Lecture 6

02/13/20

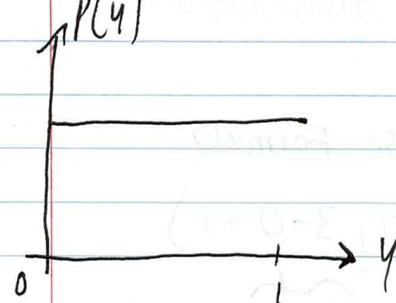
Beta Function: $B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,dt$

Parameter Space: $\alpha, \beta > 0$ $\quad$ $Supp(y) = (0,1)$

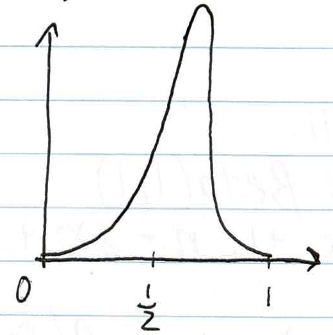
$$Y \sim Beta(\alpha, \beta) = \frac{1}{B(\alpha,\beta)} y^{\alpha-1}(1-y)^{\beta-1}$$

$E[Y] = \dfrac{\alpha}{\alpha+\beta}$ $\quad$ $Mode[Y] = \dfrac{\alpha-1}{\alpha+\beta-2}$ $\quad \alpha, \beta \geq 1$

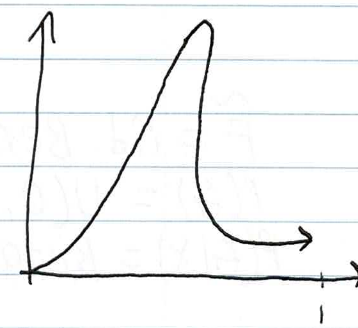
$Med[Y] = qbeta(.5, \alpha, \beta)$
(no closed form)



