Lecture 5.

Some times we want to estimate a function $f(\theta)$ of θ rather than θ itself. If f is a smooth function and $t_n(x_1, \ldots, x_n)$ is an estimate of θ with

$$E_\theta[(t_n - \theta)^2] \simeq \frac{v(\theta)}{n}$$

by Taylor expansion we saw that $f(t_n) - f(\theta) = f'(\theta)(t_n - \theta)$ and we expect

$$E_\theta[(f(t_n) - f(\theta))^2] \simeq \frac{[f'(\theta)]^2 v(\theta)}{n}$$

The Cramér-Rao inequality becomes

$$E_\theta[f(t_n)] = f_n(\theta)$$

$$\sum_{x_1,\ldots,x_n} [f(t_n) \frac{\partial \phi(\theta, x_1, \ldots, x_n)}{\partial \theta})] = f_n'(\theta)$$

$$\sum_{x_1,\ldots,x_n} [f(t_n) \frac{\partial \log \phi(\theta, x_1, \ldots, x_n)}{\partial \theta} \phi(\theta, x_1, \ldots, x_n)] = f_n'(\theta)$$

$$E_\theta[\frac{\partial \log \phi(\theta, x_1, \ldots, x_n)}{\partial \theta} \phi(\theta, x_1, \ldots, x_n)] = 0$$

$$E_\theta[(f(t_n) - f(\theta)) \frac{\partial \log \phi(\theta, x_1, \ldots, x_n)}{\partial \theta} \phi(\theta, x_1, \ldots, x_n)] = f_n'(\theta)$$

$$E_\theta[(f(t_n) - f(\theta))^2] \geq \frac{[f'(\theta)]^2}{nI(\theta)}$$

Maximum Likelihood estimate.

The maximum likelihood estimation is perhaps the most important method of estimation for parametric families. Whether it is probabilities $p(\theta, x)$ or densities $f(\theta, x)$ the likelihood function is the joint probability or density and is given by

$$L(\theta, x_1, x_2, \ldots, x_n) = \Pi_i p(\theta, x_i)$$

or

$$L(\theta, x_1, x_2, \ldots, x_n) = \Pi_i f(\theta, x_i)$$

Given the observed values $(x_1, x_2, \ldots, x_n)$ this is viewed as a function of θ and the value $\hat{\theta} = t(x_1, \ldots, x_n)$ that maximizes it is taken as the estimate of θ.

Examples.

1. Binomial. With $t = \sum x_i$ the number of heads

$$L(\theta, \mathbf{x}) = \theta^t (1-\theta)^{n-t}$$

$$\frac{d \log L}{d\theta} = 0$$

$$\frac{t}{\widehat{\theta}} = \frac{n-t}{1-\widehat{\theta}}$$

reduces to $\widehat{\theta} = \frac{t}{n}$

2. Normal family with known variance equal to 1,

$$f(\mu, x) = \frac{1}{\sqrt{2\pi}} \exp[-\frac{(x-\mu)^2}{2}]$$

$$\log L(\mu, x_1, \ldots, x_n) = -\frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2 - \frac{n}{2}\log 2\pi$$

$$\frac{\partial \log L}{\partial \mu} = \sum_{i=1}^{n}(x_i - \mu) = 0$$

$$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

3. Normal family with mean 0 but unknown variance θ.

$$f(\theta, x) = \frac{1}{\sqrt{2\pi\theta}} \exp[-\frac{x^2}{2\theta}]$$

$$\log L(\theta, x_1, \ldots, x_n) = -\frac{1}{2\theta}\sum_{i=1}^{n} x_i^2 - \frac{n}{2}\log 2\pi - \frac{n}{2}\log\theta$$

$$\frac{\partial \log L}{\partial \theta} = \frac{1}{2\theta^2}\sum_{i=1}^{n} x_i^2 - \frac{n}{2\theta} = 0$$

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

4. Gamma family.

$$f(p,x) = \frac{1}{\Gamma(p)} e^{-x} x^{p-1}$$

$$L(p, x_1, \ldots, x_n) = -n \log \Gamma(p) - \sum_{i=1}^{n} x_i + (p-1)\sum_{i=1}^{n} \log x_i$$

$$\frac{\partial \log L}{\partial p} = -n\frac{\Gamma'(p)}{\Gamma(p)} + \sum_{i=1}^{n} \log x_i = 0$$

$\hat{p}$ is the solution of the equation

$$\frac{\Gamma'(p)}{\Gamma(p)} = \frac{1}{n}\sum_{i=1}^{n} \log x_i$$

Properties of a good estimator.

