Solutions Chapter 8 - Distributions

Alexandra Hotti

October 2019

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.

Answer:

$$\mathbb{E}[X_{mild}] = \text{no trials} \cdot \text{success rate} = 10\theta_{mild} \tag{1}$$

$$\mathbb{E}[X_{severe}] = \text{no trials} \cdot \text{success rate} = 10\theta_{severe} \tag{2}$$

$$\mathbb{E}[X_{homogeneous}] = \text{no trials} \cdot \text{success rate} = 10\theta_{homogeneous} = 0.5(10\theta_{mild} + 10\theta_{severe}) = 5(\theta_{mild} + \theta_{severe}) \tag{3}$$

Problem 8.1.2.

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

Answer: So now i simply combine the two first groups and get:

$$\mathbb{E}[X_{comb}] = \text{no trials} \cdot \text{success rate} = 10 \cdot \frac{\theta_{mild} + \theta_{severe}}{2} = \tag{4}$$

$$5 \cdot (\theta_{mild} + \theta_{severe}) = 10\theta_{homogeneous} = \mathbb{E}[X_{homogeneous}] \tag{5}$$

Therefore, our combined group and the homogeneous group have the same expected value of X.

Problem 8.1.3.

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

Answer:

$$var[X_{mild}] = 10\theta_m(1 - \theta_m) \tag{6}$$

$$var[X_s] = 10\theta_s(1 - \theta_s) \tag{7}$$

$$var[X_h] = 10\theta_h(1 - \theta_h) \tag{8}$$

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?

Answer: The law of total variance gives us:

$$var(X_{combo}) = \mathbb{E}[var(X|D)] + var(\mathbb{E}[X|D]) = \mathbb{E}[var(X|D)] + \mathbb{E}(\mathbb{E}[X|D]^2) - (\mathbb{E}(\mathbb{E}[X|D]))^2 =$$
$$= (0.5 \cdot 10\theta_m(1-\theta_m) + 0.5 \cdot 10\theta_s(1-\theta_s)) + \mathbb{E}_C[(10\theta_m)^2 + (10\theta_s)^2] - \mathbb{E}_C[10\theta_m + 10\theta_s]^2 =$$
$$= 5(\theta_m(1-\theta_m) + \theta_s(1-\theta_s)) + 0.5 \cdot ((10\theta_m)^2 + (10\theta_s)^2) - (0.5 \cdot (10\theta_m + 10\theta_s))^2 =$$

$$= 5(\theta_m(1-\theta_m) + \theta_s(1-\theta_s)) + 0.5 \cdot ((10\theta_m)^2 + (10\theta_s)^2) - (10\theta_h)^2$$

Defining the relations $\theta_m = \theta_h - \epsilon$ and $\theta_m = \theta_h + \epsilon$ finally gives us:

$$var(X_c) = 10 \cdot \theta_h(1-\theta_h) + \epsilon \cdot 10(10-1) \tag{9}$$

Thus, even though the homogeneous and the combined populations have the same expectation, the new group is more dispersed.

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.

Answer:

$$\mathbb{E}[\theta] = \frac{\alpha}{\alpha + \beta} \tag{10}$$

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 the expectation equal to θ:

$$\theta = \frac{\alpha}{\alpha + \beta} \tag{11}$$

and then solving for α and β gives us

$$\alpha = \frac{\theta\beta}{1-\theta} \tag{12}$$

$$\beta = \frac{\alpha(1-\theta)}{\theta} \tag{13}$$

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?

Answer:
The variance of a beta distribution is:

$$var[\theta] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \tag{14}$$

To keep the same mean of θ we use the expression for α from eq. (12).

$$var[\theta] = \frac{\theta(1-\theta)^2}{\beta + 1 - \theta} \tag{15}$$

As $\beta \to 0$:

$$var[\theta] \to \theta(1-\theta) \tag{16}$$

As $\beta \to \infty$:

$$var[\theta] \to 0 \tag{17}$$

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?