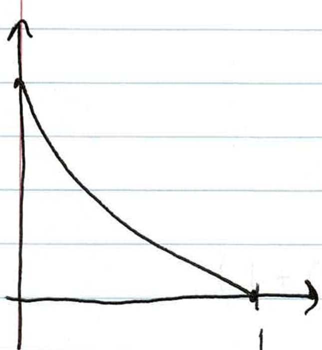
$\alpha = \beta = 1$ $\quad U(0,1)$



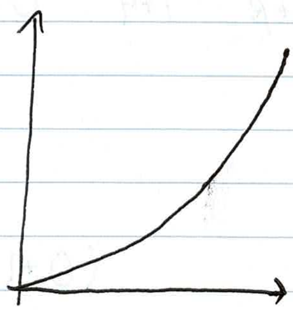
$\alpha = \beta = 100$



$\alpha = 10, \beta = 90$

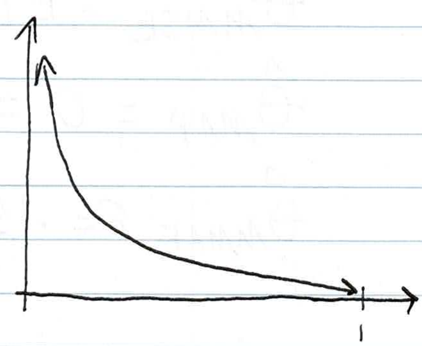


$\alpha = 1, \beta = 4$

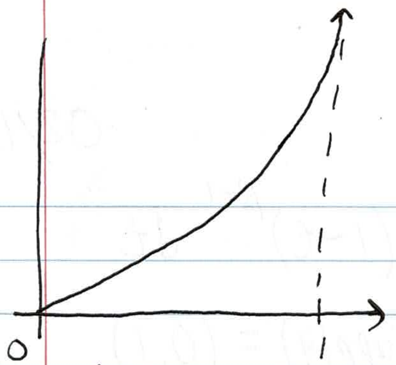


$\alpha = 4, \beta = 1$

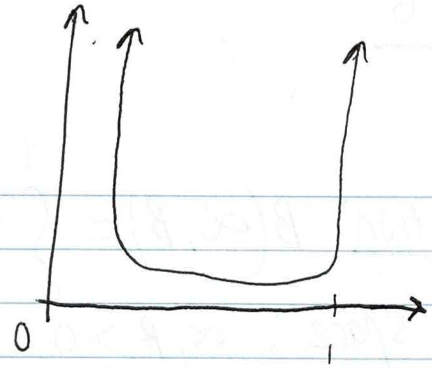
$\varepsilon > 0$
but small



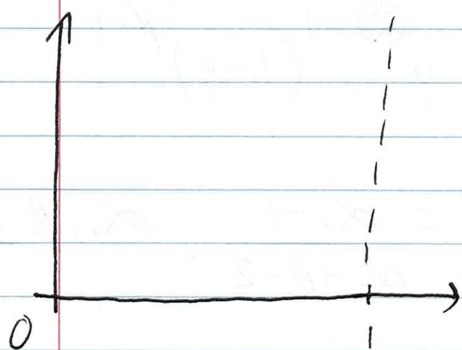
$\alpha = 1 - \varepsilon \quad \beta = 4$



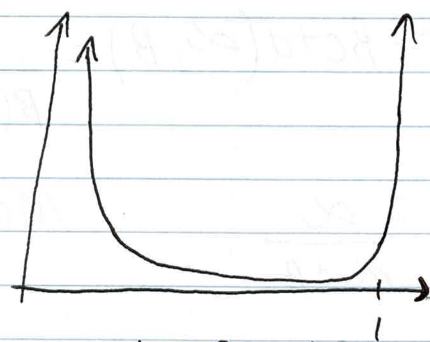
$\alpha = 4 \cdot \beta = 1 - \varepsilon$



$\alpha = \beta = 1 - \varepsilon$



$\alpha = \beta - \varepsilon$



$\alpha = \beta = \frac{1}{2}$

arcsine distribution

$\tilde{F} =$ iid Bernoulli

$P(\theta) = U(0,1) = Beta(1,1)$

$P(\theta|X) = Beta(\Sigma X_i + 1, n - \Sigma X_i + 1)$

use Formula

$(0+1, 3-0+1)$

Let $X = \langle 0, 0, 0 \rangle$ $\quad P(\theta|X) = Beta(1,4)$

$\hat{\theta}_{MMSE} = E[\theta|X] = \frac{\alpha}{\alpha+\beta} = \frac{1}{1+4} = 0.2$

$\hat{\theta}_{MAP} = 0 = \hat{\theta}_{MLE}$

$\hat{\theta}_{MMAE} \approx .159$

$(0+1, 1-0+1)$

$P(\theta|X_1) = \frac{P(X_1|\theta)P(\theta)}{P(X)} = Beta(1,2)$

$$P(\theta|X_1,X_2) = \frac{P(X_1,X_2|\theta)\overbrace{P(\theta|X_1)}^{\rightarrow Beta(1,2)}}{P(X_1,X_2)} = Beta(1,3)$$

$$P(\theta|X_1,X_2,X_3) = \frac{P(X_1,X_2,X_3|\theta)\overbrace{P(\theta|X_1X_2)}^{Beta(1,3)}}{P(X_1,X_2,X_3)} = Beta(1,4)$$

$\tilde{F}$ = iid Bernoulli $\quad X$ is n observations

$P(\theta) = Beta(\alpha, \beta)$

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{\int_0^1 P(X|\theta)P(\theta)\,d\theta}$$

cancel terms

$$= \frac{\overbrace{\left(\theta^{\Sigma x_i}(1-\theta)^{n-\Sigma x_i}\right)}^{likelihood}\cancel{\frac{1}{Beta(\alpha,\beta)}}(1-\theta)^{\beta-1}}{\int_0^1 \theta^{\Sigma x_i}(1-\theta)^{n-\Sigma x_i}\cancel{\frac{1}{Beta(\alpha,\beta)}}(1-\theta)^{\beta-1}\,d\theta}$$

