1/2

lec. 7
Feb 24

Consider the dataset $X = <0,0,0>$ $\hat{\theta}_{MLE} = 0$ ← Not a good Idea

$B := iid(\theta)$

$\theta \sim U(0,1) = Beta(1,1) \Rightarrow P(\theta|x) = Beta(\underbrace{\Sigma x_i + 1}_{\text{Alpa parameter}}, \underbrace{n - \Sigma x_i + 1}_{\text{Beta parameter}}) = Beta(1,4)$

$\hat{\theta}_{MMSE} = E[\theta|X] = \frac{\Sigma x_i + 1}{n+2} = \frac{0+1}{3+2} = \frac{1}{5} = .2$

$\hat{\theta}_{MMAE} = Med[\theta|X] = qbeta(0.5, 1, 4) = .1591$

$\hat{\theta}_{MAP} = \frac{\Sigma x_i + 1 - 1}{n + 2 - 2} = \frac{0}{3} = 0$

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

These make sense we've solved a real problem

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

$P(\theta) = U(0,1) = Beta(1,1)$, $x_1 = 0, x_2 = 0, x_3 = 0$

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

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

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

It seems that a beta prior yields a beta posterior for F: iid Bern(θ). Let's prove this generally:

F: iid Bern(θ), $P(\theta) = Beta(\alpha, \beta)$

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

$$P(X) = \int_0^1 P(X|\theta)P(\theta)\,d\theta$$

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

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

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

$P(\theta)$

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

Conjugacy: the prior and the posterior are the same random variable model. We say that "beta" is the "conjuate prior" for the "iid bernoulli likelihood model".

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

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

$$X_1, \ldots, X_n \overset{iid}{\sim} \text{Bern}(\theta) \Rightarrow X := \sum X_i \sim \text{Binom}(n, \theta)$$

with n fixed

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)} = \frac{P(X|\theta)P(\theta)}{\int_0^1 P(X|\theta)P(\theta)\,d\theta} =$$

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

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

Formula

$$\hat{\theta}_{MMSE} = \frac{x+\alpha}{n+\alpha+\beta} \qquad \hat{\theta}_{MAP} = \frac{x+\alpha-1}{n+\alpha+\beta-2}$$

$$\hat{\theta}_{MMAE} = \text{beta}(0.5, x+\alpha, n-x+\beta)$$

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

$$\text{Beta}(\alpha,\beta) \xRightarrow{x} \text{Beta}(\overset{\text{\# of Pseudo successes}}{\alpha} + \underbrace{x}_{\text{\# of Success}}, \overset{\text{\# of Pseudo failure}}{\beta} + \underbrace{n-x}_{\text{\# of failure}})$$

$\alpha+\beta = n_0 \Rightarrow$ # of pseudotrials / pseudocounts

Pseudo success and Pseudofailure are called pseudocount. Laplace's principal of indifference prior is $\theta \sim U(0,1) =$ Beta(1,1) which means $\alpha = 1$ and $\beta = 1$ which means you are pretending to see pseudotrials where 1 is a pseudosuccess and 1 is a pseudofailure.

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

Consider our MMSE Bayesian point estimate: $\hat{\theta}_{MMSE} = \frac{x+\alpha}{n+\alpha+\beta}$

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

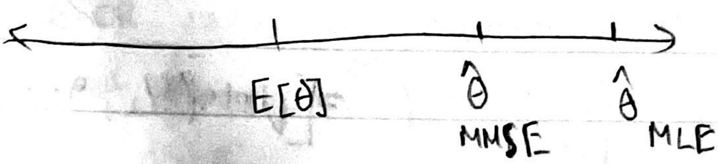
$$= \frac{n}{n+\alpha+\beta} \underbrace{\frac{x}{n}}_{\hat{\theta}_{MLE}} + \underbrace{\frac{\alpha+\beta}{n+\alpha+\beta}}_{\rho} \underbrace{\frac{\alpha}{\alpha+\beta}}_{E[\theta]}$$

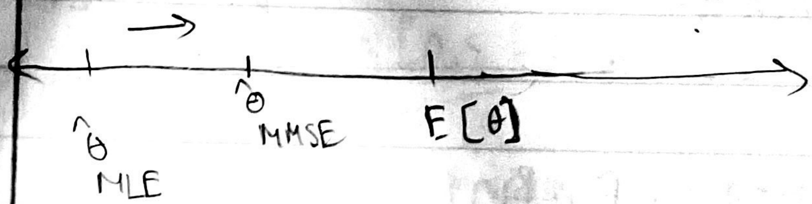
greek letter row ρ

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

linear combination of the MLE and prior mean

This means that MMSE in the "beta-binomial conjuate model" is a "Shrinkage estimator". It take the MLE and it "shrinks" it towards the prior mean.



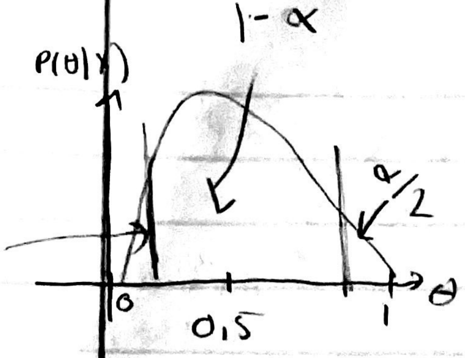


$$\lim_{n\to\infty} \rho = 0$$

When α and β are bigger # and n is small that's when the graph increases.

Thus far, we've only talked about the first goal of inference i.e point estimation. What about the second goal, confidence set (provide a region of reasonable value of θ).

$n,1, n=2, \alpha=\beta=1 \Rightarrow P(\theta|X) = Beta(2,2) \rightarrow$



such that

lets say I want a set R s.t $P(\theta \text{ in } R|X) = 1-\alpha$ where R represents the "middle of the posterior distribution". This is called the "credible region (CR) for θ at level 1-α:

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

$$= [\text{qbeta}(\alpha/2, \alpha+x, \beta+n-x), \text{qbeta}(1-\alpha/2, \alpha+x, \beta+n-x)]$$

beta binomial model