CI

?

Math 381 Lec 11 3/14/18

Uniform prior: a prior which does not have a "large" effect on posterior

$\theta \sim U(0,1) = \text{Beta}(1,1)$

$\theta | X \sim \text{Beta}(\overset{\alpha}{1} + x, \overset{\beta}{1} + n - x) \Rightarrow \hat{\theta}_{MMSE} = \frac{x+1}{n+2}$ AKA "Wilson estimate"

U(0,1) even though "uniform" yielded pseudocounts of 1 success, 1 failure. That's not "no information"!

What would NO info look like? Uniform? YES

$\theta | X \sim \text{Beta}(\overset{\alpha}{0} + x, \overset{\beta}{0} + n - x) \Rightarrow \hat{\theta}_{MMSE} = \frac{x}{n} = \hat{\theta}_{MLE}$

$\Rightarrow \theta \sim \text{Beta}(0,0)$ what's wrong? $\alpha > 0, \beta > 0$

this is an illegal prior! AKA "improper prior"

But posterior is proper if $X \neq 0$ and $x \neq n$

Tons of theory on this... some say improper priors okay, some say no.

Okay for us... except... you must be careful your posterior is proper!

$\theta \sim \text{Beta}(1,1) \Rightarrow$ indifference but successes and failures are known to be possible

$\theta \sim \text{Beta}(0,0) \Rightarrow$ successes and failures not known to be possible

AKA complete ignorance

Haldane prior (1932)

New concept:

You are trying to guess a baseball player's the B.A., θ. The sample BA is

$$\hat{\theta} = BA := \frac{\# HITS}{\# AT BATS} = \frac{x}{n}$$

with some approx's ... we use the model ..

$$\# HITS \approx Bin(\# at bats, \theta)$$

BA is the $\hat{\theta}_{MLE}$

When does $\hat{\theta}_{MLE}$ have poor performance? If n is small, lets say $n=2$

$n=2$, $x=0 \Rightarrow \hat{\theta}=0$, $x=1 \Rightarrow \hat{\theta}=0.5$, $x=2 \Rightarrow \hat{\theta}=1$ all absurd!

Solution! Shrink! Use $\theta \sim Beta(\alpha, \beta)$

$\hat{\theta}_{MMSE} = \frac{x+\alpha}{\alpha+\beta+n}$ which includes a shrinkage towards $\frac{\alpha}{\alpha+\beta}$ with weight $\frac{\alpha+\beta}{n+\alpha+\beta}$

How to pick prior?

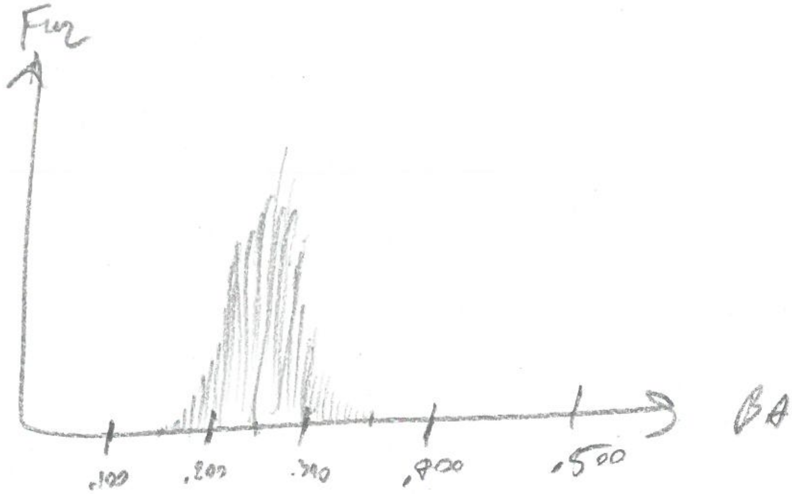
$\theta \sim Beta(1,1)$ Shrink towards 0.5 absurd!

How about look at all historical BA's for tons of players!

i.e. use prior know to build a prior $\Rightarrow$ empirical Bayes.

Here's how it works. Get many reasonable $\hat{\theta}$'s for previous players

let's say $n > 500$ at bats.



Fit a beta to the prior data. Using MLE's, $\hat{\alpha}_{MLE} = 78.7$, $\hat{\beta}_{MLE} = 224.8$

This has the strength of $n = 303.5$ at bats $\Rightarrow$ Strong!

$$\hat{\theta}_{MMSE} = \frac{x + 78.7}{n + 303.5}$$ which will perform better than $\hat{\theta}_{MLE} = \frac{x}{n}$ for n small.

