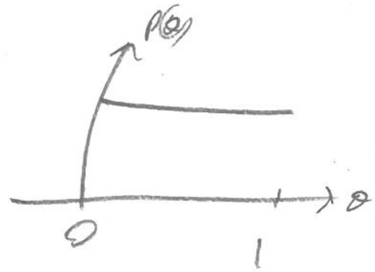
Math 341 Lec 11 3/14/19

X, Y cont r.v.'s with $Y = t(X)$ and f_X known. Find f_Y.

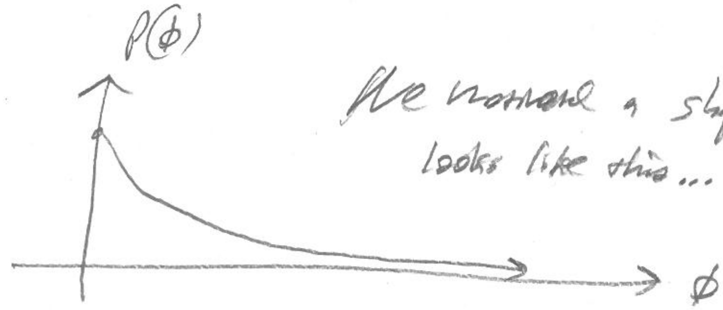
$$f_Y(y) = f_X(t^{-1}(y)) \left| \frac{d}{dy}[t^{-1}(y)] \right|$$

$$P(\theta) = U(0,1)$$



$$\phi = Odds(\theta) = \frac{\theta}{1-\theta}$$

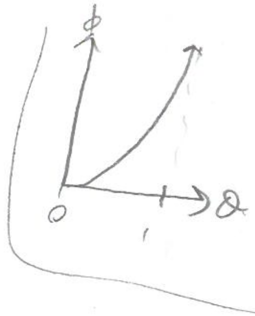
$$P(\phi) = ?$$

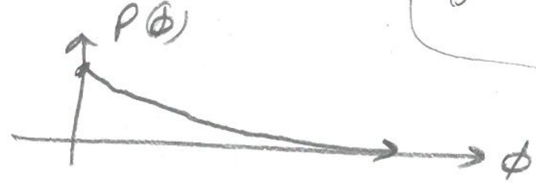


We wanted a shape that looks like this...

$$\phi(\theta) \overset{t}{=} \frac{\theta}{1-\theta} \Rightarrow \phi - \phi\theta = \theta \Rightarrow \phi = \theta + \phi\theta \Rightarrow \phi = \theta(1+\phi) \Rightarrow \theta \overset{t^{-1}}{=} \frac{\phi}{1+\phi}$$

$$f_\phi(\phi) = \underbrace{f_\theta\left(\frac{\phi}{1+\phi}\right)}_{1} \left|\underbrace{\frac{d}{d\phi}\left[\frac{\phi}{1+\phi}\right]}_{\frac{(1+\phi)(1)-(\phi)(1)}{(1+\phi)^2}}\right| = \left|\frac{1}{(1+\phi)^2}\right| = \frac{1}{(1+\phi)^2}$$





Is this a density?

$$\int_{supp(\phi)} f_\phi(\phi)\,d\phi = \int_0^\infty \frac{1}{(1+\phi)^2}\,d\phi = \left[\frac{\phi}{1+\phi}\right]_0^\infty = 1-0=1 \checkmark$$

Again...

$P(\theta)=U(0,1) \neq P(\phi)=\frac{1}{(1+\phi)^2} \Rightarrow$ If you are indifferent on the prob. scale, you are **not** indifferent on the odds scale

Fisher used this argument ~~to~~ denounce the stupidity of Bayesian stat with a uniform prior.

$$P(\theta\in[0,0.5]) = 0.5 = P(\phi\in[0,1])$$
$$P(\theta\in(0.5,1]) = 0.5 = P(\phi>1)$$

Bigger question: now that this doesn't work, is there a protocol to pick priors? Begin with $\mathcal{L}$ and produce an uninformative prior so that

$$\begin{tikzcd}
\mathcal{L} = P(X|\theta) \arrow[r, "\text{protocol}"] \arrow[d, "t"] & P(\theta) \arrow[d, leftrightarrow, "\text{transformation works}"] \\
P(X|\phi) \arrow[r, "\text{same protocol}"] & P(\phi)
\end{tikzcd}$$

In order to derive the Jeffreys procedure, we need kernels and Fisher Info.

