SDS 383D: Exercises 1

January 23, 2017

Professor Scott

Spencer Woody

Problem 1

Bayesian inference in simple conjugate families

(A) $X_1, \dots, X_N | w \overset{\text{iid}}{\sim} \text{Bernoulli}(w)$, $w \sim \text{Beta}(a, b)$. Define $Y := \sum_{i=1}^n X_i$, so $Y|w \sim \text{Binomial}(N, w)$.

$$p(y|w) = P(Y = y|w) = \binom{N}{y} w^y (1-w)^{N-y} \tag{1}$$

$$p(w) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} w^{a-1}(1-w)^{b-1} \tag{2}$$

By Bayes' Rule,

$$\begin{aligned} p(w|y) &\propto p(w)p(y|w) && (3)\\ &\propto \left(w^{a-1}(1-w)^{b-1}\right)\left(w^y(1-w)^{N-y}\right) && (4)\\ &= w^{a+y-1}(1-w)^{b+N-y-1}, && (5) \end{aligned}$$

so $w|y \sim \text{Beta}(a+y, b+N-y)$

(B) We have two independently distributed variables, $X_1 \sim \text{Gamma}(a_1, 1)$ and $X_2 \sim \text{Gamma}(a_2, 1)$. The joint distribution of X_1 and X_2 is

$$f_{X_1,X_2}(x_1, x_2) = \frac{b^{a_1+a_2}}{\Gamma(a_1)\Gamma(a_2)} x_1^{a_1-1} x_2^{a_2-1} \exp\left[-(x_1 + x_2)\right] \tag{6}$$

Then we define the transformation of variables $(X_1, X_2) \mapsto (Y_1, Y_2)$ as follows:

$$Y_1 = \frac{X_1}{X_1 + X_2} \tag{7}$$

$$Y_2 = X_1 + X_2. \tag{8}$$

We can find the joint distribution of Y_1 and Y_2 with

$$f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}(g_1(y_1, y_2), g_2(y_1, y_2)) \left|J\right|, \tag{9}$$

where $x_1 = g_1(y_1, y_2) = y_1 y_2$, $x_2 = g_2(y_1, y_2) = y_2(1 - y_1)$, and J is the determinant of the Jacobian matrix,

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix} = \begin{vmatrix} y_2 & y_1 \\ -y_2 & 1 - y1 \end{vmatrix} = y_2(1 - y_1) + y_2 y_1 = y_2. \tag{10}$$

Y_2 is the ratio of two nonnegative variables, so $|J| = |y_2| = y_2$. Now we can write (9) as

$$\begin{aligned} f_{Y_1,Y_2}(y_1, y_2) &= \frac{b^{a_1+a_2}}{\Gamma(a_1)\Gamma(a_2)} (y_1 y_2)^{a_1-1} \left[y_2(1-y_1)\right]^{a_2-1} \exp\left[-(y_1 y_2 + y_2(1-y_1))\right] y_2 && (11)\\ &= \frac{b^{a_1+a_2}}{\Gamma(a_1)\Gamma(a_2)} y_1^{a_1-1}(1-y_1)^{a_2-1} y_2^{a_1+a_2-1} \exp(-y_2). && (12) \end{aligned}$$

Therefore, $Y_1 \sim \text{Beta}(a_1, a_2)$ independent of $Y_2 \sim \text{Gamma}(a_1 + a_2, 1)$.

(C) $X_i|\theta \overset{\text{iid}}{\sim} \mathcal{N}(\theta, \sigma^2)$, $i = 1, 2, \ldots, N$ where σ^2 is *known* and $\theta \sim \mathcal{N}(m, v)$ is *unknown*. The posterior distribution of θ given $x_1, \ldots, x_N$ is

