Lecture 6 2-13-20

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

Beta function

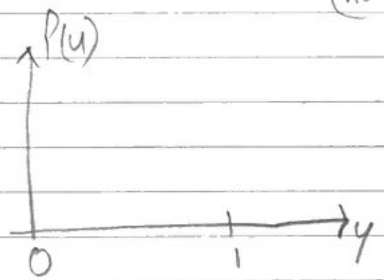
Param Space: $\alpha, \beta > 0$

$\text{Supp}[Y] = (0,1)$

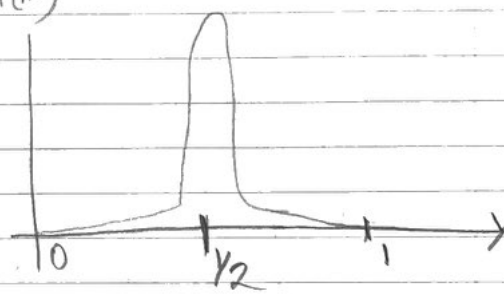
$Y \sim \text{Beta}(\alpha, \beta) = \dfrac{1}{B(\alpha,\beta)} y^{\alpha-1} (1-y)^{\beta-1}$

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

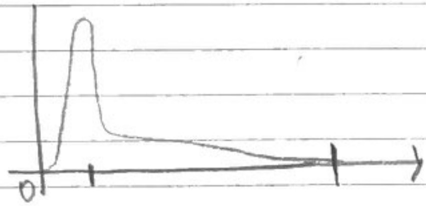
$\text{Med}[Y] = \text{qbeta}(0.5, \alpha, \beta)$ (no closed form)



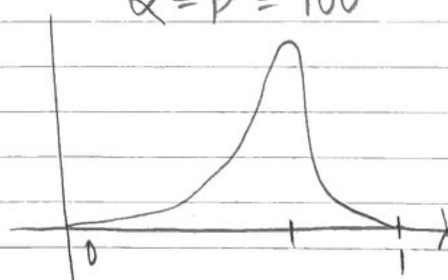
$\alpha = \beta = 1$ U(0,1)



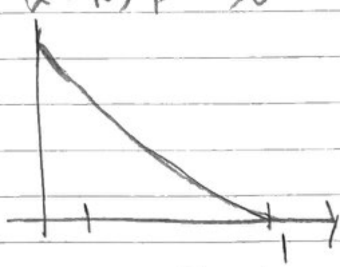
$\alpha = \beta = 100$



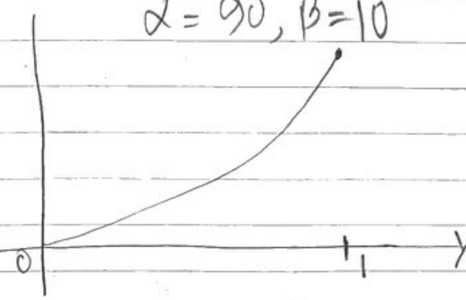
$\alpha = 10, \beta = 90$



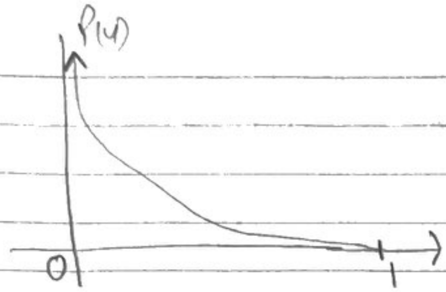
$\alpha = 90, \beta = 10$



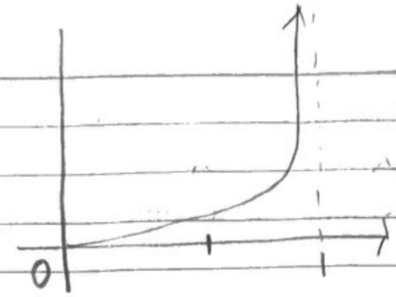
$\alpha = 1, \beta = 4$



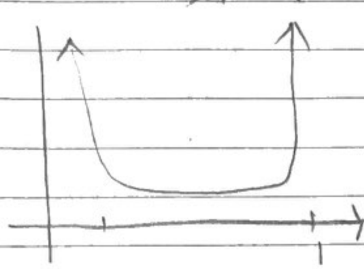
$\alpha = 4, \beta = 1$



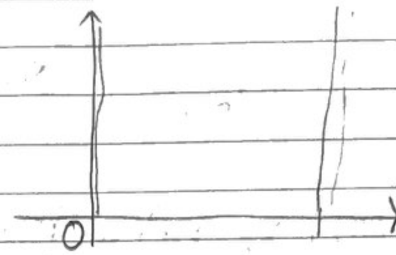
$\alpha = 1-\varepsilon,\ \beta = 4$



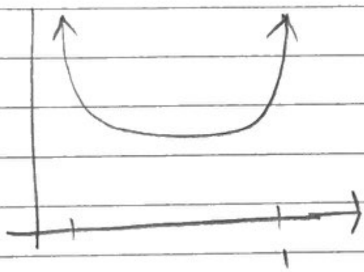
$\alpha = 4,\ \beta = 1-\varepsilon$



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

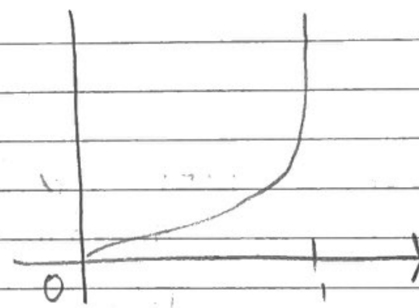


$\alpha = \beta = \varepsilon$

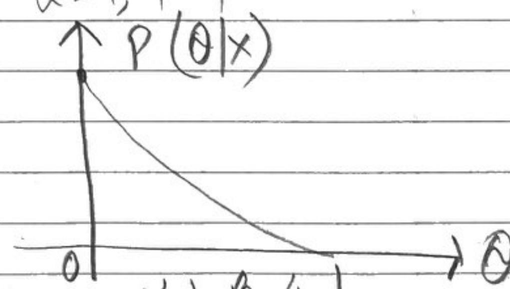


$\alpha = \beta = \frac{1}{2}$

Arcsin dist.



