Homework 5: Mathematical Statistics (MATH-UA 234)

Due 11/03 at the beginning of class on Gradescope

Problem 1. *Suppose $X_1, \dots, X_n$ are real numbers. Then the likelihood function is*

$$L_n(\theta) = \prod_{i=1}^{n} f_\theta(X_i).$$

In your own words, describe the meaning of $L_n(\theta)$ for a specific value of θ when

(a) f_θ is a probability mass function

(b) f_θ is a probability density function

Solution.

(a) When f_θ is a pmf, then $f_\theta(X_i)$ is the probability that we get $X_i' = X_i$ if we sampled a new random variable X_i' from F_θ. Thus, $L_n(\theta)$ is the probability that we get $(X_1', \dots, X_n') = (X_1, \dots, X_n)$ if we sampled new random variables $X_1', \dots, X_n'$ independently from F_θ.

Formally,

$$L_n(\theta) = \mathbb{P}[X_1' = X_1, \dots, X_n' = X_n | X_1', \dots, X_n' \sim F_\theta]$$

(b) When f_θ is a pmf, then $f_\theta(X_i)$ is no longer a probability. Rather, it is the probability of being in a small region about X_i, divided by the width of that region. Thus, $L_n(\theta)$ is the probability that $(X_1', \dots, X_n')$ lives in a small neighborhood of $(X_1, \dots, X_n)$, divided by the size (volume) of that neighborhood.

Somewhat more formally, for some small value "dx",

$$L_n(\theta) \approx \frac{\mathbb{P}[X_1' \in (X_1, X_1 + \mathrm{d}x), \dots, X_n' \in (X_n, X_n + \mathrm{d}x) | X_1', \dots, X_n' \sim F_\theta]}{(\mathrm{d}x)^n}.$$

Problem 2 (Wasserman 9.1). *In this problem, you will fill in some pieces we skipped in lecture. If you take the approach of searching for critical points, you should justify why your response is a global maximum.*

(a) Let $X_1, \dots, X_n \sim \operatorname{Ber}(p)$ where p is an unknown parameter. Show $\ell_n(p) = S\log(p) + (n - S)\log(1 - p)$, where $S = X_1 + \cdots + X_n$, and then find the maximizer of $\ell_n(p)$.

(b) Let $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ where μ and σ are unknown parameters. Show $\ell(\mu, \sigma) = c - n\log(\sigma) - nS^2/(2\sigma^2) - n(\bar{X} - \mu)^2/(2\sigma^2)$, where $\bar{X} = n^{-1}(X_1 + \cdots + X_n)$, $S^2 = n^{-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$, and c is some constant, and then find the maximizer of $\ell_n(\mu, \sigma)$.

Solution.

(a) The derivation of the log likelihood function is the same as in the book.

We can compute the derivative of the log likelihood

$$\ell_n'(p) = \frac{S}{p} - \frac{n - S}{1 - p} = \frac{np - S}{p(1 - p)}.$$

problems with a textbook reference are based on, but not identical to, the given reference

This is zero at $p = S/n$.

We have $\ell_n(0) = \ell_n(1) = -\infty$, so $\ell_n(p)$ must be maximized at the critical point rather than the boundary. Thus, the MLE is $\hat{p}_n = (X_1 + \cdots + X_n)/n$

(b) The derivation of the log likelihood function is the same as in the book, except that we maintain constants in front which result in the c term above.

We can compute the partial derivatives of the log likelihood

$$\frac{\partial \ell(\mu,\sigma)}{\partial \mu} = \frac{2n(\bar{X}-\mu)}{2\sigma^2}, \qquad \frac{\partial \ell(\mu,\sigma)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{nS^2 + n(\bar{X}-\mu)^2}{\sigma^3}.$$

Clearly

$$\frac{\partial \ell(\mu,\sigma)}{\partial \mu} = 0 \qquad \Longrightarrow \qquad \mu = \bar{X}.$$

Since this $\ell_n(\mu,\sigma)$ is a downward opening quadratic in μ, this has to be the maximizer (for any value of σ). If we plug this value in for μ, we find

$$\frac{\partial \ell(\bar{X},\sigma)}{\partial \sigma} = 0 \qquad \Longrightarrow \qquad \sigma = \pm S,$$

but since $\sigma > 0$ we should take the positive value. Moreover, $\ell_n(0) = \ell_n(\infty) = -\infty$, so the maximum must occur at the critical point.

thus, we find MLE

$$\hat{\mu}_n = \bar{X}, \qquad \hat{\sigma}_n = \left(n^{-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \right)^{1/2}.$$

Problem 3 (Wasserman 9.1). *Let $X_1, \dots, X_n \sim$ Gamma(α,β) where α and β are unknown parameters. Find the method of moments estimator for α and β.*

Solution. We will use the parametrization from the book on pages 29 and 30 which has density:

$$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} \exp(-x/\beta).$$

Thus, if $X \sim$ Gamma(α,β), from the table on page 53 we know:

$$\mathbb{E}[X] = \int x f(x) \mathrm{d}x = \alpha\beta$$
$$\mathbb{E}[X^2] = \mathbb{V}[X] + \mathbb{E}[X]^2 = \int x^2 f(x) \mathrm{d}x = \alpha\beta^2 + (\alpha\beta)^2 = (\alpha + \alpha^2)\beta^2.$$

For the method of moments, we should find α and β such that

$$\alpha\beta = \hat{m}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad (\alpha+\alpha^2)\beta^2 = \hat{m}_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2.$$

Using Mathematica, we find solution

$$\hat{\alpha}_n = -\frac{\hat{m}_1^2}{\hat{m}_1^2 - \hat{m}_2}, \qquad \hat{\beta}_n = \frac{\hat{m}_2 - \hat{m}_1^2}{\hat{m}_1}.$$

If you used the other standard parametrization, then your value for $\hat{\beta}_n$ will be the reciprocal.

