Lec 12 Math 341 3/19/10

to derive Jeffreys' protocol we need
(a) kernels
(b) Fisher Info

Jeffreys' priors

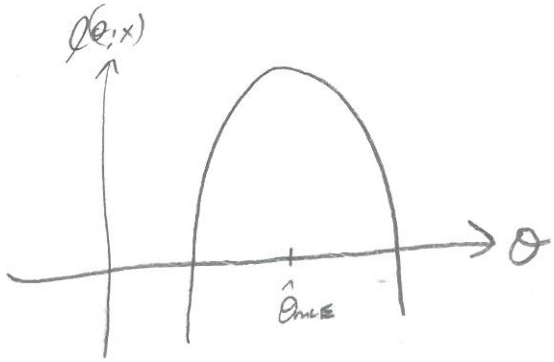
$$P(X|\theta) \xrightarrow{\text{Jeffreys' protocol}} P(\theta)$$
$$\phi = t(\theta)$$
$$P(X|\phi) \xrightarrow{\text{'' ''}} P(\phi)$$

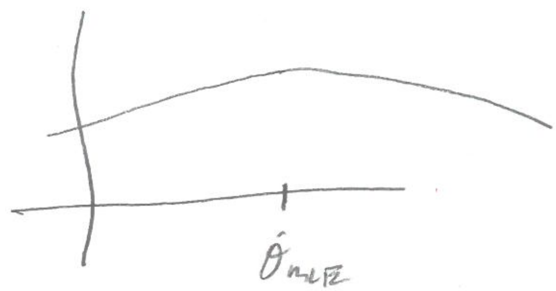
respects change of variables

$\boxed{\text{Fisher Info}}$

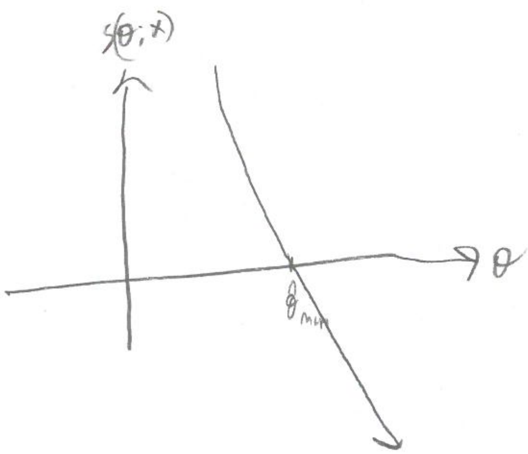
Recall likelihood and log likelihood

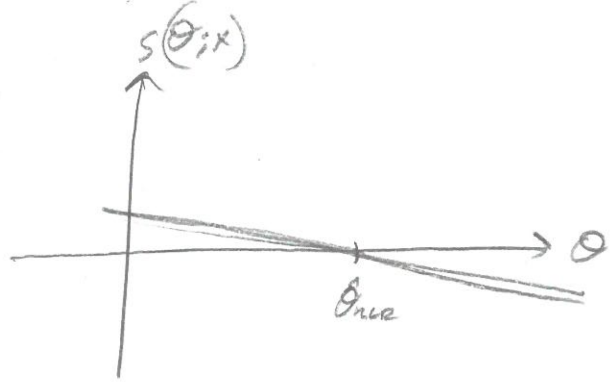
$$\mathcal{L}(\theta;x) = P(X;\theta) \qquad \ell(\theta;x) := \ln\left(\mathcal{L}(\theta;x)\right)$$





Define $s(\theta;x) := \ell'(\theta;x)$, the score function

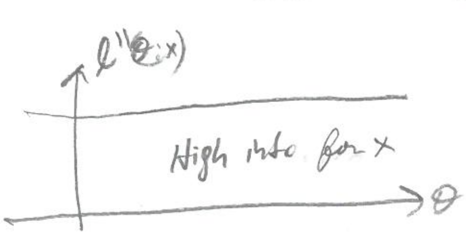


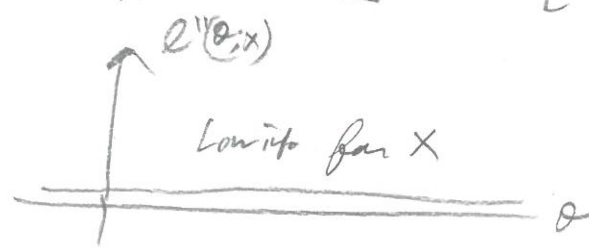


Stat 633   Stat 633

$$I(\theta) := \mathrm{Var}_X\left[s(\theta;x)\right] = \ldots = E_X\left[s(\theta;x)^2\right] = \ldots = E_X\left[-\ell''(\theta;x)\right]$$

"curvature" 2nd deriv





Fisher Information: $I(\theta) := Var_X[S(\theta;x)] = \ldots = E_X[S(\theta;x)^2] = \ldots = E[-l''(\theta;x)]$

$I(\theta)$ measures how much information is in X for a r.v.