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



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

F = iid Bern

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

$\Rightarrow P(\theta|x) = \text{Beta}(\Sigma x_i + 1,\ n - \Sigma x_i + 1)$

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

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

$\hat{\theta}_{MMSE} = 0.2, \quad \hat{\theta}_{MMAE} \approx 0.159$

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

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

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

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

ⓔ F = iid Bernoulli

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

x is n observations.

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

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

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

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

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

$$\boxed{\begin{array}{l} P(\theta) \xrightarrow{x} P(\theta|x) \\ \underbrace{\text{Beta}(\alpha, \beta)}_{\text{Prior Parameter}} \xrightarrow{x} \text{Beta}(\underbrace{\Sigma x_i + \alpha}_{\alpha'}, \underbrace{n - \Sigma x_i + \beta}_{\beta'}) \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{Posterior Parameters} \end{array}}$$

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

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

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

"Conjugacy" for a given likelihood model means to prior and the posterior have the same R.V. (different parameter).

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

(x) $f = \text{Bin}(n, \theta)$ with n fixed, known and θ unknown.

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

Let $x = \sum x_i$ is the iid Bernoulli Model

$$n - x = n - \sum x_i$$

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

\# prior successes
\# of pseudosuccess

\# prior failure
\# pseudofailure

Pseudocounts

Principle indifference

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

α β $\Rightarrow n_0 = 2$

Prior Observation

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

The principle of indifference is "Not so Indifference". Because it contains information.

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

default pt. est.

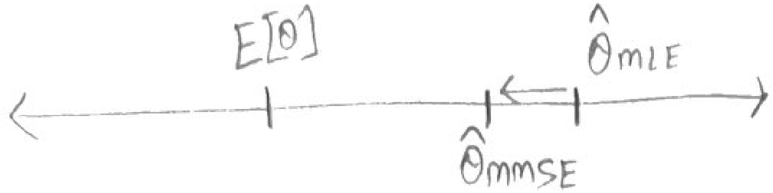
$$= \frac{x}{n+\alpha+\beta} \cdot \frac{n}{n} + \frac{\alpha}{n+\alpha+\beta} \cdot \frac{\alpha+\beta}{\alpha+\beta}$$

$$= \underbrace{\frac{n}{n+\alpha+\beta}}_{1-e} \cdot \underbrace{\left(\frac{x}{n}\right)}_{= \bar{x}} + \underbrace{\frac{\alpha+\beta}{n+\alpha+\beta}}_{e} \cdot \frac{\alpha}{\alpha+\beta}$$

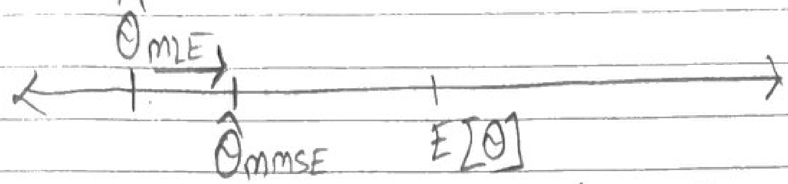
$$\boxed{\hat{\theta}_{MMSE} = (1-e)\,\hat{\theta}_{MLE} + e\,E[\theta]}$$

linear combination

"Shrinking"



"Shrinkage estimator"



α, β small $\Rightarrow$ prior uninformative

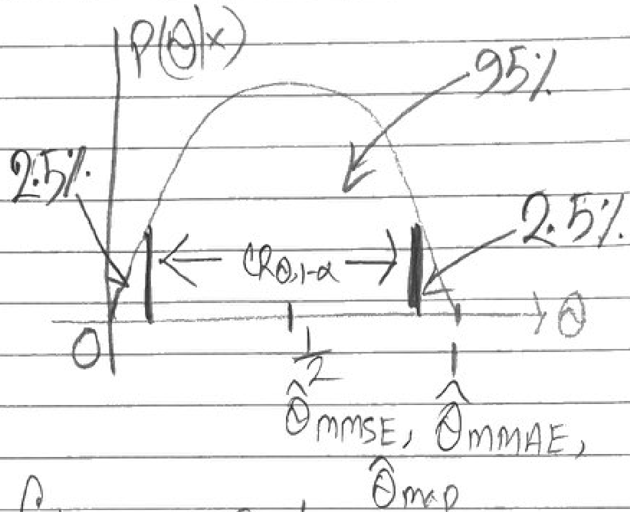
α, β large $\Rightarrow$ prior informative

(ex) F = Binomial

$x = 1, \; n = 2$

$P(\theta)$ prior indifference

$P(\theta|x) = Beta(2,2)$



2^{nd} goal of inference: Confidence 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):

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

$(\alpha > 0)$

$$P(\theta \in CR_{\theta,1-\alpha}|x) = 1-\alpha.$$