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

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

$$= Beta(\Sigma x_i+\alpha, n-\Sigma x_i+\beta)$$

$$P(\theta) \xrightarrow{x} P(\theta|x)$$

$$\underbrace{Beta(\alpha, \beta)}_{\text{Prior Parameters}} \xrightarrow{X} Beta(\underbrace{\sum x_i + \alpha}_{\alpha'}, \underbrace{n - \sum x_i + \beta}_{\beta'})$$

Posterior Parameters

$$\hat{\theta}_{MMSE} = E[\theta|X] = \frac{\sum x_i + \alpha}{n + \alpha + \beta}$$

$$\hat{\theta}_{MMAE} = Med[\theta|X] = qbeta(.5, \sum x_i + \alpha, n - \sum x_i + \beta)$$

$$\hat{\theta}_{MME} = mode[\theta|X] = \frac{\sum x_i + \alpha - 1}{n - \sum x_i + \beta - 2} \quad \text{if } \alpha + X > 1, \text{ if } \beta + n - X > 1$$

<u>Conjugacy</u> for a given likelihood model means the prior and the posterior have the same r.v. (different parameters). Beta is the conjugate prior to the iid Bernoulli likelihood.

$\tilde{F} = Bin(n, \theta)$ with n fixed, known and θ unknown

$\Downarrow$

$\binom{n}{x}\theta^x(1-\theta)^{n-x}$

Let $X = \sum x_i$ in the iid Bernoulli model

$n - X = n - \sum x_i$

$$P(\theta) = Beta(\alpha, \beta) \Rightarrow P(\theta|x) = Beta(X + \underbrace{\alpha}_{\#\text{ of successes}}, n - X + \overbrace{\beta}^{\#\text{ of failures}})$$

\# of prior success or # of pseudosuccess

\# of prior failure or # of pseudo failures

pseudocounts

$$p(\theta) = U(0,1) = \mathrm{Beta}(\overset{\alpha}{1}, \overset{\beta}{1})$$

Principle of indifference $\qquad n_0 = 2$

$$E[\theta] = \frac{1}{2}$$

The principle of indifference is "not so indifferent" because it contains information

Default Point Estimate $\leftarrow \hat{\theta}_{MMSE}$ → Shrinkage Estimator

$$\hat{\theta}_{MMSE} = E[\theta \mid x] = \frac{X+\alpha}{n+\alpha+\beta} = \frac{X}{n+\alpha+\beta}\left(\frac{n}{n}\right) + \frac{\alpha}{n+\alpha+\beta}\left(\frac{\alpha+\beta}{\alpha+\beta}\right)$$

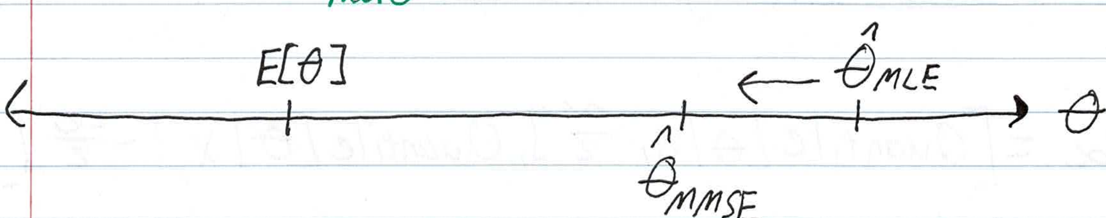
$$= \underbrace{\frac{n}{n+\alpha+\beta}}_{1-\rho} \cdot \overbrace{\frac{X}{n}}^{\bar{X}} + \underbrace{\frac{\alpha+\beta}{n+\alpha+\beta}}_{\rho \to \text{rho}} \cdot \frac{\alpha}{\alpha+\beta}$$

