Lecture - 06 02|13|2020

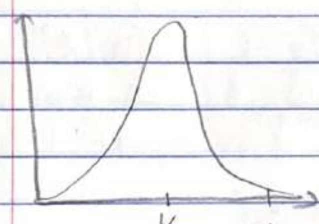
$\beta(\alpha,\beta) : \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,dt$.

beta function

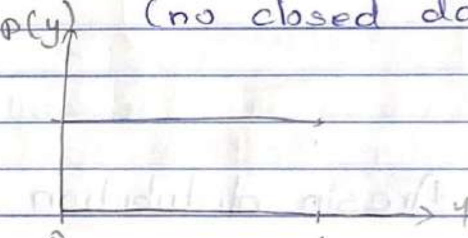
$Y \sim \beta eta(\alpha,\beta) = \frac{1}{\beta(\alpha,\beta)} y^{\alpha-1}(1-y)^{\beta-1}$

$E[Y] = \frac{\alpha}{\alpha+\beta}$ mode$[Y] = \frac{\alpha-1}{\alpha+\beta-2}$ $\alpha,\beta \geq 1$ parameter space

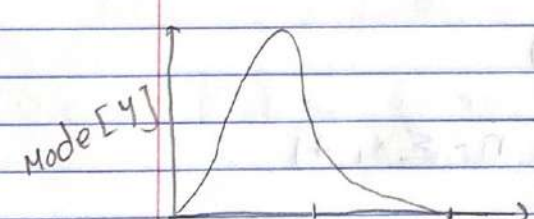
supp$[Y] = (0,1)$
med$[Y] = $ qbeta$(0.5, \alpha, \beta)$ (no closed data)



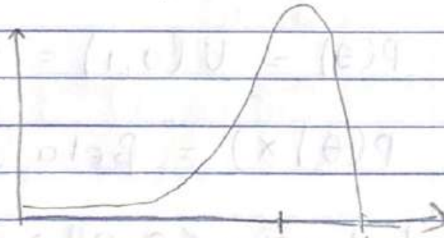
$\alpha = \beta = 100$



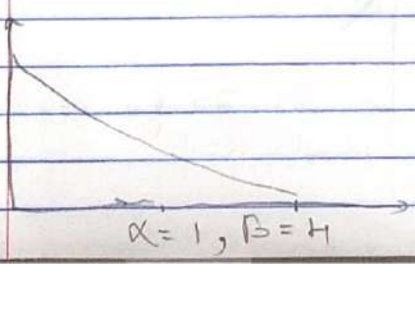
$\alpha = \beta = 1$ U(0,1)

