Preston Mui — mui@berkeley.edu
Econ-240a: Pset 1

2 Binomial-Beta Learning

1. **What is the conditional distribution of Z_N given θ?**

   Z_N is distributed with p.m.f.

$$f_{Z_N|\Theta}(z|\theta) = \begin{cases} \binom{n}{z}\theta^z(1-\theta)^{N-z} & \text{if } z \in 0, 1, \cdots, N \\ 0 & \text{otherwise} \end{cases}$$

2. **Calculate the joint distribution of Z_N and θ**

$$\begin{aligned} f_{Z_N,\Theta}(z,\theta) &= f_\Theta(\theta) f_{Z_N|\Theta}(z|\theta) \\ &= \frac{\theta^{a-1}(1-\theta)^{b-1}}{B(a,b)}\binom{n}{z}\theta^z(1-\theta)^{N-z} \text{ if } z \in 0,1,\cdots,N \text{ and } \theta \in [0,1]; 0 \text{ otherwise} \end{aligned}$$

3. **Calculate the conditional distribution of θ given Z_N. What is the mean of this distribution, and why would you call it Posterior?**

   By Bayes' Rule,

$$\begin{aligned} f_{\Theta|Z_N}(\theta|z) &= \frac{f_{\Theta,Z_N}(\theta,z)}{f_{Z_N}} \\ &= \frac{\frac{\theta^{a-1}(1-\theta)^{b-1}}{B(a,b)}\binom{n}{z}\theta^z(1-\theta)^{N-z}}{\int_0^1 \frac{\theta^{a-1}(1-\theta)^{b-1}}{B(a,b)}\binom{n}{z}\theta^z(1-\theta)^{N-z}d\theta} \\ &= \frac{\theta^{a-1}(1-\theta)^{b-1}\theta^z(1-\theta)^{N-z}}{\int_0^1 \theta^{z+a-1}(1-\theta)^{N-z+b-1}d\theta} \\ &= \frac{\theta^{z+a-1}(1-\theta)^{N-z+b-1}}{B(z+a, N-z+b)} \end{aligned}$$

   So $\Theta|Z_N$ is beta-distributed $(z+a, N-z+b)$, with a mean of $\frac{z+a}{z+a+N-z+b} = \frac{z+a}{N+a+b}$. The name is "posterior" because it is finding a distribution of θ **after**, or "post" drawing the data.

4. **Say $\alpha = \beta = 1/2$.**

   This prior assigns the bulk of the prior belief to the tails (close to 0 and 1) with low probability for θ close to 0.5, and even probability to $\theta > 0.5$ and $\theta < 0.5$. This prior is not very good because almost assuredly more than half of Cal students would intend to vote for Obama, so assigning equal probability mass to both sides of 0.5 does not make sense. Also, a belief that assigns high probability mass to *both* large support for Obama and very little support for Obama does not make sense.

5. **The Harold Jeffreys interval:** This interval uses the $\frac{\alpha}{2}$ and $1-\frac{\alpha}{2}$ quantiles of the posterior distribution from above. So, given the prior distribution on θ from above, the Harold Jeffreys interval contains the true value of θ with probability $1-\alpha$.

6. **Given a Jeffrey's interval of [0.47,0.73], what is the probability that this interval contains θ?** If I have constructed a Jeffrey's interval for some prior distribution of θ, some data sample, and some choice of α, then I would say that the interval contains θ is $1-\alpha$ with probability $1-\alpha$. Of course, in *reality*, the probability is either 0 or 1, because the "true" θ is fundamentally deterministic. This probability reflects some probabilistic belief about θ that depends on the choice of prior.

