BETA-BINOMIAL – BINOMIAL JOINT DISTRIBUTION OF THE SAMPLE ("LIKELIHOOD"), BETA PRIOR

D. Uniform Prior, Binomial Likelihood:

1. Suppose we take a random sample from a Bernoulli distribution with parameter **p.** Our joint distribution of the sample is ("Likelihood" function) is:

$$f_n(\underline{x} \mid p) = \prod_{i=1}^{n} f(x_i \mid p) = p^{X_1}(1-p)^{1-X_1} p^{X_2}(1-p)^{1-X_2} \ldots p^{X_n}(1-p)^{1-X_n}$$

$$= p^y (1 - p)^{n-y} \quad \text{and} \quad y = \sum_{i=1}^{n} X_i$$

2. Now, suppose our prior distribution of **p** is simply Uniform on 0 to 1; that is:

$$\xi(p) = \begin{cases} 1 & 0 < p < 1 \\ 0 & \text{otherwise} \end{cases}$$

3. Hence the joint distribution of the sample ***and*** **p** is

$$h(x_1, x_2, x_3, \ldots, x_n, p) = f_n(x_1, x_2, x_3, \ldots, x_n \mid p)\xi(p)$$

Or simply:

$$h(\underline{x}, p) = f_n(\underline{\mathbf{x}} \mid \theta)\, \xi(\theta) = p^y (1 - p)^{n-y} \quad \text{and} \quad y = \sum_{i=1}^{n} X_i$$

4. The marginal distribution of the sample is:

$$g_n(x_1, x_2, x_3, \ldots, x_n) = \int_p h(x_1, x_2, x_3, \ldots, x_n, p)\, dp =$$

$$\int_p p^y (1 - p)^{n-y} dp = \frac{\Gamma(y+1)\Gamma(n-y+1)}{\Gamma(n+2)}$$

This result is from the form of the Beta distribution is:

$$f(\underline{x} \mid \alpha_1, \alpha_2) = \begin{cases} \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases} \quad \alpha > 0,\ \beta > 0$$

Where α=y+1 and β=n-y+1.

5. So that the posterior distribution is:

$$\xi(\theta|\, x_1, x_2, x_3, \ldots, x_n) = \frac{f_n(x_1, x_2, x_3, \ldots, x_n \mid \theta)\xi(\theta)}{g_n(x_1, x_2, x_3, \ldots, x_n)} =$$

$$p^y(1-p)^{n-y}\frac{\Gamma(n+2)}{\Gamma(y+1)\Gamma(n-y+1)}$$

This is a Beta distribution with parameters:

$$\alpha = y+1 = \sum_{i=1}^{n} X_i + 1 \text{ and } \beta = n-y+1 = n - \sum_{i=1}^{n} X_i + 1$$

6. The expected value and variance of the Beta distribution is:

$$E(X) = \frac{\alpha}{\alpha+\beta} \text{ and VAR(X) = } \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

7. Hence, the Bayesian Estimator for the Mean and Variance is:

$$\hat{p} = \frac{\sum_{i=1}^{n} X_i + 1}{n+2} \text{ and } VAR(\hat{p}) = \frac{\left(\sum_{i=1}^{n} X_i + 1\right)\left(n - \sum_{i=1}^{n} X_i + 1\right)}{(n+2)^2(n+3)}$$

8. And the MLE for the Mean and Variance is:

$$\hat{p}_{mle} = \frac{\sum_{i=1}^{n} X_i}{n} \text{ and } VAR_{mle}(\hat{p}) = \frac{\hat{p}(1-\hat{p})}{n}$$

Note that, as the sample size increases:

$$\hat{p}_{bayes} \rightarrow \hat{p}_{mle}$$

This is also true of the variances. To see this, divide the numerator and denominator by n^2; that is:

$$VAR(\hat{p}) = \frac{\left(\sum_{i=1}^{n} X_i + 1\right)\left(n - \sum_{i=1}^{n} X_i + 1\right)}{(n+2)^2(n+3)} = \frac{\left(\hat{p}_{mle} + \frac{1}{n}\right)\left(1 - \hat{p}_{mle} + \frac{1}{n}\right)}{\left(n + 4 + \frac{4}{n}\right)\left(1 + \frac{3}{n}\right)}$$