Answer: The Variance for the number of disease cases X can be described for the combined group by a beta-binomial distribution since the variance of this distribution exceeds that of the binomial distribution (which would be a better fit for the homogeneous groups with lower variance) and since we know that the probability of success θ is drawn from a beta distribution. Thus, we get the following variance for X:

$$Var[X|\alpha, \beta, n] = \frac{n\alpha\beta(\alpha + \beta + n)}{(\alpha + \beta)^2(\alpha + \beta + 1)} \tag{18}$$

By fixing the mean in accordance with (12), we get:

$$Var[X|\alpha, \beta, n] = n\theta(1-\theta)\frac{(\alpha + \beta + n)}{(\alpha + \beta + 1)} = n\theta(1-\theta)(\frac{\alpha + \beta + 1}{\alpha + \beta + 1} + \frac{n-1}{\alpha + \beta + 1}) = \tag{19}$$

$$n\theta(1-\theta)(1 + \frac{n-1}{\alpha + \beta + 1}) = n\theta(1-\theta) + n\theta(1-\theta) \cdot \frac{n-1}{\alpha + \beta + 1} \tag{20}$$

Since $\alpha > 0$, $\beta > 0$ and $n \geq 1$, this expressing will always be greater or equal to: $Var[X|\theta, n] = n\theta(1-\theta)$. Thus, as expected the variance is over-dispersed and greater than that of the equivalent binomial distribution.

Problem 8.1.9.

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

Answer: For large values of α and β.

Political partying

Suppose that in polls for an upcoming election there are three political parties that individuals can vote for, 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?

Answer: Multinomial. Since we have multiple trials (polls), categorical outcomes and outcomes consisting of aggregated number of outcomes for each category.

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.

Answer:

First the proportion of voters for each party is assumed to be: $\{p_A, p_B, p_c\}$. If we define a r.v. Z as the aggregate number of people in our sample that will wote for a certain party: $Z_i = \sum_{j=1}^{n} X_i^j = n_i$ where $i \in \{A, B, O\}$

Thus, the likelihood becomes:

$$Pr(Z_A = n_A, Z_B = n_B, Z_C = z_C | p_A, p_B, p_C) = \frac{(n_A + n_B + n_c)!}{n_A! n_B! n_C!} p_A^{n_A} p_B^{n_B} p_C^{n_C} \tag{21}$$

The log likelihood becomes:

$$log(\mathcal{L}) = log(\frac{(n_A + n_B + n_c)!}{n_A! n_B! n_C!}) + n_A log(p_A) + n_B log(p_B) + n_C log(p_C) \tag{22}$$

We can reduce the dimensionality of the log-likelihood in eq.(24) by using the relation: $p_c = 1 - p_A - p_C$ and $n_C = n - n_A - n_B$

$$log(\mathcal{L}) = log(\frac{(n_A + n_B + n_c)!}{n_A! n_B! n_C!}) + n_A log(p_A) + n_B log(p_B) + (n - n_A - n_B) log(1 - p_A - p_C) \tag{23}$$

Differentiating wrt p_A and solving for the extreme point gives us:

$$\frac{\partial log(\mathcal{L})}{\partial p_A} = \frac{n_A}{p_A} - \frac{n - n_A - n_B}{1 - p_A - p_C} = 0 \tag{24}$$

Now we can find the maximum using lagrange multipliers, or by simply looking up the well known MLE for the multinomial distribution:

$$\hat{p_i} = \frac{n_i}{n} \tag{25}$$

So, for our data we get:

$$p_A = \frac{6}{10}, p_B = \frac{3}{10}, p_C = \frac{1}{10} \tag{26}$$

Problem 8.2.3.

Graph the likelihood in (p_A, p_B) space.

Answer:

