$\mathcal{F}$ iid Bern

Recall problems of frequentist inference. Consider $X = (0,0,0)$

$\hat{\theta}_{MLE} = \bar{x} = \frac{0}{3} = 0$ not a good idea. Saying heads is impossible.

$\theta \sim U(0,1) = \text{Beta}(1,1) \Rightarrow P(\theta|x) = \text{Beta}(\sum x_i + 1, n - \sum x_i + 1)$ $\overset{\text{Beta}(1,4) \text{ using data set}}{=}$

$\hat{\theta}_{MMSE} = E[\theta|X] =$ from last class $\frac{\alpha}{\alpha+B} = \frac{\sum x_i + 1}{n+2} \overset{\text{data}}{=} \frac{0+1}{3+2} = 0.2$

$\hat{\theta}_{MMAE} = \text{Med}[\theta|X] =$ use computer, $\underbrace{\text{qbeta}(0.5, 1, 4)}_{\text{for test.}} = .1591$

$\hat{\theta}_{MAP} = \frac{\alpha - 1}{\alpha + B - 2} = \frac{\sum x_i + 1 - 1}{n+2-2} = \frac{0}{3} = 0$ ← same as $\bar{x}$, expected in this context.

$\overset{\parallel \leftarrow \theta \sim U(0,1) \text{ principle of indifference}}{\hat{\theta}_{MLE}}$

So we have solved a real problem here. You need to believe prior is $U(0,1)$. HW. Problem #2

$P(\theta) = U(0,1) = \text{Beta}(1,1)$, $x_1 = 0$, $x_2 = 0$, $x_3 = 0$ Imagine data coming in increments

$x_1: P(\theta|X_1) = \frac{P(X_1|\theta)P(\theta)}{P(X_1)} = \text{Beta}(1,2)$

$x_2: P(\theta|X_2) = \frac{P(X_2|\theta)P(\theta|X_1)}{P(X_2)} = P(\theta|X_1, X_2) = \text{Beta}(1,3)$

$x_3: P(\theta|X_3) = \frac{P(X_3|\theta)P(\theta|X_1, X_2)}{P(X_3)} = P(\theta|X_1, X_2, X_3) = \text{Beta}(1,4)$

It seems that a beta prior leads a beta posterior for $\mathcal{F}$ iid Bern(θ)

Lets prove this generally.

MATH 341 Lec 07 2/24

X: iid Bern(θ), $p(\theta) = $ Beta(α, β)

$$p(\theta|x) = \frac{P(x|\theta)P(\theta)}{p(x)} = \frac{\theta^{\sum x_i}(1-\theta)^{n-\sum x_i} \cdot \cancel{\frac{1}{B(\alpha,\beta)}}\theta^{\alpha-1}(1-\theta)^{\beta-1}}{\int_0^1 \theta^{\sum x_i}(1-\theta)^{n-\sum x_i} \cdot \cancel{\frac{1}{B(\alpha,\beta)}}\theta^{\alpha-1}(1-\theta)^{\beta-1}\,d\theta}$$

$$p(x) = \int_0^1 p(x|\theta)p(\theta)\,d\theta$$

yields prob x

same as numerator but integrate over θ

re-write this as

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

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

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

$$\underbrace{\text{Beta}(\alpha,\beta)}_{p(\theta)} \xrightarrow{\text{data}} \underbrace{\text{Beta}\left(\alpha+\sum x_i, \beta+n-\sum x_i\right)}_{p(\theta|x)}$$

Conjugacy: the prior and the posterior are the same r.v.
we say that "~~beta~~ the beta is the "conjugate prior" for the "iid bernoulli model"

α, β are parameters of the prior distribution. Thus they are called "hyperparameters" because they are a step removed from parameters, θ, the target of our inference. They are "meta"! who specified their values? You.

MATH 341 Lec 07 2/24

We are now going to show tha E: iid Bern(θ) is the same as F: one realization of a Binomial (n,θ) with n fixed, Recall:

X — X is now the sum of X_i's

$$X_1, \dots, X_n \overset{iid}{\sim} \text{Bern}(\theta) \Rightarrow \overset{X}{\overset{\|}{\sum X_i}} \sim \text{Binomial}(n,\theta) \text{ with n fixed.}$$

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

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

both must be ≥ 1

$$\hat\theta_{MMSE} = \frac{x+\alpha}{n+\alpha+\beta} \qquad \hat\theta_{MMAE} = \text{qbeta}(0.5,\ x+\alpha,\ n-x+\beta)$$

$$\hat\theta_{MAP} = \frac{x+\alpha-1}{n+\alpha+\beta-2}$$

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

\# of pseudo successes — # of pseudo failures

α+β ← pseudo count of trials

$$\text{Beta}(\alpha,\beta) \xrightarrow{\text{data } (x)} \text{Beta}(\alpha + \underbrace{x}_{\text{num of successes}},\ \beta + \underbrace{n-x}_{\text{num of failures}})$$

α = ghost success
β = ghost experiments

Laplace's principle of indifference prior is $\theta \sim U(0,1) = \text{Beta}(1,1)$
which means $\alpha = 1$, $\beta = 1$ which means you are pretending to see
two pseudo trials where 1 is a pseudo success and 1 is a pseudo-failure.
$E[\theta] = 1/2$.

