Chapter 7: Point Estimation

Assume that we know the family of dist'ns our random sample is drawn from:

$$\mathcal{N}(\mu,\sigma^2);\ \text{bin}(n,p);\ \text{Uniform}(\theta_1,\theta_2)\ \ldots$$

However, the parameters that define the specific dist'n are not known & we wish to estimate them

$$\underset{\sim}{\theta} = (\mu,\sigma^2);\quad \theta = p \text{ or } \underset{\sim}{\theta} = (n,p);\ \underset{\sim}{\theta} = (\theta_1,\theta_2)\ \ldots$$

Definition 7.1.1 A *point estimator* is any function $W(X_1,\ldots,X_n)$ of a sample; that is, any statistic is a point estimator.

Estimat<u>or</u> = $W(X_1, X_2, \ldots, X_n)$ = Function of RVs.
Estimat<u>e</u> = $W(x_1, x_2, \ldots, x_n)$ = Realized value

Methods of finding Estimators

§7.2.1 Method of Moments

$X_1,\ldots X_n$ iid $\sim f(x|\theta_1,\ldots\theta_k)$ k unknown parameters

sample moments

$$\left\{\begin{array}{ll} m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i^1, & \mu'_1 = \mathrm{E}X^1, \\ m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2, & \mu'_2 = \mathrm{E}X^2, \\ \vdots & \\ m_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k, & \mu'_k = \mathrm{E}X^k. \end{array}\right\}$$

population moments

set sample moments equal to pop'n moments:

$$m_1 = \mu'_1(\theta_1,\ldots,\theta_k),$$
$$m_2 = \mu'_2(\theta_1,\ldots,\theta_k),$$
$$\vdots$$
$$m_k = \mu'_k(\theta_1,\ldots,\theta_k).$$

k eq'ns
k unknowns

solve for $\theta_1,\ldots,\theta_k$

Method of Moments

Advantages
- easy to find
- good properties if n large
  $\tilde{m}_r \xrightarrow{P} m_r$ by WLLN
- $E[\tilde{m}_r] = m_r$

Disadvantages
- Not always the 'best' estimate
  → Can be improved upon
- Range of estimator, may not coincide with the range of the parameter it is estimating

§ 7.2.2 Maximum Likelihood Estimators (MLEs)

Recall that if $X_1, \ldots, X_n$ are an iid sample from a population with pdf or pmf $f(x|\theta_1, \ldots, \theta_k)$, the likelihood function is defined by

$$L(\theta|\mathbf{x}) = L(\theta_1, \ldots, \theta_k|x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i|\theta_1, \ldots, \theta_k). \tag{7.2.3}$$

Definition 7.2.4 For each sample point $\mathbf{x}$, let $\hat{\theta}(\mathbf{x})$ be a parameter value at which $L(\theta|\mathbf{x})$ attains its maximum as a function of θ, with $\mathbf{x}$ held fixed. A *maximum likelihood estimator* (MLE) of the parameter θ based on a sample $\mathbf{X}$ is $\hat{\theta}(\mathbf{X})$.

Intuitively MLE is a 'reasonable' choice

→ Parameter point for which the observed sample is most likely.

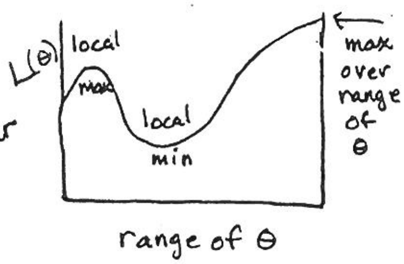
Drawbacks: 1) Finding a global max (over range of Θ) & verifying that it is a global max.
2) Numerical Sensitivity - may be sensitive to small changes in the data. i.e. slightly different samples → vastly different mles.

