F = Binomial, fixed n

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

$E[\theta] = \frac{\alpha}{\alpha+\beta}$

$var[\theta] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{\underbrace{P(X)}_{\int_\Theta P(X|\theta)P(\theta)}} = \frac{\left(\cancel{\binom{n}{x}}\theta^x(1-\theta)^{n-x}\right)\left(\cancel{\frac{1}{B(\alpha,\beta)}}\theta^{\alpha-1}(1-\theta)^{\beta-1}\right)}{\int_0^1 \cancel{\binom{n}{x}}\theta^x(1-\theta)^{n-x}\cancel{\frac{1}{B(\alpha,\beta)}}\theta^{\alpha-1}(1-\theta)^{\beta-1}\,d\theta}$$

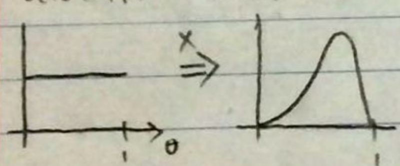
$$= \frac{\theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}}{\int_0^1 \theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}\,d\theta}$$

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

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

$$\underset{Beta(\alpha,\beta)}{\theta} \xrightarrow{x} \underset{Beta(x+\alpha,\, n-x+\beta)}{\theta|X}$$

"the beta is the conjugate prior for the binomial likelihood model"



can be computed explicitly

* (1) $\hat{\theta}_{MMSE} := E[\theta|X] = \frac{x+\alpha}{n+\alpha+\beta}$

(2) $\hat{\theta}_{MAP} := Mode[\theta|X] = \frac{x+\alpha-1}{n+\alpha+\beta-2}$ if $x+\alpha>1$ & $n-x+\beta>1$

(3) $\hat{\theta}_{MAE} := Med[\theta|X]$ → need to use computer

average over all θ

θ unknown

pulling out θ from posterior model & $P(X^*|\theta)$ "believes" the theta

posterior predictive

$$P(X^*|X) = \int_\Theta P(X^*|\theta)P(\theta|X)\,d\theta$$

future data point

$$f(x|z) = \int f(x,y|z)\,dy$$

$$= f(x|y,z)f(y|z)$$

$n^* = 1$

$$P(x^*|x) = \int_0^1 \underbrace{\theta^{x^*}(1-\theta)^{1-x^*}}_{\text{PMF}} \underbrace{\frac{1}{B(x+\alpha, n-x+\beta)} \theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}}_{\text{PDF}} d\theta$$

$$= \frac{1}{B(x+\alpha, n-x+\beta-1)} \int_0^1 \theta^{x^*+x+\alpha-1}(1-\theta)^{-x^*+n-x+\beta} d\theta$$

$$= \frac{B(x^*+x+\alpha, -x^*+n-x+\beta+1)}{B(x+\alpha, n-x+\beta-1)}$$

$$= \frac{\frac{\Gamma(x^*+x+\alpha)\Gamma(-x^*+n-x+\beta+1)}{\Gamma(n+\alpha+\beta+1)}}{\frac{\Gamma(x+\alpha)\Gamma(n-\alpha+\beta)}{\Gamma(n+\alpha+\beta)}}$$

let $x^*=1$

$$\Rightarrow p(x^*=1|x) = \frac{\frac{\Gamma(1+x+\alpha)\cancel{\Gamma(n-x+\beta)}}{\Gamma(1+\alpha+\beta)}}{\frac{\Gamma(x+\alpha)\cancel{\Gamma(n-x+\beta)}}{\Gamma(n+\alpha+\beta)}}$$

$\Gamma(x+1) = x\Gamma(x)$

$$= \frac{\frac{(x+\alpha)\Gamma(x+\alpha)}{(n+\alpha+\beta)\Gamma(n+\alpha+\beta)}}{\frac{\Gamma(x+\alpha)}{\Gamma(n+\alpha+\beta)}}$$

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

$$x^*|x \sim \text{Bern}\left(\frac{x+\alpha}{n+\alpha+\beta}\right)$$

$$\theta \xrightarrow{x} \theta|x$$

$\text{Beta}(\alpha,\beta)$ $\qquad$ $\text{Beta}(x+\alpha, n-x+\beta)$

α: # of prior successes "pseudosuccesses"

β: # of prior failures "pseudofailures"

x: # of successes in the data

$n-x$: # of failure in its data

α & β represents "pseudocounts"

* if α & β is big, then you are assuming a lot

$\theta \sim U(0,1) \Rightarrow \alpha=1, \beta=1$

↳ assuming 1 success & 1 failure

$E[\theta] = \frac{1}{1+1} = 0.5$

In a conjugate model, the prior parameters "usually" can be interpreted as pseudodata. they are as if you've seen data before.

