Lecture 9

The Haldane Prior

Given, the parametric model, $\mathcal{F} : Bin(n, \theta)$ with n fixed, $P(\theta) = Beta(\alpha, \beta) \rightarrow P(\theta|x) = Beta(\alpha + x, \beta + n - x)$ Laplace's prior, $p(\theta) = Beta(1, 1)$ gives two starting pseudo trials, such that you can say $n_0 = 2$ pseudo trials. Laplace's uniform prior was "flat" in an effort to be "objective" i.e. let the data speak for itself and not be "subjective" i.i. allow your personal biases to be part of your inferential conclusion. Laplace sort of failed his own objectivity with this principle of priors because, especially when n is small, applying these prior assumption makes a difference.

The question becomes, can we be more objective? Can we create a prior that has no part in the inferential conclusion. This would mean $n_0 = 0$. This implies, $\alpha = \beta = 0$. Theres a problem with this. The parameter space for the Beta is $\alpha, \beta > 0$, this is not a PDF since its integral over the support diverges. This makes it an *improper prior* since it is not a true random variable. But do we care? Churning through the math, we get the posterior:

$$P(\theta|x) = Beta(x, n - x)$$

This posterior is proper as long as long as $x < n$ and $x \neq 0$ which means you need to have at least one success and at least one failure in your data. If its proper, you have full Bayesian inference including, point estimates, credible region's, p-values. However, you always have $\hat{\theta}_{MMSE}$.

$$\hat{\theta}_{MMSE} = \frac{n}{x} = \hat{\theta}_{MLE}$$

Also, $p = 0$ (no shrinkage). We believe this prior was first introduced by Haldane in 1932 so we'll call it the "Haldane prior". He was an objectivist who believed the data must speak for indifference.

All material up to here will be on midterm I.

The Betabinomial Distribution

Back to probability class, we will introduce mixture/compound distributions.

$$X \sim \{^{\mathbb{N}(0,1^2 up \frac{1}{2}}_{\mathbb{N}(10,2^2) up \frac{1}{2}}$$

This is usually called a Gausian mixture model, where the model (ex. $\mathbb{N}(0, 1^2)$) is complemented by a mixture (ex. $\frac{1}{2}$). We can integrate over the joint density. You take the sum instead of the integral if θ is discrete of course.

$$P(x) = \int_{\Theta} P(x, \vec{\theta}) d\vec{\theta} = \frac{1}{\sqrt{2\pi 1^2}} e^{\frac{1}{2\times 1^2}(x-0)^2}(0.5) + \frac{1}{\sqrt{2\pi 2^2}} e^{\frac{1}{2\times 2^2}(x-10)^2}(0.5)$$

Where this density defines the density of the plot defined above, $P(x)$.

Lets attempt another example,

$$X \sim \{^{Bin(10,0.1 prb \frac{1}{4}}_{Bin(10,0.8) prb \frac{3}{4}}$$

$$P(x) = \binom{10}{x} 0.1^x 0.9^{10-x} \times \frac{1}{4} + \binom{10}{x} 0.8^x 0.2^{10-x} \times \frac{3}{4}$$

Have we seen $P(X)$ before that is the result of a margining making $P(X)$ a mixture/compound distribution? Remember Bayes' Rule,

$$P(\theta|x) = \frac{P(x|\theta)P(\theta}{P(x)}$$

We always computed P(x) as:

$$\int_{\Theta} P(x|\theta)P(\theta)d\theta$$

If we look at the beta binomial parametric model $\mathcal{F} : Bin(n, \theta)$ with n fixed, $P(\theta) = Beta(\alpha, \beta)$

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

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

$$= \frac{\binom{n}{x}}{B(\alpha, \beta)} B(\alpha + x, \beta + n - x)$$

We still have what we were working with before, but now θ is no where to be found. This is its own random variable which we call the BetaBinomial.

$$BetaBinomial(n, \alpha, \beta)$$

Say, $Y \sim BetaBinomial(n, \alpha, \beta)$

$$Supp[Y] = \{0, 1, ..., n\},\ n \in \mathbb{N}, \alpha > 0, \beta > 0$$

$$E[Y] = n(\frac{\alpha}{\alpha+\beta}),\ Var[Y] = n(\frac{\alpha\beta(\alpha+\beta+n)}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Since the beta function is not available in closed form, the PMF/CDF are not available in closed form. To compute, you need a computer. Here is the notation we'll use in this class (the R notation):
'P(Y=y) = dbetabinom(y, n, α, β)' (CDF) 'P(Y $\leq y) = pbetabinom(y, n, \alpha,$ β) (PDF)

$$Let\ \theta = \frac{\alpha}{\alpha+\beta} \rightarrow \theta\alpha + \theta\beta = \alpha \rightarrow (\theta-1)\alpha = -\theta\beta \rightarrow \beta = alpha\frac{1-\theta}{\theta}$$

This proves that $E[X] = n\theta$ is an intuitive formula for the betabinomial expectation since it is the same as the binomial expectation.

Let $\alpha \rightarrow \infty$, but keep $\theta = \frac{\alpha}{\alpha+\beta}$ constant.

$$lim_{\alpha\rightarrow\infty}Var[X] = lim_{\alpha\rightarrow\infty}n(\frac{\alpha(\alpha\frac{1-\theta}{\theta})(\alpha+\alpha\frac{1-\theta}{\theta}+n)}{(\alpha+\alpha\frac{1-\theta}{\theta})^2(\alpha+\alpha\frac{1-\theta}{\theta}+1)}$$

$$= nlim\frac{\theta(1-\theta)}{1} = n\theta(1-\theta)$$

$n\theta(1-\theta)$ is the same variance as the random variable $Bin(n, \theta)$. Furthermore, the variance of X can be written as so,

$$Var[X] = n\frac{\alpha\beta}{(\alpha+\beta)^2}\frac{\alpha+\beta+n}{\alpha+\beta+1} = n\theta(1-\theta)\frac{\alpha+\beta+n}{\alpha+\beta+1}$$

$n\theta(1-\theta)$ is the variance of the binomial, as we just defined.

$\frac{\alpha+\beta+n}{\alpha+\beta+1}$ is the over dispersion $\in (1, n)$. In statistics, over dispersion is the presence of greater variability in a data set than would be expected based on a given statistical model.