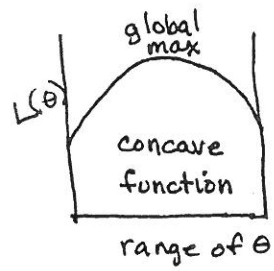
Finding MLEs: If differentiable solve:

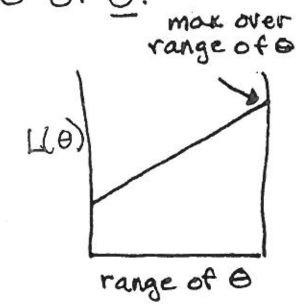
$$\frac{\partial}{\partial \theta_i} L(\theta_1, \ldots, \theta_k) = 0 \qquad i = 1, \ldots, k$$

and demonstrate resulting value is the max over the range of $\underline{\theta}$.

simple case θ = scalar







- derivative = 0 at min, max, inflection point.
- must also check extrema (i.e. check boundaries on Θ, the parameter space.)

Recall Example:

$X_1, \ldots, X_n$ iid Bernoulli (p)

$$L(p|\underline{x}) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} \qquad x_i = \begin{matrix} 1 \text{ success} \\ 0 \text{ failure} \end{matrix}$$

$$= p^{\Sigma x_i}(1-p)^{n-\Sigma x_i}$$

$$\log L(p|\underline{x}) = \Sigma x_i \log(p) + (n - \Sigma x_i)\log(1-p) \qquad \boxed{0 < \Sigma x_i < n}$$

$$\frac{\partial}{\partial p} \log L(p|\underline{x}) = \Sigma x_i / p + (n - \Sigma x_i)/(1-p) * (-1)$$

$$\frac{\Sigma x_i}{p} - \frac{(n - \Sigma x_i)}{(1-p)} = 0 \implies \hat{p} = \frac{\Sigma x_i}{n} = \frac{\# \text{ success}}{\# \text{ trials}}$$

$$\frac{\partial^2 \log L(p|\underline{x})}{\partial p^2} = \overbrace{-\frac{\Sigma x_i}{p^2}}^{<0} + \overbrace{\frac{(\Sigma x_i - n)}{(1-p)^2}}^{<0} < 0 \quad \text{for } 0 < \Sigma x_i < n$$

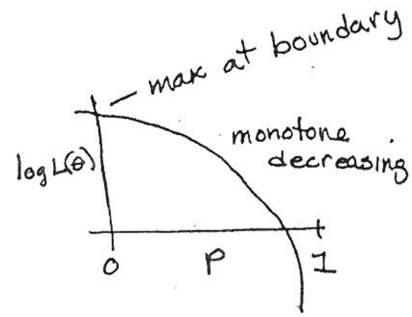
Masters
Theory III

Lecture 11/4

Example cont.

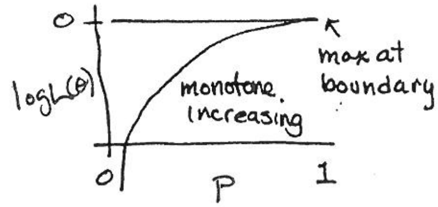
If $\Sigma x_i = 0$ $\quad \log L = n \log(1-p)$

$$\frac{\partial \log L}{\partial p} = \frac{-n}{(1-p)} = 0 \quad ?$$



If $\Sigma x_i = n$ $\quad \log L = n \log(p)$

$$\frac{\partial \log L}{\partial p} = \frac{n}{p} = 0 \quad ?$$



Finding MLEs by 'inspection'

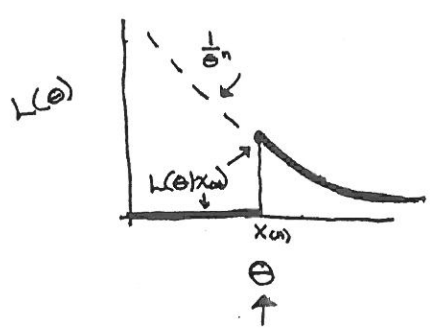
example: $X_1, \dots X_n \sim U(0, \theta)$ $\quad \theta > 0; \quad 0 \leq X_i \leq \theta$

$$f(x_i|\theta) = \frac{1}{\theta} I(\theta - x_i) \quad \text{where } I(\theta - x_i) = \begin{matrix} 1 & \theta - x_i \geq 0 \ (\theta \geq x_i) \\ 0 & \text{else} \end{matrix}$$

$$f(\underset{\sim}{x}|\theta) = \frac{1}{\theta^n} \prod_{i=1}^{n} I(\theta - x_i)$$

$$= \frac{1}{\theta^n} I(\theta - x_{(n)}) \quad \text{where } I(\theta - x_{(n)}) = \begin{matrix} 1 & \theta - x_{(n)} \geq 0 \\ 0 & \text{else} \end{matrix}$$

$$L(\theta|\underset{\sim}{x}) = \frac{1}{\theta^n} I(\theta - x_{(n)})$$



$$\hat{\theta} = x_{(n)}$$

Masters
Theory III

Lecture 11/5

CAUTION! Finding MLEs using a maximization process, susceptible to problems in the process, among them Numerical Instability.

