3/22/21 Lec 12

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

$P_J(\theta) = \text{Beta}\left(\frac{1}{2}, \frac{1}{2}\right) = \frac{1}{B(\frac{1}{2}, \frac{1}{2})} \theta^{-\frac{1}{2}} (1-\theta)^{-\frac{1}{2}}$

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

$= \frac{1}{B(\frac{1}{2}, \frac{1}{2})} \left(\frac{\phi}{1+\phi}\right)^{-\frac{1}{2}} \left(\frac{1}{\phi+1}\right)^{-\frac{1}{2}} \frac{1}{(1+\phi)^2}$

$= \frac{1}{B(\frac{1}{2}, \frac{1}{2})} \frac{\phi^{-\frac{1}{2}}}{1+\phi}$

$= \frac{1}{\pi} \phi^{-\frac{1}{2}} (1+\phi)^{-1} = F_{1,1}$ distribution.

$P(x|\phi) = \binom{n}{x} \left(\frac{\phi}{1+\phi}\right)^x \left(\frac{1}{1+\phi}\right)^{n-x}$

$= \binom{n}{x} \frac{\phi^x}{(1+\phi)^n} = \mathcal{L}(\phi; x)$

$P_J(\phi) \propto \sqrt{I(\phi)}$

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

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

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

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

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

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

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

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

$$P_J(\phi) \propto \sqrt{\frac{n}{\phi(1+\phi)^2}} \propto \frac{1}{\sqrt{\phi}} \frac{1}{1+\phi} = \phi^{-\frac{1}{2}}(1+\phi)^{-1} \propto F_{1,1}$$

$$= \frac{1}{\pi} \phi^{-\frac{1}{2}} (1+\phi)^{-1}$$

$$= F_{1,1} \text{ distribution}$$

This varifies that Jeffrey's Procedure works for the binomial model and the odds reparameterization. Let's now prove it for all models and all reparamuterization.

Given $P(X|\theta), P(X|\theta)$, assume $P_J(\theta) \propto \sqrt{I(\theta)}$,
Prove $P(\phi) \propto \sqrt{I(\phi)}$

Proof:

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

$$= \sqrt{I(\theta) \frac{d\theta}{d\phi} \frac{d\theta}{d\phi}}$$

Fact: $I(\theta) := Var_X[\ell'(\theta;X)] = \ldots = E[-\ell''(\theta;X)] = \ldots = E[\ell'(\theta;X)^2]$

$$= \sqrt{E[\ell'(\theta;X)^2] \frac{d\theta}{d\phi} \frac{d\theta}{d\phi}}$$

$$= \sqrt{E\left[\frac{d\ell}{d\theta} \frac{d\theta}{d\theta}\right] \frac{d\theta}{d\phi} \frac{d\theta}{d\phi}} = \sqrt{E\left[\frac{d\ell}{\cancel{d\theta}} \frac{d\theta}{\cancel{d\theta}} \frac{\cancel{d\theta}}{d\phi} \frac{\cancel{d\theta}}{d\phi}\right]}$$

$$= \sqrt{E_X\left[\frac{d\ell}{d\phi} \frac{d\ell}{d\phi}\right]} = \sqrt{E_X[\ell'(\phi;X)^2]} = \sqrt{I(\phi)}$$

QED

Three Principle non-informative priors AKA "Objective"
(a) Laplace / uniform
(b) Haldan
(c) Jeffrey's

