Likelihood Function for sampling from Normal Distribution

Example (vector case)

$$X_i \sim N(\mu, \sigma^2) \quad , \quad \theta = (\mu, \sigma^2)'$$

$$-\infty < \mu < \infty \; , \; \sigma^2 > 0$$

$$-\infty < x_i < \infty$$

Equivalent to model
$x_i = \mu + \epsilon_i$
ϵ_i iid $N(0, \sigma^2)$

$$f(x_i;\theta) = (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}(x_i-\mu)^2\right\}$$

Let $x_1, \ldots, x_n$ be an iid sample s.t. $X_i \sim N(\mu, \sigma^2)$

$i = 1, \ldots, n$. Then

$$L(\theta) = \prod_{i=1}^{n} (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}(x_i-\mu)^2\right\}$$

$$= (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

Remark: Let $\sigma^2 = 1$. Then

$$L(\theta) = L(\mu) = (2\pi)^{-n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

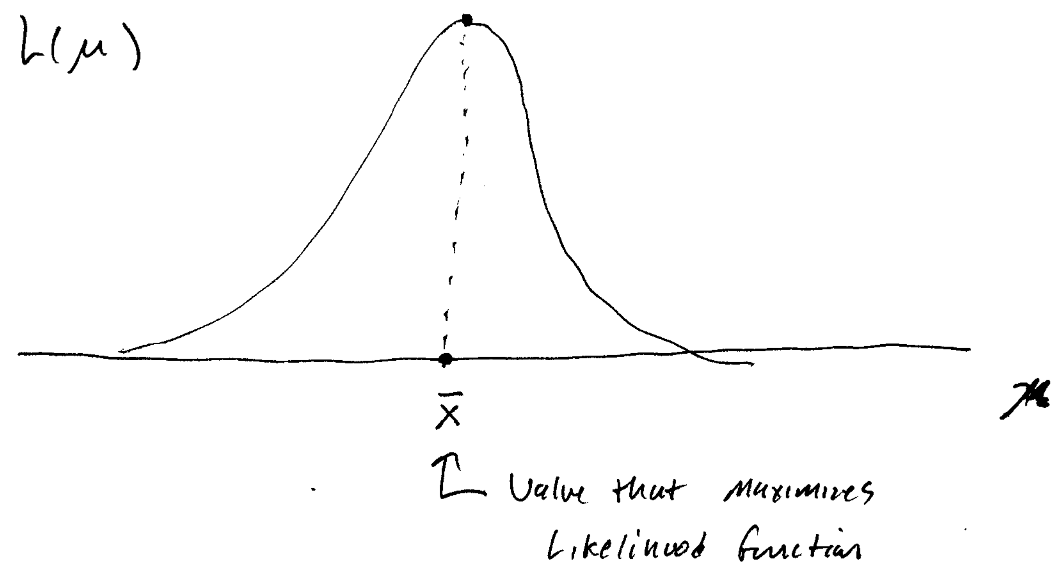
Now, $$\sum_{i=1}^{n}(x_i-\mu)^2 = \sum_{i=1}^{n}(x_i-\bar{x}+\bar{x}-\mu)^2 = \sum_{i=1}^{n}\left[(x_i-\bar{x})^2 + 2(x_i-\bar{x})(\bar{x}-\mu) + (\bar{x}-\mu)^2\right]$$

$$= \sum_{i=1}^{n}(x_i-\bar{x})^2 + n(\bar{x}-\mu)^2$$

So that

$$L(\mu) = (2\pi)^{-n/2} \exp\left\{ -\frac{1}{2}\left[\sum_{i=1}^{n} (x_i - \bar{x})^2 + n(\bar{x}-\mu)^2 \right] \right.$$

and it is clear that $L(\mu)$ is maximized at $\mu = \bar{x}$



Since the Normal Likelihood is regular we can find the MLE by maximizing the log likelihood:

$$\ln L(\theta) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

~~Here~~. Since $\theta = (\mu, \sigma^2)'$, $S(\theta)$ is a 2×1 vector:

$$S(\theta) = \begin{pmatrix} \dfrac{\partial \ln L(\theta)}{\partial \mu} \\[2ex] \dfrac{\partial \ln L(\theta)}{\partial \sigma^2} \end{pmatrix}$$ where

$$\ln L(\theta) = \frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2}(\sigma^2)^{-1}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\frac{\partial \ln L(\theta)}{\partial \mu} = \frac{-2}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)(-1) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)$$

$$\frac{\partial \ln L(\theta)}{\partial \sigma^2} = -\frac{n}{2}(\sigma^2)^{-1} + \frac{1}{2}(\sigma^2)^{-2}\sum_{i=1}^{n}(x_i-\mu)^2$$