$$
\begin{aligned}
f(\theta|x_1,\ldots,x_N) &\propto f(x_1,\ldots,x_N|\theta) f(\theta) && (13)\\
&\propto \left(\prod_{i=1}^N \exp\left[-\frac{(x_i-\theta)^2}{2\sigma^2}\right]\right) \exp\left[-\frac{(\theta-m)^2}{2v}\right] && (14)\\
&= \exp\left[-\frac{\sum_{i=1}^N (x_i-\theta)^2}{2\sigma^2} - \frac{(\theta-m)^2}{2v}\right] && (15)\\
&\propto \exp\left[-\frac{n\theta^2 - 2n\bar{x}\theta}{2\sigma^2} - \frac{\theta^2-2m\theta}{2v}\right] && (16)\\
&= \exp\left[-\frac{\theta^2 - 2\bar{x}\theta}{\frac{2\sigma^2}{n}} - \frac{\theta^2-2m\theta}{2v}\right] && (17)\\
&= \exp\left[-\frac{1}{2\frac{\sigma^2 v}{n}}\left(v\theta^2 - 2v\bar{x}\theta + \frac{\sigma^2}{n}\theta^2 - 2\frac{\sigma^2}{n}m\theta\right)\right] && (18)\\
&= \exp\left[-\frac{1}{2\frac{\sigma^2 v}{n}}\left(\left[v+\frac{\sigma^2}{n}\right]\theta^2 - 2\left[v\bar{x} + \frac{\sigma^2}{n}m\right]\theta\right)\right] && (19)\\
&= \exp\left[-\frac{1}{2\frac{\sigma^2 v}{n}\left(\frac{1}{v+\frac{\sigma^2}{n}}\right)}\left(\theta^2 - 2\frac{v\bar{x}+\frac{\sigma^2}{n}m}{v+\frac{\sigma^2}{n}}\theta\right)\right] && (20)\\
&\propto \exp\left[-\frac{1}{2\left(\frac{n}{\sigma^2}+\frac{1}{v}\right)^{-1}}\left(\theta - \frac{v\bar{x}+\frac{\sigma^2}{n}m}{v+\frac{\sigma^2}{n}}\right)^2\right] && (21)\\
&= \exp\left[-\frac{1}{2\left(\frac{n}{\sigma^2}+\frac{1}{v}\right)^{-1}}\left(\theta - \frac{\frac{\sum_{i=1}^N x_i}{\sigma^2}+\frac{m}{v}}{\frac{n}{\sigma^2}+\frac{1}{v}}\right)^2\right] && (22)\\
&= \exp\left[-\frac{1}{2\left(\frac{1}{v}+\frac{n}{\sigma^2}\right)^{-1}}\left(\theta - \frac{\frac{m}{v}+\frac{\sum_{i=1}^N x_i}{\sigma^2}}{\frac{1}{v}+\frac{n}{\sigma^2}}\right)^2\right], && (23)
\end{aligned}
$$

so

$$\theta|x_1,\ldots,x_2 \sim \mathcal{N}\left(\frac{\frac{m}{v}+\frac{\sum_{i=1}^N x_i}{\sigma^2}}{\frac{1}{v}+\frac{n}{\sigma^2}}, \left[\frac{1}{v}+\frac{n}{\sigma^2}\right]^{-1}\right). \tag{24}$$

(D) $X_i|\sigma^2 \overset{\text{iid}}{\sim} \mathcal{N}(\theta, \sigma^2)$, $i = 1, 2, \ldots, N$ where θ is *known* and $\sigma^2 \sim \text{IG}(a, b)$ is *unknown*. Let $w = \sigma^{-2}$ so

Problem 1 continued on next page…

$w \sim \text{Gamma}(a, b)$. The posterior distribution of w given $x_1, \ldots, x_N$ is

$$
\begin{aligned}
f(w|x_1,\ldots,x_2) &\propto f(x_1,\ldots,x_2|w) f(w) && (25)\\
&\propto \left(\prod_{i=1}^{N} w^{1/2} \exp\left[-\frac{w}{2}(x_i - \theta)^2 \right] \right) w^{a-1} \exp(-bw) && (26)\\
&= w^{n/2} \exp\left[-\frac{w}{2} \sum_{i=1}^{N} (x_i - \theta^2) \right] w^{a-1} \exp(-bw) && (27)\\
&= w^{a+n/2-1} \exp\left[-\left(b + \frac{\sum_{i=1}^{N}(x_i - \theta^2)}{2} \right) w \right], && (28)
\end{aligned}
$$

so

$$
\begin{aligned}
w|x_1,\ldots,x_2 &\sim \text{Gamma}\left(a + \frac{n}{2}, b + \frac{\sum_{i=1}^{N}(x_i - \theta^2)}{2} \right) && (29)\\
\sigma^2|x_1,\ldots,x_2 &\sim \text{IG}\left(a + \frac{n}{2}, b + \frac{\sum_{i=1}^{N}(x_i - \theta^2)}{2} \right) && (30)
\end{aligned}
$$

