Lecture 6

2/13 목 베이지안

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

[beta function] ↗

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

· parameter space

$\alpha, \beta > 0$.

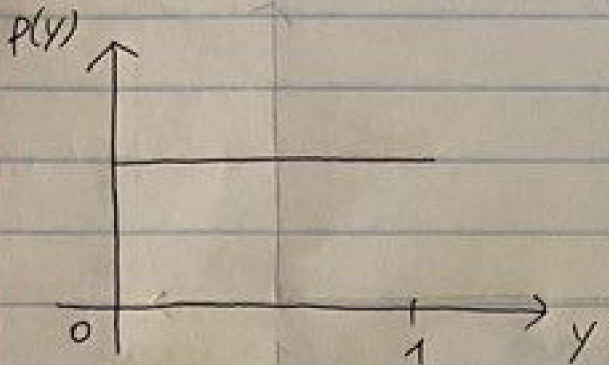
· $supp[Y] = (0, 1)$

$E(Y) = \frac{\alpha}{\alpha+\beta}$

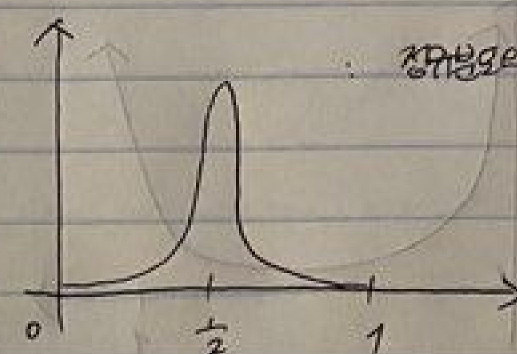
$Mode[Y] = \frac{\alpha-1}{\alpha+\beta-2}$, $\alpha, \beta \geq 1$

$Med[Y] = qbeta(0.5, \alpha, \beta)$

[shape]

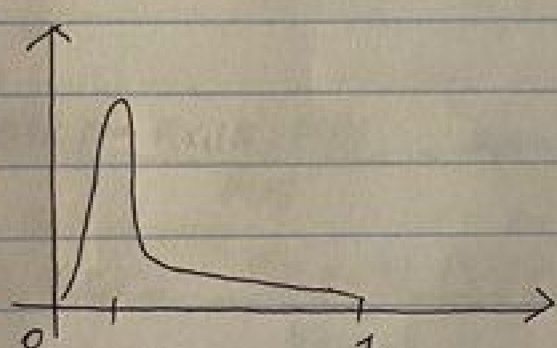


$\alpha = \beta = 1$. $U(0,1)$

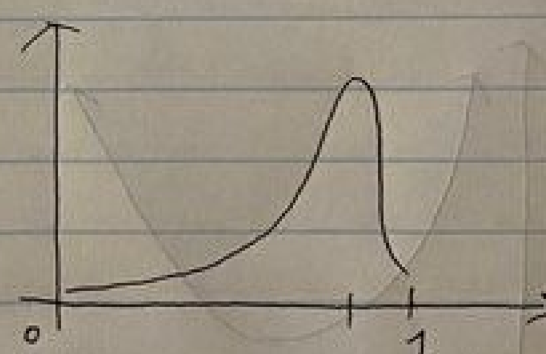


: 정규분포와 비슷한 모양.

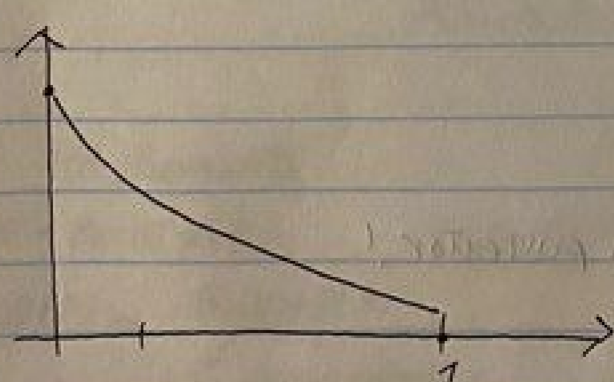
$\alpha = \beta = 100$



$\alpha = 10, \beta = 90$



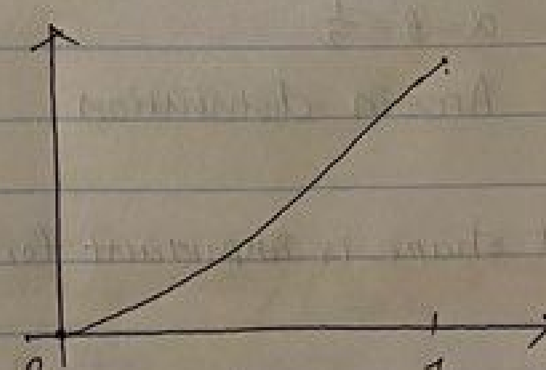
$\alpha = 90, \beta = 10$



$\alpha = 1, \beta = 4$

↓

$y = 0$.

