Lecture 6:

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

beta function.

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

$$E[Y] = \frac{\alpha}{\alpha+\beta} \qquad \text{Mode}[Y] = \frac{\alpha-1}{\alpha+\beta-2}$$

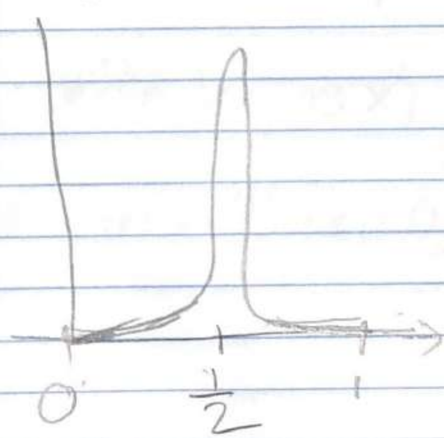
Parameter space $\alpha, \beta \geq 1$

$$\text{Sup}(y) = (0,1)$$

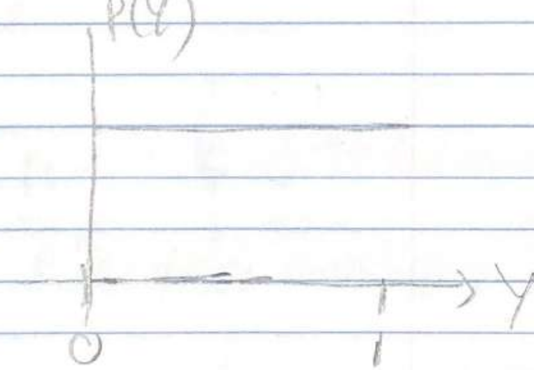
$\text{Med}[Y] = q\,\text{beta}(0.5, \alpha, \beta)$

(no closed data

(2)

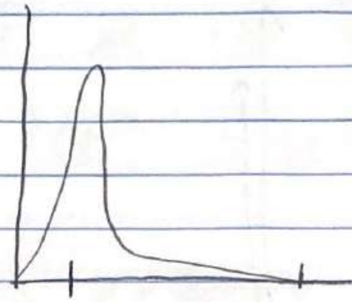


$\alpha = \beta = 100$

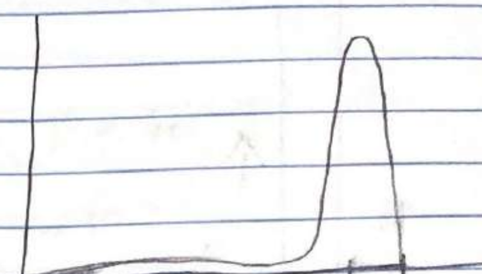


$\alpha = \beta = 1$ 'U(0,1)

