SDS 383D Ex 01: Preliminaries

January 18, 2016

Jennifer Starling

Bayesian Inference in Simple Conjugate Families

Part A

Let $x_1, ..., x_n \sim$ iid Bernoulli(w). Let $w \sim$ Beta(a,b) be the prior.

Let y be the number of successes in the sequence of n Bernoulli trials. Then $y \sim Binom(n, w)$.

We begin with the following pdfs:

- Prior is $p(w) = \frac{1}{Beta(a,b)} w^{a-1}(1-w)^{b-1}$
- Sampling model is $p(y|w) = \binom{n}{y} w^y (1-w)^{n-y}$

Then $posterior \propto sampling * prior$:

- $p(w|y) \propto w^y(1-w)^{n-y} * w^{a-1}(1-w)^{b-1}$

  $= w^{a+y-1}(1-w)^{b+(n-y)-1}$

This is the kernel of the $Beta(a+y, b+n-y)$ distribution. Therefore:

The posterior is $p(w|y) \sim Beta(a+y, b+n-y)$. $\blacksquare$

Part B

The pdf for the $gamma(a, b)$ distribution is: $p(x) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}$.

Let $x_1 \sim gamma(a_1, 1)$ and $x_2 \sim gamma(a_2, 1)$. Define $y_1 = \frac{x_1}{x_1+x_2}$ and $y_2 = x_1 + x_2$.

First, obtain the joint density of (y_1, y_2) using the standard bivariate transformation procedure (as defined in Chapter 4 of Casella and Berger).

Step 1: Obtain Transformation Equations

Find $g_1^{-1}(x_1, x_2)$ and $g_2^{-1}(x_1, x_2)$ inverse equations, and check that transformation is 1-1 and onto.

$y_1 = \frac{x_1}{x_1+x_2}$ and $y_2 = x_1 + x_2$.

- Plug second equation into first to obtain $y_1 = \frac{x_1}{x_2}$. Then $x_1 = y_1 y_2 \rightarrow g_1^{-1}(y_1, y_2) = y_1 y_2$.
- Plug previous result for x_1 into second equation.

  Then $y_2 = y_1 y_2 + x_2 \rightarrow x_2 = y_2 - y_1 y_2 \rightarrow g_2^{-1}(y_1, y_2) = y_2 - y_1 y_2$.

This transformation is 1-1 and onto, with support mapping $\{x_1 > 0, x_2 > 0\} \rightarrow \{0 < y_1 < 1, y_2 > 0\}$.

- Onto: Met since able to find unique inverse equations in part 1, above.
- 1-1. Met. Let $(y_{11}, y_{21}) = (y_{21}, y_{22})$. We can then do the algebra to show that $(x_{11}, x_{21}) = (x_{21}, y_{22})$.

Step 2: Jacobian

$$|J| = \left| \begin{pmatrix} \frac{\delta g_1^{-1}}{\delta y_1} & \frac{\delta g_1^{-1}}{\delta y_2} \\ \frac{\delta g_2^{-1}}{\delta y_1} & \frac{\delta g_1^{-1}}{\delta y_2} \end{pmatrix} \right| = \left| \begin{pmatrix} y_2 & y_1 \\ -y_2 & (1-y_1) \end{pmatrix} \right| = |y_2(1-y_1)| + y_1 y_2 = |y_2 - y_1 y_2| = |y_2| = y_2 \text{ since } y_2 > 0.$$

Therefore, $|J| = y_2$.

Step 3: Joint pdf

Since $x_1 \perp x_2$, the joint pdf of x_1 and x_2 is:

$$f_{x_1,x_2}(x_1, x_2) = f(x_1) f(x_2) = \frac{1}{\Gamma(a_1)\Gamma(a_2)} x_1^{a_1-1} x_2^{a_2-1} e^{(-x_1-x_2)}.$$

The joint pdf of y_1 and y_2 is:

$$f_{x_1,x_2}(g_1^{-1}, g_2^{-1})|J| = \frac{1}{\Gamma(a_1)\Gamma(a_2)} (y_1 y_2)^{a_1-1} (y_2 - y_1 y_2)^{a_2-1} e^{\{-y_1 y_2 - y_2(1-y_1)\}} y_2$$

$$= \frac{1}{\Gamma(a_1)\Gamma(a_2)} (y_1 y_2)^{a_1-1} (y_2 - y_1 y_2)^{a_2-1} e^{(y_2)} y_2$$

