STAT 450

Lecture 19

Today's notes
Reading for Today's Lecture:
Goals of Today's Lecture:

- Define the Fisher information matrix.
- Illustrate standard MLE theory on $N(\mu,\sigma^2)$ example.

Today's notes

So far: we have shown that for $X_1,\ldots,X_n$ an iid sample from $N(\mu,\sigma^2)$ we have

$$B \equiv \mathrm{Var}(U(X_1,\ldots,X_n;\mu,\sigma)) = \left[\begin{array}{cc} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{2n}{\sigma^2} \end{array}\right]$$

Next: compute $A = -\mathrm{E}(\partial U/\partial\theta)$. First

$$\begin{aligned}
\frac{\partial U(\mu,\sigma)}{\partial\theta} &= \left[\begin{array}{cc} \frac{\partial}{\partial\mu}\frac{\sum(X_i-\mu)}{\sigma^2} & \frac{\partial}{\partial\sigma}\frac{\sum(X_i-\mu)}{\sigma^2} \\ \frac{\partial}{\partial\mu}\left(\frac{\sum(X_i-\mu)^2}{\sigma^3}-\frac{n}{\sigma}\right) & \frac{\partial}{\partial\sigma}\left(\frac{\sum(X_i-\mu)^2}{\sigma^3}-\frac{n}{\sigma}\right) \end{array}\right] \\
&= \left[\begin{array}{cc} -\frac{n}{\sigma^2} & -\frac{2\sum(X_i-\mu)}{\sigma^3} \\ -\frac{2\sum(X_i-\mu)}{\sigma^3} & -\frac{3\sum(X_i-\mu)^2}{\sigma^4}+\frac{n}{\sigma^2} \end{array}\right]
\end{aligned}$$

Take expected values to get

$$A = \left[\begin{array}{cc} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{2n}{\sigma^2} \end{array}\right] = B$$

Definition: The **Fisher Information** is

$$\mathcal{I}_n(\theta) = -E_\theta(U'(\theta)).$$

General properties of likelihoods

In general if $f(x,\theta)$ denotes the joint density of all the data then

$$1 = \int f(x;\theta)dx$$

for all θ.

This is an identity in θ, so we can differentiate both sides with respect to θ to get

$$\begin{aligned}
0 &= \frac{\partial}{\partial\theta}\int f(x;\theta)dx \\
&= \int \frac{\partial}{\partial\theta} f(x;\theta)dx \\
&= \int \frac{\frac{\partial}{\partial\theta}f(x;\theta)}{f(x;\theta)} f(x;\theta)dx \\
&= \mathrm{E}_\theta(U(\theta))
\end{aligned}$$

The calculation above involves changing the order of differentiation and integration and is not always valid. In the irregular examples on your homework this doesn't work. It does work in the normal example I did; this is the usual outcome.

Differentiate the same identity again and get

$$\begin{aligned}
0 &= \frac{\partial}{\partial\theta}\int \frac{\frac{\partial}{\partial\theta}f(x;\theta)}{f(x;\theta)}f(x;\theta)dx \\
&= \int \frac{\partial}{\partial\theta}\frac{\frac{\partial}{\partial\theta}f(x;\theta)}{f(x;\theta)}f(x;\theta)dx \\
&= \int \frac{\frac{\partial^2}{\partial\theta^2}f(x;\theta)}{f(x;\theta)}f(x;\theta)dx \\
&\quad + \int \frac{\frac{\partial}{\partial\theta}f(x;\theta)}{f(x;\theta)}\frac{\partial}{\partial\theta}f(x;\theta)dx \\
&= \mathrm{E}_\theta(\frac{\partial U}{\partial\theta}) + \int \left(\frac{\frac{\partial}{\partial\theta}f(x;\theta)}{f(x;\theta)}\right)^2 f(x;\theta)dx
\end{aligned}$$

This gives a so-called **Bartlett identity**

$$\mathrm{E}_\theta(U(\theta)^2) = \mathrm{Var}_\theta(U(\theta)) = -\mathrm{E}_\theta(\frac{\partial U}{\partial\theta})$$

In the notation of the previous example and theory it is always the case for likelihood estimates (in "regular" models) that $A = B$ and $A^{-1}BA^{-1} = \mathcal{I}^{-1}(\theta)$.

