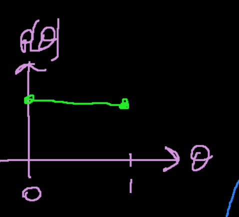
$X \sim \text{Bern}(\theta)$, $\theta = P(X=1)$, $\theta \in \Theta = (0,1)$

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

$$\phi = \overset{1:1}{t}(\theta) = \frac{\theta}{1-\theta}, \quad \phi \in (0, \infty) \Rightarrow \phi - \theta\phi = \theta$$
$$\Rightarrow \phi = \theta + \theta\phi$$
$$\Rightarrow \phi = \theta(1+\phi)$$
$$\Rightarrow \theta = t^{-1}(\theta) = \frac{\phi}{1+\phi}$$

"Odds"



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

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

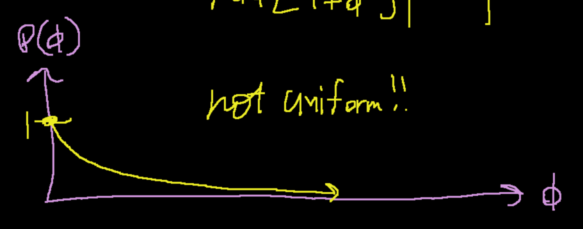
$\downarrow t$?

$P(\phi) \overset{?}{=} \text{Uniform}$ No. It is impossible to have a prior on the support (0, infinity)

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

$$= \left| \frac{d}{d\phi}\left[\frac{\phi}{1+\phi}\right] \right| = \left| \frac{(1+\phi)(1) - (\phi)(1)}{(1+\phi)^2} \right| = \frac{1}{(1+\phi)^2} = F_{2,2}$$

Fun fact: Fisher-Snedecor distr.



not uniform!!

Is this a valid density?

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

What did we prove? We proved that if you're indifferent on the probability scale then you're *not* indifferent 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 the parameter of curlyF and t(θ) = φ, a 1:1 reparameterization. Is there a procedure that can accomplish the following?

$$\begin{array}{ccc} \mathcal{F}: P(X \mid \theta) & \xrightarrow{\text{procedure}} & P(\theta) \\ t \downarrow \uparrow t^{-1} & & t \downarrow \uparrow t^{-1} \\ P(X \mid \phi) & \xrightarrow{\text{procedure}} & P(\phi) \end{array}$$

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 two more tools: (1) kernels (2) Fisher Information.

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

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

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

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

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

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

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

$$\underbrace{Y \sim N(\theta,\sigma^2)}_{f(y;\theta,\sigma^2)} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y-\theta)^2} \propto e^{-\frac{1}{2\sigma^2}(y-\theta)^2} = e^{-\frac{1}{2\sigma^2}(y^2-2y\theta+\theta^2)}$$

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

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

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

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

$\Downarrow$

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

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

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

$$\underbrace{I(\theta)}_{\text{Fisher Information}} := \text{Var}_X\left[S(\theta;x)\right] = \ldots = E_X\left[\ell''(\theta;x)\right]$$

An example Fisher information calculation: one X ~ Bern(θ).

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

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

$$I(\theta) = E_X\left[-\ell''\right] = E_X\left[\frac{X}{\theta^2} + \frac{1-X}{(1-\theta)^2}\right] = \frac{1}{\theta^2}E[X] + \frac{1}{(1-\theta)^2}(1-E[X])$$

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

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.

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

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

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

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

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

The Jeffrey's prior is Beta(1/2, 1/2)! It's amazing that it came out conjugate. Who knows what could've happened?

$$\begin{array}{ccc} P(X \mid \theta) & \xrightarrow{\text{Jeffrey's procedure}} & P_J(\theta) = \text{Beta}\left(\frac{1}{2},\frac{1}{2}\right) \\ t \downarrow \uparrow t^{-1} & & t \downarrow \uparrow t^{-1} \\ P(X \mid \phi) & \xrightarrow{\text{Jeffrey's procedure}} & P_J(\phi) = ? \end{array}$$

We will verify this using φ = t(θ) = θ / (1-θ), the "odds".
But just because it works once, doesn't mean we've proven the thm! We then need to prove it...