Chapter 8

An introduction to distributions for the mathematically uninclined

8.1 Drug trials

We suppose that we are testing the efficacy of a certain drug which aims to cure depression, across two groups, each of size 10, with varying levels of the underlying condition: *mild* and *severe*. We suppose that the success rate of the drug varies across each of the groups, with $\theta_{mild} > \theta_{severe}$. We are comparing this with another group of 10 individuals, which has a success rate equal to the mean of the other two groups $\theta_{homogeneous} = \frac{\theta_{mild}+\theta_{severe}}{2}$.

Problem 8.1.1. Calculate the mean number of successful trials in each of the three groups.

Across each of the three groups:

- $\mathbb{E}[X_{mild}] = 10\theta_{mild}$
- $\mathbb{E}[X_{severe}] = 10\theta_{severe}$
- $\mathbb{E}[X_{homogeneous}] = 10\theta_{homogeneous}$

Problem 8.1.2. Compare the mean across the two heterogeneous groups with that of the single group of 10 homogeneous people.

Taking the mean of the means for each group,

$$\mathbb{E}[X_{combined}] = \frac{1}{2} \times 10(\theta_{mild} + \theta_{severe}) \quad (8.1)$$
$$= 10\theta_{homogeneous} \quad (8.2)$$
$$(8.3)$$

In words, the mean outcome across the two groups is the same.

Problem 8.1.3. Calculate the variance of outcomes across each of the three groups.

The variance across each of the three groups is given by:

- $var(X_{mild}) = 10\theta_{mild}(1 - \theta_{mild})$
- $var(X_{severe}) = 10\theta_{severe}(1 - \theta_{severe})$
- $var(X_{homogeneous}) = 10\theta_{homogeneous}(1 - \theta_{homogeneous})$

Problem 8.1.4. How does the variance across both heterogeneous studies compare with that of a homogeneous group of the same sample size and same mean?

Here we need to use the law of total variance. This is because there are two sources of variance: that which is intra-group, and another which is between group,

$$var(X_{combined}) = \mathbb{E}[var(X|D)] + var(\mathbb{E}[X|D]), \tag{8.4}$$

where D means the depressive status of the particular subgroup.

Using this we have:

$$\begin{aligned}
var(X_{combined}) &= \mathbb{E}[var(X|D)] + \mathbb{E}\left(\mathbb{E}[X|D]^2\right) - \left(\mathbb{E}\left(\mathbb{E}[X|D]\right)\right)^2 && (8.5)\\
&= \frac{1}{2} \times 10 \times \theta_{mild}(1 - \theta_{mild}) + \frac{1}{2} 10 \times \theta_{severe}(1 - \theta_{severe}) && (8.6)\\
&+ \frac{1}{2} \times 10^2 \times \theta_{mild}^2 + \frac{1}{2} \times 10^2 \times \theta_{severe}^2 - 10^2 \times \theta_{homogeneous}^2 && (8.7)
\end{aligned}$$

Now supposing that we can write $\theta_{mild} = \theta_{homogeneous} - \epsilon$ and $\theta_{severe} = \theta_{homogeneous} + \epsilon$. We can then substitute this into the above yielding:

$$var(X_{combined}) = n\theta_{homogeneous}(1 - \theta_{homogeneous}) + \epsilon^2 n(n - 1) \tag{8.8}$$

Here $n = 10$, so the variance is greater than that of the homogeneous group. Note, the latter term disappears if $n = 1$ since there is no between-group variance!

Problem 8.1.5. Now consider the extension to a large number of trials, with the depressive status of each group unknown to the experimenter, but follows $\theta \sim beta(\alpha, \beta)$. Calculate the mean value of the beta distribution.

This can be calculated straightforwardly, and found to be $\frac{\alpha}{\alpha+\beta}$.

Problem 8.1.6. Which combinations of α and β would make the mean the same as that of a single study with success probability θ?

Setting these equal:

$$\frac{\alpha}{\alpha+\beta} = \theta \tag{8.9}$$

Rearranging this we get the following relationship:

$$\alpha = \frac{\beta\theta}{1-\theta} \tag{8.10}$$

Problem 8.1.7. How does the variance change, as the parameters of the beta distribution are changed, so as to keep the same mean of θ?

The variance of a beta distribution can be calculated as:

$$var(\theta) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \tag{8.11}$$

This can be shown to be equal to:

$$var(\theta) = \frac{\theta(1-\theta)^2}{\beta+1-\theta} \tag{8.12}$$

Therefore, as $\beta \to \infty \implies var(\theta) \to 0$.

Problem 8.1.8. How does the variance of the number of disease cases compare to that of the a single study with success probability θ?

