Lecture 6.

We have the identity

$$D^2 e^g = e^g D^2 g + e^g [Dg]^2$$

Taking $g = \log f(\theta, x)$ and differentiating $f = e^g$ twice with respect to θ we have

$$\begin{aligned}\frac{\partial^2 f(\theta,x)}{\partial\theta^2} &= \frac{\partial^2 \exp[\log f(\theta,x)]}{\partial\theta^2} \\ &= \frac{\partial^2 \log f(\theta,x)}{\partial\theta^2} f(\theta,x) + \left[\frac{\partial \log f(\theta,x)}{\partial\theta}\right]^2 f(\theta,x)\end{aligned}$$

On the other hand if we differentiate twice the identity

$$\int f(\theta,x)dx \equiv 1$$

we get

$$\frac{d}{d\theta}\int f(x,\theta)dx = 0$$

and

$$\frac{d^2}{d\theta^2}\int f(x,\theta)dx = 0$$

They can be rewritten as

$$\int \frac{\partial \log f(x,\theta)}{\partial\theta} f(\theta,x)dx = 0$$

$$\int \frac{\partial^2 \log f(x,\theta)}{\partial\theta} f(\theta,x)dx + \int \left[\frac{\partial \log f(\theta,x)}{\partial\theta}\right]^2 f(\theta,x)dx = 0$$

Therefore

$$I(\theta) = E_\theta\left[[\frac{\partial \log f(\theta,x)}{\partial\theta}]^2\right] = -E_\theta\left[\frac{\partial^2 \log f(\theta,x)}{\partial\theta^2}\right]$$

Let θ_0 be the true value of the parameter θ, i.e. we have a random sample from the population $f(\theta_0, x)$. We have n random variables $X_1, X_2, \ldots, X_n$ with joint density

$$f(\theta_0, x_1) \cdots f(\theta_0, x_n)$$

The log likelihood function is the random function $\log L$ with derivative

$$J(\theta) = \frac{\partial \log L(\theta, X_1, X_2, \ldots, X_n)}{\partial \theta} = \sum_{j=1}^{n} \frac{\partial \log f(\theta, X_j)}{\partial \theta}$$

If t_n is the MLE, then

$$0 = J(t_n) = J(\theta_0) + J'(\theta_0)(t_n - \theta_0) + o(|t_n - \theta_0|)$$

$$\sqrt{n}(t_n - \theta_0) \simeq \frac{J(\theta_0)}{\sqrt{n}}[-n^{-1}J'(\theta_0)]^{-1}$$

By the CLT, $\frac{J(\theta_0)}{\sqrt{n}}$ is asymptotically normal with mean 0 and variance $I(\theta_0)$. On the other hand, by the Law of large numbers, $n^{-1}J'(\theta_0)$ is close to its expectation

$$I(\theta_0) = -E_{\theta_0}\left[\frac{\partial^2 \log f}{\partial \theta^2}(\theta_0, X)\right] = E_{\theta_0}\left[[\frac{\partial \log f}{\partial \theta}(\theta_0, X)]^2\right]$$

Therefore $\sqrt{n}(t_n - \theta_0)$ is asymptotically normal with mean 0 and variance $[I(\theta_0)]^{-1}$.

Remark. If there are several parameters $\theta = \{\theta_i\}$, $I(\theta)$ is a matrix

$$I_{i,j}(\theta) = E_\theta[\frac{\partial \log f}{\partial \theta_i}(\theta, x)\frac{\partial \log f}{\partial \theta_j}(\theta, x)] = -E_\theta[\frac{\partial^2 \log f}{\partial \theta_i \partial \theta_j}(\theta, x)]$$

and the MLE $\sqrt{n}(t_n - \theta_0)$ is asymptotically distributed as multivariate normal with mean 0 and covariance $[I(\theta_0)]^{-1}$.

Order Statistics.. Let the n observations be ordered as $X_{(1)} < \ldots < X_{(n)}$. $X_{(i)}$ is called the i-th order statistic. If $i = [np]$ the X_i is called the p-th quantile. If $p = \frac{1}{2}$ it is called a median.

Let $\{X_i\}$ be i.i.d from a continuous distribution with $P[X \le x] = F(x)$. Then

$$P[X_{(i)} \le x] = P[\#\{j : X_j \le x\} \ge i] = \sum_{r=i}^{n} \binom{n}{r} F(x)^r (1 - F(x))^{n-r}$$

Differentiating with respect to x the density of the i-th order statistic from a sample of size n has the density,

$$\begin{aligned}
&\mathbf{f}_{i,n}(x) \\
&= \sum_{r=i}^{n} \binom{n}{r} [rF(x)^{r-1}(1-F(x))^r - (n-r)F(x)^r(1-F(x))^{n-r-1}]f(x) \\
&= i\binom{n}{i} F(x)^{i-1}(1-F(x))^{n-i} f(x) \\
&\quad + \sum_{j \ge i+1} F(x)^{j-1}(1-F(x))^{n-j} \left[j\binom{n}{j} - (n-j+1)\binom{n}{j-1} \right] f(x) \\
&= i\binom{n}{i} F(x)^{i-1}(1-F(x))^{n-i} f(x) \\
&= \frac{n!}{(i-1)!(n-i)!} F(x)^{i-1}(1-F(x))^{n-i} f(x)
\end{aligned}$$

