3/16/17

↗ how much does this score function vary over all incarnations of the data

$I(\theta) := \text{var}[S(\theta;x)]$ where $S(\theta;x) = \ell'(\theta;x)$

↳ Fisher Information

$$= \cdots = E[S(\theta;x)^2] = \cdots = E[-\ell''(\theta;x)]$$

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

$$\mathcal{L}(\theta;x) = P(x) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$$

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

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

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

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

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

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

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

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

$X_1 \sim \text{Binom}(1, .5) \Rightarrow I(\theta) = 4$

$X_2 \sim \text{Binom}(1, .01) \Rightarrow I(\theta) = 101.01$

Given $\overset{=p(x|\theta)}{F}$, pick $p(\theta)$.

reparametrization ↘

$\phi = t(\theta)$ s.t. t is 1:1, smooth

$$\left.\begin{array}{l} P(X|\theta) \xrightarrow{\text{pick}} P(\theta) \\ p(X|\phi) \xrightarrow{\text{pick}} P(\phi) \end{array}\right\}$$ but you want $P(\theta)$, $P(\phi)$ to be related via change of variables.

<u>Jeffrey's Prior</u>:

$$P(\theta) \propto \sqrt{I(\theta)}$$

$X \sim \text{Binom}(n, \theta) \Rightarrow P(\theta) \propto \sqrt{n\left(\frac{1}{\theta(1-\theta)}\right)} \propto \sqrt{\frac{1}{\theta(1-\theta)}} = \theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}}$

$\propto \text{Beta}(\frac{1}{2}, \frac{1}{2})$

↳ equidistant to Beta(0,0) & Beta(1,1)

PDF of Beta$(\frac{1}{2}, \frac{1}{2})$

$$= \underbrace{\frac{1}{\text{Beta}(\frac{1}{2}, \frac{1}{2})}}_{\pi} \theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}} = \frac{1}{\pi \underbrace{\sqrt{\theta(1-\theta)}}_{\text{arcsin distribution}}} = P(\theta)$$

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

$$P(R) \propto \sqrt{I(R)}$$

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

$$P(X|R) = \binom{n}{x}\left(\frac{R}{R+1}\right)^x \underbrace{\left(1 - \frac{R}{R+1}\right)}_{\frac{1}{R+1}}^{n-x} = \binom{n}{x}\frac{R^x}{(R+1)^n}$$

$$\ell(X|R) = \ln\binom{n}{x} + x\ln(R) - n\ln(R+1)$$

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

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

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

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

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

$$= n\frac{1}{R(R+1)^2}$$

$$\Rightarrow P(R) \propto \sqrt{n\frac{1}{R(R+1)^2}} \propto \frac{1}{\sqrt{R}} \frac{1}{R+1}$$

$$\propto \frac{1}{\pi}\frac{1}{\sqrt{R}}\frac{1}{\sqrt{R+1}} = P(\phi)$$

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

$$= \frac{1}{\pi}\left(\frac{R}{R+1}\right)^{-\frac{1}{2}}\left(\frac{1}{R+1}\right)^{-\frac{1}{2}}\frac{1}{(R+1)^2} = \frac{1}{\pi}R^{-\frac{1}{2}}(R+1)^1\frac{1}{(R+1)^2}$$

$$= \frac{1}{\pi}\frac{1}{\sqrt{R}}\frac{1}{(R+1)} = P(\phi)$$

$P(X|\theta),\ P(X|\Phi)$

$P(\theta) \propto \sqrt{I(\theta)}$

$P(\Phi) \propto \sqrt{I(\Phi)}$

$$P(\Phi) = P_\theta(\underbrace{t^{-1}(\Phi)}_{\theta}) \left| \frac{d}{d\Phi}\left[\underbrace{t^{-1}(\Phi)}_{\theta}\right]\right| \propto \sqrt{I(\Phi)}$$

$$= P_\theta(\theta)\left|\frac{d\theta}{d\Phi}\right| \propto \sqrt{I(\theta)}\left|\frac{d\theta}{d\Phi}\right| = \sqrt{I(\theta)\frac{d\theta}{d\Phi}\frac{d\theta}{d\Phi}}$$

Empirical Bayes

$$= \sqrt{E[S(\theta;x)^2]\frac{d\theta}{d\Phi}\frac{d\theta}{d\Phi}}$$

$$= \sqrt{E\left[S(\theta;x)^2\frac{d\theta}{d\Phi}\frac{d\theta}{d\Phi}\right]} \Rightarrow E\left[\left(\frac{dl}{d\Phi}\right)^2\right]$$

S

$$S = l' \Rightarrow \quad = \sqrt{E\left[\frac{dl}{\cancel{d\theta}}\frac{dl}{\cancel{d\theta}}\cdot\frac{\cancel{d\theta}}{d\Phi}\frac{\cancel{d\theta}}{d\Phi}\right]}$$

$$= \sqrt{E[S(\Phi;x)^2]} = \sqrt{I(\Phi)}$$

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

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

$$\theta|X \sim \text{Beta}(\underbrace{\alpha+x}_{\alpha'}, \underbrace{\beta+n-x}_{\beta'}) \Rightarrow \hat{\theta}_{MMSE} = \frac{\overbrace{\alpha+x}^{\alpha'}}{\underbrace{n+\alpha+\beta}_{\beta'}}$$

Batting Avg (BA), θ.

$$\hat{\theta} := BA = \frac{\#\text{ hits}}{\#\text{ at bats}} = \frac{x}{n} = \hat{\theta}_{MLE}$$