The joint pdf can be factored into functions of y_1 and y_2 as follows. We can also multiply and divide by $\Gamma(a_1 + a_2)$ to make it easy to identify the marginal densities.

$$= \frac{1}{\Gamma(a_1+a_2)} \frac{\Gamma(a_1+a_2)}{\Gamma(a_1)\Gamma(a_2)} y_1^{a_1-1} y_2^{a_1-1} y_2^{a_2-1} (1-y_1)^{a_2-1} y_2 e^{(-y_2)}$$

$$= \left[\frac{\Gamma(a_1+a_2)}{\Gamma(a_1)\Gamma(a_2)} y_1^{a_1-1} (1-y_1)^{a_2-1} \right] \left[\frac{1}{\Gamma(a_1+a_2)} y_2^{a_1+a_2-1} e^{(-y_2)} \right]$$

These are the forms of the beta and gamma densities, respectively. Therefore:

- $y_1 \sim Beta(a_1, a_2)$
- $y_2 \sim Gamma(a_1 + a_2, 1)$

We can then devise a process to generate Beta realizations. We can generate two independent gamma realizations (x_1, x_2) and calculate $y_1 = \frac{x_1}{x_1+x_2}$ to simulate the Beta realizations.

Part C

Let $x_1, ..., x_N \sim N(\theta, \sigma^2)$ where θ is unknown and σ^2 is known. The prior for θ is $\theta \sim N(m, v)$. Derive the posterior for $p(\theta | x_1, ..., x_N)$.

- Prior: $p(\theta) = \frac{1}{\sqrt{2\pi v}} exp\{-\frac{1}{2v}(\theta - m)^2\}$

- Sampling model: $p(x_1, ..., x_n | \theta, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} exp\{-\frac{1}{2\sigma^2}(x_i - \theta)^2\}$

  $= (2\pi\sigma^2)^{-n/2} exp\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta)^2\}$

Expand the summation in the exponential term to make it easier to work with:

$$\sum_{i=1}^n (x_i - \theta)^2 = \sum_{i=1}^n (x_i - \theta)(x_i - \theta) = \sum_{i=1}^n [x_i^2 - 2\theta \sum_{i=1}^n x_i + n\theta^2] = n\bar{x}^2 - 2n\theta\bar{x} + n\theta^2.$$

Then $posterior \propto sampling * prior$:

$$p(\theta | x_1, ..., x_n) \propto exp\left\{-\frac{1}{2\sigma^2}\left(n\bar{x}^2 - 2n\theta\bar{x} + n\theta^2\right)\right\} * exp\left\{-\frac{1}{2v}\left(\theta^2 - 2\theta m + m^2\right)\right\}$$

Drop all terms unrelated to θ (remember, σ^2 is known, so is okay). Combine into one exponential term.

$$= exp\left\{-\frac{1}{2}\left(\frac{n}{\sigma^2}\theta^2 + \frac{1}{v}\theta^2 - \frac{2n\bar{x}}{\sigma^2}\theta - \frac{2m}{v}\theta\right)\right\}$$

Combine the θ^2 coefficients and the θ coefficients to make this form easier to work with. Let:

- $a = \left(\frac{n}{\sigma^2} + \frac{1}{v}\right)$
- $b = \left(\frac{2n\bar{x}}{\sigma^2} + \frac{2m}{v}\right)$

This yields the equation $= exp\left\{-\frac{1}{2}\left(a\theta^2 - 2b\theta\right)\right\}$. Now we need to complete the square.

Aside: A brief refresher on completing the square.

- Begin with $ax^2 - 2bx$. Need form $x^2 - 2bx + b^2$, since this factors into $(x + b)^2$.
- Accomplish this by factoring out a to obtain $a(x^2 - 2\frac{b}{a}x)$.
- Then add and subtract $(\frac{b}{a})^2$ inside the parenthesis.

In our case, begin with $= exp\left\{-\frac{1}{2}\left(a\theta^2 - 2b\theta\right)\right\}$. Set aside the exponential; just work with the term inside to complete the square.

