170S Week 4 Discussion Notes

Colin Ni

January 26, 2025

Warm-up

What does the Poisson distribution Pois(λ) model? What does the geometric distribution Geom(p) model? Describe their supports and the permitted values of λ and p.

Solution. Here $\lambda > 0$ is the rate, and $p \in [0, 1]$ is the probability of success. Assuming that λ events are expected to occur in a given period of time, the Poisson distribution Pois(λ) expresses the probability that k events occur, where k is any nonnegative integer. Assuming that the probability of success for a single Bernoulli trial is p, the geometric distribution Geom(p) expresses the probability that it takes exactly k trials for the first success to occur, where k is any positive integer. □

More examples of MLE's

Recall. Let X be a random variable that is defined in terms of some parameters $\theta = (\theta_1, \ldots, \theta_m)$, and suppose we have a sample $X_1, \ldots, X_n$ drawn from X. To estimate θ, we find θ that maximizes the likelihood

$$L(X_1, \ldots, X_n \mid \theta) = f_X(X_1 \mid \theta) \cdots f_X(X_n \mid \theta)$$

of the sample given θ. If there is a unique set of parameters $\hat{\theta}$ that maximizes this likelihood, we call it the maximum likelihood estimator (MLE) of θ.

Example. In lecture, we saw that for Norm(μ, σ^2), the MLE's are $\widehat{\mu} = \bar{X}$ (the sample mean) and $\widehat{\sigma^2} = \frac{n-1}{n} s_X^2$ (the naive sample variance). In particular, the MLE for σ^2 is biased, because s_X^2 is unbiased by Bessel's correction (see week 2 discussion notes).

Example. Let us show that the MLE $\widehat{\lambda}$ of the parameter λ for the Poisson distribution Pois(λ) is the sample mean

$$\widehat{\lambda} = \bar{X}.$$

Maximizing the likelihood

$$\begin{aligned}
L(X_1, \ldots, X_n \mid \lambda) &= f_{\mathrm{Pois}(\lambda)}(X_1) \cdots f_{\mathrm{Pois}(\lambda)}(X_n) \\
&= \frac{\lambda^{X_1} e^{-\lambda}}{X_1!} \cdots \frac{\lambda^{X_n} e^{-\lambda}}{X_n!} \\
&= \frac{\lambda^{n\bar{X}} e^{-n\lambda}}{X_1! \cdots X_n!}
\end{aligned}$$

is equivalent to maximizing the log-likelihood

$$\ln L(X_1, \ldots, X_n \mid \lambda) = n\bar{X} \ln(\lambda) - n\lambda - \ln(X_1! \cdots X_n!).$$

If $X_1 = \cdots = X_n = 0$ (recall the support from our warm-up), then the log-likelihood is a constant function, so there is no MLE. Thus, assume that not all X_i are zero. The critical points of the log-likelihood occur where

$$0 = \frac{\mathrm{d} \ln L}{\mathrm{d}\lambda}(X_1, \ldots, X_n \mid \lambda) = \frac{n\bar{X}}{\lambda} - n.$$

This shows the log-likelihood has a unique critical point $\lambda = \bar{X}$. Moreover this is a maximum because

$$\frac{\mathrm{d}^2 \ln L}{\mathrm{d}\lambda^2}(X_1, \ldots, X_n \mid \overline{X}) = -\frac{n\bar{X}}{\bar{X}} = -n < 0,$$

where $\bar{X} > 0$ because we assumed that not all X_i are zero.

Example. Let us show that the MLE $\widehat{p}$ of the parameter p for the geometric distribution $\mathrm{Geom}(p)$ is given by

$$\widehat{p} = \frac{1}{\bar{X}}.$$

This makes sense intuitively because $\bar{X}$ is the average number of trials needed to get a success, so $\frac{1}{\bar{X}}$ estimates the probability of success. Maximizing the likelihood

$$\begin{aligned}
L(X_1, \ldots, X_n \mid p) &= f_{\mathrm{Geom}(p)}(X_1) \cdots f_{\mathrm{Geom}(p)}(X_n) \\
&= (1-p)^{X_1 - 1} p \cdots (1-p)^{X_n - 1} p \\
&= (1-p)^{n\bar{X} - n} p^n
\end{aligned}$$

is equivalent to maximizing the log-likelihood

$$\ln L(X_1, \ldots, X_n \mid p) = (n\bar{X} - n) \ln(1-p) + n \ln(p).$$

If $X_1 = \cdots = X_n = 1$, then this is $n \ln(p)$, so there is no maximal value and hence no MLE. Thus, assume that not all X_i are one, or in other words that

$\bar{X} > 1$ (recall the support from our warm-up). Then the critical points of of the log-likelihood occur where

$$0 = \frac{\mathrm{d}\ln L}{\mathrm{d}p}(X_1, \dots, X_n \mid p) = -\frac{n\bar{X} - n}{1 - p} + \frac{n}{p}.$$

This shows the log-likelihood has a unique critical point $p = \frac{1}{\bar{X}}$. To see that this is a maximum, observe that

$$\frac{\mathrm{d}^2 \ln L}{\mathrm{d}p^2}\left(X_1, \dots, X_n \,\middle|\, \frac{1}{\bar{X}}\right) = -\frac{n\bar{X} - n}{(1 - \frac{1}{\bar{X}})^2} - \frac{n}{\frac{1}{\bar{X}}^2} < 0$$

where

$$\frac{\bar{X} - 1}{(1 - \frac{1}{\bar{X}})^2} + \frac{1}{\frac{1}{\bar{X}}^2} > 0$$

because we assumed that $\bar{X} > 1$.