Kernels Let $X \sim f(x;\theta)$

$f(x;\theta) \propto \underbrace{K(x;\theta)}_{\text{kernel}}$ means $\exists c>0$ independent of x s.t. $f(x;\theta) = cK(x;\theta)$.

given $K(x;\theta)$. Locate c.

$$\int_{Supp(X)} f(x;\theta)\,d\theta = 1 \Rightarrow \int cK(x;\theta)\,d\theta = 1 \Rightarrow \int K(x;\theta) = \frac{1}{c} \Rightarrow c = \left(\int K(x;\theta)\,d\theta\right)^{-1}$$

c is called the "normalization constant" since including it, the integral $=1$.

Since the c remains the same, $K(x;\theta)$ can identify the r.v. It is the density without the normalization constant.

Can $\int K(x;\theta)\,d\theta = \infty$? No because then any c multiplied would yield f which integrates to ∞.

Can $\exists x\ K(x;\theta) < 0$? No. Since $c > 0 \Rightarrow f < 0$ which would not be a density.

You've seen this before.

$$P(\theta|x) = \frac{P(x|\theta)\,P(\theta)}{P(x)} \propto P(x|\theta)\,P(\theta)$$

Why? $P(x)$ is not a function of θ. It is constant over θ.

$$Y \sim Beta(\alpha,\beta) = \frac{1}{B(\alpha,\beta)} y^{\alpha-1}(1-y)^{\beta-1} \propto y^{a}(1-y)^{b} = K(y;a,b) \Rightarrow Y \sim Beta(a+1,b+1)$$

If you see this then you know you got a beta!

try proving beta conjugacy...

$$P(\theta|x) \propto P(x|\theta)\,P(\theta) = \binom{n}{x}\theta^{x}(1-\theta)^{n-x}\frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} \propto \underbrace{\theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}}_{K(\theta)\text{ of Beta}} \propto Beta(x+\alpha, n-x+\beta)$$

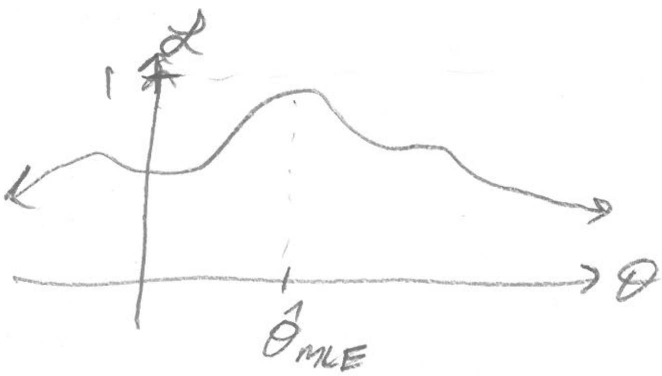
that easy!!!

$Y \sim N(\theta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y-\theta)^2} \propto e^{-\frac{1}{2\sigma^2}(y-\theta)^2}$ done?

$$e^{-\frac{1}{2\sigma^2}(y^2 - 2\theta y + \theta^2)} = e^{-\frac{1}{2\sigma^2}y^2} e^{\frac{\theta y}{\sigma^2}} \underbrace{e^{-\frac{\theta^2}{2\sigma^2}}}_{\text{doesn't change with } y.} \propto \underbrace{e^{-\frac{y^2}{2\sigma^2} + \frac{\theta y}{\sigma^2}}}_{\text{If you see this} \Rightarrow \text{Normal.}}$$

