SM Exercise 11.5.2

Kristine Villaluna

19/02/2023

SM Exercise 11.5.2:

Consider $X \sim Binom(m, p)$, given p, and assume a $Beta(\alpha, \beta)$ prior for p. As we know, the posterior distribution $\pi(p \mid x)$ is $Beta(\alpha + x, \beta + m - x)$.

a) What is the Bayes estimate of p under squared-error loss? Write this as a shrinkage estimate of the maximum likelihood estimate X/m to the prior mean.

Recall that the Bayes estimate is the mean of the posterior. As we know, the posterior distribution $\pi(p \mid x)$ is $Beta(\alpha + x, \beta + m - x)$.

Thus, we can begin by finding the expectation of $\pi(p)$ like so:

$$
\begin{aligned}
E(\pi(p)) &= \int_0^1 x \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\text{Beta}(\alpha, \beta)} dx \\
&= \frac{1}{\text{Beta}(\alpha, \beta)} \int_0^1 x^{\alpha}(1-x)^{\beta-1} \\
&= \frac{1}{\text{Beta}(\alpha, \beta)} \text{Beta}(\alpha+1, \beta) \quad \text{by integral representation of Beta} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+1+\beta)} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)} \frac{\Gamma(\alpha+1)}{\Gamma(\alpha+1+\beta)} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha+1+\beta)} \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta)(\alpha+\beta)} \frac{\Gamma(\alpha)\alpha}{\Gamma(\alpha)} \\
&= \frac{\alpha}{\alpha+\beta}
\end{aligned}
$$

Thus, we can find the expectation of the posterior to find the Bayes estimate like so:

$$
\begin{aligned}
E(\pi(p|x)) &= \frac{x+\alpha}{x+\alpha+\beta+m-x} \\
&= \frac{x+\alpha}{m+\alpha+\beta} \\
&= w\left(\frac{X}{m}\right) + (1-w)\frac{\alpha}{\alpha+\beta} \\
&= w\hat{p} + (1-w)\frac{\alpha}{\alpha+\beta}
\end{aligned}
$$

Where w is the shrinkage coefficient and $\hat{p} = \frac{X}{m}$ is the maximum likelihood estimate.

Thus, we have expressed the Bayes estimate as a shrinkage estimate of the maximum likelihood estimate X/m to the prior mean. We can see that this is a weighted average of the prior mean and MLE.

b) Show that the marginal distribution of X is

$$f(x \mid \mu, \nu) = \frac{\Gamma(\nu)}{\Gamma(\nu\mu)\Gamma\{\nu(1-\mu)\}} \binom{m}{x} \frac{\Gamma(x+\nu\mu)\Gamma\{m-x+\nu(1-\mu)\}}{\Gamma(m+\nu)}, \quad x = 0, \ldots, m,$$

where $\mu = \alpha/(\alpha+\beta)$ and $\nu = \alpha+\beta$, and deduce that

$$E(X/m) = \mu, \quad \text{var}(X/m) = \frac{\mu(1-\mu)}{m}\left(1 + \frac{m-1}{\nu+1}\right).$$

In order to find the marginal distribution, we must first find the joint density of X,P.

To begin, let us note that we have $X \sim \text{Binomial}(m,p)$:

$$f(x) = \binom{m}{x} p^x (1-p)^{m-x}, \quad x = 0, \ldots, m$$

With a Beta prior on p:

$$\pi(p) = \text{Beta}(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

Now, we can find the joint density as follows:

$$\begin{aligned}
f(x,p) &= f(p)f(x|p) \\
&= \text{Beta}(\alpha,\beta)\binom{m}{x}p^x(1-p)^{m-x} \\
&= \frac{p^{\alpha-1}(1-p)^{\beta-1}}{\text{Beta}(\alpha,\beta)}\binom{m}{x}p^x(1-p)^{m-x} \\
&= \frac{\binom{m}{x}}{\text{Beta}(\alpha,\beta)}p^{\alpha+x-1}(1-p)^{\beta+m-x-1}
\end{aligned}$$

Now, we can use this joint density to find the marginal density by solving the following integral:

$$\begin{aligned}
f(x|\mu,\nu) &= \int_0^1 \frac{\binom{m}{x}}{\text{Beta}(\alpha,\beta)}p^{\alpha+x-1}(1-p)^{\beta+m-x-1}dp \\
&= \frac{\binom{m}{x}}{\text{Beta}(\alpha,\beta)}\int_0^1 p^{\alpha+x-1}(1-p)^{\beta+m-x-1}dp \\
&= \binom{m}{x}\frac{\text{Beta}(\alpha+x,\beta+m-x)}{\text{Beta}(\alpha,\beta)}
\end{aligned}$$

Where we use the integral representation of Beta.

Recall that Beta can take the form:

$$\text{Beta}(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

Thus, we can expand the following like so:

$$
\begin{aligned}
f(x|\mu,\nu) &= \binom{m}{x}\frac{\mathrm{Beta}(\alpha+x,\beta+m-x)}{\mathrm{Beta}(\alpha,\beta)} \\
&= \binom{m}{x}\frac{\Gamma(\alpha+x)\Gamma(\beta+m-x)}{\Gamma(\alpha+\beta+m)}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}
\end{aligned}
$$

Next, as specified in the question, let

$$
\mu = \frac{\alpha}{\alpha+\beta}, \text{ and } \nu = \alpha+\beta
$$

Then,

$$
\begin{aligned}
\alpha &= \mu\nu \\
\beta &= (1-\mu)\nu
\end{aligned}
$$

Finally, we can rewrite the marginal density to get the desired result:

$$
\begin{aligned}
f(x|\mu,\nu) &= \binom{m}{x}\frac{\Gamma(\alpha+x)\Gamma(\beta+m-x)}{\Gamma(\alpha+\beta+m)}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\binom{m}{x}\frac{\Gamma(\alpha+x)\Gamma(\beta+m-x)}{\Gamma(\alpha+\beta+m)} \\
&= \frac{\Gamma(\nu)}{\Gamma(\nu\mu)\Gamma\{\nu(1-\mu)\}}\binom{m}{x}\frac{\Gamma(x+\nu\mu)\Gamma\{m-x+\nu(1-\mu)\}}{\Gamma(m+\nu)}, \quad x = 0,\ldots,m,
\end{aligned}
$$

Thus, we have shown that the marginal distribution of X is equal to the desired result.

