$$= \frac{\cancel{\binom{n}{x}}\, \theta^{x}(1-\theta)^{n-x}\, \frac{1}{\cancel{B(\alpha,\beta)}}\, \theta^{\alpha-1}(1-\theta)^{\beta-1}}{\int \cancel{\binom{n}{x}}\, \theta^{x}(1-\theta)^{n-x}\, \frac{\cancel{1}}{B(\alpha,\beta)}\, \theta^{\alpha-1}(1-\theta)^{\beta-1}\, d\theta}$$

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

$$\| $$

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

Beta "Conjugacy"

$\downarrow$ X

Beta Prior and posterior are same family

"The Beta is conjugate prior for the binomial model"

2/2/17

F = Binomial, fixed n

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

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

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{\underbrace{P(X)}_{\int_{\Theta} P(X|\theta)P(\theta)\,d\theta}}$$

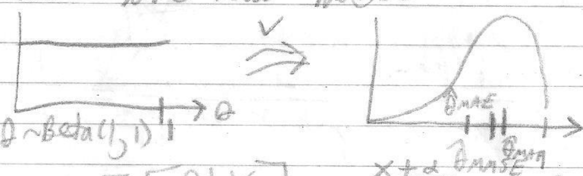
$$= \frac{\left(\binom{n}{x}\theta^{x}(1-\theta)^{n-x}\right)\left(\frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\right)}{\int_0^1 \binom{n}{x}\theta^{x}(1-\theta)^{n-x}\, \frac{1}{B(\alpha,\beta)}\, \theta^{\alpha-1}(1-\theta)^{\beta-1}\, d\theta}$$

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

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

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

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



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

② $\hat{\theta}_{MAP} := Mode[\theta | X]$

③ $\hat{\theta}_{MAE} := Med[\theta | X]$

Test Questions

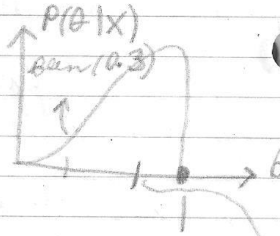
$P(X^* | X)$ ← posterior preditive distribution

↑ future data point

$n^* = 1$

find x^* distribution

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



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

$\longrightarrow f(x | y, z) f(y|z)$

$X^* | X \sim Bern($

Bern(0.8)

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

pmf $\quad$ pdf

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

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

$X^* | X \sim \text{Bern}(\quad)$

Let $x^* = 1$

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

(previous page) $\Rightarrow P(X^* = 1 \mid X) = \dfrac{\dfrac{\Gamma(1+x+\alpha)\,\cancel{\Gamma(n-x+\beta)}}{\Gamma(1+n+\alpha+\beta)}}{\dfrac{\Gamma(x+\alpha)\,\cancel{\Gamma(n-x+\beta)}}{\Gamma(n+\alpha+\beta)}}$

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

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

$W \sim \text{Bern}(\theta)$

$P(W=1) = \theta$

$P(W=0) = 1-\theta$

$$\underset{\text{Beta}(\alpha,\beta)}{\theta} \xrightarrow{x} \theta | x$$

number of prior successes "pseudosuccesses"

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

\# of successes in the data

\# of failures in the data

\# of prior failures "pseudofailures"

$X \sim \text{Binom}(15, 0.3)$

$x = 6$

α, β represent pseudocounts

$\theta \sim U(0,1)$

$\hookrightarrow \text{Beta}(1,1)$

seen 1 success & 1 failure previously

Further, $E(\theta) = \dfrac{1}{1+1} = \dfrac{1}{2} = 0.5$

"In a conjugate model, the prior parameters "usually" can be interpreted in pseudocounts.

$$\hat{\theta}_{MMSE} := E(\theta|X) = \frac{x+\alpha}{n+\alpha+\beta} = \frac{n}{n} \cdot \frac{x}{n+\alpha+\beta} + \frac{\alpha+\beta}{\alpha+\beta}\,\frac{\alpha}{n+\alpha+\beta}$$

$$= \frac{n}{n+\alpha+\beta}\hat{\theta}_{MLE} + \frac{\alpha+\beta}{n+\alpha+\beta} E(\theta)$$