3 Multivariate Normal Distribution

1. **Let C be a $K \times K$ nonsingular matrix. Show that $Z = CY$ is distributed according to** $\mathcal{N}(C\mu, C\Sigma C')$:

Because $Z = CY$ is a linear transformation, and C is invertible,

$$\begin{aligned}
f(z_1, \cdots, z_K) &= \frac{1}{|C|} f[C^{-1}Y] \\
&= (2\pi)^{-K/2}|\Sigma|^{-1/2}|C|^{-1} \exp\bigg(-\frac{1}{2}(C^{-1}y-\mu)'\Sigma^{-1}(C^{-1}y-\mu)\bigg) \\
&= (2\pi)^{-K/2}|\Sigma|^{-1/2}|C|^{-1/2}|C|^{-1/2} \exp\bigg(-\frac{1}{2}(C^{-1}y-C^{-1}C\mu)'\Sigma^{-1}(C^{-1}y-C^{-1}C\mu)\bigg) \\
&= (2\pi)^{-K/2}(|C||\Sigma||C|)^{-1/2} \exp\bigg(-\frac{1}{2}(C^{-1}(y-C\mu))'\Sigma^{-1}(C^{-1}(y-C\mu))\bigg) \\
&= (2\pi)^{-K/2}(|C\Sigma C'|)^{-1/2} \exp\bigg(-\frac{1}{2}(y-C\mu)'C^{-1'}\Sigma^{-1}C^{-1}(y-C\mu)\bigg) \\
&= (2\pi)^{-K/2}(|C\Sigma C'|)^{-1/2} \exp\bigg(-\frac{1}{2}(y-C\mu)'C'^{-1}\Sigma^{-1}C^{-1}(y-C\mu)\bigg) \\
&= (2\pi)^{-K/2}(|C\Sigma C'|)^{-1/2} \exp\bigg(-\frac{1}{2}(y-C\mu)'(C\Sigma C')^{-1}(y-C\mu)\bigg)
\end{aligned}$$

Which is a distribution with mean $C\mu$ and variance-covariance matrix $C\Sigma C'$.

2. Show that Y_1', Y_2' are independent if $\Sigma_{12} = \Sigma_{21}' = \underline{0}\underline{0}'$.

   If Y_1' and Y_2' are independent, then $f(Y_1')f(Y_2') = f(Y_1', Y_2')$:

$$\begin{aligned}
f(Y_1', Y_2') &= (2\pi)^{-K/2}|\Sigma|^{-1/2} \exp\bigg(-\frac{1}{2}(y-\mu)'\Sigma^{-1}(y-\mu)\bigg) \\
&= (2\pi)^{-(K_1+K_2)/2}(|\Sigma_{11}||\Sigma_{22}|)^{-1/2} \exp\bigg(-\frac{1}{2}(y_1-\mu_1, y_2-\mu_2)' \begin{pmatrix} (\Sigma_{11})^{-1} & \underline{0}\underline{0}' \\ \underline{0}\underline{0} & (\Sigma_{22})^{-1} \end{pmatrix} (y_1-\mu_1, y_2-\mu_2)\bigg)
\end{aligned}$$

using the following facts:

$\Sigma_{12} = \Sigma_{21} = \underline{0}\underline{0}' \implies |\Sigma| = |\Sigma_{11}||\Sigma_{22}|$

$\Sigma_{12} = \Sigma_{21} = \underline{0}\underline{0}' \implies \Sigma^{-1} = \begin{pmatrix} (\Sigma_{11})^{-1} & \underline{0}\underline{0}' \\ \underline{0}\underline{0} & (\Sigma_{22})^{-1} \end{pmatrix}$

$$\begin{aligned}
&= (2\pi)^{-(K_1+K_2)/2}(|\Sigma_{11}||\Sigma_{22}|)^{-1/2} \exp\bigg(-\frac{1}{2}\big((y_1-\mu_1)'(\Sigma_{11})^{-1}, (y_2-\mu_2)'\Sigma_{22})^{-1}\big)\big(y_1-\mu_1, y_2-\mu_2\big)\bigg) \\
&= (2\pi)^{-(K_1+K_2)/2}(|\Sigma_{11}||\Sigma_{22}|)^{-1/2} \exp\bigg(-\frac{1}{2}\big((y_1-\mu_1)'(\Sigma_{11})^{-1}(y_1-\mu_1) + (y_2-\mu_2)'\Sigma_{22})^{-1}(y_2-\mu_2)\big)\bigg) \\
&= (2\pi)^{-K_1/2}|\Sigma_{11}|^{-1/2} \exp\bigg(-\frac{1}{2}(y_1-\mu_1)'(\Sigma_{11})^{-1}(y_1-\mu_1)\bigg) \times \cdots \\
&\qquad \cdots (2\pi)^{-K_2/2}|\Sigma_{22}|^{-1/2} \exp\bigg(-\frac{1}{2}(y_2-\mu_2)'(\Sigma_{22})^{-1}(y_2-\mu_2)\bigg) = f(Y_1')f(Y_2')
\end{aligned}$$

   Where Y_1 and Y_2 are independently distributed with means μ_1, μ_2 and variance-covariance matrices Σ_{11}, Σ_{22}.