Asymptotic normality

In the normal example we had an iid sampling model. The Fisher information when we have n observations was

$$\mathcal{I}_n(\theta) = n\begin{bmatrix} \sigma^{-2} & 0 \\ 0 & 2\sigma^{-2} \end{bmatrix}$$

Notice that $\mathcal{I}_n(\theta) = n\mathcal{I}_1(\theta)$. We refer to $\mathcal{I}_1(\theta_0)$ as the information in 1 observation. In *any* iid sampling problem we will have $\mathcal{I}_n(\theta) = n\mathcal{I}_1(\theta)$.

This leads to the following theorem.

Theorem: In iid sampling

$$\sqrt{n}(\hat\theta - \theta) \Rightarrow N(0, \mathcal{I}_1^{-1}(\theta)$$

Examples, then uses

A: Exponential rate λ. We have $X - 1, \ldots, X_n$ iid with density $f(x,\lambda) = \lambda e^{-\lambda x}1(x > 0)$.

We find

$$
\begin{aligned}
L(\lambda) &= \lambda^n e^{-\lambda \sum X_i} \\
\ell(\lambda) &= n\log(\lambda) - \lambda \sum X_i \\
U(\lambda) &= \frac{n}{\lambda} - \sum X_i \\
\hat{\lambda} &= n/\sum X_i = 1/\bar{X} \\
\frac{\partial U}{\partial \lambda} &= \frac{-n}{\lambda^2} \\
\mathcal{I}(\lambda) &= n/\lambda^2 \\
\hat{\lambda} - \lambda &\approx N(0, \lambda^2/n) \\
\frac{\sqrt{n}(\hat{\lambda} - \lambda)}{\lambda} &\approx N(0,1)
\end{aligned}
$$

B: Exponential mean μ. We have $X_1, \ldots, X_n$ iid with density $f(x,\mu) = \frac{1}{\mu} e^{-x/\mu} 1(x > 0)$. We find

$$
\begin{aligned}
L(\mu) &= \frac{1}{\mu^n} e^{-\sum X_i/\mu} \\
\ell(\mu) &= -n\log(\mu) - \sum X_i/\mu \\
U(\mu) &= -\frac{n}{\mu} + \sum X_i/\mu^2 \\
\hat{\mu} &= \sum X_i/n = \bar{X} \\
\frac{\partial U}{\partial \mu} &= \frac{n}{\mu^2} - 2\sum X_i/\mu^3 \\
\mathcal{I}(\mu) &= n/\mu^2 \\
\hat{\mu} - \mu &\approx N(0, \mu^2/n) \\
\frac{\sqrt{n}(\hat{\lambda} - \lambda)}{\lambda} &\approx N(0,1)
\end{aligned}
$$

C: Cauchy with location parameter θ. We have $X_1, \ldots, X_n$ iid with density $f(x,\theta) =$