Now, we would like to deduce that

$$
E(X/m) = \mu, \quad \mathrm{var}(X/m) = \frac{\mu(1-\mu)}{m}\left(1+\frac{m-1}{\nu+1}\right).
$$

First, let us consider $E(X/m) = \mu$.

Recall that $X \sim \mathrm{Binomial}(m, P)$ where we have P. It then follows that $E(X|P) = mP$.

Now, using the total law of expectation, $E(E(X|Y)) = E(X)$, we have:

$$
\begin{aligned}
E(X/m) &= E\left(E\left(\frac{X}{m}|P\right)\right) \\
&= E\left(\frac{1}{m}mP\right) \\
&= E(P) \\
&= \frac{\alpha}{\alpha+\beta} \\
&= \mu
\end{aligned}
$$

Next, we can deduce $\mathrm{var}(X/m) = \frac{\mu(1-\mu)}{m}\left(1+\frac{m-1}{\nu+1}\right)$ using the total law of variance,

$$
\mathrm{Var}(x) = E(\mathrm{Var}(X|Y)) + \mathrm{Var}(E(X|Y))
$$

First, note that $\mathrm{Var}(X|P) = mP(1-P)$.

Then, we can find the variance of X/m as follows:

$$
\begin{aligned}
\text{Var}(X/m) &= E\left[\text{Var}\left(\frac{X}{m}|P\right)\right] + \left[E\left(\text{Var}\frac{X}{m}|P\right)\right] \\
&= E\left[\frac{1}{m^2}mP(1-P)\right] + \text{Var}\left[\frac{1}{m}mP\right] \\
&= \frac{1}{m}\left[E(P) - \text{Var}(P) - E(P)^2 + m\text{Var}(P)\right] \\
&= \frac{1}{m}\left[\frac{\alpha}{\alpha+\beta} - \frac{\alpha+\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} - \frac{\alpha^2}{(\alpha+\beta)^2} + \frac{m\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}\right] \\
&= \frac{1}{m}\left[\frac{\alpha(\alpha+\beta)(\alpha+\beta+1) - \alpha\beta - \alpha^2(\alpha+\beta+1) + m\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}\right] \\
&= \frac{1}{m}\left[\frac{\alpha^2\beta + \alpha\beta^2 + m\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}\right] \\
&= \frac{1}{m}\left[\frac{\alpha\beta(\alpha+\beta+m)}{(\alpha+\beta)^2(\alpha+\beta+1)}\right] \\
&= \frac{1}{m}\left[\frac{\alpha^2\beta + \alpha\beta^2 + m\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}\right] \\
&= \frac{1}{m}\frac{\alpha}{\alpha+\beta}\frac{\beta}{\alpha+\beta}\left[\frac{\alpha+\beta+m}{\alpha+\beta+1}\right] \\
&= \frac{\frac{\alpha}{\alpha+\beta}\frac{\beta}{\alpha+\beta}}{m}\left[\frac{\alpha+\beta+m}{\alpha+\beta+1}\right] \\
&= \frac{\mu(1-\mu)}{m}\left[\frac{\alpha+\beta+m}{\alpha+\beta+1} + \frac{m-1}{\alpha+\beta+1}\right] \\
&= \frac{\mu(1-\mu)}{m}\left(1 + \frac{m-1}{\nu+1}\right)
\end{aligned}
$$

Thus, we have shown that $\text{var}(X/m) = \frac{\mu(1-\mu)}{m}\left(1 + \frac{m-1}{\nu+1}\right)$ as required.

c) Suppose now that we have n binomial responses $X_1, \ldots, X_n$, with

$$
f(x_i \mid p_i) = \binom{m}{x_i} p_i^{x_i}(1-p_i)^{m-x_i}, \quad x_i = 0, \ldots, m.
$$

Find empirical Bayes estimates of $p_i, i = 1, \ldots, n$, using estimates of μ and ν derived from part (b).

We are now interested in the empirical Bayes estimates using the estimates of μ and ν derived from part (b). It is now empirical Bayes as we plug in the estimates of α and β, based on our data. In other words, we can try to use the n binomial responses to estimate μ and ν. Then, because $\mu = \frac{\alpha}{\alpha+\beta}$ and $\nu = \alpha + \beta$, we can estimate α and β.

Looking at this question in the Statistical Models (SM) textbook, Davison writes that the method of moments estimators based on a random sample $R_1, \ldots, R_n$ all with denominator m are:

$$
\hat{\mu} = \bar{R}, \quad \hat{\nu} = \frac{\hat{\mu}(1-\hat{\mu}) - S^2}{S^2 - \hat{\mu}(1-\hat{\mu})/m}
$$

Where $\hat{R}$ and S^2 are the sample average and variance of the R_j.

The ideal answer would incorporate something like this, where the resulting empirical Bayes estimate is only in terms of the data and m and n.