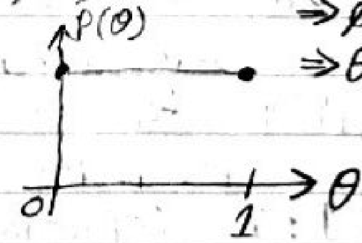
Math 341/650

3/17/21 Lec

$X \sim \text{Bern}(\theta)$, $\theta = P(X=1)$

There is another way to "parameterize" the Bernoulli. Consider:

$$\underset{\text{"odds"}}{\phi} = \underset{1:1}{f(\theta)} = \frac{\theta}{1-\theta}, \quad \phi \in (0,\infty) \Rightarrow \phi - \theta\phi = \theta \Rightarrow \phi = \theta + \theta\phi$$
$$\Rightarrow \phi = \theta(1+\phi)$$
$$\Rightarrow \theta = (f^{-1}\theta)$$
$$= \frac{\phi}{1+\phi}$$



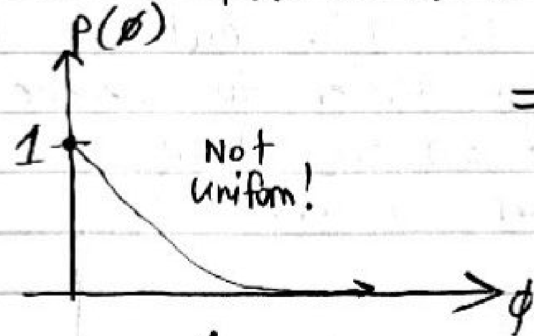
Laplace $P(\theta) = U(0,1)$

$\downarrow f$

$P(\phi) \overset{?}{=} \text{uniform}$ NO. It is impossible to have a prior on Supp$(0,\infty)$

$$X \sim \text{Bern}(\phi) = \left(\frac{\phi}{1+\phi}\right)^x \left(\frac{1}{1+\phi}\right)^{1-x} = \frac{\phi^x}{1+\phi}$$

$$P_\phi(\phi) = P_\theta(f^{-1}\theta)\left|\frac{d}{dx}[f^{-1}(\theta)]\right| = P\theta\left(\frac{\phi}{1+\phi}\right)\left|\frac{d}{d\phi}\left[\frac{\phi}{1+\phi}\right]\right|$$
$$= \left|\frac{d}{d\phi}\left[\frac{\phi}{1+\phi}\right]\right| = \left|\frac{(1+\phi)(1)-(\phi)(1)}{(1+\phi)^2}\right|$$



$$= \frac{1}{(1+\phi)^2} = F_{2,2}$$ ← Fun Fact.
$= $ Fisher-Schner distribution.

Is $\frac{1}{(1+\phi)^2}$ a valid density?

Check ① $\displaystyle\int_0^\infty \frac{1}{(1+\phi)^2}\,d\phi = \left[\frac{\phi}{1+\phi}\right]_0^\infty = 1-0 = 1$ ✓

What did we prove? We proved that if you're indifferent on the probability scale then you're *not* indiffrent on the odds scale. Fisher used this example to show how stupid Laplace's prior and to further show how stupid Bayesian stats is in general

If you change the Parameterization, yes, the inference can change.

Can we address this problem in Part? Can we do something? Can this something pick a Prior for us? Consider the following. Let θ be parameter of $\mathcal{F}$ and $f(\theta)=\phi$ at 1:1 reparameterization. Is there a procedure that can accomplish the following?

$$\begin{tikzcd} \mathcal{F} := P(X|\theta) \arrow[r, "\text{Procedure}"] \arrow[d, shift right, "f"'] & P(\theta) \arrow[d, shift right, "f"'] \\ P(X|\phi) \arrow[r, "\text{Procedure}"] \arrow[u, shift right, "f^{-1}"'] & P(\phi) \arrow[u, shift right, "f^{-1}"'] \end{tikzcd}$$

It was Harold Jeffrey's idea that solved this puzzle. The prior that is the result of the procedure is then called the "Jeffrey's Prior". In order to derive the Procedure, we need 2 more tools...

