lec 5 2/16/17 Math 341

$\mathcal{F}$ = Bernoulli, $X = \langle 0,1,1 \rangle$, $\theta \sim U(0,1)$

$$p(\theta \mid X) = 12\,\theta^2(1-\theta)$$

$p(\theta)$ $\Rightarrow$ $p(\theta \mid X)$

let's generalize . .

$\mathcal{F}$ = Bernoulli $X = X_1, \ldots, X_n$, $\theta \sim U(0,1)$

$$p(\theta \mid X) = \frac{p(X \mid \theta)\, p(\theta)}{p(X)} = \frac{p(X \mid \theta)\, p(\theta)}{\int p(X \mid \theta)\, p(\theta)\, d\theta}$$

$$p(X \mid \theta) = \prod_{i=1}^{n} p(X_i \mid \theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} = \theta^{\Sigma x_i}(1-\theta)^{n-\Sigma x_i} = \theta^{x}(1-\theta)^{n-x}$$

$\Rightarrow$ post. only dep. on Σx_i . . . let $x := \Sigma x_i$ from now on

$$\Rightarrow p(X \mid \theta) = \frac{\theta^x(1-\theta)^{n-x}}{\underbrace{\int_0^1 \theta^x(1-\theta)^{n-x}\, d\theta}_{\text{famous function}}}$$

$$B(\alpha,\beta) := \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\, dt$$

capital β from β-function

$$P(\theta|x) = \frac{1}{B(x+1, n-x+1)} \theta^x (1-\theta)^{n-x} = \text{Beta}(x+1, n-x+1)$$

A new r.v.!

$$X \sim \text{Beta}(\alpha, \beta) := \underbrace{\frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}}_{f(x)} \quad \text{Supp}[X] = [0, 1]$$

Valid PDF?

$$\int_{\text{Supp}(x)} f(x)\,dx = 1 \Rightarrow \int_0^1 \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}\,dx = \frac{1}{\cancel{B(\alpha,\beta)}} \cancel{B(\alpha,\beta)} = 1 \checkmark$$

For what values of α, β does this hold? I.e. ... what is the parameter space?

When is $B(\alpha,\beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,dt$ finite?

only if $\alpha > 0$ and $\beta > 0$ $\Rightarrow$ param space

It can be shown that

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \quad \text{where} \quad \Gamma(x) := \int_0^\infty e^{-t} t^{x-1}\,dt$$

the gamma function

where the gamma function is the extension of the factorial function to all #'s

Properties $\Gamma(x) = (x-1)! = (x-1)(x-2)! = (x-1)\Gamma(x-1)$

Using this property...

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

$$E(X) = \int_0^1 x \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}\,dx$$

$$= \int_0^1 \frac{1}{B(\alpha,\beta)} x^{\alpha}(1-\alpha)^{\beta-1}\,dx$$

$$= \frac{B(\alpha+1,\beta)}{B(\alpha,\beta)} = \frac{\dfrac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)}}{\dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}} = \frac{\dfrac{\alpha\,\cancel{\Gamma(\alpha)}\,\cancel{\Gamma(\beta)}}{(\alpha+\beta)\,\cancel{\Gamma(\alpha+\beta)}}}{\dfrac{\cancel{\Gamma(\alpha)}\,\cancel{\Gamma(\beta)}}{\cancel{\Gamma(\alpha+\beta)}}} = \boxed{\frac{\alpha}{\alpha+\beta}}$$

$Var(X)$... calc on HW (similar)

$$\text{Mode}(X) = \arg\max \left\{ \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} \right\}$$

$$= \arg\max \left\{ x^{\alpha-1}(1-x)^{\beta-1} \right\} = \arg\max \left\{ (\alpha-1)\ln(x) + (\beta-1)\ln(1-x) \right\}$$

take derivative... set $= 0$

$$x^{\alpha-1}(\beta-1)(1-x)^{\beta-2} + (1-x)^{\beta-1}(\alpha-1)x^{\alpha-2} = 0$$

$$x^{\alpha+1}(\beta-1)(1-x)^{\beta-2} + (1-x)^{\beta-1}(\alpha-1) = 0$$

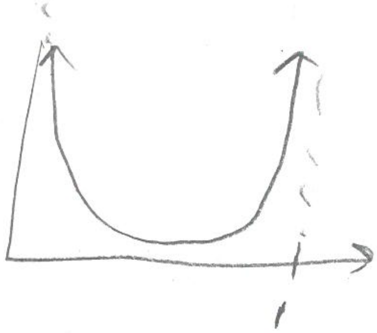
$$x^{\alpha+1}(\beta-1) + (1-x)^{\beta+1}(\alpha-1) = 0 \Rightarrow \frac{x^{\alpha+1}}{(1-x)^{\beta+1}} = \frac{\beta-1}{\alpha-1}$$

take deriv set $= 0$

$$\frac{\alpha-1}{x} - \frac{\beta-1}{1-x} = 0 \Rightarrow \frac{\alpha-1}{x} = \frac{\beta-1}{1-x} \Rightarrow \frac{1}{x} - 1 = \frac{\beta-1}{\alpha-1}$$