Problem 4 (Wasserman 9.2). *Let $X_1, \dots, X_n \sim \text{Unif}(a, b)$ where a and b are unknown parameters with $a < b$.*

(a) Find the method of moments estimator for a and b.

(b) Find the maximum likelihood estimators for a and b.

Solution.

(a) If $X \sim \text{Unif}(a, b)$, from the table on page 53 we know:

$$\mathbb{E}[X] = \frac{a+b}{2}$$
$$\mathbb{E}[X^2] = \mathbb{V}[X] + \mathbb{E}[X]^2 = \frac{(b-a)^2}{12} + \frac{(a+b)^2}{2^2} = \frac{a^2+ab+b^2}{3}.$$

For the method of moments, we should find a and b such that

$$\frac{a+b}{2} = \hat{m}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad \frac{a^2+ab_b^2}{3} = \hat{m}_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2.$$

Since a and b are symmetric up to ordering, this has multiple solutions. But we should take the ones with $\hat{a}_n \le \hat{b}_n$.

$$\hat{a}_n = \hat{m}_1 - \sqrt{3}\sqrt{\hat{m}_2 - \hat{m}_1^2}, \qquad \hat{b}_n = \hat{m}_1 + \sqrt{3}\sqrt{\hat{m}_2 - \hat{m}_1^2}.$$

(b) We know the density is

$$f_{a,b}(x) = \begin{cases} \frac{1}{b-a} & a \le x \le b \\ 0 & \text{o.w.} \end{cases}$$

Thus, we have likelihood function,

$$L_n(a,b) = \prod_{i=1}^{n} f_{a,b}(X_i) \begin{cases} \frac{1}{(b-a)^n} & \forall i : a \le X_i \le b \\ 0 & \text{o.w.} \end{cases}$$

Clearly we can make $L_n(a,b) > 0$ by just picking a small and b big. So we know the maximum must occur at $1/(b-a)^n$ for some a and b satisfying $\forall i : a \le X_i \le b$. This condition is equivalent to $a \le \min\{X_1, \dots, X_n\}$ and $b \ge \max\{X_1, \dots, X_n\}$, so in order to make $b - a$ as small as possible, we should pick $a = \min\{X_1, \dots, X_n\}$ and $b = \max\{X_1, \dots, X_n\}$.

Problem 5 (Wasserman 9.5). *Let $X_1, \dots, X_n \sim \text{Poisson}(\lambda)$ where λ is an unknown parameter.*

(a) Find the method of moments estimator for λ.

(b) Find the maximum likelihood estimators for λ.

Solution.

(a) If $X \sim \text{Poisson}(\lambda)$, then $\mathbb{E}[X] = \lambda$. The method of moments estimator is then

$$\hat{\lambda}_n = \hat{m}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

(b) We have pmf $f_\lambda(x) = \exp(-\lambda)\lambda^x/x!$, so the likelihood function is

$$L_n(x) = \prod_{i=1}^{n} \exp(-\lambda)\frac{\lambda^{X_i}}{X_i!}$$

and the log likelihood function is

$$\begin{aligned}
\ell_n(x) &= \sum_{i=1}^{n} \log\left(\exp(-\lambda)\frac{\lambda^{X_i}}{X_i!}\right) \\
&= \sum_{i=1}^{n} \log\left(\exp(-\lambda)\right) + \log\left(\lambda^{X_i}\right) - \log]\,(X_i!) \\
&= \sum_{i=1}^{n} -\lambda + X_i\log(\lambda) - \log(X_i!) \\
&= -n\lambda + S\log(\lambda) - \sum_{i=1}^{n} \log(X_i!)
\end{aligned}$$

where $S = X_1 + \cdots + X_n$.

The derivative of the log likelihood function is

$$\ell_n'(\lambda) = -n + \frac{S}{\lambda}$$

which is zero when $\lambda = S/n$.

It's clear that $\lambda(0) = \lambda(\infty) = -\infty$, so this critical point is the global maximum and the MLE is $\hat{\lambda}_n = (X_1 + \cdots + X_n)/n$.

Problem 6 (Wasserman 9.6). *Let $X_1, \dots, X_n \sim N(\theta, 1)$ where θ is an unknown parameter. Define*

$$Y_i = \begin{cases} 1 & X_i > 0 \\ 0 & X_i \le 0. \end{cases}$$

Let $\psi = \mathbb{P}[Y_1 = 1]$.

Find the maximum likelihood estimator for ψ.

Solution.

Since Y_i are always zero or one, they are Bernoulli. In particular, $Y_i \sim \mathrm{Ber}(\psi)$. Thus, using 2(a), we could write the maximum liklihood estimator (in terms of the distribution of the Ys) as

$$\hat{\psi}_n = \frac{Y_1 + \cdots + Y_n}{n}.$$

The problem was intended to write the solution in terms of the distribution of the Xs. Towards this end, we have that

$$\psi = \mathbb{P}[Y_i = 1] = \mathbb{P}[X_i \ge 0] = 1 - \mathbb{P}[X_i \le 0] = 1 - \mathbb{P}[X_i - \theta < -\theta] = 1 - \Phi(-\theta) = \Phi(\theta)$$

where we have used that $X_i - \theta \sim N(0, 1)$.

Thus, using the equivariance of the MLE, we know that

$$\hat{\psi}_n = \Phi(\hat{\theta})_n = \Phi(n^{-1}(X_1 + \cdots + X_n))$$

where we have used 2(b) to get that the MLE for θ is the sample average of the data.