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

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

$\frac{x}{n} = \bar{x} = \hat{\theta}_{MLE}$ $\qquad$ $\frac{\alpha}{\alpha+\beta} = E[\theta]$

$$= \underbrace{\frac{n}{n+\alpha+\beta}}_{1-\rho} \hat{\theta}_{MLE} + \underbrace{\frac{\alpha+\beta}{n+\alpha+\beta}}_{\rho} E[\theta]$$

$\frac{n}{n+\alpha+\beta} + \frac{\alpha+\beta}{n+\alpha+\beta} = 1 \Rightarrow$

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

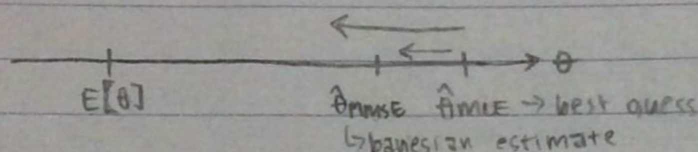
$n\uparrow \Rightarrow \rho\downarrow \Rightarrow \hat{\theta}_{MLE}$ dominates

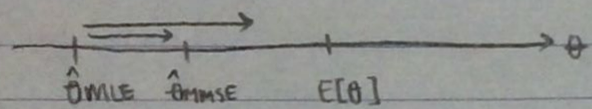
$n\downarrow \Rightarrow \rho\uparrow \Rightarrow E[\theta]$ dominates

$\lim_{n\to\infty} \rho = 0$

2/21/17 - ②

$E[\theta|X]$ is called a "Shrinkage estimator" because it "shrinks" to $E[\theta]$.





$\alpha, \beta \uparrow \Rightarrow \rho \uparrow$

Let $\theta \sim U(0,1) \Rightarrow \alpha = 1, \beta = 1$

$n=2 \Rightarrow \hat{\theta}_{MLE} = 0$

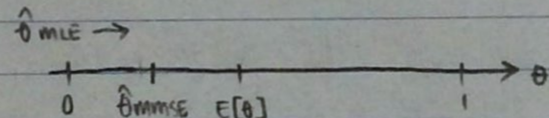
$X=0$

$\rho = 0.5$

$\rho = 0.5$

$E[\theta|X] = \cancel{(1-\rho)\hat{\theta}_{MLE}} + \rho E(\theta)$

$= 0 + \frac{1}{2}(\frac{1}{2}) = \frac{1}{4}$



if α & β are bigger it shrinks harder

$E[\theta|X] = \frac{X+1}{n+2}$ "Wilson Estimate"

↳ when $\alpha, \beta = 1$

$X=1 \Rightarrow \hat{\theta} = \bar{x} = 0.5$

$n=2$

$$CI_{\theta, 1-\alpha} = \left[\hat{\theta} \pm Z_{\frac{\alpha}{2}} SE[\hat{\theta}_{MLE}]\right]$$

$$SE[\hat{\theta}_{MLE}] = \sqrt{\frac{\theta(1-\theta)}{n}} \approx \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}$$

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

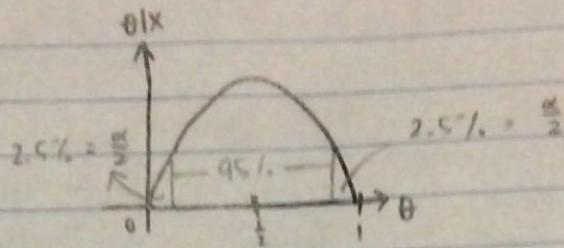
$$= [-.21, 1.21] \approx (0,1)$$

↳ absurd ↳ useless

$x=1,\ n=2$

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

$\theta|x \sim \text{Beta}(\underset{\overset{\shortparallel}{2}}{x+1},\ \underset{\overset{\shortparallel}{2}}{n-x+1})$



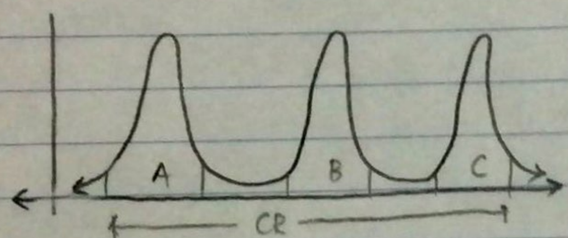
Credible Region (CR) for θ of size $1-\alpha$

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

$$= [\text{Quantile}(\text{Beta}(2,2), 2.5\%), \text{Quantile}(\text{Beta}(2,2), 97.5\%)]$$

incomplete beta $= [\text{qbeta}(.025, 2, 2),\ \text{qbeta}(.975, 2, 2)]$

$= [.094, .906]$



$CR = A \cup B \cup C$

↖ HDR (highest density region)

Disadvantages:

1 not palatable to have non-contiguous regions

2 computationally expensive (when it goes beyond 2 dimension)