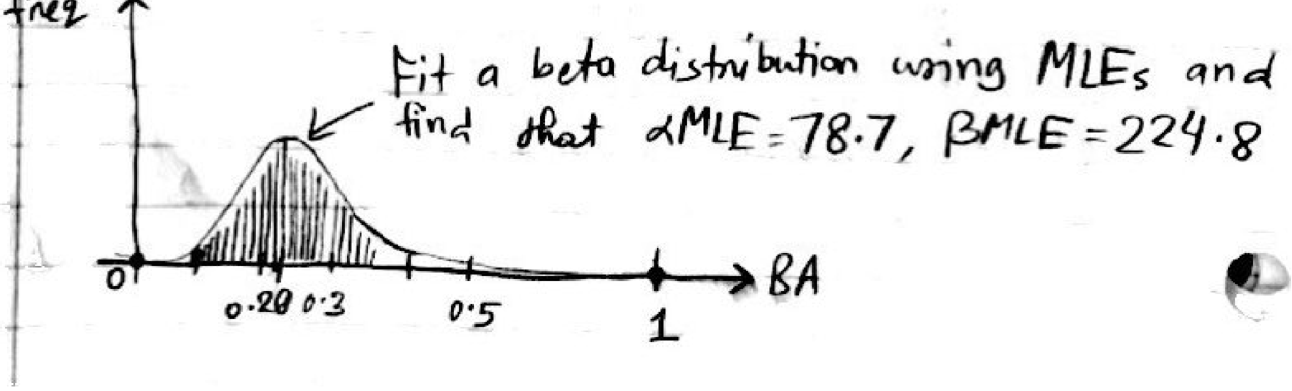
Informative Priors i.e. Subjective Priors! Imagine you're trying to infer a <u>new</u> baseball player's batting ability $\theta$, the probability he gets a hit during an at bat. The batting ability is usually inferred by the "batting average", $BA := \frac{X}{n} = \hat{\theta}_{MLE}$ where $X = \#$ of hits & $n = \#$ of relevent at bats.

The problem is the MLE is a poor estimate if $n$ is low. For example, $n = 3$, $X = 2 \Rightarrow BA = \frac{2}{3} = 0.667$ But this batting ability is impossible. In fact the higest BA ever recorded in baseball history is 0.366 by Ty Cobb.

Will Bayes estimates with uninformative priors help you here? Consider Laplace uniform Prior $\Rightarrow \hat{\theta}_{MMSE} = \frac{3}{5} = 0.600$ which is also absured!

We can solve this by using an uniformative Prior that provides an "empirical Bayes" estimate i.e. uses historical data. Here's how...

Let's Look at previous data. Let's subset on all Players that have at least 500 at bats. (arbitrary cutoff, but we have to start somewhere) If you Plot the BA's, you get something like this:



shrink hard to

Then we use this as our Prior!

$P(\theta) = \text{Beta}(78.7, 224.8) \Rightarrow E[\theta] = 0.260$

$n_0 = 303.5$

Let's use this prior to estimate $\theta$ for our new batter:

$$\hat{\theta}_{MMSE} = (1-\rho)\,\hat{\theta}_{MLE} + \rho E[\theta]$$

$$= \frac{3}{303.5+3} \cdot 0.667 + \frac{303.5}{303.5+3} \cdot 0.260$$

$$\approx 1\% \cdot 0.667 + 99\% \cdot 0.260$$

$$= 0.263$$

The use case for informative Priors is when you belive the new rv behaves like historical rv's behave. Then you can use old data to fit an empirical Bayes prior which will be informative with high Shrinkage. Then use that to do your inference.

$\mathcal{F}: \text{Bin}(n, \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$

Imagine $n \to \infty$ and $\theta \to 0$ but $\overbrace{n\theta = \lambda}^{\theta = \frac{\lambda}{n}} > 0$ but not too big. what is an approximate PMF for this binomial?

$$\lim_{n\to\infty} \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

$$= \lim_{n\to\infty} \frac{n!}{x!(n-x)!} \frac{\lambda^x}{n^x} \left(1-\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{-x}$$

$$= \frac{\lambda^x}{x!} \lim_{n\to\infty} \frac{\overbrace{n!}^{n(n-1)\cdots(n-x+1)}}{(n-x)!\,n^x} \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \lim_{n\to\infty} \frac{\lambda}{n}\right)^{-x}$$

$$= \frac{\lambda^x}{x!} \lim \frac{\overbrace{n(n-1)\cdots(n-x+1)}^{x \text{ terms}}}{\underbrace{n \cdot n \cdots n}_{x \text{ terms}}} (e^{-\lambda})(1)$$

$$= \frac{\lambda^x e^{-\lambda}}{x!} = \text{Poisson}(\lambda)$$

Poisson is an approximation of a binomial if $n$ is large & $\theta$ is small.