If f is uniform on $[0, 1]$, then

$$\mathbf{f}_{i,n}(x) = \frac{n!}{(i-1)!(n-i)!} x^{i-1}(1-x))^{n-i}$$

is a Beta$(i, n-i+1)$. We are interested in the distribution of $\sqrt{n}(X_{([np])} - a_p)$ where the p-th quantile a_p is the solution of $F(a_p) = p$. Substituting $x = a_p + \frac{\xi}{\sqrt{n}}$ and using

$$F(a_p + \frac{\xi}{\sqrt{n}}) \simeq p + \frac{\xi}{\sqrt{n}} f(a_p)$$

$$\begin{aligned}
g_{n,p}(\xi) &\simeq c_{n,p} F(a_p + \frac{\xi}{\sqrt{n}})^{[np]-1}(1 - F(a_p + \frac{\xi}{\sqrt{n}}))^{n-[np]} \\
&\simeq k_{n,p} \exp[np \log(1 + \frac{\xi f(a_p)}{p\sqrt{n}}) + n(1-p)\log(1 - \frac{\xi f(a_p)}{(1-p)\sqrt{n}})] \\
&\simeq k f(a_p) \exp[-\frac{f(a_p)^2 \xi^2}{2p(1-p)}]
\end{aligned}$$

$$\sqrt{n}(X_{[np]} - a_p) \sim N[0, \frac{p(1-p)}{f(a_p)^2}]$$

Example.

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

MLE minimizes $\sum_i |x_i - \theta|$. The minimizer is the median $X_{[\frac{n+1}{2}]}$. By symmetry $a_{\frac{1}{2}} = \theta$. $f(a_{\frac{1}{2}}) = \frac{1}{2}$. Asymptotic variance of the median is roughly $\frac{1}{n}$. The derivative $\frac{\partial \log f}{\partial \theta} = \pm 1$. Therefore $I(\theta) = 1$. Cramér-Rao bound is $\frac{1}{n}$. The second derivative $\frac{\partial^2 \log f}{\partial \theta^2}$ does not exist. But still CLT is valid.

Confidence Intervals. We want to give an interval for the unknown parameter. Instead of just providing the estimate we want to say that θ is likely to be in A where A is an interval around the estimate. We would like A to be a small interval. But A is random and we would like the probability that A contains the true parameter (confidence level) to be high. Usually 95% or 99%.

Example. We have a sample of size n from a normal distribution with an unknown mean μ and variance 1. The sample mean is $\bar{x}$. .

$$P[\bar{x} - \frac{a}{\sqrt{n}} \le \mu \le \bar{x} + \frac{a}{\sqrt{n}}] = P[-\frac{a}{\sqrt{n}} \le \bar{x} - \mu \le \frac{a}{\sqrt{n}}]$$
$$= \frac{1}{\sqrt{2\pi}} \int_a^a \exp[-\frac{x^2}{2}]dx$$

Adjust a so that the the probability is .95 or .99 as the situation warrants. Higher the probability larger the value of a and hence longer the interval. Its size is of order $\frac{1}{n}$.

Example. We have a sample of size n from a normal distribution with an unknown mean μ and variance 1. the sample mean is $\bar{x}$. We have a sample of size n from a normal distribution with an unknown mean μ and unknown variance θ. The sample mean is $\bar{x}$ and the sample variance is $s^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2$. $\frac{ns^2}{\theta}$ has χ^2 distribution with $n-1$ degrees of freedom, i.e. it has the Gamma density $f_{\alpha,p}(x)$ with $\alpha = \frac{1}{2}$ and $p = \frac{n-1}{2}$. Pick u_1, u_2 so that with $\alpha = 0.95$ or 0.99

$$\int_{u_1}^{u_2} f_{\alpha,p}(x)dx = \alpha$$

Then the interval

$$u_1 \leq \frac{ns^2}{\theta} \leq u_2$$

with probability α turns into a confidence interval

$$\frac{ns^2}{u_2} \leq \theta \leq \frac{ns^2}{u_1}$$

for θ with a confidence level of α. u_1 and u_2 are usually picked so that

$$\int_{-\infty}^{u_1} f_{\alpha,p}(x)dx = \int_{u_2}^{\infty} f_{\alpha,p}(x)dx = \frac{1-\alpha}{2}$$

Example. We have a sample of size n from a normal distribution with an unknown mean μ and variance 1. the sample mean is $\bar{x}$. We have a sample of size n from a normal distribution with an unknown mean μ and unknown variance θ. The quantity $\frac{\bar{x}-\mu}{s}\sqrt{n-1}$ has the t-distribution $f_{n-1}(t)$ with $n-1$ degrees of freedom. Pick a so that with $\alpha = 0.95$ or 0.99

$$\int_{-a}^{a} f_{n-1}(t)dt = \alpha$$

Then the interval

$$-a \leq \frac{\bar{x}-\mu}{s}\sqrt{n-1} \leq a$$

with probability α turns into a confidence interval

$$\bar{x} - \frac{as}{n-1} \leq \mu \leq \bar{x} + \frac{as}{n-1}$$

for μ with a confidence level of α.