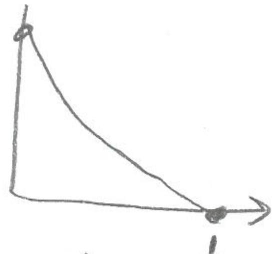
$$\Rightarrow \frac{1}{x} = \frac{\beta+\alpha-2}{\alpha-1} \Rightarrow x = \frac{\alpha-1}{\alpha+\beta-2}$$ only for $\alpha > 1, \beta > 1$ o/t local min!

Shapes of the Beta r.v. density:

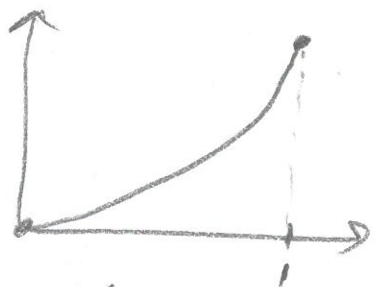


$\alpha = \beta = 0.5$

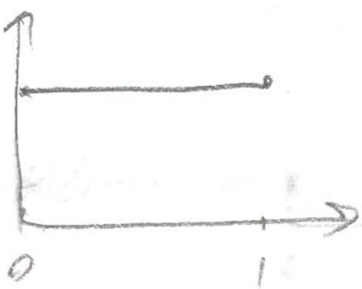
"arcsin distr"



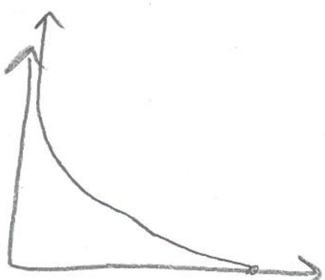
$\alpha = 1, \beta = 3$



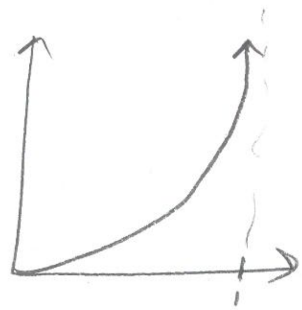
$\alpha = 5, \beta = 1$



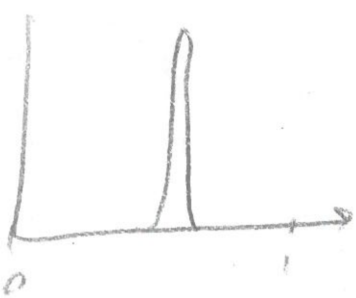
$\alpha = \beta = 1$



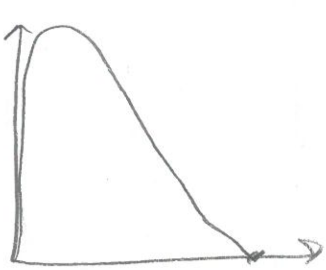
$\alpha = 0.99, \beta = 3$



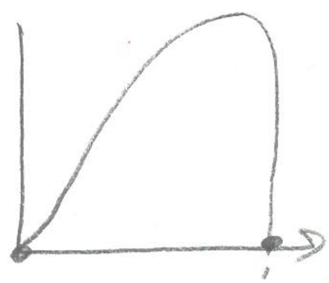
$\alpha = 5, \beta = 0.99$



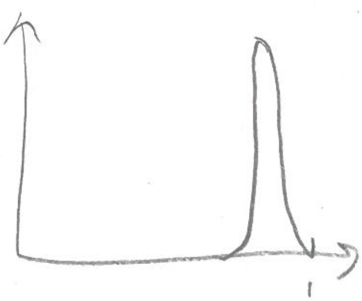
$\alpha = \beta = 100$



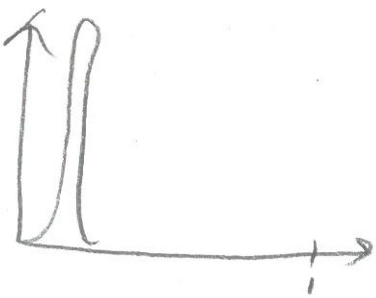
$\alpha = 1.01, \beta = 3$



$\alpha = 5, \beta = 1.01$



$\alpha = 100, \beta = 10$



$\alpha = 10, \beta = 100$

5

Let's return to Bayesian...