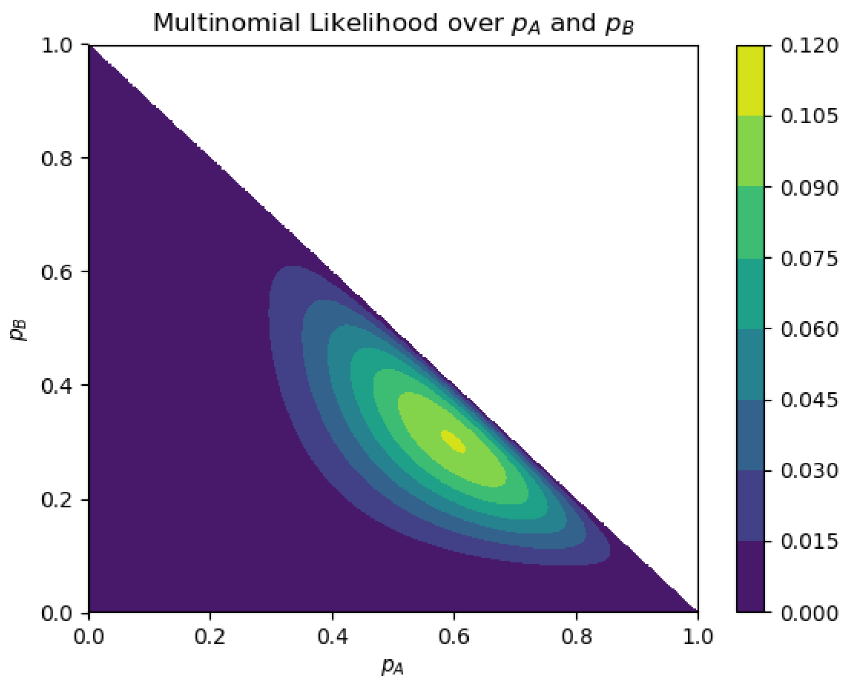

Figure 1: Multinomial likelihood over p_A and p_B.

Problem 8.2.4.

If we specify a Dirichlet(a, b, c) prior on the probability vector 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.

Answer: The pdf for the Dirichlet prior is given by:

$$p(\theta|\bar{\alpha} = \{1, 1, 1\}) = \frac{1}{B(\bar{\alpha})} \prod_{i=1}^{3} \theta_i^{\alpha_i - 1} \tag{27}$$

where B is the beta function defined as:

$$B(\bar{\alpha}) = \frac{\prod_{i=1}^{k} \Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^{k} \alpha_i)} \tag{28}$$

where $\Gamma(n) = (n-1)!$.

For $Dirichlet(1,1,1)$ I get:

$$B(\bar{\alpha}) = \frac{1}{2} \tag{29}$$

$$p(\theta|\bar{\alpha} = \{1,1,1\}) = 2\prod_{i=1}^{3}\theta_i^{\alpha_i - 1} = 2 \cdot (\theta_1^0\theta_2^0\theta_3^0) = 2 \tag{30}$$

The prior for $Dirichlet(1,1,1)$ over p_A and p_B:

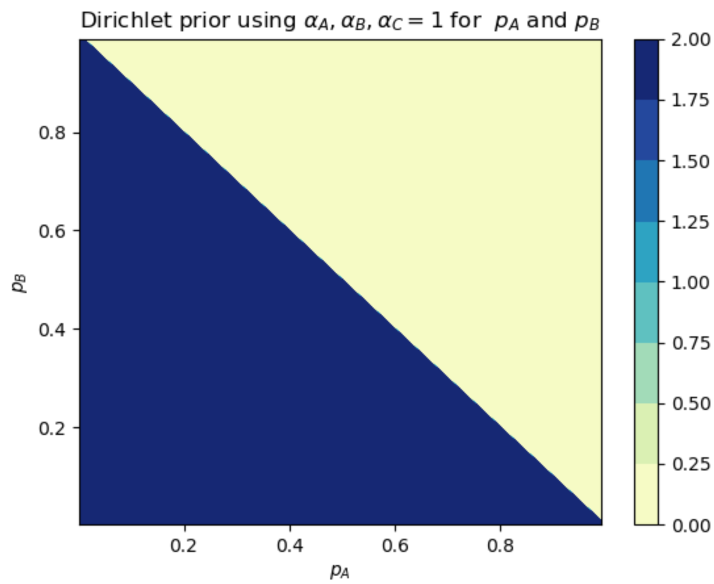

Figure 2: Dirichlet(1, 1, 1) prior over p_A and p_B.

The posterior for the Multinomial Likelihood and the conjugate prior $Dirichlet(1,1,1)$ over p_A and p_B:

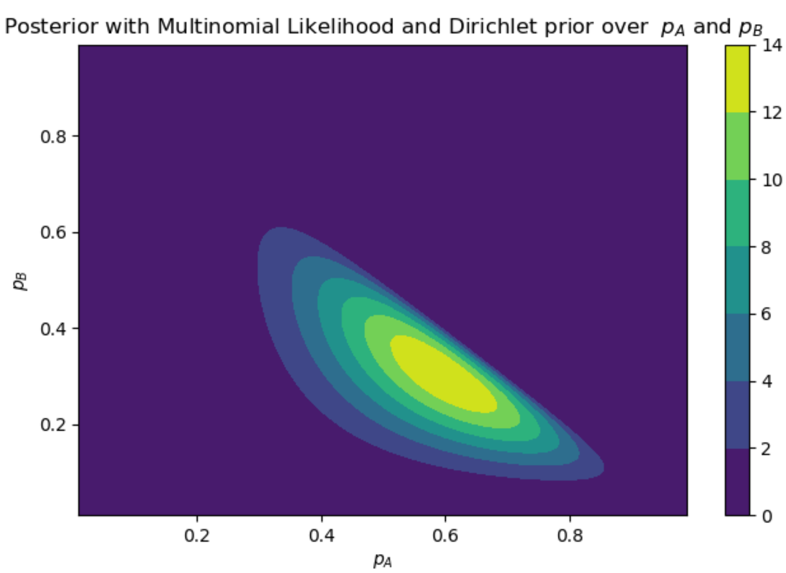

Figure 3: Dirichlet(6+1, 3+1, 1+1) posterior over p_A and p_B.