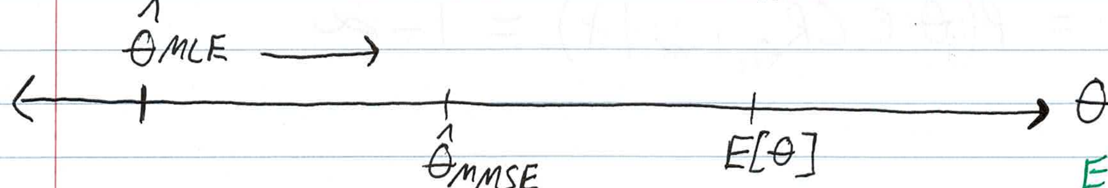
rho

Linear Combination $\vec{w} + \vec{x}$

$$\boxed{\hat{\theta}_{MMSE} = \underbrace{(1-\rho)\,\hat{\theta}_{MLE}}_{\text{If } n \text{ is large "weigh" this more}} + \underbrace{\rho\, E[\theta]}_{\text{If } n \text{ is small, "weigh" this more}}}$$

"Shrinking"



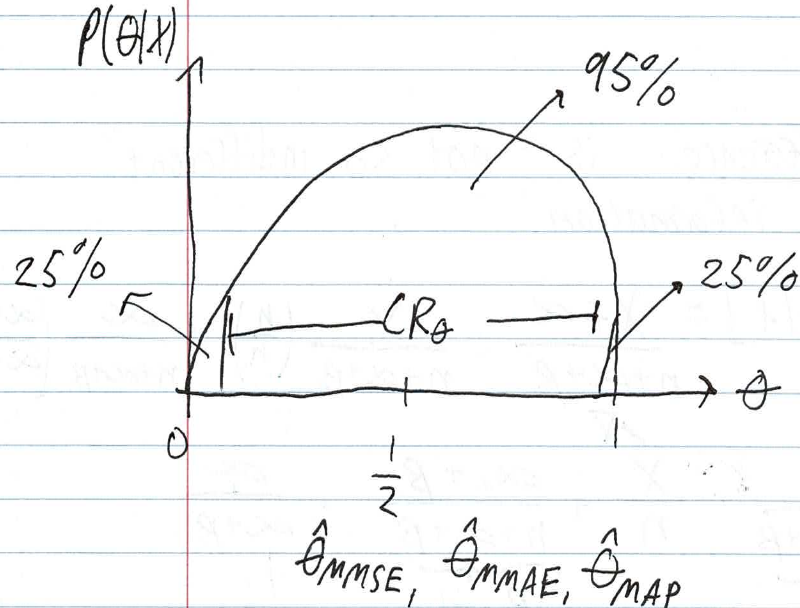


If α, β is small $\Rightarrow$ prior Uninformative → Expresses Vague info about a Variable

If α, β is large $\Rightarrow$ prior informative → Expresses definite info about a Variable

$\tilde{F}$ = Binomial $\quad X=1,\ n=2$

$P(\theta)$ prior of indifference $\Rightarrow P(\theta|x) = \text{Beta}(2,2)$



2^{nd} goal of inference: Confidence Set

$$CI_{\theta,\,95\%} = \left[0.5 \pm 2\sqrt{\frac{.5(1-.05)}{2}}\right] = [-0.2,\ 1.21]$$

i.e. nonsense $\rightarrow$ out of range of possible θ's

Bayesian Credible Ranges (CR)

$$CR_{\theta,1-\alpha} = \left[\text{Quantile}\left[\theta|X, \tfrac{\alpha}{2}\right], \text{Quantile}\left[\theta|x, 1-\tfrac{\alpha}{2}\right]\right]$$
$$= P(\theta \in CR_{\theta,1-\alpha}|x) = 1-\alpha$$