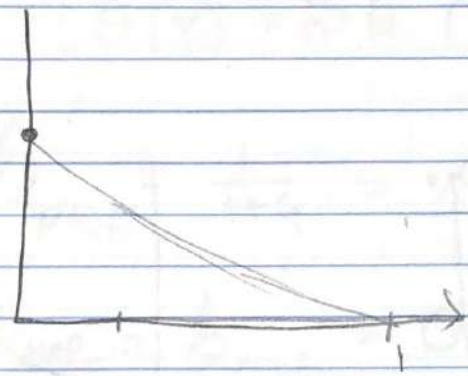
Mode[θ] =



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

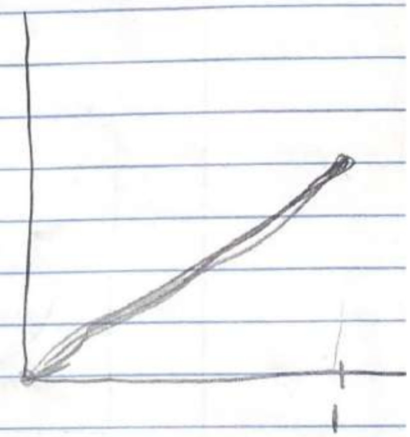


$\alpha = 90,\ \beta = 10$

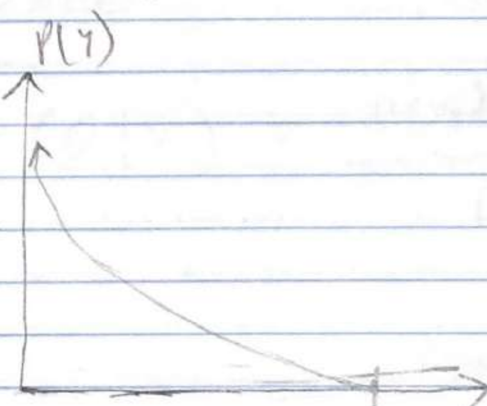


$\alpha = 1,\ \beta = 4$

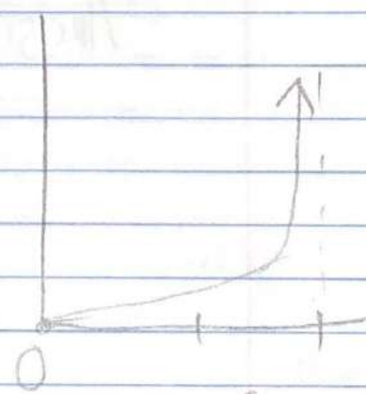
$\varepsilon > 0$ but small



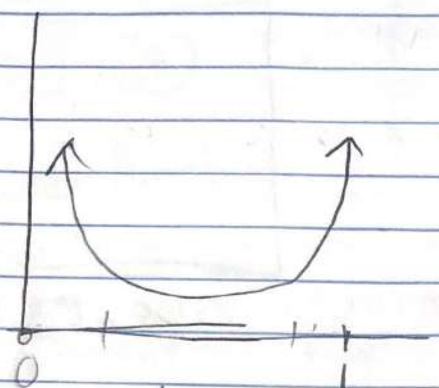
$\alpha = 4,\ \beta = 1$



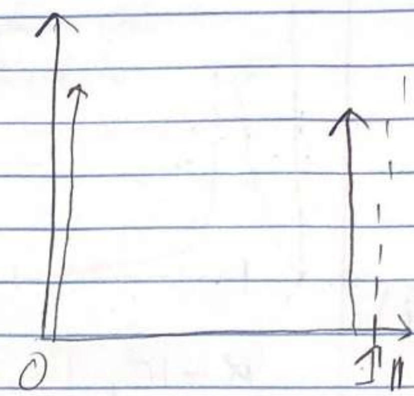
$\alpha = 1-\varepsilon,\ \beta = 4$



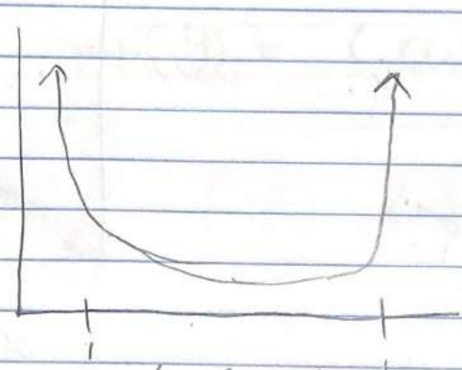
$\alpha = 4,\ \beta = 1-\varepsilon$



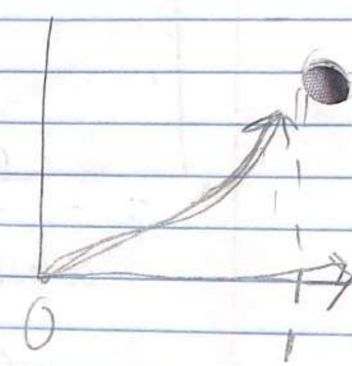
$\alpha = \beta = 1-\varepsilon$



$\alpha = \beta = \varepsilon$



$\alpha = \beta = 1/2$



Arcsin distn

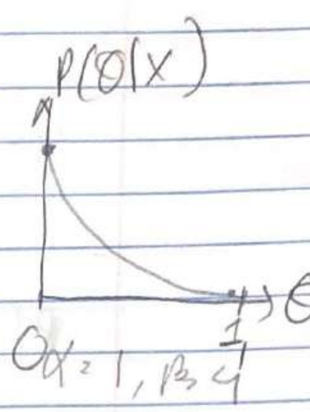
F is iid Bernoulli

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

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

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

$P(\theta|X) = Beta(1, 4)$



$\hat{\theta}_{MMSE} = \frac{1}{1+4} = .2,$

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

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

beta prior

$P(\theta|X_1) = \frac{P(X_1|\theta)P(\theta)}{P(X)}$

$= Beta(1,2) \rightarrow$ ask

Beta posterior

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

Picture

$$P(\theta|x_1,x_2,x_3) = \frac{P(x_1,x_2,x_3|\theta)\,P(\theta|x_1,x_2)}{P(x_1,x_2,x_3)}$$

$= Beta(1,4)$

F = iid Bernoulli

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

x is n observation

$$P(\theta|X) = \frac{P(X|\theta)\,P(\theta)}{\int_0^1 (X|\theta)P(\theta)d\theta} = \cancel{\theta^{\Sigma x}(1-\theta)^{n-\Sigma x_i}}$$

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

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

the "conjugate prior" for the iid Bernoulli likelihood

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

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

take prior $\qquad\qquad$ Posterior.

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

$$Beta(\alpha, \beta) \xrightarrow{X} Beta(\underbrace{\Sigma x_i + \alpha}_{\alpha'}, \underbrace{n - \Sigma x_i + \beta}_{\beta'})$$