It is possible to work out the variance of the beta-binomial distribution, and one finds it equal to:

$$var(X|n,\alpha,\beta) = \frac{n\alpha\beta(\alpha+\beta+n)}{(\alpha+\beta)^2(\alpha+\beta+1)} \tag{8.13}$$

By recognising that $\theta = \frac{\alpha}{\alpha+\beta}$, and $1-\theta = \frac{\beta}{\alpha+\beta}$, the above expression can be written as:

$$var(X|n,\alpha,\beta) = n\theta(1-\theta)\frac{\alpha+\beta+n}{\alpha+\beta+1} \tag{8.14}$$

This can then be rearranged to yield:

$$\begin{aligned} var(X|n,\alpha,\beta) &= n\theta(1-\theta)\left[1+\frac{n-1}{\alpha+\beta+1}\right] && (8.15)\\ &= n\theta(1-\theta)+\epsilon && (8.16)\\ &\geq var(X|n,\theta) = n\theta(1-\theta) && (8.17) \end{aligned}$$

Therefore the variance of this distribution exceeds that of an equivalent binomial distribution. Hence, why it is called an over-dispersed distribution.

Problem 8.1.9. Under what conditions does the variance in disease cases tend to that from a binomial distribution?

Substituting in $\alpha = \frac{\beta\theta}{1-\theta}$,

$$var(X|n,\alpha,\beta) = n\theta(1-\theta) + \frac{(n-1)(1-\theta)}{1+\beta-\theta}, \tag{8.18}$$

meaning that as $\beta \rightarrow \infty$, the variance in disease cases tends to that of a binomial distribution.

8.2 Political partying

Suppose that in polls for an upcoming election there are three political parties that individuals for which can vote denoted by $\{A, B, C\}$ respectively.

Problem 8.2.1. If we assume independence amongst those individuals that are polled then what might likelihood might be choose?

A multinomial likelihood with probabilities of voting for each party given by (p_A, p_B, p_C).

Problem 8.2.2. In a sample of 10 individuals we find that the numbers of individuals who intend to vote for each party are $(n_A, n_B, n_C) = (6, 3, 1)$. Derive and calculate the maximum likelihood estimators of the proportions voting for each party.

The likelihood for this case is of the form,

$$L(p_A, p_B, p_C|n_A, n_B, n_C) = \frac{(n_A + n_B + n_C)!}{n_A!n_B!n_C!} p_A^{n_A} p_B^{n_B} p_C^{n_C}, \tag{8.19}$$

which on taking the log becomes,

$$l(p_A, p_B, p_C|n_A, n_B, n_C) = const + n_A \log p_A + n_B \log p_B + (n - n_A - n_B) \log (1 - p_A - p_B). \tag{8.20}$$

Differentiating with respect to p_A and finding the MLE,

$$\frac{\partial l}{\partial p_A} = \frac{n_A}{p_A} - \frac{n - n_A - n_B}{1 - p_A - p_B} = 0. \tag{8.21}$$

Then solving for the MLE we find $\hat{p}_i = \frac{n_i}{n}$ where $i \in \{A, B, C\}$. So in this case we have that $(\hat{p}_A, \hat{p}_B, \hat{p}_C) = (\frac{6}{10}, \frac{3}{10}, \frac{1}{10})$.

Problem 8.2.3. Graph the likelihood in (p_A, p_B) space.

See Figure 8.1.

Problem 8.2.4. If we specify a $Dirichlet(a, b, c)$ prior on the probability vector $\boldsymbol{p} = (p_A, p_B, p_C)$ the posterior distribution for a suitable likelihood is given by a $Dirichlet(a + n_A, b + n_B, c + n_C)$. Assuming a $Dirichlet(1, 1, 1)$ prior, and for the data given find the posterior distribution and graph it in (p_A, p_B) space.

The posterior in this case is given by a $Dirichlet(7, 4, 2)$ distribution (see Figure 8.1).

Problem 8.2.5. How do the posterior means compare with the maximum likelihood estimates?

The posterior means are given by, $\hat{p}_A^{post} = \frac{7}{7+4+2} = \frac{7}{13} < \frac{6}{10} = \hat{p}_A^{MLE}$. This is because the posterior reflects both the likelihood and the prior. The latter has most weight towards equal proportions of each category.

Problem 8.2.6. How does the posterior shape change if we use a $Dirichlet(10, 10, 10)$ prior?

The probability mass shifts over towards the peak of the prior (see middle row panel of Figure 8.1).

