreover, the second derivative $\ell_n''(\mu)$ is also nicely behaved (bounded in absolute value by 2), so that $I_n(\mu) = J_n(\mu)$. Thus the asymptotic distribution of $\sqrt{n}(\widehat{\mu}_n^{MLE} - \mu)$ is $\mathrm{N}(0, 1/J_1(\mu))$ with

$$J_1(\mu) = \int_{-\infty}^{\infty} \frac{2\exp\{-(x-\mu)\}}{[1+\exp\{-(x-\mu)\}]^2} f(x;\mu)dx = \int \frac{2e^y}{(1+e^y)^4}dy = \int_0^\infty \frac{2z}{(1+z)^4} = \int_0^\infty \frac{2}{(1+z)^3}dz - \frac{2}{(1+z)^4}dz = \frac{1}{3}$$

not depending on μ.

(b) The density is symmetric and has a finite mean, so that mean has to be μ. Therefore we can take $\tilde{\mu}_n = \overline{X}_n$. It is consistent and $\sqrt{n}(\tilde{\mu}_n - \mu)$ is bounded in probability (by CLT). Thus it can be used as the starting point for a one-step approximation of the maximum likelihood estimator. The one-step estimator is $\mu_n^* = \tilde{\mu}_n - \overline{\ell}_n'(\tilde{\mu}_n)/\overline{\ell}_n''(\tilde{\mu}_n)$, which has an explicit expression as the derivatives have been calculated already. The asymptotic distribution of $\sqrt{n}(\mu_n^* - \mu)$ is the same as that of the maximum likelihood estimator, i.e., $\mathrm{N}(0, 1/J_1(\mu))$ with $J_1(\mu) = 1/3$ as before.

Exercise 4 Let $X_1, \dots, X_n$ be a sample from the Gamma distribution with known a (rate) and unknown p (shape) (the density is $f(x; a, p) = \frac{a^p}{\Gamma(p)} x^{p-1} e^{-ax} 1_{(0,\infty)}(x)$). Find an estimator $\tilde{p}$ of p and use it to construct a one-step estimator p^*. Find the asymptotic distribution of $\sqrt{n}(p^* - p)$.

Solution 4 We have $\mathbb{E}X = p/a$, so we can take $\tilde{p} = a\overline{X}$. The one-step estimator is $\widehat{p} = \tilde{p} - \frac{\ell_n'(\tilde{p})}{\ell_n''(\tilde{p})}$ where

$$\begin{aligned} \ell_n'(p) &= n \log a - \tfrac{n\Gamma'(p)}{\Gamma(p)} + \sum_{i=1}^n \log X_i, \\ \ell_n''(p) &= -n \tfrac{\Gamma''(p)\Gamma(p) - \Gamma'(p)^2}{\Gamma(p)^2}. \end{aligned}$$

The asymptotic distribution of $\sqrt{n}(p^* - p)$ is the same as that of the maximum likelihood estimator which is $\mathrm{N}(0, 1/J_1(p))$ with

$$J_1(p) = \tfrac{\Gamma''(p)\Gamma(p) - \Gamma'(p)^2}{\Gamma(p)^2}.$$

The regularity conditions hold here, since we have an exponential family. It is also easy to check them directly — the MLE is consistent either by the corollary to Cramér theorem or the law of large numbers. The second derivative does not depend on x at all.

Exercise 5 Let $X_1, \ldots, X_n$ be a sample from a Weibull distribution with density

$$f(x; a, p) = apx^{p-1}\exp\{-ax^p\}1_{(0,\infty)}(x),$$

where the parameter a is known to be 1 while the value of $p > 0$ is unknown. Suppose however that we have falsely assumed that the distribution is $\Gamma(1, p)$ (density $\frac{1}{\Gamma(p)}x^{p-1}e^{-x}1_{(0,\infty)}(x)$) with parameter $p > 0$ estimated by the maximum likelihood estimator $\widehat{p}_n^{MLE}$. What does $\widehat{p}_n^{MLE}$ estimate? Find its asymptotic distribution.
Hint: you may use without proof that $\int_0^\infty e^{-y}\log y dy = -\gamma \approx .577$ (the Euler–Mascheroni constant).

Solution 5 The regularity conditions hold, since we have an exponential family. Specifically here the second and third derivatives of ℓ_1 do not depend on x at all. Since the first derivative with respect to p is

$$\overline{\ell}_n'(p) = \frac{1}{n}\sum_{i=1}^n \log X_i - \frac{\Gamma'(p)}{\Gamma(p)}$$

and using the properties of Γ'/Γ, the MLE is such that this is equal to zero and converges by the law of large numbers to the value p^* such that

$$\frac{\Gamma'(p^*)}{\Gamma(p^*)} = \mathbb{E}_G \log X = \int_0^\infty p_0 x^{p_0-1}e^{-x^{p_0}}\log x dx = \frac{1}{p_0}\int_0^\infty dy e^{-y}\log y = \frac{-\gamma}{p_0}.$$

Here G denotes the true Weibull distribution and g the corresponding density, with p_0 the true Weibull parameter.