Estimate likelihood parameter, θ. $\underline{x}$ held constant, but the data are measured with error. How do small changes in the data affect the MLE?

i.e. $\hat{\theta}_1$ based on $L(\theta \mid \underline{x}_1)$

vs $\hat{\theta}_2$ based on $L(\theta \mid \underline{x}_1 + \varepsilon)$ for small ε.

Intuitively, $\hat{\theta}_1$ + $\hat{\theta}_2$ should be close.

Example 7.2.13 (continuation of example 7.2.2)

bin(k,p) where both k + p are unknown.
MLE is messy, but it exists (example 7.2.9)

5 realizations of the data:

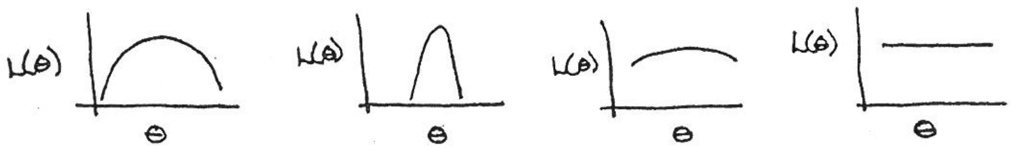
$(16, 18, 22, 25, 2\underline{7})$ $\hat{k} = 99$

$(16, 18, 22, 25, 2\underline{8})$ $\hat{k} = 190$!!

Such occurrences happen when the likelihood function is very flat in the neighborhood of its maximum or when there is no finite maximum.

Masters
Theory III

Lecture 11/6



Which Likelihood ft'n gives us the most precise info about θ?

hmmm. Which estimate has smallest variance? (more later)

MLEs need not be unique

danger, 2 sufficient stats, 1 parameter

example: $X_1, \ldots, X_n$ iid $U[\theta - 1/2, \theta + 1/2]$

$$f(x_i \mid \theta - 1/2, \theta + 1/2) = \frac{1}{(\theta + 1/2) - (\theta - 1/2)} = 1 \cdot I_{(\theta - 1/2 < x_i < \theta + 1/2)}$$

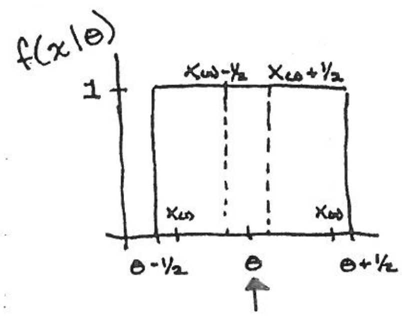
$$L(\theta \mid \underline{x}) = f(\underline{x} \mid \theta) = I_{(X_{(1)} \geq \theta - 1/2 \,;\, X_{(n)} \leq \theta + 1/2)}$$

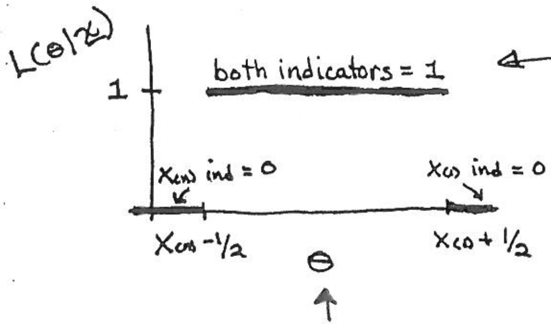
$$= I_{(\theta \leq X_{(1)} + 1/2)} \, I_{(X_{(n)} - 1/2 \leq \theta)}$$

$$= I(X_{(1)} - \theta + 1/2) \, I(\theta + 1/2 - X_{(n)})$$

where $I(X_{(1)} - \theta + 1/2) = \begin{cases} 1 & X_{(1)} - \theta + 1/2 > 0 \\ 0 & \text{else} \end{cases}$

$I(\theta + 1/2 - X_{(n)}) = \begin{cases} 1 & \theta + 1/2 - X_{(n)} > 0 \\ 0 & \text{else} \end{cases}$





any value between $X_{(n)} - 1/2$ + $X_{(1)} + 1/2$ has same likelihood + is the max.

