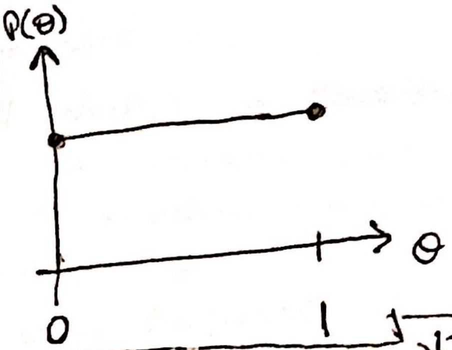
$\theta \in \Theta = (0,1)$

$X \sim \text{Bern}(\theta)$, $\theta = P(X=1)$. There is another way to "parameterize" the Bern.

Consider: $\phi = t(\theta) = \dfrac{\theta}{1-\theta}$ $\quad \phi \in (0, \infty)$ this is called "odds".

$\uparrow$ 1:1

we care because, Laplace says $\boxed{P(\theta) = U(0,1)}$ ⊛ $\xrightarrow{t}$




$P(\phi) \stackrel{?}{=}$ uniform, No Impossible to have prior on support $(0, \infty)$

$P_\phi(\phi) = P_\theta(t^{-1}(\theta))$ $\quad \boxed{\theta = t^{-1}(\theta) = \phi^{-1} = \dfrac{\phi}{1+\phi}}$

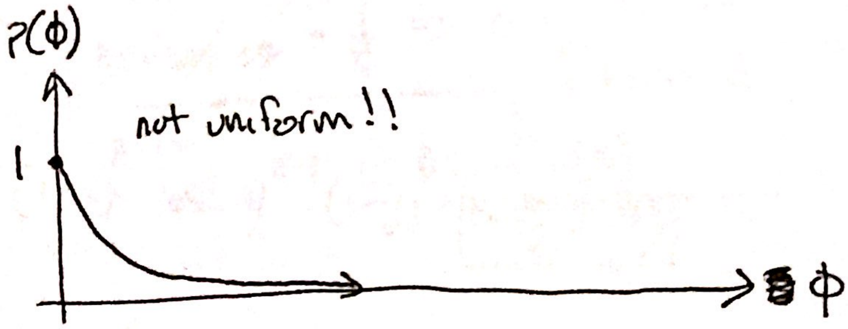
$X \sim \text{Bern}(\theta) = \left(\dfrac{\phi}{1+\phi}\right)^x \left(\dfrac{1}{1+\phi}\right)^{1-x} = \dfrac{\phi^x}{1+\phi}$

PMF of Bern in "odds".

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

⊛ 1 = follows from $U(0,1)$

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




not uniform!!

Is this a real density?

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

this is the $F_{2,2}$ dist Fisher-Snecedor distribution

What did we prove here? If you are indifferent on the probability scale then you are not indifferent on the odds scale. Fisher used this example to show how stupid Laplace prior is and to further show how stupid bayesian statz is in general.

If you change the parameterization the inference can change.

Can we address this problem somehow? Can this something pick a prior for us.

Let $\theta$ be the parameter of $\mathcal{F}$ and $t(\theta) = \phi$ a 1-1 reparameterization

Is there a procedure that can accomplish the following:

$$\begin{tikzcd}
\mathcal{F} := P(X|\theta) \arrow[r, "\text{procedure}"] \arrow[d, shift right, "t"'] & P(\theta) \arrow[d, shift right, "t"'] \\
P(X|\phi) \arrow[r, "\text{procedure}"] \arrow[u, shift right, "t^{-1}"'] & P(\phi) \arrow[u, shift right, "t^{-1}"']
\end{tikzcd}$$

It was Harold Jeffrey's idea that solved this. The prior that is the result of the procedure is then called the "Jeffrey's Prior".

In order to derive the procedure we need more tools. Back to probability land!

(1) Kernels

(2) Fisher information.