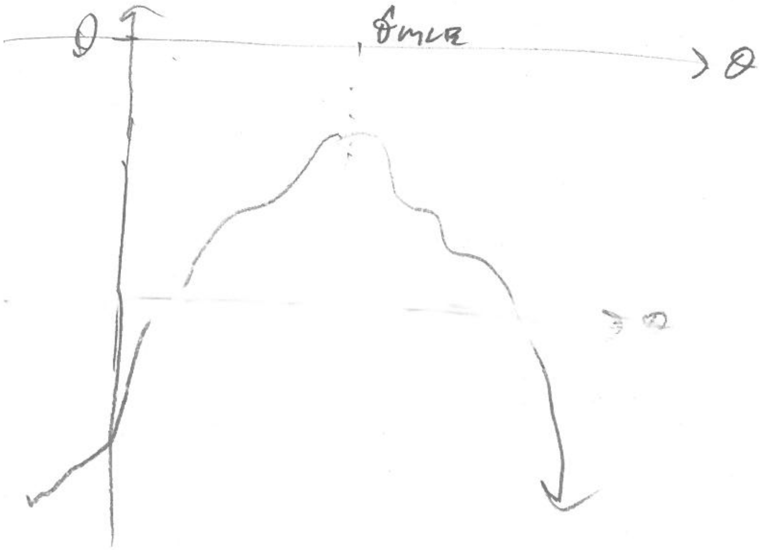
Fisher Info

Recall $\mathcal{L}(\theta; x) = P(x; \theta)$



$s(\theta; x) := \ell'(\theta; x)$

$\ell(\theta; x) := \ln(\mathcal{L})$



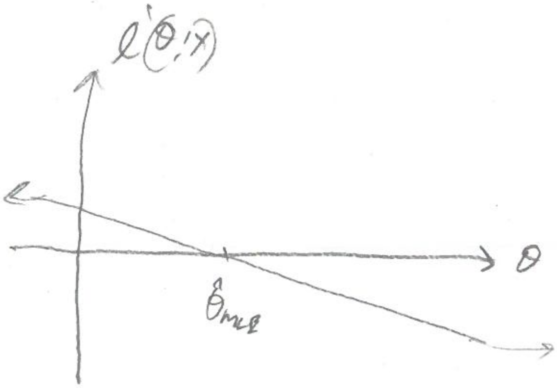
Define $I(\theta) := Var_x[s(\theta; x)] = \dots$

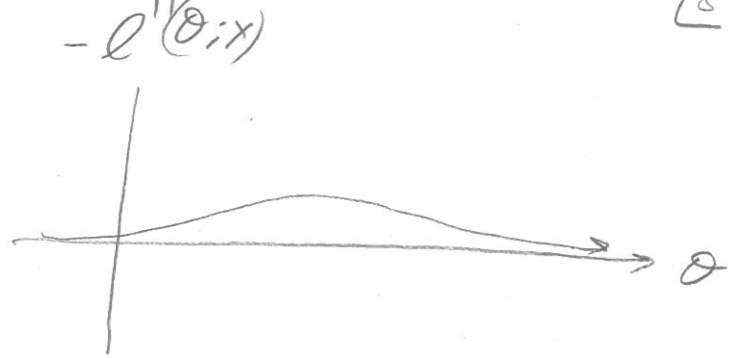
Fisher Information

$= E_x[s(\theta; x)^2]$

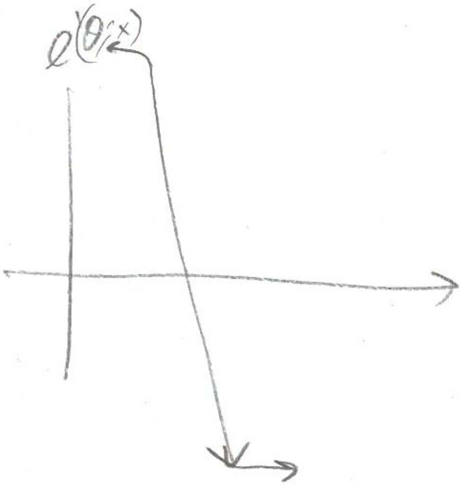
$= \dots$

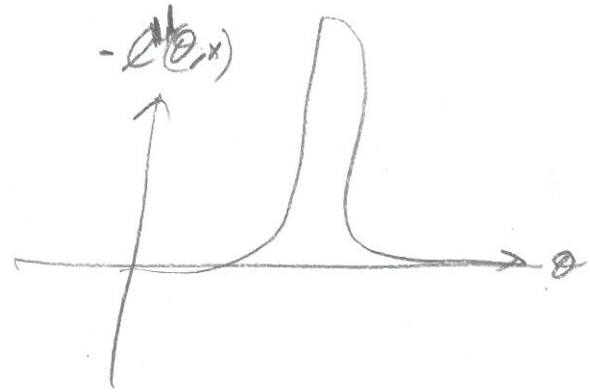
$= E_x[-\ell''(\theta; x)]$





$I(\theta) = E_x[-l''(\theta;x)]$ will be small for all θ.





$I(\theta) = E_x[-l''(\theta;x)]$ will be large

It is a degree of how much information X has in it about θ on average.

e.g.

$$p(x;\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$$

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

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

$(-1)(-1)(-1)$

$$-l''(\theta;x) = -\left(-\frac{x}{\theta^2} - \frac{n-x}{(1-\theta)^2}\right) = \frac{x}{\theta^2} + \frac{n-x}{(1-\theta)^2}$$

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

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

$I\left(\frac{1}{2}\right) = 4$ If θ changes by $\pm\delta$ from 0.5, it takes a large n to figure out the change.

$I\left(\frac{1}{100}\right) = 101.01$ If θ changes by $\pm\delta$ from 0.01, it doesn't take as many samples to figure it out.