Steps

① Get old data

② Fit conjugate prior dist to it using MLE's

③ Use the fit hyperparameters for inference

Done with beta, binomial ... on to new model..

$$\mathcal{F} = \text{Geometric} \Rightarrow X|\theta \sim \text{Geom}(\theta) := (1-\theta)^x \theta$$

x represents # failures

We see n trials .. $X = \{X_1, \ldots, X_n\}$

$$P(X;\theta) = \prod_{i=1}^{n} (1-\theta)^x \theta = (1-\theta)^{\sum x_i} \theta^n$$

$\text{Supp}[X] = \{0, 1, \ldots\} = \mathbb{N}_0$

Param space $\Theta = (0,1)$

$E[X] = \sum_{x=0}^{\infty} x(1-\theta)^x \theta = \ldots = \frac{1}{\theta} - 1$

$\theta \uparrow \Rightarrow x \downarrow$, $\theta \downarrow \Rightarrow x \uparrow$

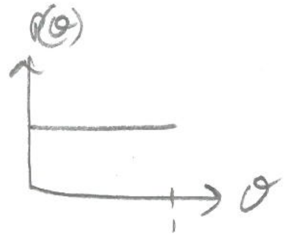
Math 241 exercise

$\hat{\theta}_{MLE} = \frac{1}{1+\bar{x}}$ ~~we did this before~~

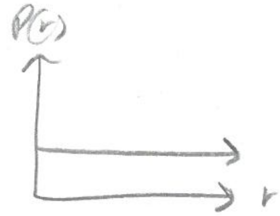
Another problem

$\theta \sim U(0,1)$ means every prob equally likely

What if I consider the odds, $r := \frac{\theta}{1-\theta}$. Am I indifferent on this scale?



$\theta \in (0,1)$



$r \in (0,\infty)$

No... not possible!

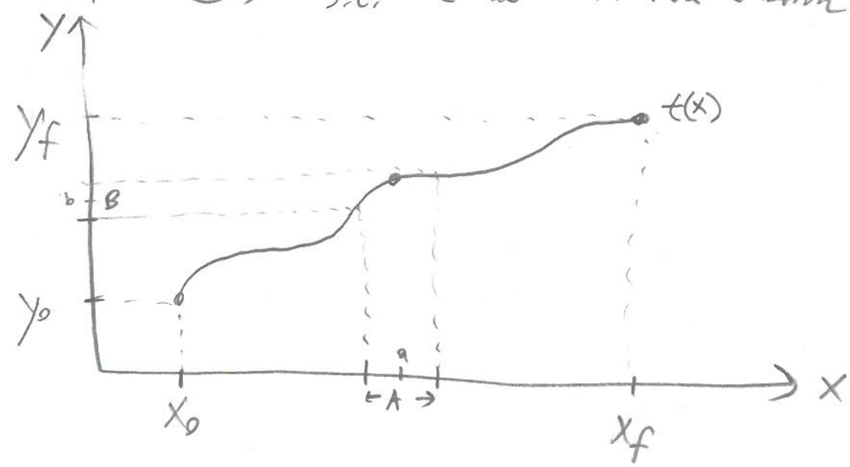
the principle of indifference has a problem!

What is PDF of R?

Math 621 covers transformation of variables.

Imagine r.v.'s X, Y with densities f_X, f_Y with f_X known, f_Y unknown

Suppose $Y = t(X)$ s.t. t is an invertible function



$Supp[X] = [x_0, x_f]$, $Supp[Y] = [y_0, y_f]$

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

$P(Y \in B) \approx f_Y(b) B$

$P(X \in A) = P(Y \in B)$

$\Rightarrow f_X(a) A = f_Y(b) B$ Note $A > 0, B > 0$

let area be small

$f_X(a)|dx| = f_Y(b)|dy|$ s.t. $b = t(a)$ or $a = t^{-1}(b)$

Solve for $f_Y(y)$ $\Rightarrow f_Y(b) = f_X(a) \left|\frac{dx}{dy}\right|$

let $b = y$ the usual dummy variable

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

What is PDF of R? $\theta \sim U(0,1) \Rightarrow f_\theta(\theta)=1$ $\quad R - RX = X \Rightarrow R = X + RX \Rightarrow R = (1+R)X \Rightarrow$