Masters
Theory III

Lecture- 11/7

MLEs (in general) are functions of sufficient statistics

Proof: $L(\theta|\underset{\sim}{x}) = f(\underset{\sim}{x}|\theta) = g(T(x)|\theta)\, h(\underset{\sim}{x})$

$L(\theta|\underset{\sim}{x}) \propto g(T(x)|\theta)$ we can drop $h(\underset{\sim}{x})$, since it is constant wrt θ.

Exception: When MLE is not unique.

Recall:

Theorem 6.2.6 (Factorization Theorem) *Let $f(\mathbf{x}|\theta)$ denote the joint pdf or pmf of a sample $\mathbf{X}$. A statistic $T(\mathbf{X})$ is a sufficient statistic for θ if and only if there exist functions $g(t|\theta)$ and $h(\mathbf{x})$ such that, for all sample points $\mathbf{x}$ and all parameter points θ,*

$$f(\mathbf{x}|\theta) = g(T(\mathbf{x})|\theta)h(\mathbf{x}). \quad (6.2.3)$$

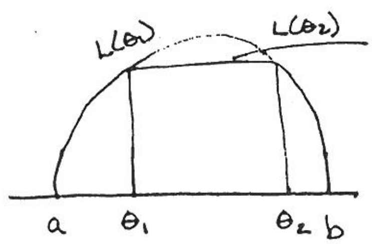
Brief Review of Concave Functions

Concave Function: A function $L(\theta)$ is strictly concave on an interval (a,b) if for any θ_1 & $\theta_2 \in (a,b)$ and any $\gamma \in [0,1]$ then

$$L(\gamma\theta_1 + (1-\gamma)\theta_2) > \gamma L(\theta_1) + (1-\gamma)L(\theta_2).$$

If $\gamma = 1/2$

$$L\left(\tfrac{1}{2}(\theta_1+\theta_2)\right) > \tfrac{1}{2}\left[L(\theta_1) + L(\theta_2)\right]$$



line joining 2 points lies underneath curve.

→ An equivalent condition is that $L'(\theta_1) > L'(\theta_2)$ $\forall \theta_1 < \theta_2$.
  → derivative is always decreasing

→ → A sufficient condition for concavity is $L''(\theta) < 0$ $\forall \theta$.

→ → If concave ft'n has a maximum, then it is unique.

Masters

Theory III

Lecture - 11/8

☆ <u>Theorem</u>: The loglikelihood associated with a one parameter exponential family in natural form is concave in the natural parameter.

Proof relies on $Var(t_j(X)) = -\frac{\partial^2}{\partial \eta_j^2} \log c^*(\eta)$

(similar to homework #3.32 ~~- last semester~~ first homework.)

$$* \quad f(x|\eta) = h(x)c^*(\eta)\exp\left(\sum_{i=1}^{k} \eta_i t_i(x)\right).$$

3.32 (a) If an exponential family can be written in the form $*$ show that the identities of Theorem 3.4.2 simplify to

$$E(t_j(X)) = -\frac{\partial}{\partial \eta_j} \log c^*(\eta),$$

$$Var(t_j(X)) = -\frac{\partial^2}{\partial \eta_j^2} \log c^*(\eta).$$

appendix

Proof: 1 parameter exponential family in natural form is concave (unique max)

exponential family $f(x|\theta) = c(\theta)h(x)\exp[t(x)w(\theta)]$

natural form $f(x|\eta) = c^*(\eta)h(x)\exp[t(x)\eta]$ $(\eta = w(\theta))$

let $c^{**}(\eta) = \log(c^*(\eta))$ & $h^*(x) = \log(h(x))$

then $L(\eta|\underline{x}) = \exp[\eta t(x) + c^{**}(\eta) + h^*(x)]$ — another form of a natural param exp family.