(E) $X_i \sim \mathcal{N}(\theta, \sigma_i^2)$, $i = 1, 2, \ldots, n$ where each $X_i \perp\!\!\!\perp X_j, i \neq j$ is observed once and has a *known* unique variance σ_i^2 and $\theta \sim \mathcal{N}(m, v)$ is *unknown*. The posterior distribution of θ is

$$
\begin{aligned}
f(\theta|x_1,\ldots,x_N) &\propto f(x_1,\ldots,x_N|\theta) f(\theta) && (31)\\
&\propto \left(\prod_{i=1}^{N} \exp\left[-\frac{(x_i - \theta)^2}{2\sigma_i^2} \right] \right) \exp\left[-\frac{(\theta - m)^2}{2v} \right] && (32)\\
&= \exp\left[-\frac{1}{2} \left(\sum_{i=1}^{n} \frac{(\theta - x_i)^2}{\sigma_i^2} + \frac{(\theta - m)^2}{v} \right) \right] && (33)\\
&\propto \exp\left[-\frac{1}{2} \left(\sum_{i=1}^{N} \frac{1}{\sigma_i^2} \cdot \theta^2 - 2 \sum_{i=1}^{N} \frac{x_i}{\sigma_i^2} \cdot \theta + \frac{1}{v}\theta^2 - 2\frac{m}{v}\theta \right) \right] && (34)\\
&= \exp\left[-\frac{1}{2} \left(\left[\frac{1}{v} + \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \right] \theta^2 - 2 \left[\frac{m}{v} + \sum_{i=1}^{N} \frac{x_i}{\sigma_i^2} \right] \theta \right) \right] && (35)\\
&= \exp\left[-\frac{1}{2\left(\frac{1}{v} + \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \right)^{-1}} \left(\theta^2 - 2 \left[\frac{\frac{m}{v} + \sum_{i=1}^{N} \frac{x_i}{\sigma_i^2}}{\frac{1}{v} + \sum_{i=1}^{N} \frac{1}{\sigma_i^2}} \right] \theta \right) \right] && (36)\\
&\propto \exp\left[-\frac{1}{2\left(\frac{1}{v} + \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \right)^{-1}} \left(\theta - \frac{\frac{m}{v} + \sum_{i=1}^{N} \frac{x_i}{\sigma_i^2}}{\frac{1}{v} + \sum_{i=1}^{N} \frac{1}{\sigma_i^2}} \right)^2 \right], && (37)
\end{aligned}
$$

so,

$$
\theta|x_1,\ldots,x_N \sim \mathcal{N}\left(\frac{\frac{m}{v} + \sum_{i=1}^{N} \frac{x_i}{\sigma_i^2}}{\frac{1}{v} + \sum_{i=1}^{N} \frac{1}{\sigma_i^2}}, \left(\frac{1}{v} + \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \right)^{-1} \right). \quad (38)
$$

Problem 1 continued on next page. . .

(F) $X|\sigma^2 \sim \mathcal{N}(0,\sigma^2)$, $w = \frac{1}{\sigma^2} \sim \text{Gamma}(a,b)$. The marginal distribution of X is

$$\begin{aligned}
f(x) &= \int_0^\infty f(x,w)dw && (39)\\
&= \int_0^\infty f(x|w)f(w)dw && (40)\\
&\propto \int_0^\infty w^{1/2} \exp\left(-\frac{w}{2}x^2\right) w^{a-1} \exp\left(-bw\right) dw && (41)\\
&= \int_0^\infty w^{a-1/2} \exp\left[-\left(b+\frac{x^2}{2}\right) w\right] dw \quad * \text{kernel of Gamma}\left(a+\frac{1}{2}, b+\frac{x^2}{2}\right) && (42)\\
&= \frac{\Gamma\left(a+\frac{1}{2}\right)}{\left(b+\frac{x^2}{2}\right)^{a+1/2}} && (43)
\end{aligned}$$

Problem 2

The multivariate normal distribution

Basics

(A) Here we prove two properties of the covariance of a vector of random variables. First, note that $E(Ax+b) = A\mu + b$.

