lec12

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

$$p_J(\theta) = Beta(\tfrac{1}{2}, \tfrac{1}{3}) = \frac{1}{B(\frac{1}{2}, \frac{1}{3})} \theta^{-\frac{1}{2}} (1-\theta)^{-\frac{1}{2}}$$

$$p_J(\phi) = p_J(t^{-1}(\phi)) \left| \frac{d}{d\phi} \left[\frac{\phi}{1+\phi} \right] \right| = \frac{1}{B(\frac{1}{2}, \frac{1}{2})} \left(\frac{\phi}{1+\phi}\right)^{-\frac{1}{2}} \left(\frac{1}{1+\phi}\right)^{-\frac{1}{2}} \cdot \left(\frac{1}{1+\phi}\right)^2$$

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

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

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

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

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

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

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

$I(\phi) = E_x\left[- \quad\right] = E\left[\frac{x}{\phi^2} - \frac{n}{(1+\phi)^2}\right] = \frac{1}{\phi^2}E[X] - \frac{n}{(1+\phi)^2}$

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

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

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

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

This verifies that Jeffrey procedure works for the Binomial mode and the odd reparameterization. Let's now prove it for all models all reparameterizations:

Given $P(X|\theta)$, $t \Rightarrow P(X|\phi)$ assume $P_J(\theta) \propto \sqrt{I(\theta)}$, prove $P(\phi) \propto \sqrt{I(\phi)}$, proof:

$P_J(\theta)$

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

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

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

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

We have three principaled non-informative priors AkA "objective"

(a) Laplace / uniform

(b) Haldance

(c) Jeffreys

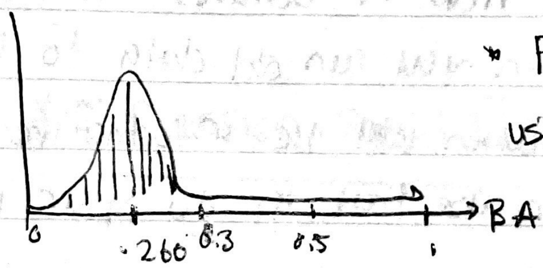
Informative priors i.e. Subjective priors! Imagine you are trying to infer a new baseball player's batting ability θ, the priobability he gets a hit during an at bat. The batting ability is usually inferred by the "batting average", $BA = \frac{X}{n} = \hat{\theta}_{MLE}$ where X is the # of hits and n is the # of relevant at bats

The problem is the MLE is a poor estimate if n is low. For example, $n=3$, $X=2 \Rightarrow BA = 2/3 = .667$. But this batting ability is impossible. In fact the highest BA ever recorded in baseball history is .366 by Cobb.

Will Bayes estimate with uninformative priors help you here?
No! consider Laplace uniform prior => $\hat{\theta}_{MMSE} = 3/5 = 0.600$
which is also absurd.

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

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



* Fit a beta distribution using MLE's and find that

alphaMLE = 78.7

and betaMLE = 224.8

Then we use this as our prior!

$P(\theta) = \text{Beta}(78.7, 224.8) \Rightarrow$

$E[\theta] = .260$

$n_0 = 303.5$

Shrink hard to this

Let's use this prior to estimate θ for our new batter:

$$\hat{\theta}_{MMSE} = (1-e)\hat{\theta}_{MLE} + eE[\theta]$$

$$= \frac{3}{303.5+3} \times .667 + \frac{303.5}{303.5+3} \times .260$$

$$= 1\% \times .667 + 99\% \times .260 = .263$$

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

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

Imagine $n \to \infty$ and θ goes to 0, but $n\theta = \lambda > 0$ but not too big. What is the an approximate PMF for this binomial?

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

$$\frac{\lambda^x}{x!} \lim \frac{n!}{(n-x)!\,n^x} \quad \lim \frac{n!}{(n-x)!\,n^x} \quad \lim\left(1-\frac{\lambda}{n}\right)^n \left(1-\lim \frac{\lambda}{n}\right)^{-x}$$

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

Poisson is an approximation of a binomial if n is large and θ is small.

$\lambda \in (0, \infty)$, $Y \sim$ Poission, $\text{supp}[Y] = \{0, 1, 2 \ldots\} = \mathbb{N}_0 =$

lec 13