Problem 8.2.5.

How do the posterior means compare with the maximum likelihood estimates?

Answer: The posterior means compared to the MLE estimates:

$$\mathbb{E}[p_A] = \frac{7}{13} < \frac{6}{10} \tag{31}$$

$$\mathbb{E}[p_B] = \frac{4}{13} > \frac{3}{10} \tag{32}$$

$$\mathbb{E}[p_c] = \frac{2}{13} > \frac{1}{10} \tag{33}$$

So, we can see that the value for p_A has decreased while p_B and p_C have increased. Since we have incorporated a uniform prior on the likelihood the posterior is shifting towards a equal probability weighting between our parameters.

Problem 8.2.6.

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

Answer: Here we are still giving a prior with equal weight given to p_A, p_B and p_C. However, it is no longer uniform and we are using higher weights, meaning that our pre-data belief is more assertive. Therefore, the posterior gets a stronger peak and smaller variance. Also, we see that the posterior peak is located inbetween the peaks of the prior and the likelihood.

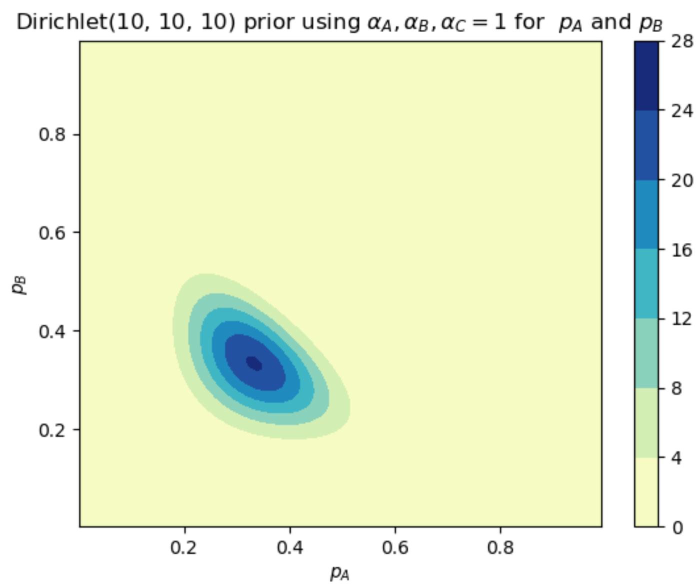

Figure 4: Dirichlet(10, 10, 10) prior over p_A and p_B.

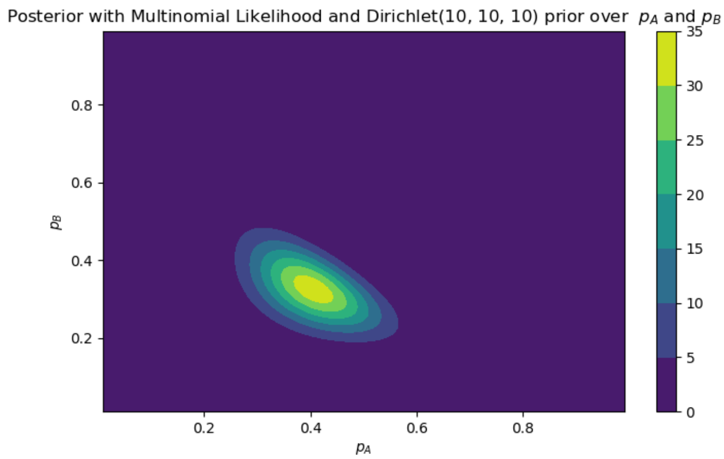

Figure 5: Dirichlet(10+3, 10+6, 10+1) posterior over p_A and p_B.

Problem 8.2.7.

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

Since we increase the amount of data the likelihoods variance around its peak becomes smaller and the uncertainty decreases. See Figure 7.

