3/9/17

$X|\theta \sim \text{Binom}(n, \theta)$

$\theta \sim \text{Beta}(\alpha, \beta)$

$\theta|X \sim \text{Beta}(\underbrace{\alpha+x}_{\alpha'}, \underbrace{\beta+n-x}_{\beta'})$

$X^*|X \sim \text{BetaBinom}(n^*, \alpha', \beta')$, $\alpha' = \alpha+x$, $\beta' = \beta+n-x$ $\quad := \binom{n^*}{x^*} \dfrac{B(\overbrace{\alpha+x+x^*}^{\alpha'}, \overbrace{\beta+n-x+n^*-x^*}^{\beta'})}{}$

$X \sim \text{BetaBinom}(n, \alpha, \beta)$ $\qquad W \sim \text{Bern}(\theta)$

if $n^* = 1$ $\qquad P(W=1) = \theta$

$$X^*|X = \frac{B(\alpha+x+x^*, \beta+n-x+1-x^*)}{B(\alpha+x, \beta+n-x)}$$

$P(x^*|x)$ is the distribution of future x^* given data x $= \int_\Theta P(x^*|\theta) P(\theta|x)\, d\theta$
↳ (posterior predictive distribution) $\qquad$ binom $\quad$ beta

(prior predictive distribution)
$P(x)$ is the distribution of data observed $\qquad = \int_\Theta P(x|\theta) P(\theta)\, d\theta$

$p(X|\{\ \}) = \int P(X|\theta) P(\theta|\{\ \})\, d\theta$ ⇒ I saw nothing
nothing yet

prior $\qquad$ posterior
nothing $\qquad$ begin $\qquad$ data

Laplace Prior
$\theta \sim U(0,1) \to \text{Beta}(1,1)$ ⇒ uninformative – "indifferent prior" (easily swamped)
$\theta|X\ \text{Beta}(1+x, 1+n-x)$

What is the most uninformative prior?
$\theta \sim \text{Beta}(0,0)$ ⟶ not a true PDF, illegal
$\alpha, \beta > 0$
(improper prior / Haldane Prior) (1932)

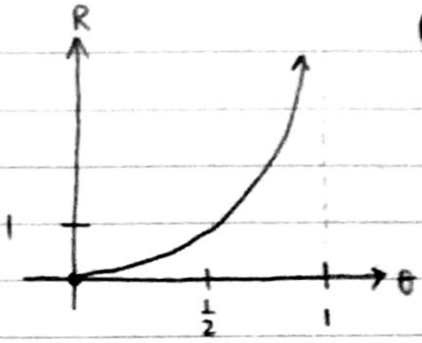
imagine $\theta \sim \text{Beta}(0,0)$ is legal...
$\theta|X \sim \text{Beta}(x, n-x) \Rightarrow \hat{\theta}_{MMSE} = \dfrac{x}{n} = \hat{\theta}_{MLE}$
↳ no "shrinkage"
$\theta \sim \text{Beta}(0,0)$ ⇒ complete ignorance

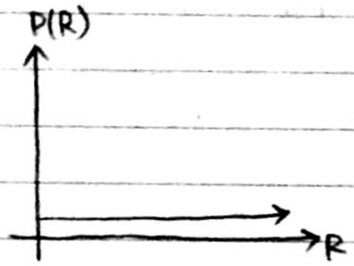
this posterior could be improper if $x = 0$ or $x = n$
$\Downarrow \alpha = 0 \qquad \Downarrow \beta = 0$
☆ using "improper" prior can cause improper posterior

"Reparameterization"

$$\text{Odds}(\theta) = \frac{\theta}{1-\theta} \qquad \theta \in (0,1)$$

$$R = \text{Odds}(0.9) = \frac{.9}{1-.9} = 9 \qquad R \in (0, \infty)$$



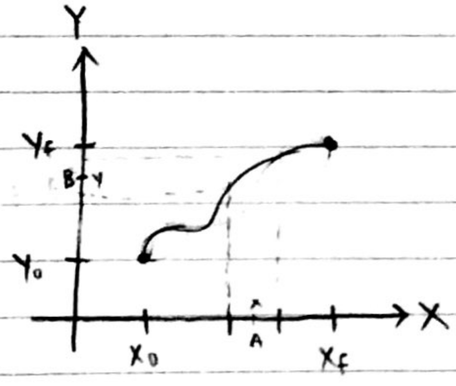


Laplace's prior on odds goes to infinity so, it doesn't work.