$\bar{X} = \hat{\theta}_{MLE}$

"prior expectation" $E(\theta)$

$$\underbrace{\frac{n}{n+\alpha+\beta}}_{1-\rho} + \underbrace{\frac{\alpha+\beta}{n+\alpha+\beta}}_{\rho} = 1 \qquad \frac{n}{n+\alpha+\beta}\hat{\theta}_{MLE} + \frac{\alpha+\beta}{n+\alpha+\beta}E(\theta)$$

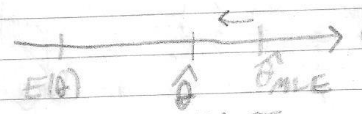
$$= (1-\rho)\hat{\theta}_{MLE} + \rho\, E(\theta)$$

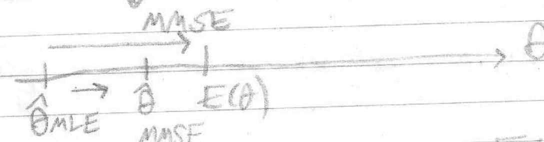
$\lim_{n\to\infty} \rho = 0 \qquad \alpha,\beta\uparrow = \rho\uparrow$

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

$n\downarrow = \rho\uparrow \Rightarrow E(\theta)$ dominates

$E(\theta|X)$ is called a "shrinkage estimator" because it "shrinks" to $E(\theta)$



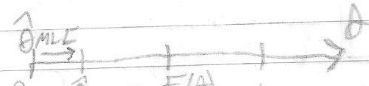


$\theta \sim U(0,1) \Rightarrow \alpha = 1, \beta = 1 \qquad E(\theta) = 0.5$

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

$x = 0$

$\rho = 0.5$



$$E(\theta|X) = (1-\rho)\overset{\to 0}{\hat{\theta}_{MLE}} + \rho E(\theta) = \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4}$$

$E(\theta|X) = \frac{x+1}{n+2}$ Wilson Estimate

$x = 1, n = 2 \Rightarrow \hat{\theta} = \bar{x} = 0.5 \qquad \sqrt{\frac{\theta(1-\theta)}{n}} \approx \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}$

$CI_{\theta, 95\%}$

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

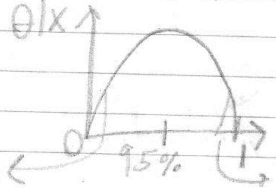
$$CI_{\theta, .95\%} = \left[0.5 \pm 2\sqrt{\frac{0.5(1-0.5)}{2}}\right] = [-.21, 1.21]$$

Bayesian CLT

Jim Berger

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

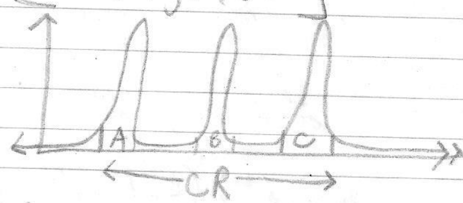
$\theta | X \sim Beta(\underbrace{x+1}_{=2}, \underbrace{n-x+1}_{=2})$



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

* $CR_{\theta, 1-\alpha} := [\,Quantile[\theta|X, \frac{\alpha}{2}],$
$Quantile[\theta|X, 1-\frac{\alpha}{2}]\,]$

$CR_{\theta,1-\alpha} = [Quantile(Beta(2,2)\ 2.5\%),$
$Quantile(Beta(2,2), 97.5\%)]$
$= [qbeta(.025, 2, 2), qbeta(.975, 2, 2)]$
$= [.094, .906]$

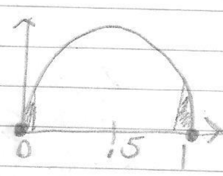


CR = $A \cup B \cup C$
HDR (highest density region)

Disadvantages
① Not adaptable to have non-continuous regions
② Computationally expensive

2/23/17

F = Binomial
$\theta \sim U(0,1), \alpha = 5\%$
$n=2, x=1$
$\theta | X \sim Beta(2,2)$



$= [.094, .906]$