Maximum likelihood estimation (part 2)

Lecture 14a (STAT 24400 F24)

Distribution of the MLE

Recall: if $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f(\cdot \mid \theta_0)$ and $\widehat{\theta}$ is the MLE, then for n large, the asymptotic distribution of the MLE is (by Fisher's Theorem)

$$\widehat{\theta} \;\approx\; \mathrm{N}\left(\theta_0, \frac{1}{n\mathcal{I}(\theta_0)}\right)$$

if we assume some regularity conditions (smoothness of $\log(f)$ as a function of θ).

How does this variance of MLE compare with other estimators?

Cramer–Rao inequality

if $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f(\cdot \mid \theta)$ then $\mathbb{E}(\widehat{\theta}) = \theta$, for every $\theta \in \Theta$

Theorem: (Cramer-Rao) For any *unbiased* estimator $\widehat{\theta}$ (to estimate the true θ_0)

$$\mathrm{Var}(\widehat{\theta}) \;\geq\; \frac{1}{n\mathcal{I}(\theta_0)}.$$

with respect to $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f(\cdot \mid \theta_0)$

How does this compare to the MLE?

- The MLE is (asymptotically) unbiased
- The MLE's variance is (asymptotically) $\frac{1}{n\mathcal{I}(\theta_0)}$

Is the MLE optimal?

- Not necessarily always;
  there may be settings where we're willing to trade bias for variance.

MLE confidence intervals

To construct a MLE based confidence interval for θ_0,
we use the more practical approximation of the MLE $\widehat{\theta} \approx \mathrm{N}\left(\theta_0, \frac{1}{n\mathcal{I}(\widehat{\theta})}\right)$

$$\sqrt{n\mathcal{I}(\widehat{\theta})} \cdot (\widehat{\theta} - \theta_0) \;\to\; \mathrm{N}(0,1) \qquad \text{(in distribution, as } n \to \infty\text{)}$$

$$\Rightarrow \quad \mathbb{P}\left(\left|\sqrt{n\mathcal{I}(\widehat{\theta})} \cdot (\widehat{\theta} - \theta_0)\right| > z_{\alpha/2}\right) \;\approx\; \alpha$$

$$\Rightarrow \quad \mathbb{P}\left(\theta_0 \;\in\; \widehat{\theta} \;\pm\; z_{\alpha/2} \cdot \frac{1}{\sqrt{n\mathcal{I}(\widehat{\theta})}}\right) \;\approx\; 1 - \alpha$$

So, after observing the data and calculating the interval

$$\widehat{\theta} \pm z_{\alpha/2} \cdot \frac{1}{\sqrt{n\mathcal{I}(\widehat{\theta})}} \;=\; \left(\widehat{\theta} - z_{\alpha/2} \cdot \frac{1}{\sqrt{n\mathcal{I}(\widehat{\theta})}} \;,\; \widehat{\theta} + z_{\alpha/2} \cdot \frac{1}{\sqrt{n\mathcal{I}(\widehat{\theta})}}\right)$$

we have approximately $(1 - \alpha)$ confidence that θ_0 lies in this interval.

Example: Normal mean μ (σ^2 known)

$X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathrm{N}(\mu, \sigma^2)$ for unknown $\mu \in \mathbb{R}$ (σ^2 is known)

- The MLE is $\widehat{\mu} = \bar{X}$
- The Fisher information is $\mathcal{I}(\mu) = \frac{1}{\sigma^2}$
- Therefore, $\widehat{\mu} \approx N\left(\mu_0, \frac{1}{n\mathcal{I}(\mu_0)}\right) = N\left(\mu_0, \frac{\sigma^2}{n}\right)$ for large n (μ_0 the true mean).
- Since σ^2 is known, an approximate $(1-\alpha)$ confidence interval for μ_0 is:

$$\mu_0 \in \ \widehat{\mu} \ \pm \ z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

(in fact, we know this distribution and conf. int. are *exact* for this case)

Example: Exponential rate parameter λ

$X_1, \dots, X_n \overset{\text{iid}}{\sim} \text{Exponential}(\lambda)$ for unknown $\lambda > 0$

- The MLE is $\widehat{\lambda} = \frac{1}{\bar{X}}$
- The Fisher information is $\mathcal{I}(\lambda) = \frac{1}{\lambda^2}$
- Therefore, $\widehat{\lambda} \approx \mathrm{N}\left(\lambda_0, \frac{\lambda_0^2}{n}\right)$
- To construct C.I., use the more useful approximation $\widehat{\lambda} \approx \mathrm{N}\left(\lambda_0, \frac{\hat{\lambda}^2}{n}\right)$.
- An approximate $(1-\alpha)$ confidence interval for the true λ_0 is:

$$\lambda_0 \ \in \ \widehat{\lambda} \ \pm \ z_{\alpha/2} \cdot \frac{\widehat{\lambda}}{\sqrt{n}}$$

Discussion: Different roles of the two $\hat{\lambda}$'s.