$$\log L(\eta|\underline{x}) = \eta t(x) + c^{**}(\eta) + h^*(x)$$

$$\frac{\partial}{\partial \eta} \log L(\eta|\underline{x}) = t(x) + \frac{\partial}{\partial \eta} c^{**}(\eta)$$

Masters
Theory III

Lecture 11/9

Proof continued:

$$\frac{\partial}{\partial\eta}\log L(\eta|\underline{x}) = t(x) + \frac{\partial}{\partial\eta}c^{**}(\eta)$$

$$\frac{\partial^2}{\partial\eta^2}\log L(\eta|\underline{x}) = \frac{\partial^2}{\partial\eta^2}c^{**}(\eta) = \frac{\partial^2}{\partial\eta^2}\log c^*(\eta)$$

Therefore

$$\frac{\partial^2}{\partial\eta^2}\log L(\eta|\underline{x}) = \frac{\partial^2}{\partial\eta^2}\log c^*(\eta)$$

from hw#3.32

$$\frac{\partial^2}{\partial\eta^2}\log L(\eta|\underline{x}) = \frac{\partial^2}{\partial\eta^2}\log c^*(\eta) = -\text{Var}(t(x))$$
$$< 0 \quad \forall \theta$$

$\rightarrow$ Concave.

Nice Property

Theorem 7.2.10 (Invariance property of MLEs) *If $\hat\theta$ is the MLE of θ, then for any function $\tau(\theta)$, the MLE of $\tau(\theta)$ is $\tau(\hat\theta)$.*

proof C&B page 320.

examples

Assume $X_1,\dots,X_n$ iid $N(\mu,\sigma^2)$ $\quad \hat\mu = \bar X$

Want MLE for $\mu^2 = \bar X^2 = (\hat\mu)^2$

Assume $X_1,\dots,X_n$ iid binomial (n,p) $\quad \hat p = \Sigma x_i/n$

MLE for $p(1-p) = \frac{\Sigma x_i}{n}\left(1-\frac{\Sigma x_i}{n}\right) = \hat p(1-\hat p)$

MLE for $\sqrt{p(1-p)} = \sqrt{\frac{\Sigma x_i}{n}\left(1-\frac{\Sigma x_i}{n}\right)} = \sqrt{\hat p(1-\hat p)}$

Masters
Theory III

Lecture 11/10

examples cont.

Assume $X_1, \ldots, X_n \sim U(\theta_1, \theta_2)$

ICBST $\hat{\theta}_1 = X_{(1)}$ $\quad \hat{\theta}_2 = X_{(n)}$

MLE of $\theta_2 - \theta_1 = X_{(n)} - X_{(1)} = \hat{\theta}_2 - \hat{\theta}_1$

Note: Invariance property holds even if θ is a vector
MLE of $(\theta_1, \theta_2, \ldots \theta_K) = (\hat{\theta}_1, \hat{\theta}_2, \ldots \hat{\theta}_K)$

Then MLE of $\tau(\theta_1, \ldots \theta_K) = \tau(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_K)$

Corollary to Theorem ☆

i) If $\frac{\partial}{\partial \eta} \operatorname{Log} L(\eta | x)$ has a solution in the interior of the natural parameter space, it is the unique MLE.

ii) The likelihood eq'n $\frac{\partial}{\partial \eta} \log L(\eta | x) = 0$ is equivalent to setting $t(x) = E[t(x)] \rightarrow$ solve for η

iii) Because functions of MLEs are MLEs, if the sol'n to $\frac{\partial}{\partial \theta} \log L(\theta | x) = 0$ exists then $\hat{\theta}$ is unique.