$\frac{1}{\pi[1+(x-\theta)^2]}$. We find

$$
\begin{aligned}
L(\theta) &= \frac{1}{\pi^n \prod[1+(X_i-\theta)^2]} \\
\ell(\theta) &= -n\log(\pi) - \sum \log[1+(X_i-\theta)^2] \\
U(\theta) &= 2\sum \frac{X_i-\theta}{1+(X_i-\theta)^2} \\
\hat{\theta} &= \text{no closed form} \\
\frac{\partial U}{\partial \theta} &= 2\sum \frac{(X_i-\theta)^2-1}{[1+(X_i-\theta)^2]^2} \\
\mathcal{I}(\theta) &= -2n\mathrm{E}_\theta\left[\frac{(X_i-\theta)^2-1}{[1+(X_i-\theta)^2]^2}\right] \\
&= -2n\int_{-\infty} \infty \frac{(x-\theta)^2-1}{\pi[1+(x-\theta)^2]^3}dx \\
&= -2n\int_{-\pi/2}^{\pi/2} \frac{\tan^2(\phi)-1}{\pi[1+\tan^2(\phi)]^2}d\phi \\
&= -\frac{2n}{\pi}\int_{-\pi/2}^{\pi/2}[\sin^2(\phi)-\cos^2(\phi)]\cos^2(\phi)d\phi \\
&= \frac{2n}{\pi}\int_{-\pi/2}^{\pi/2}\cos(2\phi)(1+\cos(2\phi))d\phi/2 \\
&= \frac{n}{\pi}\left[\left.\frac{\sin(2\phi)}{2}\right|_{-\pi/2}^{\pi/2} + \int_{-\pi/2}^{\pi/2}(1+\cos(4\phi))d\phi/2\right] \\
&= \frac{n}{\pi}\left[0+\frac{\pi}{2}+\frac{1}{8}\sin(4\phi)|_{-\pi/2}^{\pi/2}\right] \\
&= \frac{n}{2} \\
\hat{\theta}-\theta &\sim N(0, 2/n) \\
\sqrt{n/2}(\hat{\theta}-\theta) &\sim N(0,1)
\end{aligned}
$$

D: Uniform$[0,\theta]$ We have $X_1,\ldots,X_n$ iid with density $f(x,\theta)=\frac{1}{\theta}1(0\le x\le\theta)$. We find

$$L(\theta)=\frac{1}{\theta^n}1(\theta>\max(X_i))$$
$$\ell(\theta)=\begin{cases}-n\log(\theta) & \theta>\max(X_i)\\ -\infty & \text{otherwise}\end{cases}$$
$$U(\theta)=\begin{cases}-\frac{n}{\theta} & \theta>\max(X_i)\\ \text{undefined} & \text{otherwise}\end{cases}$$
$$\hat\theta=\max(X_i)$$

•Likelihood equations have no root

•MLE is not asymptotically normal

•Family is **irregular**

This family has the feature that the **support** of the density, namely $\{x: f(x;\theta)>0\}$, depends on θ. In such families it is common for the standard mle theory to fail.

Uses and extensions

Confidence Intervals: We can base confidence intervals on one of several forms. For this section I will assume that θ is a scalar (one dimensional) parameter and use $'$ to denote a derivative with respect to the parameter. There are 3 standard versions of the normal approximation:

$$\sqrt{\mathcal{I}_n(\theta)}(\hat\theta-\theta)\Rightarrow N(0,1)$$
$$\sqrt{\mathcal{I}_n(\hat\theta)}(\hat\theta-\theta)\Rightarrow N(0,1)$$
$$\sqrt{-U'(\hat\theta)}(\hat\theta-\theta)\Rightarrow N(0,1)$$

Each of these quantities may be used to derive confidence intervals for θ by finding the collection of all θ for which the quantity is smaller than some critical point.

The second and third quantities are of the form

$$\frac{\hat\theta-\theta}{\hat\sigma_{\hat\theta}}$$

If such a quantity is standard normal then

$$P\left(\left|\frac{\hat\theta-\theta}{\hat\sigma_{\hat\theta}}\right|\le z_{\alpha/2}\right)\approx 1-\alpha$$

so

$$P(\hat\theta-\hat\sigma_{\hat\theta}z_{\alpha/2}\le\theta\le\hat\theta+\hat\sigma_{\hat\theta}z_{\alpha/2})\approx 1-\alpha$$

This means that

$$\hat\theta\pm\hat\sigma_{\hat\theta}z_{\alpha/2}$$

is an approximate level $1-\alpha$ confidence interval for θ.