Example. The MLE $\widehat{b}$ of the parameter b for the uniform distribution Unif$(0, b)$ is given by the sample maximum (or in other words the nth order statistic)

$$\widehat{b} = X_{(n)}.$$

Indeed, the likelihood is

$$\begin{aligned} L(X_1, \dots, X_n \mid b) &= f_{\mathrm{Unif}(0,b)}(X_1) \cdots f_{\mathrm{Unif}(0,b)}(X_n) \\ &= \frac{1}{b}\mathbf{1}\{X_1 \in [0, b]\} \cdots \frac{1}{b}\mathbf{1}\{X_1 \in [0, b]\} \\ &= \frac{1}{b^n}\mathbf{1}\{X_{(n)} \in [0, b]\}, \end{aligned}$$

where we are being careful to take into account the support of Unif$(0, b)$, and by inspection, this is uniquely maximized at $\widehat{b} = X_{(n)}$.

Remark. The following may help for Exercise 6.4-5 (which is on the homework). You should get that the MLE for part (a) is $\widehat{\theta} = \frac{1}{2}\bar{X}$ and that the MLE for part (b) is $\widehat{\theta} = \frac{1}{3}\bar{X}$. For part (c), here is how to find θ that minimizes

$$|X_1 - \theta| + \cdots + |X_n - \theta|.$$

Consider the order statistics $X_{(1)} \leq \cdots \leq X_{(n)}$, and furthermore consider the intervals $[X_{(k)}, X_{(n-k)}]$ for $1 \leq k \leq \frac{n}{2}$. We have

$$|X_{(k)} - \theta| + |X_{(n-k)} - \theta| = |X_{(n-k)} - X_{(k)}|$$

if and only if θ is in the interval $[X_{(k)}, X_{(n-k)}]$, and otherwise it is strictly larger. Thus θ minimizes the sum if and only if it is in all of these intervals. When n is odd, this is equivalent to θ being the sample median, and when n is even, this is equivalent to θ being anywhere between the two middle X_i values. On your homework, you can state this result and draw a picture illustrating this argument (you do not need to make this argument precise).

More examples of method of moments estimators

Recall. Again let X be a random variable that is defined in terms of some parameters $\theta = (\theta_1, \dots, \theta_m)$. Suppose the first m moments $M_k = \mathbb{E}(X^k) = g_k(\theta_1, \dots, \theta_m)$ are functions of the parameters $\theta_1, \dots, \theta_m$ of X, and consider the estimators

$$\widehat{M_k} = \frac{1}{n}(X_1^k + \cdots + X_n^k)$$

of these m moments. If there exists a unique solution $\widetilde{\theta_1}, \dots, \widetilde{\theta_m}$ to the system of equations

$$\begin{aligned}
\widehat{M_1} &= g_1(\widetilde{\theta_1}, \dots, \widetilde{\theta_m}) \\
&\vdots \\
\widehat{M_m} &= g_K(\widetilde{\theta_1}, \dots, \widetilde{\theta_m}),
\end{aligned}$$

then we call it the *method of moments estimator* of $\theta_1, \dots, \theta_m$.

Example. Let us find the method of moments estimator for the parameter p for the geometric distribution $\mathrm{Geom}(p)$. Since there is only one parameter, we only need the first moment $M_1 = \mathbb{E}(X) = \frac{1}{p}$. Our system of equations is the single equation

$$\bar{X} = \widehat{M_1} = \frac{1}{\widetilde{p}},$$

so the method of moments estimator $\widetilde{p} = \frac{1}{\bar{X}}$ exists and agrees with the MLE.

Example. Let us find the method of moments estimator the random variable X with pdf

$$f_X(x) = \begin{cases} \frac{4}{\theta^2}x & 0 < x \leq \frac{\theta}{2} \\ -\frac{4}{\theta^2}x + \frac{4}{\theta} & \frac{\theta}{2} < x \leq \theta \\ 0 & \text{otherwise.} \end{cases}$$

Here θ is a parameter $\theta \in (0, 2]$, and X is supported on $[0, \theta]$. Once again, we only have one parameter, so we only need the first moment $M_1 = \mathbb{E}(X)$. Trust (but verify):

$$\begin{aligned}
\mathbb{E}(X) &= \int_0^\theta x f_X(x)\,\mathrm{d}x \\
&= \int_0^{\frac{\theta}{2}} \frac{4}{\theta^2}x^2\,\mathrm{d}x + \int_{\frac{\theta}{2}}^{\theta} -\frac{4}{\theta^2}x^2\,\mathrm{d}x + \int_{\frac{\theta}{2}}^{\theta} \frac{4}{\theta}x\,\mathrm{d}x \\
&= \frac{\theta}{6} - \frac{7\theta}{6} + \frac{3\theta}{2} \\
&= \frac{\theta}{2}.
\end{aligned}$$

Thus our system of equations is the single equation

$$\bar{X} = \widehat{M_1} = \frac{\widetilde{\theta}}{2},$$

so the method of moments estimator exists and is $\widetilde{\theta} = 2\bar{X}$.

Example. The method of moments estimator of the parameter b for the uniform distribution $\mathrm{Unif}(0, b)$ is given by

$$\widetilde{b} = 2\bar{X}$$

because $M_1 = \frac{b}{2}$ and so

$$\bar{X} = \widehat{M_1} = \frac{\widetilde{b}}{2}.$$

In contrast, recall that we found the the MLE to be $\widehat{b} = X_{(n)}$.