↑ uniqueness property holds whether or not natural parameter space.

example: $X_1, \ldots, X_n$ iid $\exp(\beta)$ $\quad f(x|\beta) = (1/\beta) e^{-x/\beta}$ $\quad \beta > 0$, $0 \le x < \infty$.

exponential family: $f(\underline{x}|\beta) = (1/\beta) \exp[(-1/\beta) \Sigma x_i]$ $\left\{ \begin{array}{ll} h(x) = 1 & t(x) = -\Sigma x_i \\ c(\beta) = 1/\beta & w(\beta) = (1/\beta) \end{array} \right.$

natural parameter : $f(\underline{x}|\eta) = \eta \exp[-\eta \Sigma x_i]$ $\left\{ \begin{array}{ll} h(x) = 1 & t(x) = \Sigma x_i \\ c^*(\eta) = \eta & \end{array} \right.$

set $t(x) = E_\eta[t(x)]$ $\quad t(x) = -\Sigma x_i$ $\quad E_\eta[t(x)] = -n/\eta$

$-\Sigma x_i = -n/\eta \Rightarrow \hat{\eta} = n/\Sigma x_i$

since ftns of mles are also mles: $\beta = 1/\eta$ $\quad \hat{\beta} = 1/\hat{\eta}$ $\quad \hat{\beta} = \bar{X}$

Masters
Theory III

Lecture 11/11

Finding MLEs when $\underset{\sim}{\theta}$ is a vector. - more general than C+Bs approach

$\theta = (\theta_1, \ldots \theta_K)$

MLE $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_K)$ obtained by maximizing $L(\theta | \underset{\sim}{x})$
or $\log L(\theta | \underset{\sim}{x})$.

Find $\underset{\sim}{\hat{\theta}}$ $\begin{cases} \text{1. by inspection} \\ \text{2. by computer search} \quad \text{- more later} \\ {}^*\text{3. using calculus} \end{cases}$

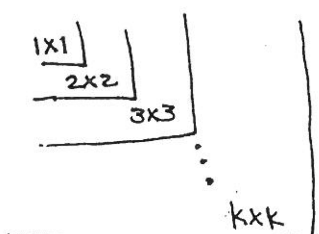
Using calculus to find MLEs when θ is a vector:

- Set first partial derivatives $= 0$

$$\frac{\partial}{\partial \theta_i} \log(\underset{\sim}{\theta} | \underset{\sim}{x}) = 0 \quad i = 1, 2, \ldots K \quad \begin{cases} K \text{ eq'ns} \\ K \text{ unknown} \end{cases}$$

- Must Check that the solution is a Max!
  - → Check that the 2^{nd} derivative matrix (Hessian) is negative definite.
  - → Easiest way to show negative definite is to show that the negative of the matrix is positive definite

positive definite



$\boxed{\text{show all determinants are} > 0}$

If determinant $= 0$ can't be sure if min, max or inflection point.

$K=2$
show negative of 2^{nd} der matrix is pos definite

$$-\begin{bmatrix} \partial^2 \log L(\underline{\theta}|\underline{x}) / \partial\theta_1^2 & \partial^2 \log L(\underline{\theta}|\underline{x}) / \partial\theta_1 \partial\theta_2 \\ \partial^2 \log L(\underline{\theta}|\underline{x}) / \partial\theta_1 \partial\theta_2 & \partial^2 \log L(\underline{\theta}|\underline{x}) / \partial\theta_2^2 \end{bmatrix}$$

$$-\frac{\partial^2 \log L(\underline{\theta}|\underline{x})}{\partial\theta_1^2} > 0$$

and

$$\left(-\frac{\partial^2 \log L(\underline{\theta}|\underline{x})}{\partial\theta_1^2}\right)\left(-\frac{\partial^2 \log L(\underline{\theta}|\underline{x})}{\partial\theta_2^2}\right) - \left(\frac{\partial^2 \log L(\underline{\theta}|\underline{x})}{\partial\theta_1 \partial\theta_2}\right)^2 > 0$$

example: $X_1, \ldots, X_n$ iid $N(\mu, \sigma^2)$

$$\log L(\theta|\underline{x}) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2}\Sigma(x_i-\mu)^2/\sigma^2$$

$$\frac{\partial \log L(\theta|\underline{x})}{\partial\mu} = -\frac{1}{2}\cdot 2 \cdot \frac{\Sigma(x_i-\mu)}{\sigma^2}(-1) = 0$$ set equal to zero

$$\frac{\Sigma(x_i-\mu)}{\sigma^2} = 0$$

$$\Sigma x_i - n\mu = 0 \qquad \boxed{\hat{\mu} = \frac{\Sigma x_i}{n}}$$

$$\frac{\partial \log L(\theta|\underline{x})}{\partial\sigma^2} = -\frac{n}{2}\frac{1}{(2\pi\sigma^2)}(2\pi) - \frac{1}{2}\Sigma(x_i-\mu)^2(-1)(\sigma^2)^{-2}$$