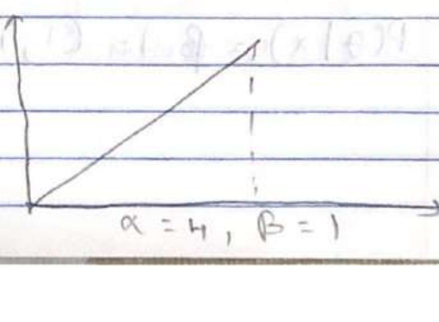


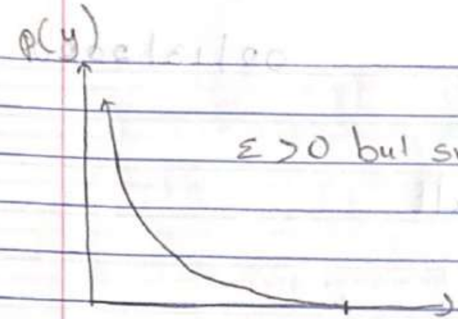
$\alpha = 10,\ \beta = 90$



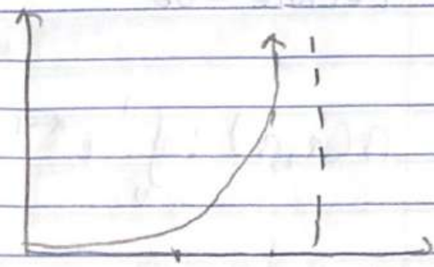
$\alpha = 90,\ \beta = 10$


$\alpha = 1,\ \beta = 4$

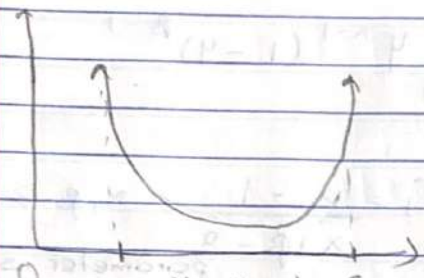

$\alpha = 4,\ \beta = 1$



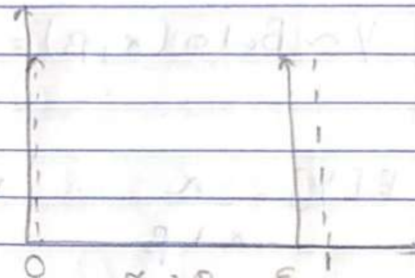
$\alpha = 1-\varepsilon,\ \beta = 3$



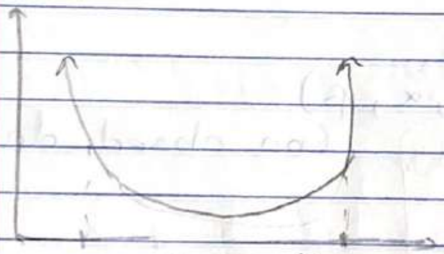
$\alpha = 4,\ \beta = 1-\varepsilon$



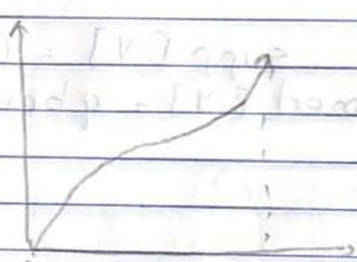
$\alpha = \beta = 1-\varepsilon$



$\alpha = \beta = \varepsilon$



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



Arcsin distribution.

F = iid Bernoulli

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

$\Longrightarrow \quad P(\theta \mid x) = \text{Beta}(\sum x_i + 1,\ n - \sum x_i + 1$

Exam Question.

Let $x = \langle 0, 0, 0 \rangle$

$P(\theta \mid x) = \text{Beta}(1, 4)$

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

$\hat{\theta}_{MMSE} = 0.2, \quad \hat{\theta}_{MAP} = \hat{\theta}_{MLE} = 0$

$\hat{\theta}_{MMAE} \approx 0.159$

$$P(\theta | x_1) = \frac{P(x_1|\theta)\,P(\theta)}{P(x_1)} = \text{Beta}(1,2)$$

$$P(\theta | x_1, x_2) = \frac{P(x_1, x_2|\theta)\,P(\theta|x_1)}{P(x_1, x_2)} = \text{Beta}(1,3)$$

$$p(\theta | x_1, x_2, x_3) = \frac{P(x_1 x_2 x_3|\theta)\,P(\theta|x_1, x_2)}{P(x_1 x_2 x_3)} = \text{Beta}(1,4)$$

T = iid Bernoulli

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

$$P(\theta|x) = \frac{P(x|\theta)\,P(\theta)}{\int_0^1 P(x|\theta)\,P(\theta)\,d\theta} = \frac{\left(\theta^{\Sigma x_i}(1-\theta)^{n-\Sigma x_i}\right)\left(\frac{1}{\beta(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\right)}{\int_0^1 \theta^{\Sigma x_i}(1-\theta)^{n-\Sigma x_i}\frac{1}{\beta(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\,d\theta}$$

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

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

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

$$p(\theta) \xrightarrow{x} p(\theta \mid x)$$

$$\underbrace{Beta(\alpha, \beta)}_{\text{prior parameter}} \xrightarrow{x} Beta(\underbrace{\Sigma x_i + \alpha}_{\alpha'}, \underbrace{n - \Sigma x_i + \beta}_{\beta'})$$

Posterior parameter

$$\hat{\theta}_{MMSE} = E[\theta \mid x] = \frac{\Sigma x_i + \alpha}{n + \alpha + \beta}$$

$$\hat{\theta}_{MMAE} = Beta[\theta \mid x] = qbeta(0.5, \Sigma x_i + \alpha, n - \Sigma x_i + \beta)$$

$$\hat{\theta}_{MAP} = Mode[\theta \mid x] = \frac{\Sigma x_i + \alpha - 1}{n + \alpha + \beta - 2} \quad \text{if } \Sigma x_i + \alpha \geq 1, \; n - \Sigma x_i + \beta \geq 1$$

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

"Conjugacy" for a given likelihood model means the prior and the posterior have the same r.v (different parameter)

Beta is the "conjugate prior" for the iid Bernoulli likelihood.

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

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

Let x now be the same of x_i.