X has more information about θ if $\theta = 0.01$ than about θ if $\theta = 0.5$.

633 material...

the Jeffreys's proposal: a means to...

Back to the issue.. Consider $P(\theta) \propto \sqrt{I(\theta)}$, $P(X|\theta) \longrightarrow P(\theta)$

$$P(\theta) \propto \sqrt{n \frac{1}{\theta(1-\theta)}} \propto \theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}} \propto Beta\left(\tfrac{1}{2}, \tfrac{1}{2}\right) = \frac{1}{\underbrace{B(\frac{1}{2},\frac{1}{2})}_{\pi}} \theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}}$$

Does this solve the problem. Try $P(X|\phi) \rightarrow P(\phi)$ using Jeffreys's proposal.

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

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

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

$$-l''(\phi; x) = -\left(-\frac{x}{\phi^2} + \frac{n}{(\phi+1)^2}\right) = \frac{x}{\phi^2} - \frac{n}{(\phi+1)^2}$$

$$E[\ \ ] = \frac{1}{\phi^2}E(X) - \frac{n}{(\phi+1)^2} = n\left(\frac{\overset{=\frac{\phi}{\phi+1}}{\theta}}{\phi^2} - \frac{1}{(\phi+1)^2}\right) = n\left(\overset{\frac{\phi+1}{\phi+1}}{\frac{1}{\phi(\phi+1)}} - \overset{\frac{\phi}{\phi}}{\frac{1}{(\phi+1)^2}}\right)$$

$$= n\left(\frac{(\phi+1) - \phi}{\phi(\phi+1)^2}\right) = n\frac{1}{\phi(\phi+1)^2}$$

$$P(\phi) \propto \sqrt{I(\phi)} = \sqrt{n\frac{1}{\phi(\phi+1)^2}} \propto \phi^{-\frac{1}{2}}(1+\phi)^{-1} = k(\phi)$$

$$C = \left(\int_{\mathrm{Supp}(\phi)} k(\phi)\, d\phi \right)^{-1} = \left(\int_0^\infty \phi^{-\frac{1}{2}} (1+\phi)^{-1}\, d\phi \right)^{-1} = \pi^{-1}$$

Calculus exercise

$$\Rightarrow P(\phi) = \frac{1}{\pi} \phi^{-\frac{1}{2}} (1+\phi)^{-1}$$

Now do $P(\phi)$ and $P(\theta)$ respect the change of variables?

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

$$= \frac{1}{\pi} \left(\frac{\phi}{\phi+1}\right)^{-\frac{1}{2}} \left(\frac{1}{\phi+1}\right)^{-\frac{1}{2}} \left| \frac{1}{(\phi+1)^2} \right|$$

$$= \frac{1}{\pi} \phi^{-\frac{1}{2}} (\phi+1)(\phi+1)^{-2} = \frac{1}{\pi} \phi^{-\frac{1}{2}} (1+\phi)^{-1} \quad \text{YES!}$$

We have a protocol that picks specific priors directly from likelihoods that is the same regardless of parameterization of the likelihood!

How is this possible? $\phi = t(\theta)$

$P(X|\theta) \longrightarrow P(\theta) \propto \sqrt{I(\theta)}$

$P(X|\phi) \longrightarrow P(\phi) \propto \sqrt{I(\phi)}$ with $P(\phi) = P_\theta(t^{-1}(\phi)) \left| \frac{d}{d\phi}[t^{-1}(\phi)] \right|$

Assume $P(\theta) \propto \sqrt{I(\theta)}$, prove $P(\phi) \propto \sqrt{I(\phi)}$

$$P(\phi) = P_\theta(\theta) \left| \frac{d\theta}{d\phi} \right| \propto \sqrt{I(\theta)} \left| \frac{d\theta}{d\phi} \right| = \sqrt{I(\theta) \frac{d\theta}{d\phi} \frac{d\theta}{d\phi}} = \sqrt{E[\ell'(\theta;x)^2] \frac{d\theta}{d\phi} \frac{d\theta}{d\phi}}$$

$$= \sqrt{E\left[\frac{d\ell}{\cancel{d\theta}} \frac{d\ell}{\cancel{d\theta}} \frac{\cancel{d\theta}}{d\phi} \frac{\cancel{d\theta}}{d\phi} \right]} = \sqrt{E\left[\frac{d\ell}{d\phi} \frac{d\ell}{d\phi} \right]} = \sqrt{E[\ell'(\phi;x)^2]} = \sqrt{I(\phi)} \checkmark$$

Jeffreys used Fisher's own information against him!