- Factor a out, to obtain $-\frac{a}{2}\left(\theta^2 - 2\frac{b}{a}\theta\right)$.

- Add and subtract $(\frac{b}{a})^2$ inside the parenthesis to get $-\frac{a}{2}\left(\theta^2 - 2\frac{b}{a}\theta + (\frac{b}{a})^2 - (\frac{b}{a})^2\right)$.

- The added and subtracted terms are not functions of θ, so we can drop the $-(\frac{b}{a})^2$ term, leaving $-\frac{a}{2}\left(\theta^2 - 2\frac{b}{a}\theta + (\frac{b}{a})^2\right)$.
- This factors into $-\frac{a}{2}(\theta - \frac{b}{a})^2$.

Plug the exponential term back into the full equation: $exp\left\{-\frac{a}{2}(\theta - \frac{b}{a})^2\right\}$.

This has the form of a normal distribution, with mean $\frac{b}{a}$ and variance $\frac{1}{a}$, ie precision equals a. Therefore:

The posterior $p(\theta|x_1,...,x_n) \sim N\left[\left(\frac{2n\bar{x}}{\sigma^2} + \frac{2m}{v}\right), \left(\frac{n}{\sigma^2} + \frac{1}{v}\right)^{-1}\right]$ ■

KEY NOTE:

Also can be more intuitively written as:

- Mean $= \left(\frac{m}{v} + \frac{\sum_{i=1}^{n}}{\sigma^2}\right) / \left(\frac{1}{v} + \frac{n}{\sigma^2}\right)$
- Variance $= 1/(\frac{1}{v} + \frac{n}{\sigma^2}) = (\frac{1}{v} + \frac{n}{\sigma^2})^{-1}$

This is important because the mean is a precision-weighted average of the prior mean and the sample mean of the data.

The precision is additive. It is often easier to work with precisions than variances.

Part D

Let $x_1,...,x_n \sim N(\theta, \sigma^2)$ where θ is known and σ^2 is unknown. Will express σ^2 in terms of precison $w = \frac{1}{\sigma^2}$. Find the posterior $p(w|x_1,...,x_n)$.

- Prior: $w \sim Gamma(a,b)$, ie $\sigma^2 \sim IG(a,b)$, so $p(w) = \frac{b^a}{\Gamma(a)} w^{a-1}e^{-bw}$. (IG is inverse Gamma.)

- Sampling model: $p(x_1,...,x_n|\theta,w) = \prod_{i=1} n \left(\frac{w}{2\pi}\right)^{\frac{1}{2}} exp\left\{-\frac{w}{2}(x_i-\theta)^2\right\}$

  $= \frac{w^{n/2}}{(2\pi)^{n/2}} exp\left\{-\frac{w}{2}\sum_{i=1}^{n}(x_i-\theta)^2\right\}$

Then $posterior \propto sampling * prior$:

$$p(w|x_1,...,x_n) \propto w^{a+\frac{n}{2}-1} exp\left\{-w\left(b + \frac{\sum_{i=1}^{n}(x_i-\theta)^2}{2}\right)\right\}$$

This is the form of the gamma distribution, so the posterior for w is $p(w|x_1,...,x_n) \sim Gamma\left(a + \frac{n}{2}, b + \frac{\sum_{i=1}^{n}(x_i-\theta)^2}{2}\right)$.

Equivalently, the posterior for σ^2 is Inverse Gamma (IG), with the same parameters. ■

Part E

Let $x_1,...,x_n \sim N(\theta, \sigma_i^2)$ where θ is common for all x_i and is unknown. Variances are unique for each x_i and are known. The prior is $\theta \sim N(m,v)$. Derive the posterior for $p(\theta|x_1,...,x_n)$.