1. Consistency.

$$P_\theta[|t_n - \theta| \geq \delta] \to 0$$

Enough if $E_\theta[|t_n - \theta|^2] \to 0$.

2. Efficiency

The variance $E_\theta[|t_n - \theta|^2]$ must be as small as possible. If the Cramér-Rao bound is approached it is good. Asymptotically efficient.

$$nE_\theta[(t_n - \theta)^2] \to \frac{1}{I(\theta)}$$

3. It is good to know the asymptotic distribution of t_n. A central limit theorem of the form

$$P[\sqrt{n}(t_n - \theta)\sqrt{I(\theta)} \leq x] \to \Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} \exp[-\frac{y^2}{2}]dy$$

will be useful.

4. If there is a sufficient statistic MLE is a function of it.

Theorem. If $f(\theta, x)$ is nice then the MLE satisfies 1,2 and 3.

Explanation. Why does it work? Consider the function $\log p(\theta, x)$ as a function of θ and compute its expectation under a particular value θ_0 of θ.

$$\begin{aligned}
E_{\theta_0}[\log p(\theta, x)] - E_{\theta_0}[\log p(\theta_0, x)] &= E_{\theta_0}\left[\log \frac{p(\theta, x)}{p(\theta_0, x)}\right] \\
&\le \log E_{\theta_0}\left[\frac{p(\theta, x)}{p(\theta_0, x)}\right] \\
&= \log \sum_x p(\theta, x) \\
&= 0
\end{aligned}$$

If the sample is from the poulation with $\theta = \theta_0$ by the law of large numbers the function

$$\frac{1}{n} \log L(\theta, x_1, x_2, \ldots, x_n) \simeq E_{\theta_0}[\log p(\theta, x)]$$

has its maximum at θ_0. Therefore $L(\theta, x_1, x_2, \ldots, x_n)$ is likely to have its maximum close to θ_0.

More Examples.

$$f(\theta, x) = \frac{1}{\theta}; \quad 0 \le x \le \theta$$

$$f(\theta, x_1, \ldots, x_n) = \frac{1}{\theta^n}$$

θ wants to be as small as possible. But $\theta \ge x_i$ for every i.

$$t_n(x_1, \ldots, x_n) = \max\{x_1, \ldots, x_n\}$$

Multiparameter families

1.

$$f(\mu, \theta, x) = \frac{1}{\sqrt{2\pi\theta}} \exp[-\frac{(x-\mu)^2}{2\theta}]$$

$$\log L(\theta, x_1, \ldots, x_n) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\theta - \frac{1}{2\theta}\sum_{i=1}^{n}(x_i-\mu)^2]$$

$$\frac{\partial \log L}{\partial\theta} = 0, \quad \frac{\partial \log L}{\partial\mu} = 0$$

$$\theta = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2; \quad \sum_{i=1}^{n}(x_i-\mu) = 0$$

$$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\hat{\theta} = s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2$$

2. Multivariate Normal families. $x = \{x_1, \ldots, x_d\}, \mu = \mu_1, \ldots, \mu_d \in R^d$. $A = \{a_{r,s})\}$ is a symmetric positive definite $d \times d$ matrix.

$$f(\mu, A, x) = (\frac{1}{\sqrt{2\pi|A|}})^d \exp[-\frac{< x, A^{-1}x >}{2}]$$

$$\int_{R^d} x_r f(\mu, A, x)dx = \mu_r$$

$$\int_{R^d}(x_r-\mu_r)(x_s-\mu_s)f(\mu, A, x)dx = a_{r,s}$$

$$\hat{\mu}_r = \bar{x}_r = \frac{1}{n}\sum_{i=1}^{n} x_{i,r}$$

$$\hat{a}_{r,s} = \frac{1}{n}\sum_{i=1}^{n}(x_{i,r}-\mu_r)(x_{i,s}-\mu_s) = \frac{1}{n}\sum_{i=1}^{n} x_{i,r}x_{i,s} - \bar{x}_r\bar{x}_s$$