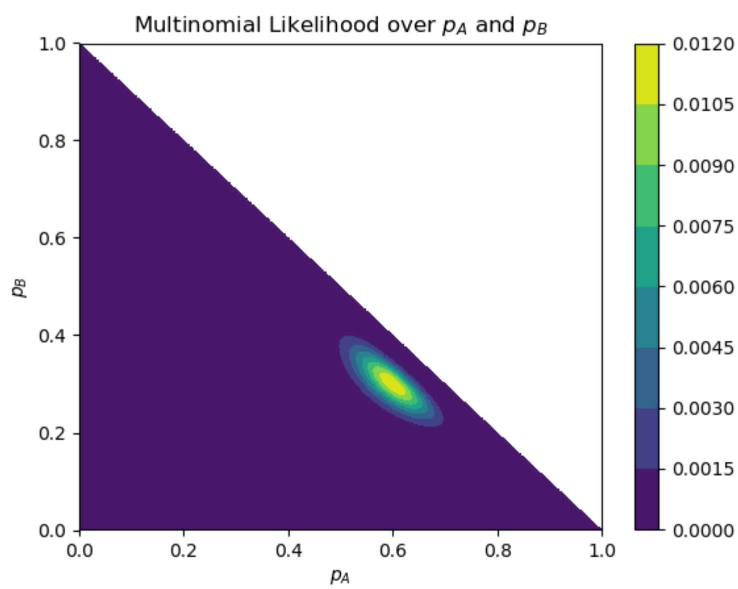

Figure 6: Likelihood over p_A and p_B when the amount of data is increased.

Since the likelihood becomes less uncertain, the posterior variance decreases and shifts towards the likelihood peak.

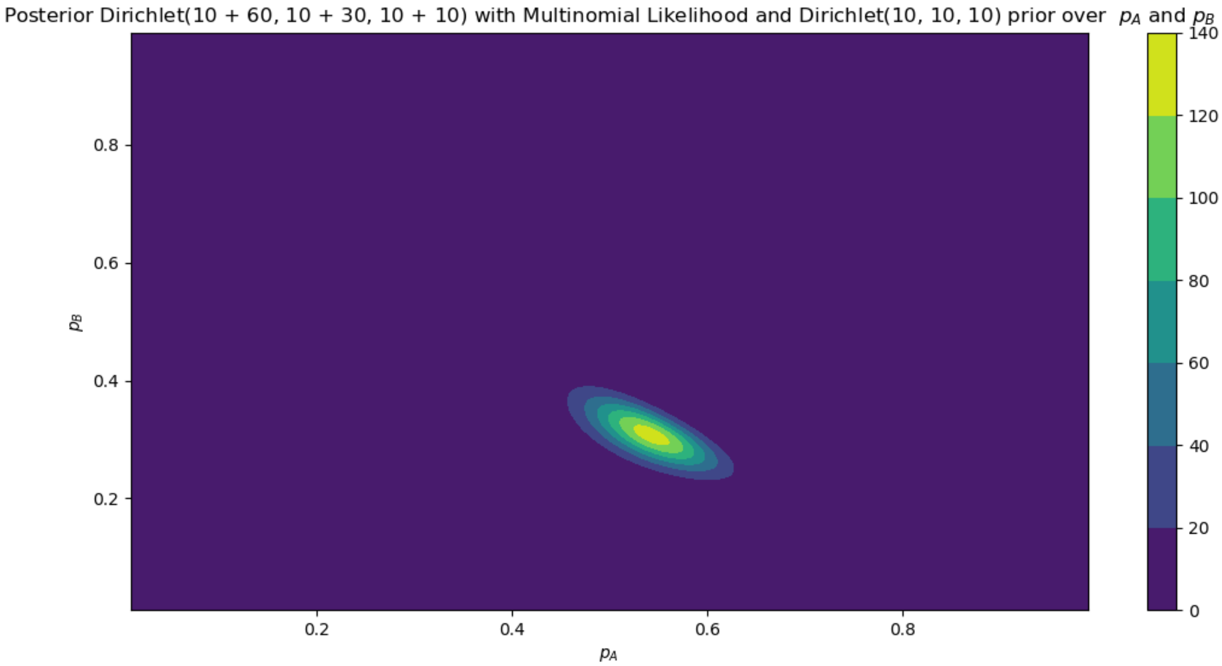

Figure 7: Dirichlet(10 + 60, 10 + 30, 10 + 10) Posterior over p_A and p_B when the amount of data is increased.