- Prior: $p(\theta) = \frac{1}{\sqrt{2\pi v}} exp\{-\frac{1}{2v}(\theta - m)^2\}$

- Sampling model: $p(x_1,...,x_n|\theta,\sigma_1^2,...,\sigma_n^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_i^2}} exp\{-\frac{1}{2\sigma_i^2}(x_i-\theta)^2\}$

We can drop the constant of proportionality from the sampling model since it does not depend on θ.

$$\propto exp\left\{-\frac{1}{2}\sum_{i=1}^n \frac{1}{\sigma_i^2}(x_i-\theta)^2\right\}$$

Then $posterior \propto sampling * prior$:

$$p(\theta|x_1,...,x_n) \propto exp\left\{-\frac{1}{2v}(\theta-m)^2 - \frac{1}{2}\sum_{i=1}^n \frac{1}{\sigma_i^2}(x_i-\theta)^2\right\}$$

$$= exp\left\{-\frac{1}{2}\left[\frac{(\theta-m)^2}{v} + \sum_{i=1}^n \frac{1}{\sigma_i^2}(x_i-\theta)^2\right]\right\}$$

$$= exp\left\{-\frac{1}{2}\left[\frac{\theta^2}{v} - \frac{2m}{v}\theta + \frac{m^2}{v} + \sum_{i=1}^n \frac{1}{\sigma_i^2}(x_i-\theta)(x_i-\theta)\right]\right\}$$

$$= exp\left\{-\frac{1}{2}\left[\frac{\theta^2}{v} - \frac{2m}{v}\theta + \frac{m^2}{v} + \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} - 2\theta\sum_{i=1}^n \frac{x_i}{\sigma_i^2} + \theta^2\sum_{i=1}^n \frac{1}{\sigma_i^2}\right]\right\}$$

The two terms $\frac{m^2}{v}$ and $\sum_{i=1}^n \frac{x_i^2}{\sigma_i^2}$ can be dropped since they do not depend on θ.

$$= exp\left\{-\frac{1}{2}\left[\frac{\theta^2}{v} - \frac{2m}{v}\theta - 2\theta\sum_{i=1}^n \frac{x_i}{\sigma_i^2} + \theta^2\sum_{i=1}^n \frac{1}{\sigma_i^2}\right]\right\}$$

Group the θ^2 terms and the θ terms.

$$= exp\left\{-\frac{1}{2}\left[\theta^2\left(\frac{1}{v} + \sum_{i=1}^n \frac{1}{\sigma_i^2}\right) - 2\theta\left(\frac{m}{v} + \sum_{i=1}^n \frac{x_i}{\sigma_i^2}\right)\right]\right\}$$

Then, as before, we can use a and b to facilitate completing the square. Let:

- $a = \left(\frac{1}{v} + \sum_{i=1}^n \frac{1}{\sigma_i^2}\right)$
- $b = \left(\frac{m}{v} + \sum_{i=1}^n \frac{x_i}{\sigma_i^2}\right)$

Then we have $exp\left\{-\frac{1}{2}[a\theta^2 - 2b\theta]\right\}$.
We can repeat the process from Part C to complete the square. Working with just the inside term of the exponential expression:

- $= a\theta^2 - 2b\theta = -\frac{a}{2}[\theta^2 - 2\frac{b}{a}\theta + \frac{b^2}{a^2} - \frac{b^2}{a^2}] = -\frac{a}{2}(\theta - \frac{b}{a})^2$

Plugging back into the exponential, we have $exp\left\{-\frac{a}{2}(\theta-\frac{b}{a})^2\right\} = exp\left\{-\frac{1}{2(1/a)}(\theta-\frac{b}{a})^2\right\}$.

This is the form of the normal density. Therefore, the posterior is distributed as follows:

$$p(\theta|x_1,...,x_n) \sim N\left(\frac{1}{v} + \sum_{i=1}^n \frac{1}{\sigma_i^2}, \frac{\frac{m}{v}+\sum_{i=1}^n \frac{x_i}{\sigma_i^2}}{\frac{1}{v}+\sum_{i=1}^n \frac{1}{\sigma_i^2}}\right) \blacksquare$$

Part F

Let $(x|\sigma^2) \sim N(0, \sigma^2)$ with prior $\frac{1}{\sigma^2} \sim Gamma(a, b)$, as in part D. Show the marginal of x is Student's t. (Note: this is for a single observation, not $x_1, ..., x_n$.)