Example: Bernoulli & Binomial

$X_1, \dots, X_n \overset{\text{iid}}{\sim} \text{Bernoulli}(p)$ for unknown $p \in (0,1)$

- The MLE is $\widehat{p} = \bar{X}$
- The Fisher information is $\mathcal{I}(p) = \frac{1}{p(1-p)}$
- Therefore, $\widehat{p} \approx \mathrm{N}\left(p_0, \frac{p_0(1-p_0)}{n}\right)$,
- More usefully, $\widehat{p} \approx \mathrm{N}\left(p_0, \frac{\hat{p}(1-\hat{p})}{n}\right)$, and an approx. $(1-\alpha)$ conf. int. is:

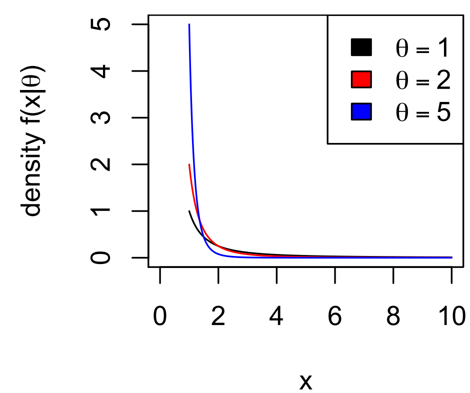
$$p_0 \in \ \widehat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

Discussion: Again, note the different roles of the $\hat{p}$'s.

Example: MLE and MoM

$X_1, \dots, X_n \overset{\text{iid}}{\sim} f(\cdot \mid \theta)$ for parameter $\theta > 0$ where the density is

$$f(x \mid \theta) = \frac{\theta}{x^{\theta+1}} \cdot \mathbb{1}_{x \geq 1}$$



Example cont. (calculate MLE)

- Calculate the MLE:

  The log-likelihood function given the data is

$$\sum_i \log f(X_i \mid \theta) = \sum_i \log\left(\frac{\theta}{X_i^{\theta+1}}\right) = n\log\theta - (\theta+1)\sum_i \log(X_i)$$

  Set the derivative to 0:

$$\frac{\partial}{\partial\theta}\left[n\log\theta - (\theta+1)\sum_i \log(X_i)\right] = \frac{n}{\theta} - \sum_i \log(X_i) = 0$$

$$\Rightarrow \qquad \widehat{\theta} = \frac{n}{\sum_i \log(X_i)}$$

Example cont. (calculate fisher information and C.I.)

- Fisher information:

$$-\frac{\partial^2}{\partial\theta^2}\left[\log f(X \mid \theta)\right] = -\frac{\partial^2}{\partial\theta^2}\left[\log(\theta) - (\theta+1)\log(X)\right] = \frac{1}{\theta^2}$$

$$\Rightarrow \qquad \mathcal{I}(\theta) = \mathbb{E}\left(-\frac{\partial^2}{\partial\theta^2}\left[\log f(X \mid \theta)\right]\right) = \frac{1}{\theta^2}$$

- Therefore, the MLE $\widehat{\theta} = \frac{n}{\sum_i \log(X_i)}$ satisfies

$$\widehat{\theta} \approx \mathrm{N}\left(\theta_0, \frac{\theta_0^2}{n}\right)$$

  and $\widehat{\theta} \approx \mathrm{N}\left(\theta_0, \frac{\widehat{\theta}^2}{n}\right)$, and an approximate $(1-\alpha)$ conf. int. for θ_0 is

$$\widehat{\theta} \pm z_{\alpha/2} \cdot \frac{\widehat{\theta}}{\sqrt{n}}.$$

Example cont. (calculate MoM)

Compare to the MoM estimator

- Calculate the first moment:

$$\mathbb{E}(X) = \int_{x\geq 1} x \cdot \frac{\theta}{x^{\theta+1}}\,\mathrm{d}x = \frac{\theta}{\theta-1} \qquad \leftarrow \text{ exists for } \theta > 1$$

  The Method of Moments set

$$\frac{\theta}{\theta-1} = \bar{X}$$

$$\Rightarrow \quad \text{MoM estimate} \quad \widehat{\theta} = \frac{\bar{X}}{\bar{X}-1}$$

Example cont. (MLE vs MoM)

An empirical comparison of $\widehat{\theta}_{MLE} = \frac{n}{\sum_i \log(X_i)}$ and $\widehat{\theta}_{MoM} = \frac{\bar{X}}{\bar{X}-1}$