① kernels ② Fisher Information

Kernels

$$f(x;\theta) \propto k(x;\theta) \Rightarrow \exists c \in \mathbb{R} \quad f(x;\theta) = c\,k(x;\theta)$$

This is also valid for PMF's as well but I'll use the f notation. Also means that k and f are 1:1 because they differ only by c.

$$\int_{Supp[X]} f(x;\theta)dx = 1 \Rightarrow \int c\,k(x;\theta)dx = 1 \Rightarrow \int k(x;\theta)\,d\theta = \frac{1}{c}$$

$$\Rightarrow c = \frac{1}{\int_{Supp[X]} k(x;\theta)\,dx}$$

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

$\mathcal{F}: Bin(n,\theta)$, n fixed, $P(\theta) = Beta(\alpha,\beta)$

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

$$= \binom{n}{x}\theta^x(1-\theta)^{n-x}\frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$$\propto \theta^x(1-\theta)^{n-x}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$$= \theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1} \propto Beta(x+n, n-x+\beta)$$

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

$f(Y;\theta,\sigma^2)$

$$= e^{-\frac{1}{2\sigma^2}(Y^2-2Y\theta+\theta^2)}$$

$$= e^{-\frac{Y^2}{2\sigma^2}} \cdot e^{\frac{Y\theta}{\sigma^2}} \cdot e^{-\frac{\theta^2}{2\sigma^2}} \propto \underbrace{e^{\frac{-Y^2}{2\sigma^2}} e^{\frac{Y\theta}{\sigma^2}}}_{K(Y;\theta,\sigma^2)}$$

$$C = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-\theta^2}{2\sigma^2}}$$

Fisher Information $X = \langle X_1, \ldots, X_n \rangle$

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

$\Downarrow$

$l(\theta;X) := \ln(\mathcal{L}(\theta;X))$

$S(\theta;X) := \frac{d}{d\theta}[l(\theta;X)]$

[$\hat{\theta}_{MLE}$ is derived by $S(\theta;X) \overset{set}{=} 0$ solve for θ]

$$I(\theta) := Var_X[S(\theta;X)] = \ldots = E_X[-l''(\theta;X)]$$

Fisher Information

An example fisher information calculation:

one $X \sim Bern(\theta)$

$$\mathcal{L}(\theta;x) = \theta^x(1-\theta)^{1-x} \Rightarrow l(\theta;x) = x\ln\theta + (1-x)\ln(1-\theta)$$

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

$$I(\theta)=E_x\left[-\frac{x}{\theta^2}-\frac{1-x}{(1-\theta)^2}\right]=E_x\left[\frac{x}{\theta^2}+\frac{1-x}{(1-\theta)^2}\right]$$

$$=\frac{1}{\theta^2}E[x]+\frac{1}{(1-\theta)^2}\left(1-E[x]\right)$$

$$=\frac{1}{\theta^2}\theta+\frac{1}{(1-\theta)^2}(1-\theta)$$

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

Thm: The Jeffrey's prior is $P_J(\theta)\propto\sqrt{I(\theta)}$

Let's see this work first and then provide a proof for the thm later.

Ex: $\text{Bin}(n,\theta)\Rightarrow \mathcal{L}(\theta;x)=\binom{n}{x}\theta^x(1-\theta)^{n-x}$

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

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

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

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

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

$$=\frac{1}{\theta^2}n\theta+\frac{1}{(1-\theta)^2}(n-n\theta)$$

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

$$P_J(\theta)\propto\sqrt{\frac{n}{\theta(1-\theta)}}\propto\sqrt{\frac{1}{\theta(1-\theta)}}=\theta^{\overbrace{-\frac{1}{2}}^{=\alpha-1}}(1-\theta)^{\overbrace{-\frac{1}{2}}^{=\beta-1}}\propto \text{Beta}\left(\frac{1}{2},\frac{1}{2}\right)$$

The Jeffrey's prior is $\text{Beta}\left(\frac{1}{2},\frac{1}{2}\right)$! It's amazing that it came out conjugate. Who knows what could've happened.

$$\begin{tikzcd}
P(x|\theta) \arrow[r, "\text{Jeffrey's Procedure}"] \arrow[d, "f"'] & P_J(\theta) = \text{Beta}\left(\frac{1}{2}, \frac{1}{2}\right) \arrow[d, shift right, "f"'] \\
P(x|\phi) \arrow[r, "\text{Jeffrey's Procedure}"] & P_J(\phi) = ? \arrow[u, shift right, "f^{-1}"']
\end{tikzcd}$$

We will verify this using $\phi = f(\theta) = \frac{\theta}{(1-\theta)}$, the "Odds".

But just because it worked once, doesn't mean we've proven the Thm! We then need to prove it!