The marginal of x is $p(x) = \int_\Theta p(x|\sigma^2)p(\sigma^2)\delta\sigma^2$.

$$p(x) = \int_0^\infty (2\pi\sigma^2)^{-1/2} e^{-\frac{1}{2\sigma^2}x^2} * \frac{b^a}{\Gamma(a)} (\sigma^2)^{-a-1} e^{-b/\sigma^2}$$

$$= 2^{-1/2}\pi^{-1/2} \frac{b^a}{\Gamma(a)} \int_0^\infty (\sigma^2)^{-a-\frac{1}{2}-1} e^{-\frac{1}{\sigma^2}[\frac{x^2}{2}+b]} \delta\sigma^2$$

The integral has the form of the Inverse Gamma pdf for $IG(a+\frac{1}{2}, \frac{x^2}{2} - b)$. This integral is missing the constant of proportionality. If $1 = c * int$, then $int = 1/c$. So the integral term is equal to $\Gamma(a+\frac{1}{2})$. Plugging this in:

$$= \frac{1}{\sqrt{2\pi}} \frac{b^a}{\Gamma(a)} \Gamma(a+\tfrac{1}{2}) (\tfrac{x^2}{2} + b)^{-(a+\frac{1}{2})}$$

This is the form of the Student's t distribution. CHECK ON EXTRA B's HANGING AROUND. ■

The Multivariate Normal Distribution

Part A

In matrix notation, $cov(x) = E\left\{(x-\mu)(x-\mu)^T\right\}$ where μ is the mean vector whose ith component is $E(x_i)$.

(1): Prove $cov(x) = E(xx^T) - \mu\mu^T$.

Begin with the definition of $cov(x)$: $cov(x) = E\left\{(x-\mu)(x-\mu)^T\right\}$

$= E\left\{(x-\mu)(x^T-\mu^T)\right\}$, then expand the terms

$= E\left(xx^T - 2x\mu^T + \mu\mu^T\right)$

$= E(xx^T) - E(2x\mu^T) + E(\mu\mu^T)$, by linearity of expectations

$= E(xx^T) - 2\mu^T E(x) + \mu\mu^T$, since $E(c) = c$ and $E(cx) = cE(x)$ for a constant c

$= E(xx^T) - 2\mu\mu^T + \mu\mu^T$, since μ is a vector whose ith component is $E(x_i)$, and $E(x) = \mu$

$= E(xx^T) - \mu\mu^T$ ■

(2): Prove $cov(Ax+b) = Acov(x)A^T$ for matrix A and vector b.

Begin with the definition of covariance: $cov(Ax+b) = E\left\{\left[(Ax+b) - E(Ax+b)\right]\left[(Ax+b) - E(Ax+b)\right]^T\right\}$

$= E\left\{\left[Ax+b-AE(x)-b\right]\left[Ax+b-AE(x)-b\right]^T\right\}$

$= E\left\{(Ax-A\mu)(Ax-A\mu)^T\right\}$, since $E(x) = \mu$ and the bs cancel

$= E\left\{(Ax-A\mu)(x^TA^T-\mu^TA^T)\right\}$, by distributing the transpose

$= E\left\{A(x-\mu)(x^T-\mu^T)A^T\right\}$, by pulling A and A^T out of parenthesis

$= E\left\{A(x-\mu)(x-\mu)^TA^T\right\}$, by pulling out transpose

$= AE\left\{(x-\mu)(x-\mu)^T\right\}A^T$, by pulling constants A and A^T out of expectation

$= Acov(x)A^T$, since $cov(x) = E\left\{(x-\mu)(x-\mu)^T\right\}$ ■

Part B

Let z be a random vector $z = (z_1, ..., z_p)^T$, with iid $z_i \sim N(0,1)$. Derive the pdf and mgf of z, in vector notation.

