March 17
lec 11

$X \sim Ber(\theta)$, $\theta = P(X=1)$

There is another way to "parameterize" the Bernoulli.

1-1 function

consider

$\phi = t(\theta) = \frac{\theta}{1-\theta}, \phi \in [0, \infty) \Rightarrow \phi - \theta\phi = \theta$

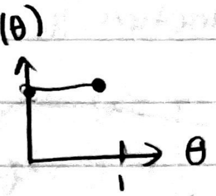
$\phi = \theta + \theta\phi$

$\phi = \theta(1+\phi)$

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

"odds"

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



$\xrightarrow{t} p(t) \stackrel{?}{=}$ uniform

No, it is impossible to have a prior on the support (0, infinity)

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

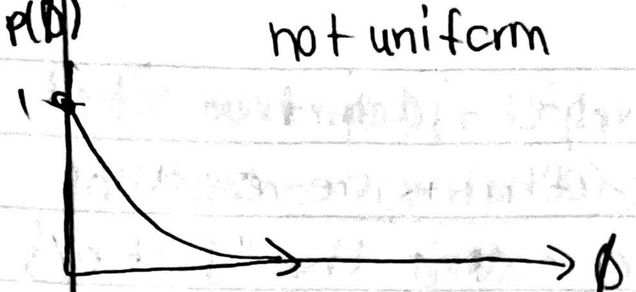
$P_{\phi}(\phi) = P_{\theta}(t^{-1}(\theta)) \left| \frac{d}{d\phi}[t^{-1}\theta) \right| = P\theta\underbrace{\left(\frac{\phi}{1+\phi}\right)}_{1} \left| \frac{d}{d\phi}\left[\frac{\phi}{1+\phi}\right] \right|$

$= 1 \cdot \left| \frac{d}{d\phi}\left[\frac{\phi}{1+\phi}\right] \right|$ ← quotient rule

Fun fact: Fisher Snedecor distribution

$= \left| \frac{1+\cancel{\phi(1)} - \cancel{\phi(1)}}{(1+\phi)^2} \right| = \frac{1}{1+\phi^2}$ → F_{22}

→



not uniform

Is this a vaild density?

$$\int_0^\infty \frac{1}{(1+\phi)^2}\,d\phi$$

$$= \left[\frac{\phi}{1+\phi}\right]_0^\infty = 1 - 0 = 1$$

what did we prove? We proved that if you're 'indifferent' on the probability scale then you're *not* '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 parametrization, yes, the inference can change.

Can we address this problem in parts? can we do something? Can this something pick a prior for us? Consider the following. Let θ be the parameter of curly F and $t(\theta) = \phi$, a lil reparametrization. Is there a procedure that can accomplish the following?

$$\begin{tikzcd}
\mathcal{F}: \; P(X|\theta) \arrow[r, "\text{procedure}"] \arrow[d, shift right, "t"'] & P(\theta) \arrow[d, shift right, "t"'] \\
P(X|\phi) \arrow[r, "\text{if you run it w/ this}"'] \arrow[u, shift right, "t^{-1}"'] & P(\phi) \arrow[u, shift right, "t^{-1}"']
\end{tikzcd}$$

+

It was Harlod Jeffray's idea that solved this puzzle. The prior that is the result of the procedure is then called the "Jeffray's prior". In order to derive the procedure, we need two more tools:
(1) Kernel (2) Fisher Information

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

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

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

$$\Downarrow$$

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

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

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

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

$$\downarrow$$

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

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

Ex. of kernel

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

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

$$= e^{\frac{-y}{2\sigma^2}} e^{\frac{y\theta}{\sigma^2}} e^{-\frac{\theta^2}{2\sigma^2}} \propto e^{\frac{-y}{2\sigma^2}} e^{\frac{y\theta}{\sigma^2}}$$

$$= K(y; \theta)$$

$$\downarrow$$

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

Fisher Information $X \langle x_1, \ldots x_n \rangle$

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

$$\downarrow$$

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

$$s(\theta; x) = \frac{d}{d\theta}[\ell(\theta; x)] \rightarrow$$

Scanned with CamScanner

$S(\theta;x) \overset{set}{=} 0$ solve for θ is equal to

$\hat{\theta}_{MLE}$ is derivated by:

$I(\theta) = Var_X[S(\theta;x)] = E_X - \ell''(\theta;x)$

This is Fisher Information.

An example Fisher Information calculation: $X \sim Bern(\theta)$

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

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

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

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

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

Thm: the Jeffrey prior $P_J(\theta) \propto \sqrt{I(\theta)}$

$\rightarrow$

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

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

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

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

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

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

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

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

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

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

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

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

But just b/c it works once, doesn't mean we've proven the thm. we the need to prove it . . . .

midterm 1 ↑

lec 12