Solving $S(\hat{\theta}_{MLE})=0$ gives the "normal equations"

$$\frac{\partial \ln L(\hat{\theta}_{MLE})}{\partial \mu} = \frac{1}{\hat{\sigma}^2_{MLE}}\sum_{i=1}^{n}(x_i-\hat{\mu}_{MLE}) = 0$$

$$\frac{\partial \ln L(\hat{\theta}_{MLE})}{\partial \sigma^2} = -\frac{n}{2}(\hat{\sigma}^2_{MLE})^{-1} + \frac{1}{2}(\hat{\sigma}^2_{MLE})^{-2}\sum_{i=1}^{n}(x_i-\hat{\mu}_{MLE}) = 0$$

The first equation gives

$$\sum_{i=1}^{n} x_i = n\hat{\mu}_{MLE} \Rightarrow \hat{\mu}_{MLE} = \frac{\sum_{i=1}^{n} x_i}{n} =$$

and the 2nd equation gives

$$\frac{n}{2}\left(\hat{\sigma}^2_{MLE}\right)^{-1} = \frac{1}{2}\left(\hat{\sigma}^2_{MLE}\right)^{-2}\sum_{i=1}^{n}(x_i-\hat{\mu}_{MLE})^2$$

$$\Rightarrow \hat{\sigma}^2_{MLE} = \frac{\sum_{i=1}^{n}(x_i-\hat{\mu}_{MLE})^2}{n} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}$$

Concentrating the Likelihood Function

Notice that solving

$$\frac{\partial \ln L(\theta)}{\partial \sigma^2} = -\frac{n}{2}(\sigma^2)^{-1} + \frac{1}{2}(\sigma^2)^{-2}\sum_1^n (x_i-\mu)^2 = 0$$

for σ^2 gives an expression in terms of μ:

$$\sigma^2(\mu) = \frac{1}{n}\sum_1^n (x_i-\mu)^2$$

Any value of $\sigma^2(\mu)$ defined this way satisfies the F.O.C. $\frac{\partial \ln L(\theta)}{\partial \sigma^2} = 0$.

If we substitute this value $\sigma^2(\mu)$ into the log-likelihood function for θ we get a "concentrated log-likelihood" function that is only a function of μ:

$$\ln L(\mu, \sigma^2(\mu)) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2(\mu))$$

$$-\frac{1}{2}(\sigma^2(\mu))^{-1}\sum_1^n (x_i-\mu)^2$$

$$= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\left(\frac{1}{n}\sum_1^n (x_i-\mu)^2\right) - \frac{1}{2}\left(\frac{1}{n}\sum_1^n (x_i-\mu)^2\right)^{-1}$$

$$* \sum_1^n (x_i-\mu)^2$$

$$= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\left(\frac{1}{n}\Sigma_1^n(x_i-\mu)^2\right)$$

$$-\frac{n}{2}$$

$$= -\frac{n}{2}(\ln(2\pi)+1) - \frac{n}{2}\ln\left(\frac{1}{n}\Sigma_1^n(x_i-\mu)^2\right)$$

$$\equiv \ln L^c(\mu) = \ln L(\mu, \sigma^2(\mu))$$

= concentrated likelihood function for μ

Now we can find the maximum likelihood estimate for μ by maximizing the concentrated log-likelihood function $\ln L^c(\mu)$

$$0 = \frac{\partial \ln L^c(\mu)}{\partial \mu} = -\frac{n}{2}\left(\frac{1}{\frac{1}{n}\Sigma_1^n(x_i-\mu)^2}\right) * \frac{2}{n}\cdot\Sigma_1^n(x_i-\mu)(\ $$

$$\Rightarrow \quad \frac{\Sigma_1^n(x_i-\mu)}{\frac{1}{n}\Sigma_1^n(x_i-\mu)^2} = 0$$

which is satisfied if $\mu = \bar{x}$ provided not all of the x_i are the same.