Consider our MMSE Bayesian point estimate: $\hat\theta_{MMSE} = \frac{x+\alpha}{n+\alpha+\beta}$ play a game:

$x = \sum x_i$

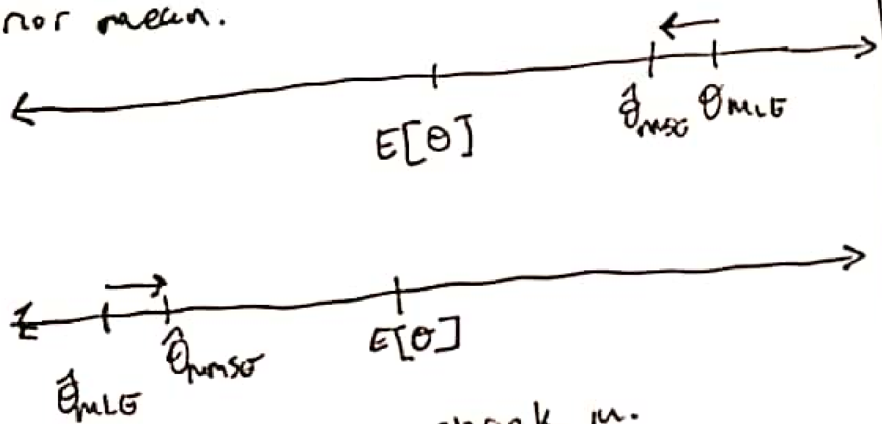
$$= \frac{x}{n+\alpha+\beta}\cdot\frac{n}{n} + \frac{\alpha}{n+\alpha+\beta}\cdot\frac{\alpha+\beta}{\alpha+\beta} = \underbrace{\frac{n}{n+\alpha+\beta}}_{1-\rho}\ \underbrace{\boxed{\frac{x}{n}}}_{\hat\theta_{MLE}} + \underbrace{\frac{\alpha+\beta}{n+\alpha+\beta}}_{\rho}\cdot\underbrace{\frac{\alpha}{\alpha+\beta}}_{E[\theta]}$$

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

Linear combination.

this means that the MMSE in the "beta-binomial conjugate model" is a "shrinkage estimator". It takes the MLE and "shrinks" it towards the prior mean.

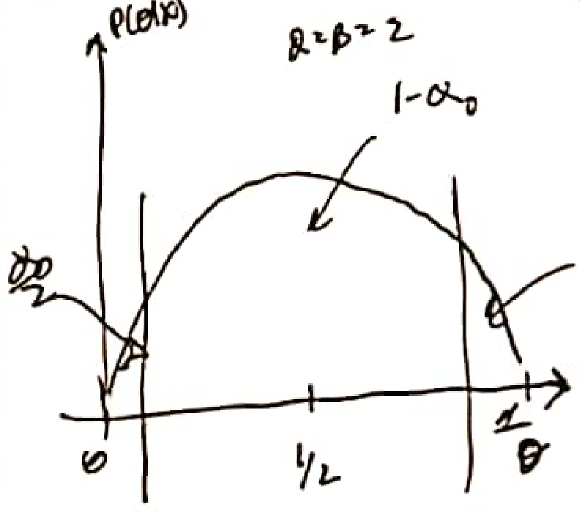
What is $\lim_{n \to \infty} \rho = 0$



the stronger the ρ the stronger you shrink in.
ρ is high when α, β are large and/or n is small.
say $n=3 \to 60\%$ shrink

Thus far we have only discussed point estimation. what about provide a region of reasonable values of theta? confidence sets.

$x=1, n=2, \alpha=\beta=1 \Rightarrow P(\theta|x) = \text{Beta}(2,2)$



Lets say I want a set R s.t. $P(\theta \in R|x) = 1-\alpha_0$ where R represents the ~~model~~ middle of the posterior distribution.

This is called the credible region for θ at level $1-\alpha_0$

$$CR_{\theta, 1-\alpha_0} := \left[\text{Quantile}[\theta|x, \tfrac{\alpha_0}{2}], \text{Quantile}[\theta|x, 1-\tfrac{\alpha_0}{2}]\right]$$

beta-binomial model

$$\left[\text{qbeta}\left(\tfrac{\alpha_0}{2}, \alpha+x, \beta+n-x\right), \text{qbeta}\left(1-\tfrac{\alpha_0}{2}, \alpha+x, \beta+n-x\right)\right]$$