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:

$$f_n(x | p) = \prod_{i=1}^{n} f(x_i | p) = p^{x_i} (1-p)^{1-x_i} = p^X (1-p)^{n-Y}$$

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

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

Or simply:

$$h(x,p) = \frac{f_n(x | \theta) \xi(\theta)}{\Gamma(\alpha) \Gamma(\beta)} = p^X (1-p)^{n-Y}$$

4. The marginal distribution of the sample is:

$$g_n(x_1, x_2, x_3, \ldots, x_n) = \int_{p} h(x_1, x_2, x_3, \ldots, x_n, p) dp = \frac{\Gamma(y+1) \Gamma(n-y+1)}{\Gamma(n+2)}$$

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

$$f(x | \alpha, \beta) = \begin{cases} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \Gamma(\beta)} x^{a-1} (1-x)^{b-1} & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$
Where $\alpha = y+1$ and $\beta = n-y+1$.

5. So that the posterior distribution is:

$$
\zeta(\theta | x_1, x_2, x_3, \ldots, x_n) = \frac{f_n(x_1, x_2, x_3, \ldots, x_n | \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:

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

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

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

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

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

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

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

Note that, as the sample size increases:

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

This is also true of the variances. To see this, divide the numerator and denominator by $n^2$; that is:
\[
VAR(\hat{p}) = \frac{\left(\sum_{i=1}^{n} X_i + 1\right)\left(n - \sum_{i=1}^{n} X_i + 1\right)}{(n+2)^2(n+3)} = \frac{\hat{p}_{\text{mle}} + \frac{1}{n}}{n+4 + \frac{4}{n}} \left[1 - \hat{p}_{\text{mle}} + \frac{1}{n}\right]
\]

So that, as the sample size increases:

\[
VAR_{\text{bayes}}(\hat{p}) \rightarrow VAR_{\text{mle}}(\hat{p})
\]

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

\[
f_n(x | p) = \prod_{i=1}^{n} f(x_i | p) = p^{x_i}(1-p)^{1-x_i} p^{x_i}(1-p)^{1-x_i} \ldots p^{x_i}(1-p)^{1-x_i}
\]

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

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

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 \( p \):

\[
h(x_1, x_2, \ldots, x_n, p) = f_n(x | p) \xi(p) = p^y(1-p)^{n-y} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1} = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{y+\alpha-1}(1-p)^{n-y+\beta-1}
\]

To get the marginal distribution of the sample we need to integrate out \( p \).

\[
g_n(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(y+\alpha)\Gamma(n-y+\beta)}{\Gamma(n+\alpha+\beta)}
\]

\[
\int_{0}^{1} \frac{\Gamma(n+\alpha+\beta)}{\Gamma(y+\alpha)\Gamma(n-y+\beta)} p^{y+\alpha-1}(1-p)^{n-y+\beta-1} dp = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(y+\alpha)\Gamma(n-y+\beta)}{\Gamma(n+\alpha+\beta)}
\]

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

\[
\xi(p \mid x) = \frac{\Gamma(\alpha^* + \beta^*)}{\Gamma(\alpha^*) \Gamma(\beta^*)} p^{\alpha^* - 1} (1 - p)^{\beta^* - 1}
\]

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

\[
\xi_2(p) = \xi_1(p \mid x) = \frac{\Gamma(\alpha^* + \beta^* + 1)}{\Gamma(\alpha^*) \Gamma(\beta^*)} p^{\alpha^* - 1} (1 - p)^{\beta^* - 1}
\]

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

\[
f_{n_2}(x_2 \mid p) = p^{y_2} (1 - p)^{n_2 - y_2}
\]

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

\[
\xi_2(p \mid x_2) = \frac{\Gamma(\tilde{\alpha} + \tilde{\beta})}{\Gamma(\tilde{\alpha}) \Gamma(\tilde{\beta})} p^{\tilde{\alpha} - 1} (1 - p)^{\tilde{\beta} - 1}
\]

where

\[
\tilde{\alpha} = y_2 + \alpha^* = y_2 + y_1 + \alpha \quad \text{and} \quad \tilde{\beta} = n_2 - y_2 + \beta^* = n_2 - y_2 + n_1 - y_1 + \beta = n_1 + n_2 - y_1 - y_2 + \beta
\]

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

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

\[ E(X) = \frac{\sum_{k=1}^{m} y_k}{\sum_{k=1}^{m} n_k} = \bar{y} = \hat{p} \]