1.

$$\begin{aligned}
\text{cov}(x) &= E\left((x-\mu)(x-\mu)^T\right) && (44)\\
&= E\left((x-\mu)(x^T-\mu^T)\right) && (45)\\
&= E\left(xx^T - x\mu^T - \mu x^T + \mu\mu^T\right) && (46)\\
&= E(xx^T) - E(x)\mu^T - \mu E(x^T) + \mu\mu^T && (47)\\
&= E(xx^T) - \mu\mu^T - \mu\mu^T + \mu\mu^T && (48)\\
&= E(xx^T) - \mu\mu^T && (49)
\end{aligned}$$

2.

$$\begin{aligned}
\text{cov}(Ax+b) &= E\left((Ax+b-(A\mu+b))(Ax+b-(A\mu+b))^T\right) && (50)\\
&= E\left((Ax-A\mu)(Ax-A\mu)^T\right) && (51)\\
&= E\left((Ax-A\mu)\left(x^TA^T-\mu^TA^T\right)\right) && (52)\\
&= E\left(Axx^TA - Ax\mu^TA^T - A\mu x^TA^T + A\mu\mu^TA^T\right) && (53)\\
&= E\left(Axx^TA^T\right) - E\left(Ax\mu^TA^T\right) - E\left(A\mu x^TA^T\right) + \left(A\mu\mu^TA^T\right) && (54)\\
&= AE\left(xx^T\right)A^T - A\mu\mu^TA^T - A\mu\mu^TA^T + A\mu\mu^TA^T && (55)\\
&= AE\left(xx^T\right)A^T - A\mu\mu^TA^T && (56)\\
&= A\left(E\left(xx^T\right) - \mu\mu^T\right)A^T && (57)\\
&= A\text{cov}(x)A^T && (58)
\end{aligned}$$

Problem 2 continued on next page. . .

(B) Define the vector $z = (z_1, \ldots, z_p)$ where $z_i \overset{\text{iid}}{\sim} \mathcal{N}(0,1), i = 1, 2, \ldots, p$. Because each component is iid, the joint probability density function (PDF) for z is

$$f(z) = \prod_{i=1}^{p} (2\pi)^{-1/2} \exp\left(-z_i^2/2\right) \tag{59}$$
$$= (2\pi)^{-p/2} \exp\left(-z^T z/2\right). \tag{60}$$

For each component z_i, the moment generating function (MGF) is

$$M_{z_i}(t_i) = E\left(\exp\left(t_i z_i\right)\right) \tag{61}$$
$$= \int_{-\infty}^{+\infty} \exp\left(t_i z_i\right) \cdot f(z_i) dz_i \tag{62}$$
$$= \int_{-\infty}^{+\infty} \exp\left(t_i z_i\right) \cdot (2\pi)^{-1/2} \exp\left(-z_i^2/2\right) dz_i \tag{63}$$
$$= \int_{-\infty}^{+\infty} (2\pi)^{-1/2} \exp\left(-z_i^2/2 + t_i z_i\right) dz_i \tag{64}$$
$$= \exp\left(t_i^2/2\right) \int_{-\infty}^{+\infty} (2\pi)^{-1/2} \exp\left(-\left[z_i - t_i\right]^2/2\right) dz_i \tag{65}$$
$$= \exp\left(t_i^2/2\right). \tag{66}$$

The MGF for the full vector z is

$$M_z(t) = E\left(\exp\left(t^T z\right)\right) \tag{67}$$
$$= \prod_{i=1}^{p} E\left(\exp\left(t_i z_i\right)\right) \tag{68}$$
$$= \prod_{i=1}^{p} \exp\left(t_i^2/2\right) \tag{69}$$
$$= \exp\left(\sum_{i=1}^{p} t_i^2/2\right) \tag{70}$$
$$= \exp\left(t^T t/2\right) \tag{71}$$

(C) We are trying to show that $x = (x_1, \ldots, x_p)$ is a multivariate normal distribution with mean vector μ and covariance matrix Σ. Let $a^T x = z \sim \mathcal{N}(m, v)$ (z is now a scalar random variable). Let t be a scalar, a is a vector of length p, and $b = ta$ is also a vector of length p. The MGF of z is

$$M_z(t) = E\left(\exp\left(tz\right)\right) \tag{72}$$
$$= E\left(\exp\left(ta^T x\right)\right) \tag{73}$$
$$= E\left(\exp\left(bx\right)\right) \tag{74}$$
$$= M_x(b) = \exp\left(mt + vt^2/2\right), \tag{75}$$

by the MGF definition of the univariate normal distribution. We can solve for m and v in terms of μ and Σ by using the first and second moments of z. The first moment of z is equal to $E(z) = m$, and can also be expressed as

$$E(z) = E\left(a^T x\right) \tag{76}$$
$$= a^T E(x) \tag{77}$$
$$= a^T \mu = m. \tag{78}$$

