3/9/17

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

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

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

$X^* \mid X \sim \text{BetaBinom}(n^*, \underset{\alpha + x}{\overset{\shortparallel}{\alpha'}}, \underset{\beta + n - x}{\overset{\shortparallel}{\beta'}})$

$$:= \binom{n^*}{x^*} B(\overbrace{\alpha + x}^{\alpha'} + x^*, \overbrace{\beta + n - x}^{\beta'} + n^* - x^*)$$

if $n^* = 1$ $\qquad B(\underbrace{\alpha + x}_{\alpha'}, \underbrace{\beta + n - x}_{\beta'})$

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

$W \sim \text{Bern}(\theta)$

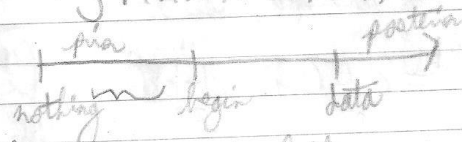
$P(W = 1) = \theta$

$P(X^* \mid X)$ is the distn. of future x^* given data x (posterior predictive distn.)

$$= \int_{\Theta} P(X^* \mid \theta) P(\theta \mid X) \, d\theta$$

$P(X)$ is the distn. of data observed $= \int_{\Theta} P(X \mid \theta) P(\theta) d\theta$

$P(X \mid \{\ \}) = \int P(X \mid \theta) P(\theta \mid \{\ \}) d\theta$ (prior predictive distribution)



$U(0,1)$ is like 1 success, 1 failure

$\frac{1}{B(\alpha, \beta)} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} = \frac{\Gamma(0)}{\Gamma(0)\Gamma(0)}$ $\Gamma(0)$ is infinity

$\Gamma(t) = \int_0^\infty \frac{e^{-t}}{t} dt$

$X \sim BetaBinom(n, \alpha, \beta)$

Laplace Prior

$\theta \sim U(0,1) = Beta(1,1)$ 1 success, 1 failure

$\theta | X \sim Bern(1+x, 1+n-x)$

→ uninformative prior b/c easily swapped

"indifferent prior"

What is the most uninformative prior?

$\theta \sim Beta(0,0)$ this is "illegal"

⇒ Improper Prior ← not a pdf

Haldane prior (1932)

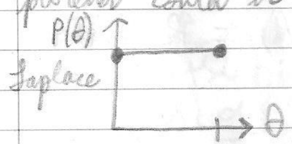
$\theta | X \sim Beta(x, n-x) \Rightarrow \hat{\theta}_{MMSE} = \frac{X}{n} = \hat{\theta}_{MLE}$

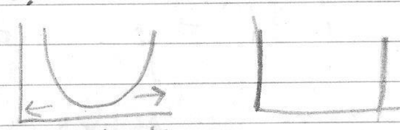
Shrinkage is 0

$\theta \sim Beta(0,0)$ complete ignorance

This posterior could be improper if $x=0$ or $x=n$.

Be careful when using "improper" priors as your posterior could be improper.



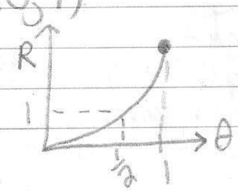


$\alpha = \beta = \frac{1}{2}$

"reparametrization"

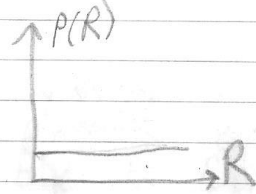
$Odds(\theta) = \frac{\theta}{1-\theta}$ $\quad Odds(0.9) = \frac{0.9}{1-0.9} = 9$

$\theta \in (0,1)$ $\quad R \in (0, \infty)$



Laplace, prior on odds

Jeffreys priors

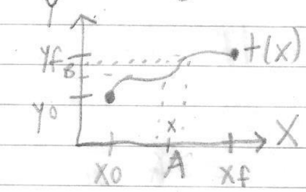


Proper priors converge

r.v.'s X, Y are related by a 1:1 monotonic transformation

$Y = f(X)$, $X = f^{-1}(Y)$

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



$P(X \in A) \approx f_X(x) A$

$\| P(Y \in B) \approx f_Y(y) B$

$\Downarrow$

$f_X(x)|dx| = f_Y(y)|dy|$

$\Rightarrow f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right|$

$\Rightarrow \boxed{f_Y(y) = f_X(f^{-1}(y)) \left|\frac{d}{dy}\left[f^{-1}(y)\right]\right|}$

Transformation of r.v.'s

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

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

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

$\left.\begin{array}{l}\theta \sim U(0,1) \\ \theta \sim Beta(0,0)\end{array}\right\}$ Uninformative

A protocol to pick priors given F.

Under a reparameterization $\phi = f(\theta)$ what if I had a protocol which allowed me to pick prior: $P(X|\theta) \xrightarrow{pick} P(\theta)$, $P(X|\phi) \xrightarrow{pick} P(\phi)$ such that we have:

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

$\Rightarrow$ Jeffreys prior

1. Kernels
2. Fisher Information

<u>Kernels</u>

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

$$f(x;\theta) \propto g(x;\theta) \quad \text{"kernel"}$$

$$f(x;\theta) = \underbrace{\frac{1}{c}}_{\lambda(\theta)} g(x;\theta)$$

$$\int f(x)dx = 1 \Rightarrow \int g(x)dx = \int c f(x) dx = c \underbrace{\int f(x)dx}_{1}$$

$$\Rightarrow \underbrace{c = \int g(x)dx}$$

<u>Note:</u> f, g are 1:1

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

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

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

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

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

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

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

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

Fisher Information

Recall likelihood

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

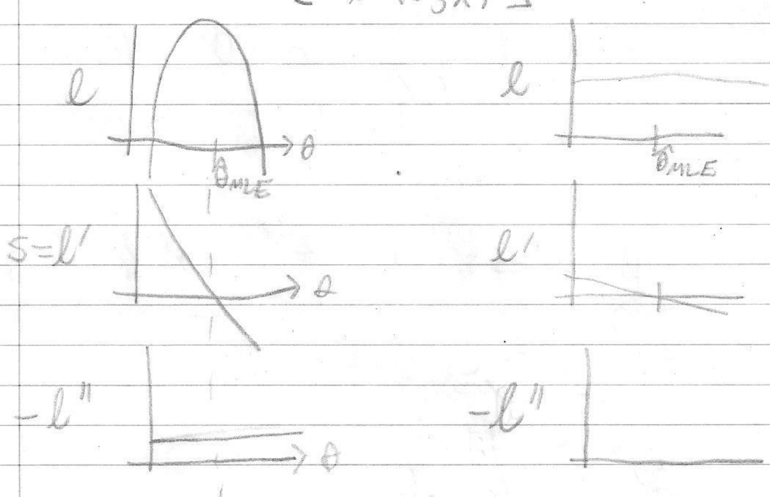
$$l(\theta;x) = \ln(\mathcal{L}(\theta;x))$$

Define the score function

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

Fisher Information $\quad I(\theta) = Var_X[S(\theta;X)] = \ldots = E_X[S(\theta;X)^2]$

$$\ldots = E_X[-l''(\theta;x)]$$



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

$X \sim Binom(n,\theta)$. Calc $I(\theta)$

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

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

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

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

$$= \frac{n\theta}{\theta^2} + \frac{n-n\theta}{(1-\theta)^2} = n\left(\frac{1}{\theta} + \frac{1}{1-\theta}\right) = \boxed{n\,\frac{1}{\theta(1-\theta)}}$$ Fisher for Binomial