We saw before for $X_1, \ldots, X_n$ from $\mathcal{F}$ = Bernoulli, only the sum of the x's mattered. Let's consider just X as a binomial

$$X \sim \text{Binom}(n, \theta) := \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

let $\theta \sim U(0,1)$

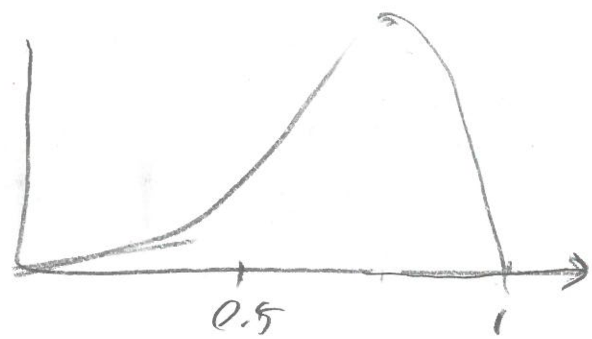
Same!

$$P(\theta \mid x) = \frac{P(x \mid \theta)\, P(\theta)}{P(x)} = \frac{\cancel{\binom{n}{x}} \theta^x (1-\theta)^{n-x}}{\int_0^1 \cancel{\binom{n}{x}} \theta^x (1-\theta)^{n-x}} = \text{Beta}(\underbrace{x+1}_{\alpha'}, \underbrace{n-x+1}_{\beta'})$$

$$P(x) = \int_\Theta P(x \mid \theta) P(\theta)\, d\theta$$

if $n=10$, $x=7$

$\theta \mid x \sim \text{Beta}(8, 4)$



$$\hat{\theta}_{MAP} = \frac{\alpha' - 1}{\alpha' + \beta' - 2} = \frac{8-1}{8+4-2} = \frac{7}{10} = 0.7$$

What about the average $\theta \mid x$?

$$E[\theta \mid x] = \frac{\alpha'}{\alpha' + \beta'} = \frac{8}{8+4} = 0.66$$

Avg

$$E[\theta] = \frac{\alpha}{\alpha+\beta} = 0.5$$

what's this? Prior mean!!

What about the median $\theta \mid x$?

$\text{Med}[\theta \mid x]$ ... no closed form... in R... qbeta(0.5, 8, 4) ≈ 0.676...

Three different means of pt. estimation.

$\hat{\theta}_{MAP} = \text{argmax}\{P(\theta|x)\}$ posterior mode

$\hat{\theta}_{MMSE} = E[\theta|x]$ posterior expectation / mean

$\hat{\theta}_{MAE} = \text{Med}[\theta|x]$ posterior median

turns out... $\hat{\theta}_{MMSE}$ minimizes squared error loss

$= \text{argmin}\ E[(\theta - \hat{\theta}_{MMSE})^2]$

$\hat{\theta}_{MAE}$ minimizes abs err loss

$= \text{argmin}\ E[|\theta - \hat{\theta}_{MAE}|]$

We will be using all 3. Default is $\hat{\theta}_{MMSE} = E[\theta|x]$

but $\hat{\theta}_{MAP}$ is easiest to get! No need to compute $P(x)$!

New idea... previously...

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

Why not let

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

where I choose α, β to reflect my prior information!

Forget how I chose α, β for now...

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)} = \frac{\left(\cancel{\binom{n}{x}}\, \theta^{x}(1-\theta)^{n-x}\right)\left(\frac{1}{\cancel{B(\alpha,\beta)}}\, \theta^{\alpha-1}(1-\theta)^{\beta-1}\right)}{\int_0^1 \left(\cancel{\binom{n}{x}}\, \theta^{x}(1-\theta)^{n-x}\right)\left(\frac{1}{\cancel{B(\alpha,\beta)}}\, \theta^{\alpha-1}(1-\theta)^{\beta-1}\right) d\theta}$$

$$= \frac{\theta^{x+\alpha-1}\,(1-\theta)^{n-x+\beta-1}}{\int_0^1 \theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}\, d\theta}$$

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

$$= \text{Beta}(x+\alpha,\, n-x+\beta)$$ Another Beta!

prior beta, posterior beta $\Rightarrow$ conjugacy Seems like a "natural" choice!

Def: P is the "conjugate prior" for F if the posterior also has same distribution

$\Rightarrow$ Beta is the conjugate prior for the binomial (or the Bernoulli)

$\hat{\theta}_{MAP} = \dfrac{x+\alpha-1}{\alpha+\beta+n-2}$, $\hat{\theta}_{MMSE} = \dfrac{x+\alpha}{n+\alpha+\beta}$, $\hat{\theta}_{MAE}$ ... use computer $= \text{qbeta}(0.5, x+\alpha, n-x+\beta)$

---

prediction distr.

After seeing n Bernoulli trials / one Binomial trial. what is the distr. of the next observation (one only)?

$P(X^*|x)$?