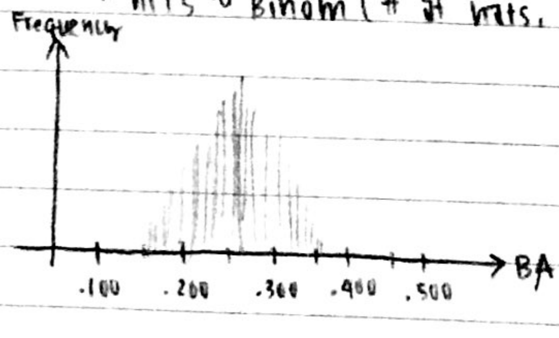
$n=2$

$x=0 \Rightarrow BA=0$

$x=1 \Rightarrow BA=0.5$ } absurd

$x=2 \Rightarrow BA=1$

$\#\text{ hits} \sim \text{Binom}(\#\text{ at bats}, \theta)$



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

$\hat{\alpha}_{MLE} = 78.7$

$\hat{\beta}_{MLE} = 224.8$

$\hat{\alpha}_{MLE} + \hat{\beta}_{MLE} = 303.5$

$$\Rightarrow \hat{\theta}_{MMSE} = \frac{x+78.7}{n+303.5}$$

Empirical Bayes:

① Get old data

② Fit prior to old data

③ use this fits hyperparameters for inference.

$$\underset{\substack{\uparrow \\ \text{x represents} \\ \text{\# of failures}}}{}$$

F = Geometric $\Rightarrow x|\theta \sim (1-\theta)^x \theta$ $\qquad$ $Supp[X] = \{0, 1, \dots\} = \mathbb{N}_0$

$\Theta = (0,1)$

$E[X] = \frac{1}{\theta} - 1$

$\theta \uparrow \Rightarrow x \downarrow$

$\theta \downarrow \Rightarrow x \uparrow$

$x_1|\theta, \dots, x_n|\theta \overset{iid}{\sim} Geo(\theta)$. $\qquad P(x|\theta) = \prod_{i=1}^{n} (1-\theta)^{x_i} \theta = (1-\theta)^{\Sigma x_i} \theta^n$

$P(\theta|x) \propto P(x|\theta)P(\theta) = \underbrace{(1-\theta)^{\Sigma x_i} \theta^n}_{\text{kernel for beta}} P(\theta)$

let $P(\theta) = Beta(\alpha, \beta) = \propto \theta^n (1-\theta)^{\Sigma x_i} \theta^{\alpha-1} (1-\theta)^{\beta-1}$

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

$\Rightarrow$ Beta is also the conjugate prior for the geometric random variable.

$x_1|\theta, \dots, x_n|\theta \overset{iid}{\sim} Geo(\theta)$

$\theta \sim Beta(\alpha, \beta)$

$\theta|x_1, \dots, x_n \sim Beta(\underbrace{\alpha+n}_{\alpha'}, \underbrace{\beta + \Sigma x_i}_{\beta'})$

hyperparameters:

α : pseudo # of trials

β : in the α pseudotrials, this is the sum total of failures ("sum total of pseudofailures")

$\hat{\theta}_{mmse} = \frac{\alpha+n}{\alpha+n+\beta+\Sigma x_i}$

$\hat{\theta}_{MAE} = qbeta(0.5, \alpha+n, \beta+\Sigma x_i)$

$\hat{\theta}_{MAP} = \frac{\alpha+n-1}{\alpha+n+\beta+\Sigma x_i - 2}$ where appropriate

Haldane: $\theta \sim Beta(0,0)$ $\quad \alpha = 0, \beta = 0$

↳ complete ignorance ; saw nothing in prior

Laplace: $\theta \sim U(0,1) = Beta(1,1)$ $\quad \alpha = 1, \beta = 1 \Rightarrow$ when $n \uparrow$, it is uninformative

↳ indifferent priors — "no special privilege to any particular value of θ

Jeffrey's: $\theta \sim Beta(0, \frac{1}{2})$ $\quad \alpha = 0, \beta = \frac{1}{2}$

↳ improper — opened the door so that θ is not degenerate

3/16/17 - (3)

$\mathcal{L}(\theta; x) = (1-\theta)^{\sum x_i} \theta^n$

$\ell(\theta, x) = \sum x_i \ln(1-\theta) + n \ln(\theta)$

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

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

$$
\begin{aligned}
I(\theta) = E[-\ell''(\theta; x)] &= E\left[\frac{\sum x_i}{(1-\theta)^2} + \frac{n}{\theta^2}\right] \\
&= \frac{E[\sum x_i]}{(1-\theta)^2} + \frac{n}{\theta^2} \\
&= \frac{n E[x]}{(1-\theta)^2} + \frac{n}{\theta^2} \\
&= n\left(\frac{\frac{1}{\theta} - 1}{(1-\theta)^2} + \frac{1}{\theta^2}\right) \\
&= n\left(\frac{\frac{\cancel{1-\theta}}{\theta}}{(1-\theta)^{\cancel{2}}} + \frac{1}{\theta^2}\right) \\
&= n\left(\frac{1}{\theta(1-\theta)} + \frac{1}{\theta^2}\right) \\
&= n\left(\frac{1}{\theta^2(1-\theta)}\right)
\end{aligned}
$$

$P(\theta) \propto \sqrt{I(\theta)} = \sqrt{n \frac{1}{\theta^2(1-\theta)}} \propto \theta^{-1} (1-\theta)^{-\frac{1}{2}} \propto Beta(0, \frac{1}{2})$

↳ although your prior is improper, once you get any data, you are proper.