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

$$n - x = n - \Sigma x_i$$

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

$$\Longrightarrow P(\theta \mid x) = \text{Beta}(\underbrace{x}_{\#\text{ of success}} + \alpha,\ \underbrace{n - x}_{\#\text{ of failure}} + \beta)$$

α: # of prior success, # of pseudo success

β: # of prior failure, # of pseudo failure

Pseudo counts.

Principle indifference

$$P(\theta) \doteq U(0,1) = \text{Beta}(1,1) \quad \#\text{ prior observation}$$

$$\underset{\text{success}}{\alpha} \quad \underset{\text{failure}}{\beta} \Longrightarrow n_0 = 2.$$

expectation

$$E(\theta) = \tfrac{1}{2}$$

The principle of indifference is "not so indifference" because in contains information.

default pt. estimate

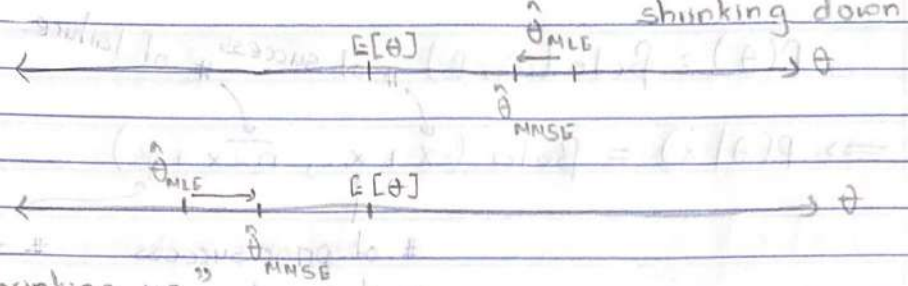
$$\hat{\theta}_{MMSE} = E[\theta \mid x] = \frac{x + \alpha}{n + \alpha + \beta}$$

$$\hat{\theta}_{MMSE} = \frac{x + \alpha}{n + \alpha + \beta} = \frac{x}{n + \alpha + \beta} \cdot \frac{n}{n} + \frac{\alpha}{n + \alpha + \beta} \cdot \frac{\alpha + \beta}{\alpha + \beta}$$

Shrinking estimator

$$= \underbrace{\frac{n}{n + \alpha + \beta}}_{(1-\rho)} \cdot \underbrace{\frac{x}{n}}_{\hat{\theta}_{MLE}} + \underbrace{\frac{\alpha + \beta}{n + \alpha + \beta}}_{\rho} \cdot \underbrace{\frac{\alpha}{\alpha + \beta}}_{\text{prior expectation}}, \qquad \frac{x}{n} = \bar{x}$$

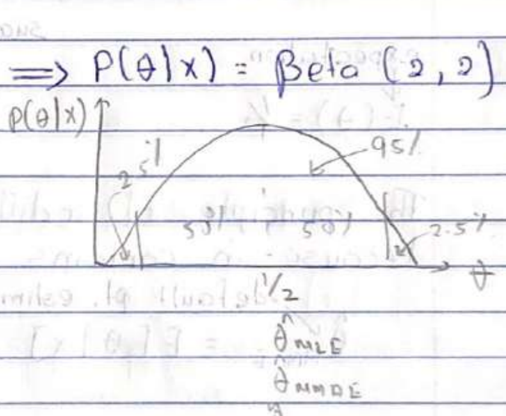
$$\hat{\theta}_{MMSE} = (1 - \rho)\,\hat{\theta}_{MLE} + \rho\, E[\theta] \qquad \underbrace{\sum x_i}_{\vec{w} \cdot \vec{x}} \Rightarrow \text{linear combination}$$



α, β small $\Longrightarrow$ prior uninformative
α, β large $\Longrightarrow$ prior informative
(relatively to n)

F = Binomial
$x = 1,\ n = 2$ $\Longrightarrow P(\theta|x) = \text{Beta}(2,2)$

$P(\theta)$ prior indifference



2^{nd} goal of inference: Confident set

$$CI_{\theta, 95\%} = \left[0.5 \pm 2\sqrt{\frac{0.5(1-0.5)}{2}}\right] = [-0.21, 1.21]$$

i.e. nonsense

Baysian Credible Regions (CR)

$$CR_{\theta, 1-\alpha} = [\text{Quanide}[\theta|x, \alpha/2], \text{Quanide}[\theta|x, 1-\alpha/2]]$$

$\alpha > 0$ $= P(\theta \in CR_{\theta|1-\alpha} \mid x) = 1 - \alpha$