Note derivative wrt σ^2 not σ.

$$\frac{n}{2\sigma^2} = \frac{\Sigma(x_i-\mu)^2}{2(\sigma^2)^2}$$ set equal to zero

$$\frac{(\sigma^2)^2}{\sigma^2} = \frac{\Sigma(x_i-\mu)^2}{n} \qquad \hat{\mu} = \bar{X}$$

$$\boxed{\hat{\sigma}^2 = \frac{\Sigma(x_i-\bar{X})^2}{n}}$$

Masters
Theory III

Lecture 11/13

$$\hat{\mu} = \Sigma x_i/n = \overline{X}$$
$$\hat{\sigma}^2 = \Sigma(x_i - \overline{X})^2/n$$

Same as Method of Moments

Note: $\hat{\sigma}^2$ is biased !?
$E[S^2] = \sigma^2$ where $S^2 = \frac{\Sigma(X_i - \overline{X})^2}{(n-1)}$

Check that $\hat{\mu}$ & $\hat{\sigma}^2$ maximize $\log L(\theta|\underline{x})$

$$\frac{\partial^2 \log L(\theta|\underline{x})}{\partial \mu^2} = \frac{\partial}{\partial \mu}\left(\frac{1}{\sigma^2}\left[\Sigma x_i - n\mu\right]\right)$$
$$= -n/\sigma^2$$

$$\frac{\partial^2 \log L(\theta|\underline{x})}{\partial \mu \, \partial \sigma^2} = \frac{\partial}{\partial \sigma^2}\left(\frac{1}{\sigma^2}\Sigma(x_i - \mu)\right)$$
$$= \frac{-1}{(\sigma^2)^2}\Sigma(x_i - \mu)$$

$$\frac{\partial^2 \log L(\theta|\underline{x})}{(\partial \sigma^2)^2} = \frac{\partial}{\partial \sigma^2}\left(\frac{-n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\Sigma(x_i - \mu)^2\right)$$
$$= \frac{n}{2(\sigma^2)^2} + \frac{1}{2}\Sigma(x_i - \mu)^2(-2)(\sigma^2)^{-3}$$

2nd derivative matrix :

$$\begin{bmatrix} -n/\sigma^2 & -1/\sigma^4\,\Sigma(x_i - \mu) \\ -1/\sigma^4\,\Sigma(x_i - \mu) & \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\Sigma(x_i - \mu)^2 \end{bmatrix}$$

Show matrix is negative definite [negative of matrix is positive definite]

$$-\frac{\partial^2 \log L(\theta|x)}{\partial\theta_i \partial\theta_j} = \begin{bmatrix} +n/\sigma^2 & +1/\sigma^4 \Sigma(x_i-\mu) \\ +1/\sigma^4 \Sigma(x_i-\mu) & -\frac{n}{2\sigma^4}+\frac{1}{\sigma^6}\Sigma(x_i-\mu)^2 \end{bmatrix}$$

$$1\times1 \text{ det} = n/\sigma^4 > 0$$

$$2\times2 \text{ det} = \left[\frac{-n^2}{2\sigma^6} + \frac{n}{\sigma^8}\Sigma(x_i-\mu)^2 - \left(\frac{1}{\sigma^4}\Sigma(x_i-\mu)\right)^2\right]$$

$$= \frac{1}{\sigma^6}\left[\frac{-n^2}{2} + \frac{n}{\sigma^2}\Sigma(x_i-\mu)^2 - \frac{1}{\sigma^2}\left(\Sigma(x_i-\mu)\right)^2\right]$$

evaluate at
$\mu = \bar{x}$
$\sigma^2 = \hat{\sigma}^2$

$$= \frac{1}{\hat{\sigma}^6}\left[\frac{-n^2}{2} + \frac{n}{\hat{\sigma}^2}\Sigma(x_i-\bar{x})^2 * \frac{n}{n} - \frac{1}{\hat{\sigma}^2}\left(\Sigma(x_i-\bar{x})\right)^2\right]$$

$$= \frac{1}{\hat{\sigma}^6}\left[\frac{-n^2}{2} + \frac{n^2}{\hat{\sigma}^2}\hat{\sigma}^2 - \frac{1}{\hat{\sigma}^2}\left(\Sigma x_i - n\bar{x}\right)^2\right]$$

$$= \frac{1}{\hat{\sigma}^6}\left[\frac{-n^2}{2} + \frac{2n^2}{2} - 0\right]$$

$$= \frac{1}{\hat{\sigma}^6}\left[\frac{n^2}{2}\right] > 0$$

1×1 det > 0
2×2 det > 0 $\Rightarrow$ 2^{nd} derivative matrix is negative definite