Prior parameters $\qquad\qquad$ Posterior parameter

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

posterior expectation

$$\hat{\theta}_{MMSE} = E[\theta | X] = \frac{\Sigma x_i + \alpha}{n + \alpha + \beta} \quad (\text{[illegible]})$$

$$\hat{\theta}_{MMAE} = Med[\theta | X] = qBeta(0.5, \Sigma x_i + \alpha, n - \Sigma x_i + \beta)$$

$\hat{\Theta}_{MAP} = \text{Mode}[\Theta | X]$

$$= \frac{\Sigma x_i + \alpha - 1}{n + \alpha + \beta - 2} \quad \text{if } \Sigma x_i + \alpha \geq 1, \quad n - \Sigma x_i + \beta \geq 1$$

$F = \text{Bin}(n, \Theta)$, with n fixed, known and Θ unknown

$\text{PDF} = \binom{n}{x} \Theta^x (1-\Theta)^{n-x}$

Let X now be the sum of x_i

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

$n - X = n - \Sigma x_i$

$P(\Theta) = \text{Beta}(\alpha, \beta) \Rightarrow P(\Theta | X)$

$$= \text{Beta}(\underbrace{X + \alpha}_{\text{number of success}}, \underbrace{n - X + \beta}_{\text{\# of failure}})$$

$$\beta\left(\underset{\downarrow}{x+\alpha},\ \ n-x+\underset{\Downarrow}{\beta}\right)$$

\# prior failure

\# prior successes

\# of psedu failure

\# of psedu "

together psedu counts

Principle indifference.

$$P(\theta) \overset{\uparrow}{=} U(0,1) = \text{Beta}(\underset{\substack{\uparrow\\ \alpha \\ \downarrow \\ 1 \text{ fake } S}}{1}, \underset{\substack{\uparrow\\ \beta \\ \downarrow \\ 1 \text{ fake } F}}{1})$$

\# Prior observation, $n_0 = 2$

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

I am waiting 50% boy 2.

Principle indifference is "not so indifference" because it contains information.

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

Default point estimate.

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

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

$$= \underbrace{\frac{n}{n+\alpha+\beta}}_{1-e}\overbrace{\frac{X}{n}}^{\hat{\theta}_{MLE}} + \underbrace{\frac{\alpha+\beta}{n+\alpha+\beta}}_{e}\frac{\alpha}{\alpha+\beta}$$

e: row

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

n is large then

linear combination.