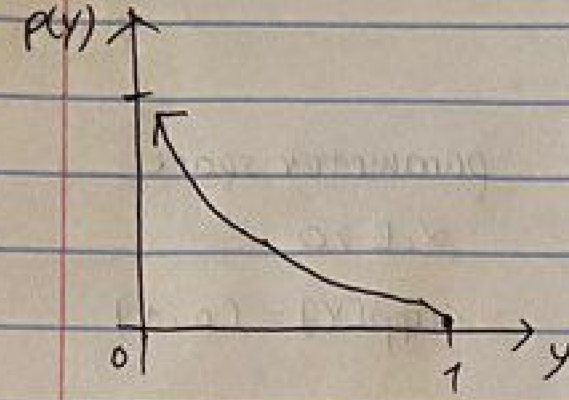


$\alpha = 4$ $\beta = 1$.

↓

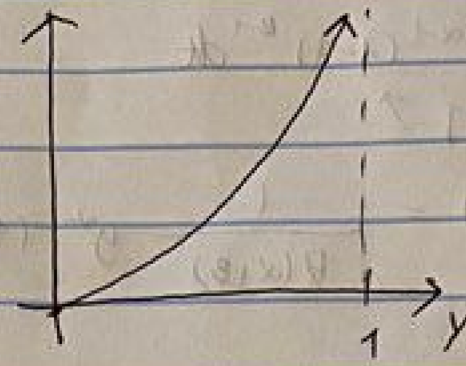
$y = 0$.

$\varepsilon > 0$ but small

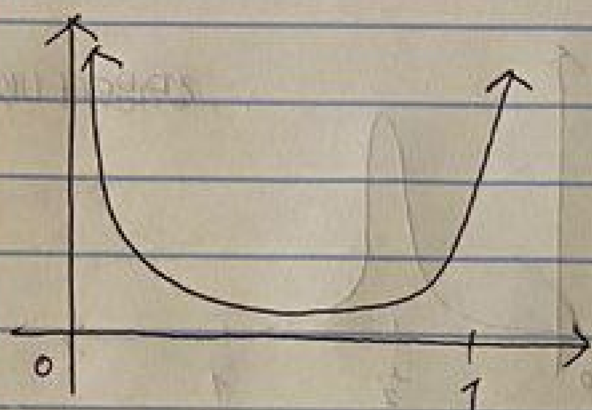


$\alpha = 1-\varepsilon$, $\beta = 4$

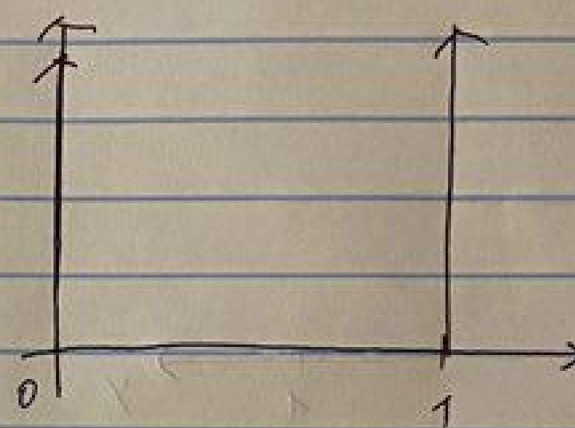
no mode[Y]



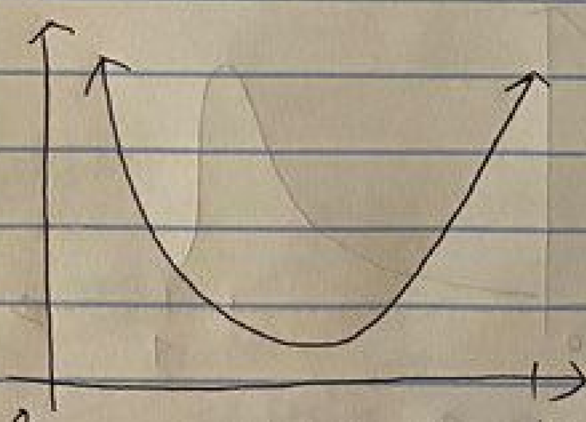
$\alpha = 4$, $\beta = 1-\varepsilon$



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



$\alpha = \beta = \varepsilon$



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

Arcsin distribution

* shape is important for the posterior!

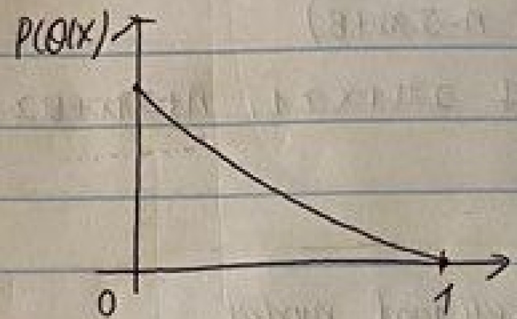
F = iid Bernoulli

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

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

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

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



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

- $\hat{\theta}_{MMSE} = 0.2$, $\hat{\theta}_{MMAE} \approx 0.159$
- $\hat{\theta}_{MAP} = \hat{\theta}_{MLE} = 0$

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

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

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

F = iid Bernoulli

$P(\theta) = Beta(\alpha,\beta)$, X is n observations

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

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

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

★

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

$\underbrace{Beta(\alpha, B)}_{\text{prior parameters}} \xrightarrow{X} Beta(\underbrace{\Sigma x_i + \alpha}_{\alpha'}, \underbrace{n - \Sigma x_i + B}_{B'})$

~~prior~~ parameters

posterior parameters.