Since G satisfies $E_G|\log g(X)| < \infty$, the hope is that the MLE will converge to the value of p that minimises

$$KL(G, \Gamma(1, p)) = \mathbb{E}_G \log \frac{p_0 x^{p_0-1}\exp(-x^{p_0})\Gamma(p)}{x^{p-1}e^{-x}} = C(p_0) + \mathbb{E}_G \log\Gamma(p) - (p-1)\log X + X$$

$$= \tilde{C}(p_0) + \log\Gamma(p) - (p-1)\mathbb{E}_G \log X.$$

The value that maximises this satisfies $\partial \log\Gamma(p)/\partial p = E_G \log X$, and is therefore p^*. Hence the MLE is consistent for the p^* corresponding to the KL projection and the theorem applies. (We could have also used Cramér's theorem and its corollary to show consistency.) Thus $\sqrt{n}(\widehat{p}_n^{MLE} - p^*) \to N(0, I_1^G(p)/[J_1^G(p)]^2)$ with

$$J_1^G(p) = -\mathbb{E}_G\overline{\ell}_n''(p) = \frac{\Gamma''(p)\Gamma(p) - [\Gamma'(p)]^2}{\Gamma^2(p)}$$

and $I_1^G(p) = var_G(\log X)$.

Exercise 6 Let $X_1, \ldots, X_n$ be a sample from the Weibull distribution with the density $f(x; \lambda, p) = (\lambda p)\,(\lambda x)^{p-1}\, e^{-(\lambda x)^p}$ for $x > 0$, where $\lambda > 0$ and $p > 0$ are unknown parameters. Suppose that we thought that the distribution was exponential with the density $g(x; \lambda) = \lambda e^{-\lambda x} 1(x > 0)$ and calculated the maximum likelihood estimator $\widehat{\lambda}_n^{MLE}$. What does $\widehat{\lambda}_n^{MLE}$ estimate and what are its asymptotic properties? Compare with the maximum likelihood estimator of λ computed under the correct model specification when the parameter p of the Weibull distribution is known.

Solution 6 A simple calculation shows the MLE is unique and satisfies $\widehat{\lambda}_n^{MLE} = 1/\overline{X}_n$, and thus converges to the $1/\mathbb{E}_G(X)$. We have

$$\begin{aligned}\ell_n(\lambda) &= n \log \lambda - \lambda n \overline{X}_n \\ \ell_n'(\lambda) &= \tfrac{n}{\lambda} - n\overline{X}_n \\ \ell_n''(\lambda) &= -\tfrac{n}{\lambda^2}\end{aligned}$$

so that $J_1^G(\lambda) = \lambda^{-2}$, where G =Weibull(λ_0, p_0) is the true distribution of the data. The regularity conditions hold here since ℓ_n'' does not depend on x and is smooth in λ. We can also see this directly using the computation

$$\mathbb{E}_G(X) = \int_0^\infty \lambda_0 p_0 x (\lambda_0 x)^{p_0 - 1} e^{-(\lambda_0 x)^{p_0}}\, dx = \int_0^\infty \lambda_0^{-1} y^{1/p_0} e^{-y} dy = \tfrac{\Gamma(1+1/p_0)}{\lambda_0}.$$

so that $E_G \ell_1'(\lambda)$ vanishes at $\lambda^* = 1/\mathbb{E}_G(X)$. Moreover

$$E_G(X^2) = \int_0^\infty \lambda_0^{-2} y^{2/p_0} e^{-y} dy = \tfrac{\Gamma(1+2/p_0)}{\lambda_0^2},$$

so that

$$I_1^G(\lambda^*) = var_G(\ell_1') = \tfrac{\Gamma(1+2/p_0) - \Gamma^2(1+1/p_0)}{\lambda_0^2}.$$

Therefore $\sqrt{n}(\widehat{\lambda}_n^{MLE} - \lambda^*) \to N(0, I_1^G/[J_1^G]^2)$ with

$$J_1^G = J_1^G(\lambda^*) = 1/(\lambda^*)^2 = E_G^2 X = \tfrac{\Gamma^2(1+1/p_0)}{\lambda_0^2}.$$

Note that the value λ^* is consistent with KL projection question on week 3.

It is easily verified that if $X \sim F_{\lambda,p}$ and p is known, then $X^p \sim \text{Exp}(\lambda^p)$. The maximum likelihood estimator of λ is thus $(n/\sum_{i=1}^n X_i^p)^{1/p}$, consistent and asymptotically normal: $\sqrt{n}\,((n/\sum_{i=1}^n X_i^p)^{1/p} - \lambda) \xrightarrow{d} \text{N}(0, \lambda^2/p^2)$.

Note that if $p = 1$, then the Weibull distribution becomes exponential and we recover the results for the MLE of an exponential distribution.