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

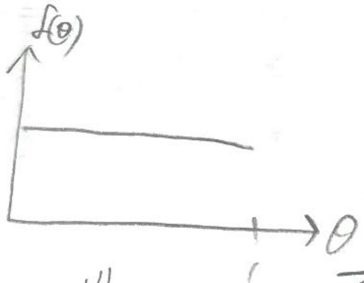
quotient rule

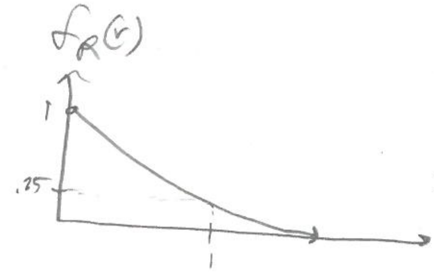
reparameterization!

$$f_R(r) = \underbrace{f_X\left(t^{-1}(r)\right)}_{=1} \left|\frac{d}{dr}\left[t^{-1}(r)\right]\right| = \left|\frac{(r+1)(1) - (r)(1)}{(r+1)^2}\right| = \frac{1}{(r+1)^2}$$

is this a density? $\displaystyle\int_{Supp(R)} f_R(r)\,dr = 1 \Rightarrow \int_0^\infty \frac{1}{(r+1)^2}\,dr = \left[\frac{r}{r+1}\right]_0^\infty = 1$ ✓

Note $f_R(r) \neq f_\theta(\theta) \Rightarrow$ If you are indifferent about θ, you are <u>not</u> indifferent about odds or any other monotone transformation of θ

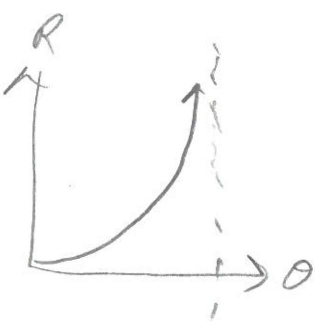




Fisher used this to show Bayes is stupid!

Is this a problem if the prior is discrete & finite?

NO



$\theta \in [0, 0.5] \Rightarrow R \in [0,1]$

$\theta \in [0.5, 1) \Rightarrow R \in [1, \infty)$

~~The problem: but $f_\theta(\theta = 0.1) = 1$~~

~~$f_R(\theta = 0.1) \stackrel{?}{=}$~~

~~$= \frac{1}{(\frac{1}{9}+1)^2} = 0.81 \neq 1$~~

~~Off question: is there any way out of this?~~

~~Is there a prior $\theta \sim p(\theta)$ s.t. any reparameterization will yield the same prior density?~~

$$\theta \sim U(0,1) \overset{\text{Beta}(1,1)}{=} \quad \text{Laplace prior}$$
$$\theta \sim \text{Beta}(0,0) \quad \text{Haldane prior}$$

} Uniformne, Vague, Flat, Diffuse

$\Downarrow$

don't affect inference too much.

Strategy
protocol

Is there a <u>way</u> to choose an uniformne prior that would be the same under reparameterizations?

Likelihood Model

$$p(X|\theta) \xrightarrow{\text{"Strategy"}} \text{pick } p(\theta)$$

and make a reparameterization:

$\phi = t(\theta)$: s.t. t is 1:1 and monotonic

$$p(X|\phi) \xrightarrow{\text{"Strategy"}} \text{pick } p(\phi)$$

wouldn't it be nice if

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

As in the strategy itself would ensure this "invariance" wouldn't break?

This is the strategy Jeffreys invented $\approx$ 1930'S.

Before we get there, we need two prereqs.

1. "Kernels"
2. Fisher Information

Kernels) Recall ...

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

why? $P(x)$ not a function of θ!

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

this means by def that $\exists c \in \mathbb{R}$ s.t.

$$f(x;\theta) = \frac{1}{c}\, g(x;\theta)$$

Given $g(x)$, How to find c?

Note: $\displaystyle\int_{supp(x)} f(x)\,dx = 1$

$$\int_{supp(x)} g(x)\,dx = \int c f(x)\,dx = c\underbrace{\int f(x)\,dx}_{1} = \boxed{c = \int g(x)\,dx}$$

Note: $\int g(x)\,dx < \infty$ ... and ... $\int g(x)\,dx > 0$

Note: $g(x)$, $f(x)$ are 1:1

r.v.'s can be identified by their kernels

So let's see this

kernel

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

$$\propto \text{Beta}(x+\alpha,\, n-x+\beta)$$

why?

$$\theta \sim \text{Beta}(\alpha,\beta) := \frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} \propto \theta^{\alpha-1}(1-\theta)^{\beta-1} = \underbrace{\theta^{9}(1-\theta)^{6}}$$

the kernel of the beta

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

$$\propto \frac{1}{x!(n-x)!}\left(\frac{\theta}{1-\theta}\right)^{x}$$ ← kernel of the binomial!!