Let's see this for $X \sim Binom(n,\theta)$ for fixed n

$$\mathcal{L}(\theta;x) = P(X;\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$$

$$l(\theta;x) = \ln\left(\binom{n}{x}\right) + x\ln(\theta) + (n-x)\ln(1-\theta)$$

$$l'(\theta;x) = \frac{x}{\theta} - \frac{n-x}{1-\theta}$$

$$l''(\theta;x) = -\frac{x}{\theta^2} - \frac{n-x}{(1-\theta)^2} \quad (-1)\cdot(-1)\cdot(-1)$$

recall $E[aX+c] = aE(X)+c$

$$I(\theta) = E_X[-l''(\theta;x)] = E\left[\frac{x}{\theta^2} + \frac{n-x}{(1-\theta)^2}\right] = \frac{E(X)}{\theta^2} + \frac{n-E(X)}{(1-\theta)^2} = \frac{n\cancel{\theta}}{\theta^{\cancel{2}}} + \frac{\overbrace{n-n\theta}^{n\cancel{(1-\theta)}}}{(1-\theta)^{\cancel{2}}} = n\left(\frac{1}{\theta}+\frac{1}{1-\theta}\right)$$

$$= n\left(\frac{1}{\theta(1-\theta)}\right)$$

Not a function of X! X is averaged over...

If $\theta = \frac{1}{2}$, $n=1$. How much info? $I\left(\frac{1}{2}\right) = 4$ ← the r.v. does not have too much info about θ on average

If $\theta = \frac{1}{100}$, $n=1$ .... $I\left(\frac{1}{100}\right) = 101.01$ ← the r.v. has a ton of info

Why?

Why should there be a mult. function of n? More data ⇒ more info. Because binomial is n bernoullis... more bernoulli draws ⇒ more info

---

Back to the issue... CONSIDER:

What if $p(\theta) \propto \sqrt{I(\theta)}$ AKA the Jeffreys prior

For $P(X|\theta) = \text{Bin}(n,\theta) \Rightarrow I(\theta) = n \frac{1}{\theta(1-\theta)}$

$$\Rightarrow P(\theta) \propto \sqrt{n \frac{1}{\theta(1-\theta)}} \propto \underbrace{\theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}}}_{\substack{\text{Kernel of} \\ \text{the} \\ \text{Beta} \\ \text{density!}}} \propto \text{Beta}\left(\tfrac{1}{2},\tfrac{1}{2}\right) = \frac{1}{\underbrace{B\left(\frac{1}{2},\frac{1}{2}\right)}_{\pi =}} \theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}}$$

$$= \frac{1}{\pi\sqrt{\theta(1-\theta)}}$$

"Arcsine distr."

Jeffreys prior is an uninformative ($\alpha+\beta=1$ i.e. small)

Does it do its job? Lets reparametrize to odds...

$P(\theta|x) \xrightarrow{\text{scalar}} P(\theta)$

$\phi = t(\theta)$

$P(\phi|x) \xrightarrow{\text{scalar}} P(\phi)$

does

$P(\phi) = P_\theta(t^{-1}(\phi))\left|\frac{d}{d\phi}(t^{-1}(\phi))\right|$

?

let $R = t(\theta) = \frac{\theta}{1-\theta}$ and $\theta = t^{-1}(R) = \frac{R}{R+1}$

$$P(X|R) = \binom{n}{x}\left(\frac{R}{R+1}\right)^x \underbrace{\left(1-\frac{R}{R+1}\right)}_{\frac{1}{R+1}}^{n-x} = \binom{n}{x}\frac{R^x}{(R+1)^n}$$

$P(R) = ???$

$$\ell(R;x) = \ln\binom{n}{x} + x\ln(R) - n\ln(R+1)$$

$$\ell'(R;x) = \frac{x}{R} - \frac{n}{R+1}$$

$$\ell''(R;x) = -\frac{x}{R^2} + \frac{n}{(R+1)^2}$$

$I(R) =$

$$E[-\ell''(R;x)] = \frac{E(x)}{R^2} - \frac{n}{(R+1)^2} = \frac{n\frac{\cancel{R}}{R+1}}{R^{\cancel{2}}} - \frac{n}{(R+1)^2} = n\left(\frac{1}{R(R+1)} - \frac{1}{(R+1)^2}\right) = n\frac{1}{R(R+1)^2}$$

$$P(R) \propto \sqrt{n\frac{1}{R(R+1)^2}} = \underbrace{\frac{1}{\sqrt{R}}\frac{1}{R+1}}_{g(R)}$$

$$\int_0^\infty g(R)\,dR = \pi \Rightarrow P(R) = \frac{1}{\pi}\frac{1}{\sqrt{R}}\frac{1}{R+1}$$

Now use change of var's!

$$P(R) = P_\theta(t^{-1}(R))\left|\frac{d}{dR}\left[t^{-1}(R)\right]\right| = \frac{1}{\pi}\left(\frac{R}{R+1}\right)^{-\frac{1}{2}}\underbrace{\left(1-\frac{R}{R+1}\right)}_{\frac{1}{R+1}}^{-\frac{1}{2}}\left|\frac{1}{(R+1)^2}\right| = \frac{1}{\pi}\frac{1}{\sqrt{R}}\cancel{\sqrt{R+1}}\frac{1}{(R+1)^{\cancel{2}}}$$

//

How is this possible? Given...