<u>Kernels</u> $f(x;\theta) \propto K(x;\theta) \Rightarrow \exists\, c \in \mathbb{R}\ \ f(x;\theta) = c\,K(x;\theta)$

this is also valid for pmf's. This also means that K and f are 1-1, because they differ only by c. Remember

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

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

---

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

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

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

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

$$Y \sim N(\theta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y-\theta)^2} \propto e^{-\frac{1}{2\sigma^2}(y-\theta)^2} \longleftarrow \text{good test question}$$

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y-\theta)^2} = f(y;\theta,\sigma^2)$$

$$e^{-\frac{1}{2\sigma^2}(y-\theta)^2} = e^{-\frac{1}{2\sigma^2}(y^2-2y\theta+\theta^2)} = e^{-\frac{y^2}{2\sigma^2}} e^{\frac{y\theta}{\sigma^2}} e^{-\frac{\theta^2}{2\sigma^2}} \propto e^{-\frac{y^2}{2\sigma^2}} e^{\frac{y\theta}{\sigma^2}} = k(y;\theta;\sigma^2)$$

$$\boxed{\text{All the things you deleted are } c \qquad c = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\theta^2}{2\sigma^2}}}$$

**Fisher Information:** $X = (x_1, \ldots, x_n)$

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

$$\Downarrow$$

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

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

Set $S(\theta;X) = 0$ and solve for $\theta$ we get $\hat{\theta}_{MLE}$. Seen this Before.

Log $\to$ 2 derivatives $\to$ expectation to get Fisher Information.

Have not seen: $I(\theta) := Var_X[S(\theta;X)] = E_X[\ell''(\theta;X)]$ = Fisher Information

Example of Fisher Information

$n=1$

$X \sim Bern(\theta)$

$$\mathcal{L}(\theta:X) = \theta^x(1-\theta)^{1-x}$$

$$\Rightarrow \ell(\theta;X) = x\ln(\theta) + (1-x)\ln(1-\theta)$$

$$\Rightarrow \ell'(\theta;X) = \frac{x}{\theta} - \frac{1-x}{1-\theta} \Rightarrow$$

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

$$\Rightarrow I[\theta] = E_X[- \quad ]$$

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

$E[X \sim Bern(\theta)] = \theta$

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

The Jeffrey's prior $P_J(\theta) \propto \sqrt{I(\theta)}$. ← (Theorem)

$X : \text{Bin}(n, \theta)$. Get $I(\theta)$. $\mathcal{L}(\theta; x) = \binom{n}{x} \theta^x (1-\theta)^{n-x} \Rightarrow \ell(\theta; x) = \ln\left(\binom{n}{x}\right) + x\ln(\theta) + n-x \ln(1-\theta$

$$\ell'(\theta;x) = \frac{x}{\theta} - \frac{n-x}{1-\theta} \Rightarrow \ell''(\theta;x) = \underbrace{\frac{-x}{\theta^2} - \frac{n-x}{(1-\theta)^2}}. \quad I(\theta) = E_x[-\square]$$

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

$$E[X \sim \text{Bin}(n,\theta)] = n\theta \Rightarrow \frac{1}{\theta^2} n\theta + \frac{1}{(1-\theta)^2}(n - n\theta) = n\left(\frac{1}{\theta} + \frac{1}{1-\theta}\right) = \boxed{\frac{n}{\theta(1-\theta)}} = I(\theta)$$

So we have $I(\theta)$.

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

Jeffrey prior is $\text{Beta}(\frac{1}{2}, \frac{1}{2})$. Came out conjugate.

$$\begin{tikzcd}
P(X|\theta) \arrow[r, "\text{Jeffrey}"] \arrow[d, shift right, "t"'] & P_J(\theta) = \text{Beta}(\frac{1}{2}, \frac{1}{2}) \arrow[d, shift right, "t"'] \\
P_X(\phi) \arrow[r, "\text{Jeffrey}"] \arrow[u, shift right, "t^{-1}"'] & P_J(\phi) = ? \arrow[u, shift right, "t^{-1}"']
\end{tikzcd}$$

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

But just because it works once doesn't mean we have proven the theorem