Problem 8.2.7. How does the posterior shape change if we use a $Dirichlet(10, 10, 10)$ prior but have data $(n_A, n_B, n_C) = (60, 30, 10)$?

The probability mass shifts over towards the peak of the likelihood (see bottom row panel of Figure 8.1).

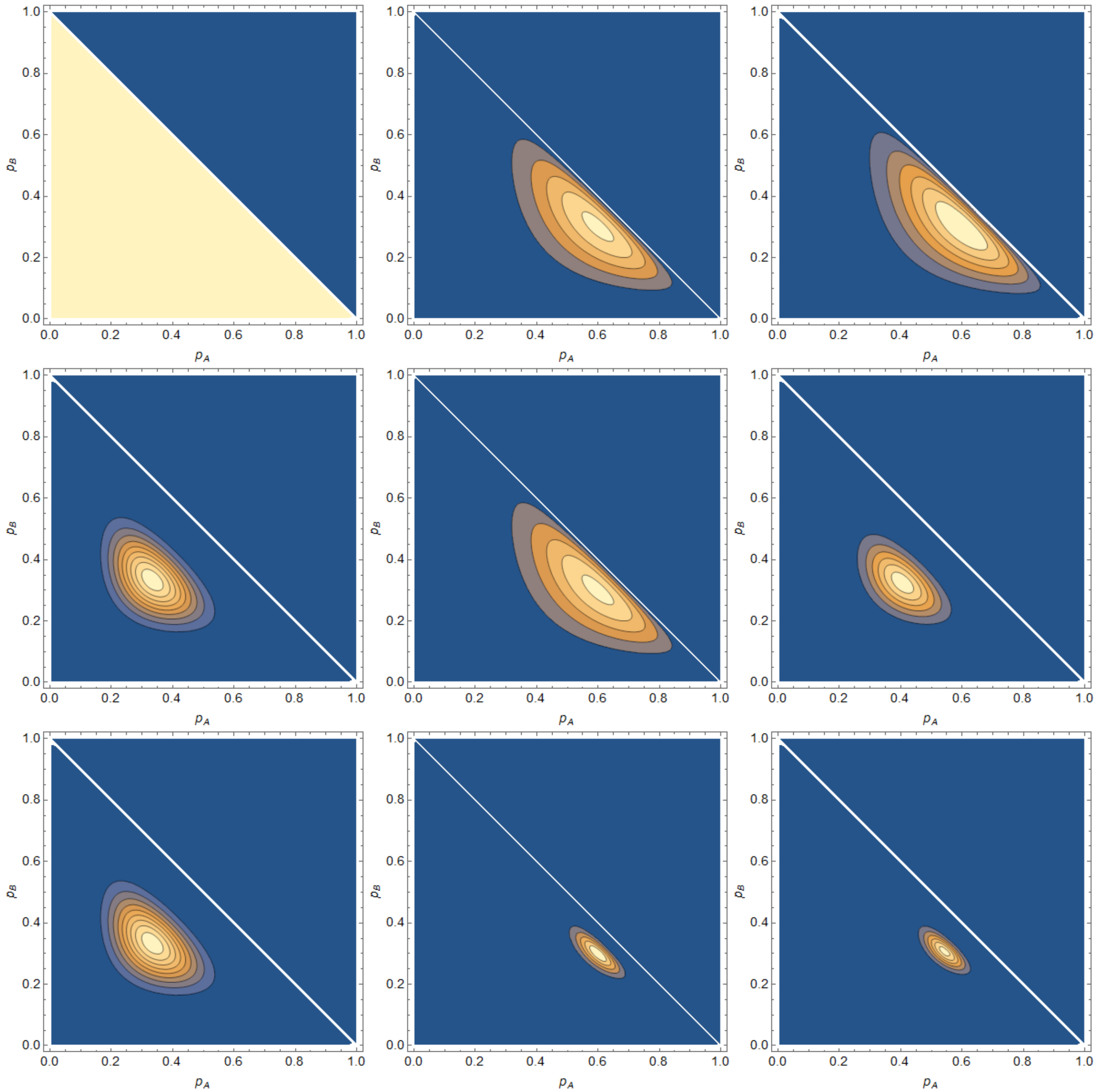

Figure 8.1: The priors (left), likelihoods (middle) and posteriors (right) for: Top, a $Dirichlet(1, 1, 1)$ prior and $(n_A, n_B, n_C) = (6, 3, 1)$; Middle: a $Dirichlet(10, 10, 10)$ prior and $(n_A, n_B, n_C) = (6, 3, 1)$; Bottom: a $Dirichlet(10, 10, 10)$ prior and $(n_A, n_B, n_C) = (60, 30, 10)$.

Bibliography