$\mathcal{F}$: iid Bern(θ)

Consider the dataset x = <0,0,0>. $\hat{\theta}_{MLE} = 0$

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

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

$$\hat{\theta}_{MMAE} = \text{Med}[\theta \mid x] = \text{qbeta}(0.5, 1, 4) = .1591\ldots$$

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

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

$\overset{\theta \sim U(0,1)}{=}$ $\hat{\theta}_{MLE}$

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

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

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

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

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

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

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

$$P(x) = \int_0^1 P(x \mid \theta)\, P(\theta)\, d\theta$$

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

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

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

Conjugacy: the prior and the posterior are the same rv 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 F: iid Bern(θ) is the same as F: one realization of a Binomial(n, θ) with n fixed. Recall:

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

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

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

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

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

pseudocounts

pseudosuccesses, # pseudo-failures

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

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

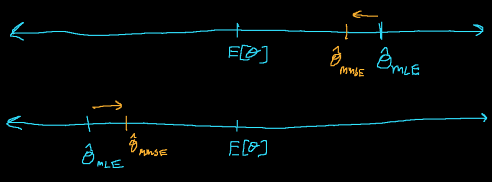
Laplace's principle of indifference prior is θ ~ U(0, 1) = Beta(1, 1) which means α = 1 and β = 1 which means you are pretending to see 2 pseudotrials where 1 is a pseudosuccess and 1 is a pseudo-failure. E[θ] = 0.5.

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

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

linear combination of the MLE and prior mean

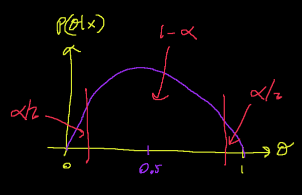
This means that the MMSE in the "beta-binomial conjugate model" is a "shrinkage estimator". It takes the MLE and it "shrinks" it towards the prior mean.



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

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).

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



Let's say I want a set R s.t. P(θ in R | x) = 1 - α_0 where R represents the "middle of the posterior distribution". This is called the "credible region" (CR) for θ at level 1 - α_0:

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

beta-binomial model

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