3. From part 1, $Z = CY$ is distributed normally with mean $C\mu$ and variance-covariance matrix $C\Sigma C'$:

$$C\mu = \begin{pmatrix} I_{K_1} & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I_{K_2} \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2 \\ \mu_2 \end{pmatrix}$$

$$\begin{aligned}
C\Sigma C' &= \begin{pmatrix} I_{K_1} & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I_{K_2} \end{pmatrix} \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \begin{pmatrix} I_{K_1} & 0 \\ (-\Sigma_{12}\Sigma_{22}^{-1})' & I_{K_2} \end{pmatrix} \\
&= \begin{pmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & \Sigma_{12} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{22} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \begin{pmatrix} I_{K_1} & 0 \\ -(\Sigma_{22}^{-1})'\Sigma_{12}' & I_{K_2} \end{pmatrix} \\
&= \begin{pmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & 0 \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \begin{pmatrix} I_{K_1} & 0 \\ -\Sigma_{22}^{-1}\Sigma_{21} & I_{K_2} \end{pmatrix} \\
&= \begin{pmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & 0 \\ \Sigma_{21} - \Sigma_{22}\Sigma_{22}^{-1}\Sigma_{21} & \Sigma_{22} \end{pmatrix} \\
&= \begin{pmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & 0 \\ 0 & \Sigma_{22} \end{pmatrix}
\end{aligned}$$

4. Let E be the matrix $\begin{pmatrix}0 & I_{K-P}\end{pmatrix}$, which is the $(K-P)\times P$ zero matrix concatenated with the $(K-P)\times(K-P)$ identity matrix. From previous results we know that $\begin{pmatrix}D\\E\end{pmatrix}Y$ is distributed $\mathcal{N}\left(\begin{pmatrix}D\\E\end{pmatrix}\mu, \begin{pmatrix}D\\E\end{pmatrix}\Sigma\begin{pmatrix}D\\E\end{pmatrix}'\right)$.

   The mean of $\begin{pmatrix}D\\E\end{pmatrix}Y$ is

$$\begin{pmatrix}D_{11} & D_{12}\\ 0 & I_{K-P,K-P}\end{pmatrix}\mu = \begin{pmatrix}D\mu\\ \mu_2\end{pmatrix}$$

   and the variance of $\begin{pmatrix}D\\E\end{pmatrix}Y$ is

$$\begin{aligned}
&\begin{pmatrix}D_{11} & D_{12}\\ 0 & I_{K-P,K-P}\end{pmatrix}\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}\begin{pmatrix}D_{11}' & 0\\ D_{12}' & I_{K-P,K-P}\end{pmatrix}\\
&= \begin{pmatrix}D_{11}\Sigma_{11}+D_{12}\Sigma_{21} & D_{11}\Sigma_{12}+D_{12}\Sigma_{22}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}\begin{pmatrix}D_{11}' & 0\\ D_{12}' & I_{K-P,K-P}\end{pmatrix}\\
&= \begin{pmatrix}D_{11}\Sigma_{11}D_{11}'+D_{12}\Sigma_{21}D_{11}'+D_{11}\Sigma_{12}D_{12}'+D_{12}\Sigma_{22}D_{12}' & D_{11}\Sigma_{12}+D_{12}\Sigma_{22}\\ \Sigma_{21}D_{11}'+\Sigma_{22}D_{12}' & \Sigma_{22}\end{pmatrix}\\
&= \begin{pmatrix}D\Sigma D' & D_{11}\Sigma+D_{12}\Sigma_{22}\\ \Sigma_{21}D_{11}'+\Sigma_{22}D_{12}' & \Sigma_{22}\end{pmatrix}
\end{aligned}$$

   Taking the first P elements of $\begin{pmatrix}D\mu\\ \mu_2\end{pmatrix}$ and the upper-left $P\times P$ elements of $\begin{pmatrix}D\Sigma D' & D_{11}\Sigma+D_{12}\Sigma_{22}\\ \Sigma_{21}D_{11}'+\Sigma_{22}D_{12}' & \Sigma_{22}\end{pmatrix}$ shows that Z is distributed normally with mean $D\mu$ and variance-covariance $D\Sigma D'$.

