Lec 07 2/24/21

Find Bern(θ)

Consider the dataset $x = \langle 0, 0, 0 \rangle$ $\quad \hat{\theta}_{MLE} = 0$

$$\theta \sim U(0,1) = \text{Beta}(1,1) \Rightarrow P(\theta \mid X) = \text{Beta}\left(\sum x_i + 1, n - \sum x_i + 1\right) = \text{Beta}(1,4)$$

Extraction of the posterior

$$\hat{\theta}_{MMSE} = E[\theta \mid X] = \frac{\sum x_i + 1}{n+2} = \frac{0+1}{3+2} = 0.2$$

$$\hat{\theta}_{MMAE} = \text{Med}[\theta \mid X] = \text{qbeta}(0.5, 1, 4) = 0.159$$

makes sense... so we've solved a real problem.

principle of indifference

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

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

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

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

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

$$x_2: \quad P(\theta | x_2) = \frac{P(x_2|\theta)P(\theta|x_1)}{P(x_2)} = P(\theta|\underbrace{x_1, x_2}) = \text{Beta}(1,3)$$

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

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

$$\mathcal{F}: \text{iid Bern}(\theta), \quad P(\theta) = \text{Beta}(\alpha, \beta)$$

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

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

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

Conjugacy: the prior and the posterior are the same r.v. model. We say that the "beta" is the "conjugate prior" for the "iid Bernoulli likelihood model."

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

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

Recall: $X_1, \ldots, X_n \overset{iid}{\sim} \text{Bern}(\theta) \Rightarrow \sum_{i=1}^{n} X_i \sim \text{Binomial}(n, \theta)$ w/ n fixed.

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

$$\hat{\theta}_{MMAE} = \text{qbeta}(0.5, x+\alpha, n-\alpha+\beta) \Leftarrow \text{Beta}(x+\alpha, n-x+\beta) \Rightarrow \hat{\theta}_{MMSE} = \frac{x+\alpha}{n+\alpha+\beta}$$

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

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

$$\text{Beta}(\alpha, \beta) \overset{X}{\Rightarrow} \text{Beta}\overbrace{(\underbrace{\alpha}_{\text{pseudo successes}} + \underbrace{x}_{\#\text{ successes}}, \underbrace{\beta}_{\text{pseudo failures}} + \underbrace{n - x}_{\text{failures}})}^{\text{pseudo counts}}$$

$$\underbrace{\alpha + \beta}_{\text{pseudo trials}} = n_0$$

Laplace's principle of indifference prior is $\theta \sim U(0,1) = \text{Beta}(1,1)$ which means $\alpha = 1$ and $\beta = 1$ which means you are pretending to see 2 pseudotrials where 1 is a pseudosuccess and 1 is a pseudofailure. $E[\theta] = 0.5$

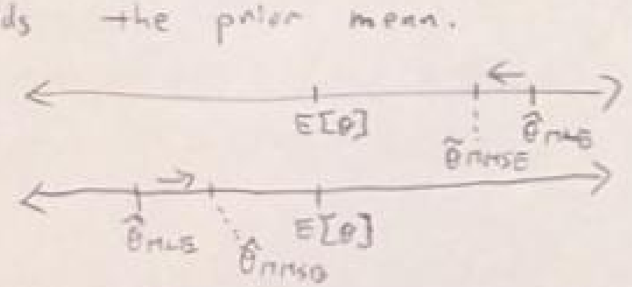
Consider our MMSE Bayesian point estimate: $\hat{\theta}_{MMSE} = \frac{x+\alpha}{n+\alpha+\beta}$

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

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

linear combination of the MLE and prior mean.

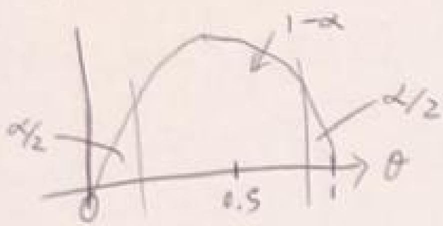
This means that the MMSE in the "beta-binomial conjugate model" is a "shrinkage estimator." It takes the MLE and it "shrinks" it towards the prior mean.



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

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

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



Let's say I wanted a set R such that $P(\theta \text{ in } R|X) = 1-\alpha$, where R represents the "middle of the posterior distribution." This is called the "credible region" (CR) for θ at level $1-\alpha_0$:

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

$$\xrightarrow{\text{beta-binomial model}} = \left[\text{qbeta}(\alpha_0/2,\ \alpha+x,\ \beta+n-x),\ \text{qbeta}\left(1-\frac{\alpha_0}{2},\ \alpha+x,\ \beta+n-x\right)\right]$$