(1): Pdf of z

Since z_i are independent, the join pdf is the product of the individual pdfs.

$$p(z) = \prod_{i=1}^{z}(2\pi)^{-1/2}e^{-z_i^2/2} = (2\pi)^{-p/2}exp\left\{\frac{-\sum_{i=1}^{n} z_i^2}{2}\right\}$$

In vector form, $\sum_{i=1}^{n} z_i^2 = z^T z$, so we can rewrite the pdf in vector notation.

$$p(z) = (2\pi)^{-p/2}exp\left\{-\frac{1}{2}z^T z\right\} \blacksquare$$

(2): Mgf of z

The definition of the mgf of a random variable vector is $M_x(t) = E(e^{t^T x})$ in vector notation (pg 3, note 5).

$$M_z(t) = E(e^{t^T z}) = \int_{-\infty}^{\infty} e^{t^T z} p(z)\delta z$$

$$= \int_{-\infty}^{\infty} e^{t^T z}(2\pi)^{-p/2}exp\left\{-\frac{1}{2}z^T z\right\}\delta x$$

$$= \int_{-\infty}^{\infty}(2\pi)^{-p/2}exp\left\{-\frac{1}{2}z^T z + t^T z\right\}\delta x$$

Now let's look at just the exponential term: $z^T z + t^T z$. Again, we will need to complete the square.

> This term is the vector version of the 1-dimensional $z^2 + bz$, where $b = t^T$, so we can add and subtract b^2, ie $t^T t$, to complete the square.

The exponential term rearranges as follows.

$$-\frac{1}{2}\left(z^T z + t^T z + t^T t - t^T t\right) = -\frac{1}{2}\left(z^T z + 2(\frac{t^T z}{2}) + t^T t - t^T t\right) = -\frac{1}{2}(z+\frac{t}{2})^T(z+\frac{t}{2}) - \frac{1}{2}t^T t$$

Plugging this result back into the full mgf function, we obtain:

$$M_z(t) = \int_{-\infty}^{\infty}(2\pi)^{-p/2}exp\left\{-\frac{1}{2}(z+\frac{t}{2})^T(z+\frac{t}{2}) - \frac{1}{2}t^T t\right\}\delta x$$

The last term in the exponential, $-\frac{1}{2}t^T t$, can factor out of the integral since $exp\left\{-\frac{1}{2}t^T t\right\}$ is not a function of z.

$$M_z(t) = exp\left\{-\frac{1}{2}t^T t\right\}\int_{-\infty}^{\infty}(2\pi)^{-p/2}exp\left\{-\frac{1}{2}(z+\frac{t}{2})^T(z+\frac{t}{2})\right\}\delta x$$

The integral is now the pdf for the multivariate normal $N(-\frac{t}{2})$, and so integrates to 1. Therefore:

$M_z(t) = exp\left\{-\frac{1}{2}t^T t\right\}$ is the standard multivariate MGF. $\blacksquare$

Part C

Part D

Let z have a standard multivariate normal distribution. Define the random vector $x = Lz + \mu$ for (pxp) matrix L of full column rank. Prove that x is multivariate normal.

Let $x = Lz + \mu$ as described above.

Note that the MGF of z is $M_z(t) = exp\left\{-\frac{1}{2}t^T t\right\}$, from Part B.

$M_x(t) = E(e^{t^T x})$, by definition (Part B)

$= E\left(e^{t^T(Lz+\mu)}\right)$, by subbing in definition of x

$= E\left(e^{t^T Lz + t^T \mu}\right)$, by expanding the product term

$= E\left(e^{t^T Lz} e^{t^T \mu}\right)$, by separating the exponential terms

$= e^{t^T \mu} E\left(e^{t^T Lz}\right)$, since $e^{t^T \mu}$ doesn't depend on z

Now rearrange the exponent term to have form of the mgf above, by pulling out transpose.

$= e^{t^T \mu} E\left(e^{t^T Lz}\right)$

$= e^{t^T \mu} E\left(e^{(L^T t)^T z}\right)$, by pulling out transpose

Now $E\left(e^{(L^T t)^T z}\right)$ has the form of $M_z(s)$, as defined from Part B, where $M_z(s) = exp\left\{-\frac{1}{2}s^T s\right\}$.

$M_x(t) = e^{t^T \mu} * e^{\frac{(L^T t)^T (L^T t)}{2}} = e^{t^T \mu} * e^{\frac{tLL^T t}{2}}$

$M_x(t) = exp\left\{t^T \mu + \frac{tLL^T t}{2}\right\}$

This is the mgf of the multivariate normal distribution. Therefore, x is multivariate normal, with mean μ and $\Sigma = LL^T$.

$x \sim N(\mu, \Sigma = LL^T)$. ■

Appendix: R Code