$p(X|\theta)$, $p(X|\phi)$, $\phi = t(\theta)$, $\theta = t^{-1}(\phi)$

Under Jeffreys strategy,

$p(\theta) \propto \sqrt{I(\theta)}$ and $p(\phi) \propto \sqrt{I(\phi)}$

Show that:

$$p(\phi) = p_\theta(\overset{\theta}{\overset{\shortparallel}{t^{-1}(\phi)}}) \left| \frac{d}{d\phi}[\overset{\theta}{\overset{\shortparallel}{t^{-1}(\phi)}}] \right| \propto \sqrt{I(\phi)}$$

$$= p_\theta(\theta) \left|\frac{d\theta}{d\phi}\right|$$

$$\propto \sqrt{I(\theta)} \left|\frac{d\theta}{d\phi}\right|$$

$$= \sqrt{I(\theta)\left(\frac{d\theta}{d\phi}\right)^2}$$

$$= \sqrt{E\left[S(\theta;X)^2\right] \frac{d\theta}{d\phi} \cdot \frac{d\theta}{d\phi}}$$

$$= \sqrt{E\left[\frac{d\ell}{d\theta} \cdot \frac{d\ell}{d\theta} \cdot \frac{d\theta}{d\phi} \cdot \frac{d\theta}{d\phi}\right]}$$

$$= \sqrt{E\left[\left(\frac{d\ell}{d\phi}\right)^2\right]}$$

$$= \sqrt{E[S(\phi;X)^2]}$$

$$= \sqrt{I(\phi)}$$

$\theta \sim \text{Beta}(0,0)$ Haldane

$\theta \sim \text{Beta}(\frac{1}{2},\frac{1}{2})$ Jeffreys

$\theta \sim \text{Beta}(1,1)$ Laplace

all uninformative

+'s, -'s to each...

up to you...

Jeffreys used Fisher's own information against him!!

Not covered in class

~~$X_1, \ldots, X_n \mid \theta \overset{iid}{\sim} \text{NegBin}(r, \theta)$ fixed (r)~~

~~$\theta \sim \text{Beta}(\alpha, \beta)$~~

~~$\theta \mid X_1, \ldots, X_n \sim \text{Beta}(rn + \alpha, \sum x_i + \beta)$~~

~~$n^* = 1, \quad X^* \mid X \sim \text{BetaNegBin}$~~

Done with beta-binomial.

Note: geometric or neg bin. is skipped this year

$$X \sim \text{Bin}(n, \theta) := \binom{n}{x} \theta^x (1-\theta)^{n-x} \quad \text{Supp}(X) = \{0, \ldots, n\}, \quad \theta \in \Theta = (0,1)$$

if $n \to \infty$, $\theta \to 0$ s.t. $\lambda = n\theta$

$$\lim_{n \to \infty} \frac{n!}{x!(n-x)!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} = \frac{\lambda^x}{x!} \lim \frac{n \cdot n \cdots}{(n-x)(n-x-1)\cdots 1 \cdot n \cdot n \cdots} \to 1$$

$$\cdot \lim \left(1 - \frac{\lambda}{n}\right)^n \cdot \cancel{\lim \left(1 - \frac{\lambda}{n}\right)^{-x} \to 0} = \frac{\lambda^x e^{-\lambda}}{x!}$$

$\Downarrow$

$$X \sim \text{Poisson}(\theta) := \frac{e^{-\theta} \theta^x}{x!} \quad \text{Supp}(X) = \{0, \ldots\} = \mathbb{N}_0, \ \theta \in \Theta \in (0, \infty)$$

different θ!!!

$$E[X] = \theta$$

$$\text{Var}(X) = \theta$$

~~Apply practice $P(X \mid \theta) \propto \frac{\theta^x}{x!}$~~

$$P(\theta \mid X) \propto P(X \mid \theta) P(\theta) = \frac{e^{-\theta} \theta^x}{x!} P(\theta) \propto \overbrace{e^{-\theta} \theta^x}^{k(\theta \mid X)} \underset{P(\theta)}{\overset{k(\theta)}{\propto}}$$

$$c! = \frac{1}{b^{a+1}} \Gamma(a+1)$$

If $P(\theta)$ was conjugate, should have the form ...

$$k(\theta) = e^{-b\theta} \theta^a$$

$$\int_0^\infty e^{-b\theta} \theta^a \, d\theta = \int_0^\infty \left(\frac{t}{b}\right)^a e^{-t} \frac{dt}{b} = \frac{1}{b^{a+1}} \int_0^\infty t^{(a+1)-1} e^{-t} \, dt$$

let $t = b\theta \quad \theta = \frac{t}{b} \quad dt = b \, d\theta \Rightarrow d\theta = \frac{dt}{b}$

Integrate and see $\quad P(\theta) = \dfrac{b^{a+1}}{\Gamma(a+1)} e^{-b\theta} \theta^a \quad$ this is called the gamma distr.

Usually its parametrized via

$$\theta \sim \text{Gamma}(\alpha, \beta) := \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta\theta} \theta^{\alpha - 1}$$