So that, as the sample size increases:

$$VAR_{bayes}(\hat{p}) \rightarrow VAR_{mle}(\hat{p})$$

D. Conjugate Priors (Part 1) – Binomial Joint Distribution of the Sample (“Likelihood function”) and Beta Prior Distribution – *Bayesian Computation With R* example of Beta-Binomial

$$f_n(\underline{x} \mid p) = \prod_{i=1}^{n} f(x_i \mid p) = p^{X_1}(1-p)^{1-X_1} p^{X_2}(1-p)^{1-X_2} \ldots p^{X_n}(1-p)^{1-X_n}$$

$$= p^y (1 - p)^{n-y} \qquad \text{and} \qquad y = \sum_{i=1}^{n} X_i$$

$$\xi(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1 - p)^{\beta-1}, \; 0 < p < 1, \; \alpha, \beta > 0$$

Recall that the joint distribution of the sample and **p** is equal to the product of the joint distribution of the sample (“likelihood function”) and the prior distribution of **p**:

$$h(x_1, x_2, \ldots, x_n, p) = f_n(\underline{\mathbf{x}} \mid p)\, \xi(p) =$$

$$p^y (1 - p)^{n-y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1 - p)^{\beta-1} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{y+\alpha-1}(1 - p)^{n-y+\beta-1}$$

To get the marginal distribution of the sample we need to integrate out **p**.

$$g_n(\underline{x}) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(y+\alpha)\Gamma(n-y+\beta)}{\Gamma(n+\alpha+\beta)} *$$

$$\int_0^1 \frac{\Gamma(n+\alpha+\beta)}{\Gamma(y+\alpha)\Gamma(n-y+\beta)} p^{y+\alpha-1}(1 - p)^{n-y+\beta-1} dp = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(y+\alpha)\Gamma(n-y+\beta)}{\Gamma(n+\alpha+\beta)}$$

And the Posterior distribution is:

$$\xi(p\,|\,\underline{x}) = \frac{h_n(\underline{x}, p)}{g_n(\underline{x})} = \frac{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{y+\alpha-1}(1-p)^{n-y+\beta-1}}{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(y+\alpha)\Gamma(n-y+\beta)}{\Gamma(n+\alpha+\beta)}} =$$

$$\frac{\Gamma(n+\alpha+\beta)}{\Gamma(y+\alpha)\Gamma(n-y+\beta)} p^{y+\alpha-1}(1-p)^{n-y+\beta-1}$$

This a Beta distribution with $\alpha^* = y + \alpha$ and $\beta^* = n - y + \beta$, so the posterior is:

$$\xi(p\,|\,\underline{x}) = \frac{\Gamma(\alpha^*+\beta^*)}{\Gamma(\alpha^*)\Gamma(\beta^*)} p^{\alpha^*-1}(1-p)^{\beta^*-1}$$

The mean of the posterior is:

$$E(X) = \hat{p} = \frac{\alpha^*}{\alpha^* + \beta^*} = \frac{y+\alpha}{\alpha+\beta+n}$$

If we take a second sample and use the posterior as our new prior then

$$\xi_2(p) = \xi_1(p\,|\,\underline{x}_1) = \frac{\Gamma(\alpha^*+\beta^*)}{\Gamma(\alpha^*)\Gamma(\beta^*)} p^{\alpha^*-1}(1-p)^{\beta^*-1}$$

and the joint distribution of the sample is (“likelihood function”) for the second sample

is:

$$f_{n_2}(\underline{x}_2\,|\,p) = p^{y_2}(1-p)^{n_2-y_2}$$

where the subscript gives the sample number. The posterior is the Beta distribution

$$\xi_2(p\,|\,\underline{x}_2) = \frac{\Gamma(\tilde{\alpha}+\tilde{\beta})}{\Gamma(\tilde{\alpha})\Gamma(\tilde{\beta})} p^{\tilde{\alpha}-1}(1-p)^{\tilde{\beta}-1}$$

where

$\tilde{\alpha} = y_2 + \alpha^* = y_2 + y_1 + \alpha$ and

$\tilde{\beta} = n_2 - y_2 + \beta^* = n_2 - y_2 + n_1 - y_1 + \beta = n_1 + n_2 - y_1 - y_2 + \beta$

and

$$E(X) = \hat{p} = \frac{\tilde{\alpha}}{\tilde{\alpha} + \tilde{\beta}} = \frac{y_1 + y_2 + \alpha}{\alpha + \beta + n_1 + n_2}$$

As the total sample size gets large this converges to the MLE estimator:

$$E(X) = \frac{\sum_{k=1}^{m} y_k}{\sum_{k=1}^{m} n_k} = \overline{y} = \hat{p}$$