$$\ln f(\theta|x_i) = \frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma^2) - \frac{1}{2}(\sigma^2)^{-1}(x_i-\mu)^2$$

Note : The score for an observation is

$$S(\theta|x_i) = \begin{pmatrix} \dfrac{\partial \ln f(\theta|x_i)}{\partial \mu} \\[2ex] \dfrac{\partial \ln f(\theta|x_i)}{\partial \sigma^2} \end{pmatrix}$$

and

$$\frac{\partial \ln f(\theta|x_i)}{\partial \mu} = -\frac{2}{2}(\sigma^2)^{-1}(x_i-\mu)(-1)$$

$$= (\sigma^2)^{-1}(x_i-\mu)$$

$$\frac{\partial \ln f(\theta|x_i)}{\partial \sigma^2} = -\frac{1}{2}(\sigma^2)^{-1} + \frac{1}{2}(\sigma^2)^{-2}(x_i-\mu)^2$$

Hence it follows that

$$S(\theta|\underset{\sim}{x}) = \sum_{i=1}^{n} S(\theta|x_i)$$

The Hessian for an observation is

$$H(\theta|x_i) = \frac{\partial^2 \ln L(\theta|x_i)}{\partial\theta\,\partial\theta'} = \begin{bmatrix} \dfrac{\partial^2 \ln f(\theta|x_i)}{\partial \mu^2} & \dfrac{\partial^2 \ln f(\theta|x_i)}{\partial \mu\,\partial \sigma^2} \\[2ex] \dfrac{\partial^2 \ln f(\theta|x_i)}{\partial \sigma^2\,\partial \mu} & \dfrac{\partial^2 \ln f(\theta|x_i)}{\partial (\sigma^2)^2} \end{bmatrix}$$

$$\frac{\partial \ln f(\theta|x_i)}{\partial \mu^2} = -(\sigma^2)^{-1}$$

$$\frac{\partial \ln f(\theta|x_i)}{\partial \mu\, \partial \sigma^2} = -(\sigma^2)^{-2}(x_i-\mu)$$

$$\frac{\partial \ln f(\theta|x_i)}{\partial \sigma^2\, \partial \mu} = -(\sigma^2)^{-2}(x_i-\mu)$$

$$\frac{\partial \ln f(\theta|x_i)}{\partial (\sigma^2)^2} = +\frac{1}{2}(\sigma^2)^{-2} - (\sigma^2)^{-3}(x_i-\mu)^2$$

Now

$$I(\theta|x_i) = -E\left[H(\theta|x_i)\right]$$

$$= \begin{bmatrix} (\sigma^2)^{-1} & E[(x_i-\mu)](\sigma^2)^{-2} \\ E[(x_i-\mu)](\sigma^2)^{-2} & -\frac{1}{2}(\sigma^2)^{-2} + (\sigma^2)^{-3}E[(x_i-\mu)^2] \end{bmatrix}$$

Now

$$E[(x_i-\mu)] = 0$$

$$E\left[\left(\frac{(x_i-\mu)^2}{\sigma^2}\right)\right] = 1 \qquad \left(E[\chi^2(1)] = 1\right)$$

and so

$$I(\theta|x_i) = \begin{bmatrix} (\sigma^2)^{-1} & 0 \\ 0 & \frac{1}{2}(\sigma^2)^{-2} \end{bmatrix}$$

Hence

$$I(\theta|\underset{\sim}{x}) = \sum_{i=1}^{n} I(\theta|x_i)$$

$$= \begin{bmatrix} n(\sigma^2)^{-1} & 0 \\ 0 & \frac{n}{2}(\sigma^2)^{-2} \end{bmatrix}$$

Notes

(1) $I(\theta|\underset{\sim}{x}) \longrightarrow \infty$ as $n \to \infty$

(2) $I(\theta|\underset{\sim}{x})$ is larger the smaller is σ^2

(3) $I(\theta|\underset{\sim}{x})$ only depends on σ^2 and not μ.

(4) $I(\theta|\underset{\sim}{x})$ is block diagonal in θ