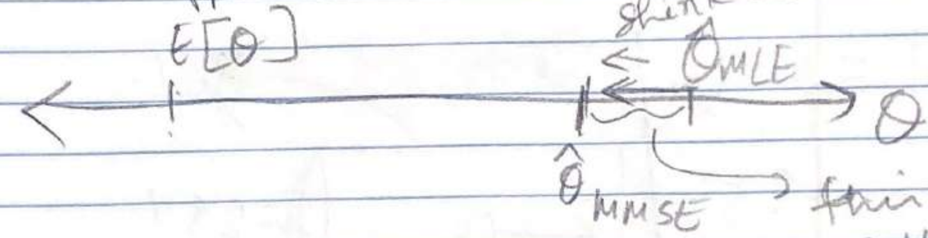
$$\hat{\theta}_{MMSE} = (1-\rho)\hat{\theta}_{MLE} + \rho\, E[\theta]$$

n is small then

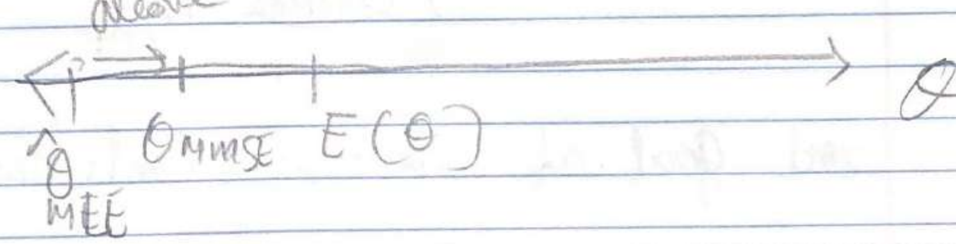
%: posterior   %: prior

Shrinking estimator.

Prior expectation.



this pulling is called shrinking



α, β is small $\Rightarrow$ Prior uninformative

α, β are large $\Rightarrow$ Prior informative.
(relatively to n)

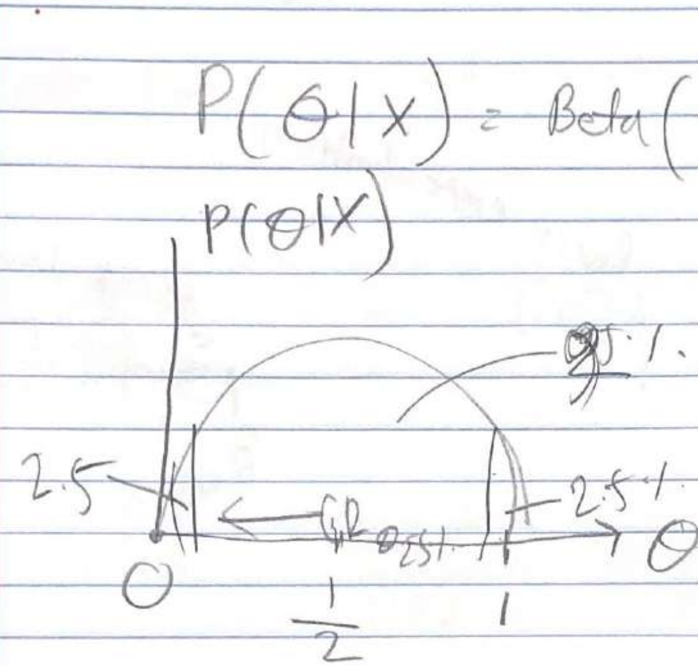
New topic

$\tilde{F}$ = Binomial

$X=1,\ n=2$

$P(\theta)$ Prior indifference. $\alpha, \beta, (1,1)$

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



$\hat{\theta}_{MMSE}$, $\hat{\theta}_{MMAE}$, $\hat{\theta}_{MAP}$

2nd goal of inference: Confidence set;

$$CI_{\theta, 95\%} = \left[0.5 \pm 2\sqrt{\frac{0.5(1-.5)}{2}}\right] = [-0.21, 1.21]$$

i.e. nonsense.

Basian Credible Regions (CR)

$$CR_{\theta,1-\alpha} = \left[\text{Quantile}\left[\theta | X, \alpha/2\right], \; \text{Quantile}\left[\theta | X, 1-\alpha/2\right] \right]$$

$$\alpha > 0$$

$$= P\left(\theta \in CR_{\theta,1-\alpha} \,\middle|\, X\right) = 1-\alpha$$