5. Let $C = \begin{pmatrix}I_{K_1} & -\Sigma_{12}\Sigma_{22}^{-1}\\ 0 & I_{K_2}\end{pmatrix}$ from question 3, where I found that $(CY)_1$ and $(CY)_2$ are independent, where $(CY)_1$ is the first K_1 elements of CY and $(CY)_2$ the last K_2 elements of CY. In question 4 I found that $(CY)_1$ is distributed $\mathcal{N}\left(\mu_1-\Sigma_{12}\Sigma_{22}^{-1}\mu_2, \Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$. Since $(CY)_1$ and $(CY)_2$ are independent, and $(CY)_2 = Y_2$, this is also the conditional distribution of $(CY)_1$ given Y_2. That is,

$$\begin{aligned}
\left((Y_1-\Sigma_{12}\Sigma_{22}^{-1}Y_2)|Y_2=y_2\right) &\sim \mathcal{N}\left(\mu_1-\Sigma_{12}\Sigma_{22}^{-1}\mu_2, \Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)\\
\left(Y_1|Y_2=y_2\right) &\sim \Sigma_{12}\Sigma_{22}^{-1}y_2+\mathcal{N}\left(\mu_1-\Sigma_{12}\Sigma_{22}^{-1}\mu_2, \Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)\\
\left(Y_1|Y_2=y_2\right) &\sim \mathcal{N}\left(\mu_1-\Sigma_{12}\Sigma_{22}^{-1}(y_2-\mu_2), \Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)
\end{aligned}$$

6. Let $D = \begin{pmatrix}I_{K\times K}^{(1)} & I_{K\times K}^{(2)} & \cdots & I_{K\times K}^{(N)}\end{pmatrix}$; that is, D is constructed by concatenating N $K\times K$ identity matrices. Then,let $\mathbb{Y} = \begin{pmatrix}Y_1 & Y_2 & \cdots & Y_N\end{pmatrix}'$; that is, $\mathbb{Y}$ is constructed by stacking the N independent random draws of Y_i stacked together. Note that since each Y_i is independently, identically, and normally distributed with mean μ and variance-covariance matrix Σ, $\mathbb{Y}$ is normally distributed with mean $\mu_{\mathbb{Y}}$ and $\Sigma_{\mathbb{Y}}$ such that

$$\mu_{\mathbb{Y}} = \begin{pmatrix}\mu^{(1)}\\ \mu^{(2)}\\ \vdots\\ \mu^{(N)}\end{pmatrix}$$

$$\Sigma_{\mathbb{Y}} = \begin{pmatrix}\Sigma^{(1)} & 0_{K\times K} & \cdots & 0_{K\times K}\\ 0_{K\times K} & \Sigma^{(2)} & \cdots & 0_{K\times K}\\ \vdots & \vdots & \ddots & \vdots\\ 0_{K\times K} & \cdots & 0_{K\times K} & \Sigma^{(N)}\end{pmatrix}$$

   where $\mu^{(i)} = \mu$ and $\Sigma^{(i)} = \Sigma$ for each $i = 1, \cdots, N$.

Then, using the result from problem 4, I can construct $\begin{pmatrix} D \\ E \end{pmatrix}$, an $NK \times NK$ square matrix such that the first K elements of $\begin{pmatrix} D \\ E \end{pmatrix} \mathbb{Y}$, $D\mathbb{Y}$, are distributed $\mathcal{N}\Big(D\mu_{\mathbb{Y}}, D\Sigma_{\mathbb{Y}}D'\Big)$.