$\hat{\mu}$ + $\hat{\sigma}^2$ are MLEs.

Exponential Families - Review

$$f(x|\underset{\sim}{\theta}) = h(x)\, c(\underset{\sim}{\theta}) \exp\left(\sum_{i=1}^{k} w_i(\underset{\sim}{\theta})\, t_i(x)\right)$$

An exponential family is often reparameterized as

$$f(x|\underset{\sim}{\eta}) = h(x)\, c^*(\underset{\sim}{\eta}) \exp\left(\sum_{i=1}^{k} \eta_i\, t_i(x)\right)$$

$h(x)$, $t(x)$, same as original

$$\eta_i = w_i(\theta) \quad i=1,\dots,k$$

→ Natural parameter space.

example: Find natural parameter space for $N(\mu, \sigma^2)$ family.

$$f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{\mu^2}{2\sigma^2}\right\} \exp\left\{\frac{-x^2}{2\sigma^2} + \frac{\mu x}{\sigma^2}\right\}$$

$h(x) = 1$

$c(\underset{\sim}{\theta}) = c(\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-\mu^2}{2\sigma^2}\right) \quad -\infty < \mu < \infty, \sigma > 0$

$w_1(\theta) = 1/\sigma^2 \qquad t_1(x) = -x^2/2$

$w_2(\theta) = \mu/\sigma^2 \qquad t_2(x) = x$

natural form: $\eta_1 = 1/\sigma^2 \quad \eta_2 = \mu/\sigma^2 \quad \sigma = \sqrt{\frac{1}{\eta_1}} \quad \mu = \eta_1 \eta_2$

$$f(x|\eta_1,\eta_2) = \frac{\sqrt{\eta_1}}{\sqrt{2\pi}} \exp\left(\frac{-\eta_2^2}{2\eta_1}\right) \exp\left(-\frac{\eta_1 x^2}{2} + \eta_2 x\right) \quad \begin{array}{l}\eta_1 > 0 \\ -\infty < \eta_2 < \infty\end{array}$$

Homework # 3.32 $j=1$

Show $E[t(X)] = -\frac{\partial}{\partial\eta}\log(c^*(\eta))$

$$\int c^*(\eta)\, h(x) \exp[\eta\, t(x)]\, dx = 1$$

and $\frac{d}{d\eta}\int c^*(\eta)\, h(x) \exp[\eta t(x)]\, dx = 0$

- one property of exponential families, is that we can interchange $\int$ & $\frac{\partial}{\partial\theta}$ order. } see C&B §2.4

$$\therefore \int \frac{\partial}{\partial\eta}[c^*(\eta)\, h(x) \exp[\eta t(x)]\, dx = 0$$

$$0 = \int \left[\frac{\partial}{\partial\eta} c^*(\eta)\right] h(x) \exp[\eta\, t(x)]\, dx$$

$$+ \int c^*(\eta)\, h(x) \left[t(x) \exp[\eta\, t(x)]\right] dx$$

$$0 = \frac{\partial}{\partial\eta} c^*(\eta) \int h(x) \exp[\eta\, t(x)]\, dx$$

$$+ E[t(X)]$$

$$= \frac{\partial}{\partial\eta} c^*(\eta) \frac{1}{c(\eta)} \overbrace{\int c(\eta)\, h(x) \exp[\eta\, t(x)]\, dx}^{1}$$

$$+ E[t(X)]$$

$$E[t(X)] = -\frac{\partial}{\partial\eta} c^*(\eta) \Big/ c^*(\eta) = -\frac{\partial}{\partial\eta}\log c^*(\eta) \quad //$$