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

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

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

"Conjugency" for a ________ likelihood model

the prior and the posterior have the same random variable. (different parameters). Beta is the "conjugate prior" for the iid Bernoulli likelihood.

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

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

pseudo = fake.

Let X Σx_i is the iid Bernoulli model.

$n - X = n - \Sigma x_i$

$P(\theta) = Beta(\alpha, B) \Rightarrow P(\theta|X) = Beta(\underset{\uparrow}{X} + \alpha, \underset{\uparrow}{n - X} + B)$

# of prior successes / # of pseudosuccesses / $\boxed{\text{pseudo count}}$.

# of prior failures / # of pseudofailures.

#of success

# of failures.

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

prior observations.

principle indifference

α B $\Rightarrow n_0 = 2$.

1 fake success / 1 fake failure.

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

The principle of indifference is "not ______ indifference" because it contains information.

↙ default point estimation.

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

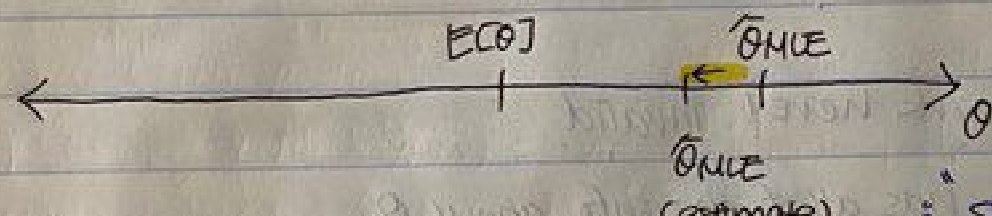
$$= \underbrace{\frac{n}{n+\alpha+B}}_{1-\varrho} \underbrace{\left(\frac{x}{n}\right)}_{=\bar{x}} + \underbrace{\frac{\alpha+B}{n+\alpha+B}}_{\varrho} \cdot \underbrace{\frac{\alpha}{\alpha+B}}_{\text{prior expectation}}$$

$$= (1-\varrho) \cdot \hat{\theta}_{MLE} + \varrho \; E(\theta)$$

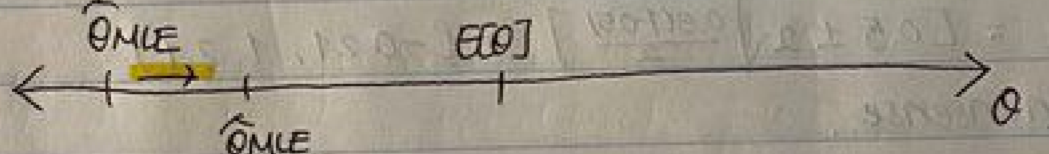
$\Rightarrow$ $\underline{\hat{\theta}_{MMSE}} = (1-\varrho)\,\hat{\theta}_{MLE} + \varrho E[\theta]$: linear combination.

"shrinkage estimation"

% of what believe after + % of what believe before



: "shrinking"



: "shrinking"

α, B small $\Rightarrow$ prior uninformate.

α, B large $\Rightarrow$ prior informate.

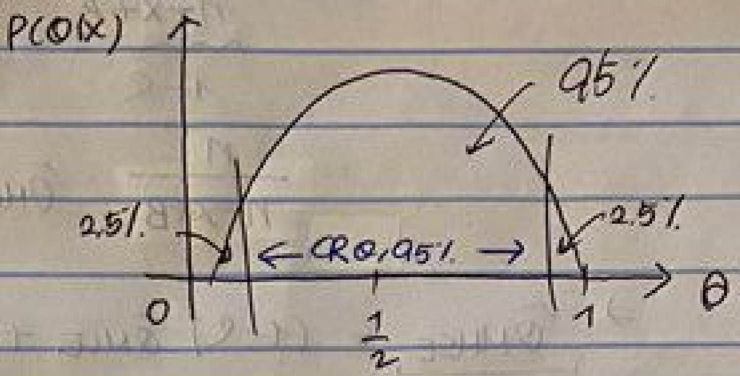
$\mathcal{F}$ = Binomial

$x = 1,\ n = 2$

failure + 1 failure.

1 pseudosuccess + 1 real success

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



$\shortparallel$

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

[2nd Goal of Inference] : confidence set.

↳ doesn't work here! Invalid.

posterior has all of the info. about θ.

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

i.e. nonsense.

[Bayesian Credible Regions] (CR)

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

$$\alpha > 0 \quad = P(\theta \in CR_{\theta,1-\alpha} \mid X) = 1-\alpha.$$