Let Y_{ij} be the jth element of the ith random vector. Then, examine $D\mathbb{Y}$:

$$D\mathbb{Y} = \begin{pmatrix} \sum_{i=1}^N Y_{i1} & \sum_{i=1}^N Y_{i2} & \cdots & \sum_{i=1}^N Y_{iK} \end{pmatrix}' = \sum_{i=1}^N Y_i$$

That is, $D\mathbb{Y}$ is the sum of the N random draws of Y_i. This sum is distributed with mean:

$$D\mu_{\mathbb{Y}} = \begin{pmatrix} \sum_{i=1}^N \mu_1 & \sum_{i=1}^N \mu_2 & \cdots & \sum_{i=1}^N \mu_K \end{pmatrix}' = N\mu$$

and variance:

$$\begin{aligned} D\Sigma_{\mathbb{Y}}D' &= \begin{pmatrix} I^{(1)}_{K\times K} & I^{(2)}_{K\times K} & \cdots & I^{(N)}_{K\times K} \end{pmatrix} \begin{pmatrix} \Sigma^{(1)} & 0_{K\times K} & \cdots & 0_{K\times K} \\ 0_{K\times K} & \Sigma^{(2)} & \cdots & 0_{K\times K} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{K\times K} & \cdots & 0_{K\times K} & \Sigma^{(N)} \end{pmatrix} \begin{pmatrix} I^{(1)}_{K\times K} & I^{(2)}_{K\times K} & \cdots & I^{(N)}_{K\times K} \end{pmatrix}' \\ &= \begin{pmatrix} \Sigma^{(1)} & \Sigma^{(2)} & \cdots & \Sigma^{(N)} \end{pmatrix} \begin{pmatrix} I^{(1)}_{K\times K} & I^{(2)}_{K\times K} & \cdots & I^{(N)}_{K\times K} \end{pmatrix}' = N\Sigma \end{aligned}$$

That is, $\sum_{i=1}^N Y_i$ is distributed $\mathcal{N}\Big(N\mu, N\Sigma\Big)$. Using the result from part 1,

$$\begin{aligned} &\overline{Y_i} = \frac{1}{N}\sum_{i=1}^N Y_i = \frac{1}{N} I_{N\times N} \sum_{i=1}^N Y_i \sim \mathcal{N}\Big(\frac{1}{N} I_{N\times N} N\mu, (\frac{1}{N} I_{N\times N}) N\Sigma (\frac{1}{N} I_{N\times N})'\Big) = \mathcal{N}\Big(\mu, \frac{1}{N}\Sigma\Big) \\ &\implies \overline{Y_i} - \mu \sim \mathcal{N}\Big(0, \frac{1}{N}\Sigma\Big) \\ &\implies \sqrt{N}(\overline{Y_i} - \mu) \sim \mathcal{N}\Big(\sqrt{N}0, \sqrt{N}^2 \frac{1}{N}\Sigma\Big) = \mathcal{N}\Big(0, \Sigma\Big) \end{aligned}$$

7. Because Σ is a symmetric and positive definite matrix, there exists a unique and symmetric $\sqrt{\Sigma}$ such that $\sqrt{\Sigma}\sqrt{\Sigma} = \Sigma$. Then, examine W:

$$\begin{aligned} W &= N(\bar{Y} - \mu)'\Sigma^{-1}(\bar{Y} - \mu) \\ &= \sqrt{N}(\bar{Y} - \mu)'\Sigma^{-1}\sqrt{N}(\bar{Y} - \mu) \end{aligned}$$

Let $Z \sim \mathcal{N}(0, I)$; that is, Z is a standard multivariate normal. Above we found that $\sqrt{N}(\bar{Y} - \mu) \sim \mathcal{N}(0, \Sigma)$; using the result from question 1 and the aforementioned properties of Σ, we can say that $\sqrt{\Sigma}Z \sim \mathcal{N}(0, \Sigma)$. So, W can be rewritten as

$$\begin{aligned} W &= (\sqrt{\Sigma}Z)'\Sigma^{-1}(\sqrt{\Sigma}Z) \\ &= Z'\sqrt{\Sigma}'(\sqrt{\Sigma}\sqrt{\Sigma})^{-1}\sqrt{\Sigma}Z \\ &= Z'\sqrt{\Sigma}\sqrt{\Sigma}^{-1}\sqrt{\Sigma}^{-1}\sqrt{\Sigma}Z = Z'Z \end{aligned}$$

which is distributed as χ^2_K.