Random Variables X, Y are related by a 1-1 monotonic transformation

$$Y = t(X), \quad X = t^{-1}(Y)$$

we know $f_X(x)$ the PDF of X. we want $f_Y(y)$



$$P(X \in A) \approx f_X(x)A \Rightarrow f_X(x)|dx| = f_Y(y)|dy|$$

$$P(Y \in B) \approx f_Y(y)B \Rightarrow f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right|$$

$$\Rightarrow \underbrace{f_Y(y) = f_X(t^{-1}(y))\left|\frac{d}{dy}[t^{-1}(y)]\right|}_{\text{transformation of random variables}}$$

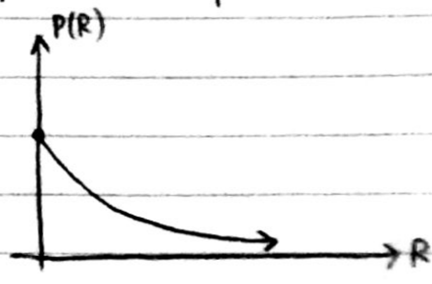
$$R = t(\theta) = \frac{\theta}{1-\theta}$$

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

$$f_R(r) = f_\theta(t^{-1}(r))\left|\frac{d}{dr}[t^{-1}(r)]\right| = f_\theta\left(\frac{r}{r+1}\right)\left|\frac{d}{dr}\left[\frac{r}{r+1}\right]\right|$$

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

↳ because probability density always equal 1 because Laplace's prior



$\theta \sim U(0,1)$ } uninformative
$\theta \sim Beta(0,0)$

A protocol to pick prior given $\mathcal{F}$.

Under a reparametrization $\Phi = t(\theta)$. What if I had a protocol which allowed me to pick priors:

$$P(X|\theta) \xrightarrow{\text{pick}} P(\theta), \quad P(X|\Phi) \xrightarrow{\text{pick}} P(\Phi)$$

Such that we have:

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

$\Rightarrow$ Jeffrey's Prior

Kernels

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

$f(x;\theta) \propto g(x;\theta)$ "kernel"

$$f(x,\theta) = \underbrace{\frac{1}{c}}_{h(\theta)} g(x;\theta) \Rightarrow \int f(x)\,dx = 1 \Rightarrow \int g(x)\,dx = \int c\, f(x)\,dx = c\underbrace{\int f(x)\,dx}_{1}$$

$$\Rightarrow c = \int g(x)\,dx$$

NOTE: f, g are 1:1

$X|\theta \sim \text{Binom}(n,\theta)$, $\theta \sim \text{Beta}(\alpha,\beta)$

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

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

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

$$\theta \sim \text{Beta}(\alpha,\beta) = \frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} \propto \underbrace{\theta^a(1-\theta)^b}_{\text{kernels of the beta}}$$

$$X|\theta \sim \text{Binom}(n,\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$$

$$= \frac{n!}{x!(n-x)!}\theta^x(1-\theta)^n \cdot (1-\theta)^{-x}$$

$$\propto \frac{1}{x!(n-x)!}\left(\frac{\theta}{1-\theta}\right)^x$$

Fisher Information

Recall: likelihood, log-likelihood,

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

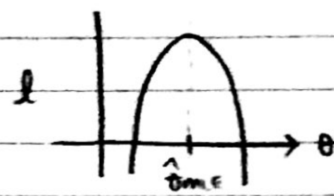
$\ell(\theta;x) := \ln(\mathcal{L}(\theta;x))$

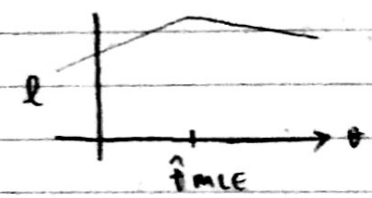
Define the score function:

$S(\theta;x) = \ell'(\theta;x)$

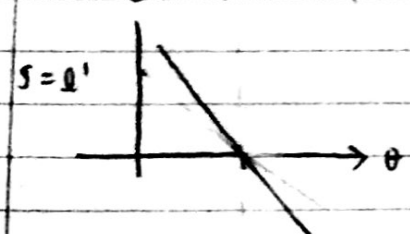
$I(\theta) = \mathrm{Var}_x[S(\theta;x)] = \cdots = E_x[S(\theta;x)^2]$

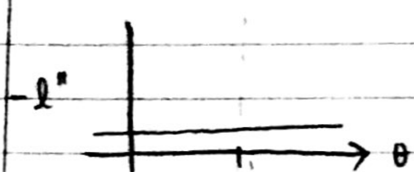
Fisher information $= \cdots = E_x[-\ell''(\theta;x)]$

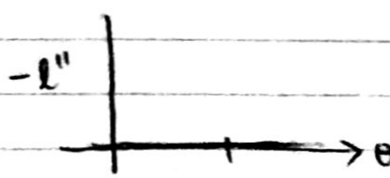




↳ tells a lot about θ (see clear "winner")







☆ ↳ $I(\theta)$ measures the information in X about θ

$X \sim \mathrm{Binom}(n,\theta)$. Calculate $I(\theta)$

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

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

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

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

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