Stat 543 2-7-05

Correction: "obviously" in the Poisson example I should have written that one should really check that $\overline{X}$ maximizes $L(\lambda)$ and is not just a solution to $\frac{d}{d\lambda}\log L(\lambda) = 0$

Example $X_1, X_2, \ldots, X_n$ iid $N(\mu, \sigma^2)$

$Y_i = X_i$ rounded to nearest integer

discrete with marginal pmf

$$f(y \mid \mu, \sigma^2) = \Phi\left(\frac{y + .5 - \mu}{\sigma}\right) - \Phi\left(\frac{y - .5 - \mu}{\sigma}\right)$$

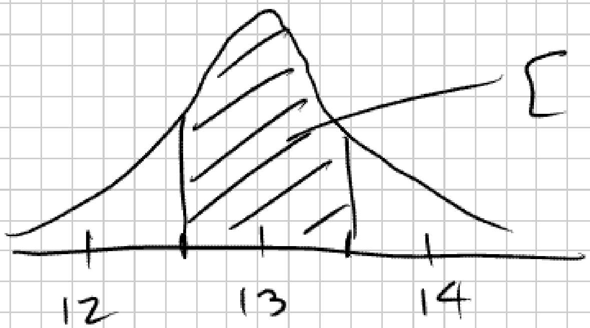
for integer y

The joint pmf

$$f(y \mid \mu, \sigma^2) = \prod_{i=1}^{n} \left[\Phi\left(\frac{y_i + .5 - \mu}{\sigma}\right) - \Phi\left(\frac{y_i - .5 - \mu}{\sigma}\right) \right]$$

Suppose all $y_i = 13$

$$L(\mu, \sigma^2) = f(y \mid \mu, \sigma^2) = \left[\Phi\left(\frac{13.5 - \mu}{\sigma}\right) - \Phi\left(\frac{12.5 - \mu}{\sigma}\right) \right]^n$$



[ ]

with $\mu \in (12.5, 13.5)$

letting $\sigma \to 0$

will make $[\quad] \to 1$

$L(\mu, \sigma^2) \to 1$

$L(\mu, \sigma^2) \le 1$ — strictly speaking, there is no MLE — even if I allowed $\sigma = 0$ the MLE would not be unique

Both in terms computations typically employed to "maximize" a likelihood and in terms of theory available to describe maximum likelihood a slightly indirect approach is common

Suppose $\Theta \subset \mathbb{R}^k$ — if $L(\theta)$ is differentiable then a necessary condition for θ^* to maximize $L(\theta)$ is that all 1st partials of

$$l(\theta) = \log L(\theta)$$

are 0 at θ^* — i.e.

$$\left.\frac{\partial}{\partial\theta_1}\ell(\theta)\right|_{\theta=\theta^*} = 0$$

$$\vdots$$

$$\left.\frac{\partial}{\partial\theta_k}\ell(\theta)\right|_{\theta=\theta^*} = 0$$

likelihood equations

score function

$$\nabla\ell(\theta^*) = 0$$

why the log? maximizing ℓ and L are equivalent and in iid models the logarithm will turn products into sums which makes things more tractable

<u>Example</u> $X \sim Bi(5, p)$

$$L(p) = \binom{5}{x} p^{x} (1-p)^{5-x}$$

$$\ell(p) = \log\binom{5}{x} + x \log p + (5-x)\log(1-p)$$

$$\ell'(p) = \frac{x}{p} - \frac{(5-x)}{1-p}$$

score function

setting $\ell'(p) = 0$ is a way to look for a maximizer of $L(p)$

One context in which the likelihood equations are especially nice is that of an exponential family – for $X_1, X_2, \ldots, X_n$ iid (B+D form)

$$\underline{L(\eta)} = f(x|\eta) = \left[\prod_{i=1}^{n} h(x_i)\right] \exp -nA(\eta) \exp \sum_{j=1}^{k} \eta_j \left(\sum_{i=1}^{n} T_j(x_i)\right)$$

and $\underline{\ell(\eta)} = \log f(x|\eta)$

$$= \sum \log h(x_i) - nA(\eta) + \sum_{j=1}^{k} \eta_j \left(\sum_{i=1}^{n} T_j(x_i)\right)$$

So

$$\frac{\partial}{\partial \eta_j} \log f(\underline{x} \mid \eta) = -n \frac{\partial}{\partial \eta_j} A(\eta) + \sum_{i=1}^{n} T_j(x_i)$$

page 59 of B+D by Corollary 1.6.1 : if $\mathcal{E}$ has nonempty interior this is $E_\eta T_j(X)$

So the jth of the likelihood equations is

$$\underbrace{-n \frac{\partial}{\partial \eta_j} A(\eta)}_{E_\eta T_j(X)} + \sum_{i=1}^{n} T_j(x_i) = 0$$

jth likelihood equation

$$E_\eta T_j(X) = \frac{1}{n} \sum_{i=1}^{n} T_j(x_i)$$

i.e. the plan is to set theoretical means of $T_j(X)$ equal to empirical/sample means and try to solve

one must worry some in an application whether a solution to the likelihood equations is really a maximizer, whether it is unique, etc. — one result given in B+D in this direction

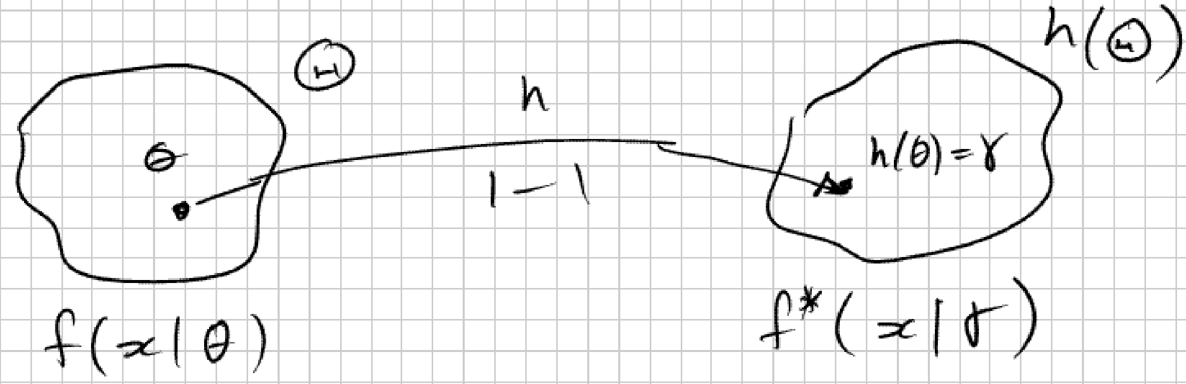
Corollary 2.3.2 If the equations

$$E_\eta T_j(X) = \frac{1}{n}\sum_{i=1}^{n} T_j(x_i) \quad j=1,\ldots,k$$

have a solution $\eta \in$ interior of $\mathcal{E}$, it is the unique MLE of η

??? -- what if I don't like η but prefer θ

(problem 2.2.16a)



"clearly" if $\hat{\theta}$ maximizes $f(x|\theta)$ over Θ
then $\hat{\gamma} = h(\hat{\theta})$ maximizes $f^*(x|\gamma)$ over $h(\Theta)$