Note that $m^2 = (a^T\mu)^2 = a^T\mu\mu^T a$. Next, the second moment of z is equal to $E(z^2) = \text{var}(z) + E(z)^2 = v + m^2$, which can also be expressed as

$$\begin{aligned}
E(z^2) &= E\left(z \cdot z\right) && (79)\\
&= E\left(a^T x x^T a\right) && (80)\\
&= a^T E(xx^T)a && (81)\\
&= a^T(\text{cov}(x) + \mu^T\mu)a && (82)\\
&= a^T(\Sigma + \mu^T\mu)a && (83)\\
&= a^T\Sigma a^+ a^T\mu^T\mu a = v + m^2 = v + a^T\mu\mu^T a && (84)\\
&\Rightarrow v = a^T\Sigma a && (85)
\end{aligned}$$

Now we return to the (75) to write the MGF of x as

$$\begin{aligned}
M_x(b) &= \exp\left(mt + vt^2/2\right) && (86)\\
&= \exp\left(ta^T\mu + t^2 a^T\Sigma a^T/2\right) && (87)\\
&= \exp\left(ta^T\mu + (ta^T)\Sigma(ta)/2\right) && (88)\\
&= \exp\left(b^T\mu + b^T\Sigma b/2\right) && (89)\\
&\text{Q.E.D.} && (90)
\end{aligned}$$

(D) The p-length vector $z \sim \mathcal{N}_p(0, I_p)$ follows the standard multivariate normal distribution. We will full prove that the vector $x = Lz + \mu$, where L is a $p \times p$ matrix of full column rank, is multivariate normal. The MGF of x is,

$$\begin{aligned}
M_x(t) &= E\left(\exp\left(t^T x\right)\right) && (91)\\
&= E\left(\exp\left(t^T(Lz + \mu)\right)\right) && (92)\\
&= E\left(\exp(t^T Lz + t^T\mu)\right) && (93)\\
&= \exp\left(t^T\mu\right) E\left(\exp(t^T Lz)\right) && (94)\\
&= \exp\left(t^T\mu\right) M_z(t^T L) && (95)\\
&= \exp\left(t^T\mu\right) \exp\left[\frac{1}{2}\left(t^T L\right) I_p \left(t^T L\right)^T\right] && (96)\\
&= \exp\left(t^T\mu + t^T LL^T t/2\right). && (97)
\end{aligned}$$

Therefore x follows a multivariate normal distribution with mean vector μ and covariance matrix LL^T, $x \sim \mathcal{N}_p\left(\mu, LL^T\right)$.

(E) By definiton, $x = Lz + \mu$ is an affine transformation of a vector of standard normal random variables, z. To generate random numbers from $x \sim \mathcal{N}_p(\mu, \Sigma)$, first perform the Cholesky decomposition of Σ to obtain a lower triangle matrix L such that $\Sigma = LL^T$, generate p iid scalar normal random numbers to make the z vector, and finally compute $x = Lz + \mu$.

(F) Before we begin, let us first show that

$$\det\left(L^{-1}\right) = \left[\det\left(\Sigma\right)\right]^{-1/2}.$$

This may be shown by

$$\begin{aligned}
L^{-1}L &= I \\
\det\left(L^{-1}L\right) &= \det(I) \\
\det\left(L^{-1}\right)\det(L) &= 1 \\
\det\left(L^{-1}\right) &= [\det(L)]^{-1}
\end{aligned}$$

and

$$\begin{aligned}
\Sigma &= LL^T \\
\det(\Sigma) &= \det\left(LL^T\right) \\
\det(\Sigma) &= \det(L)\det\left(L^T\right) \\
\det(\Sigma) &= \det(L)^2 \\
[\det(\Sigma)]^{1/2} &= \det(L) \\
[\det(\Sigma)]^{-1/2} &= [\det(L)]^{-1} \\
[\det(\Sigma)]^{-1/2} &= \det\left(L^{-1}\right).
\end{aligned}$$

Now we can derive the PDF of the multivariate normal $x \sim \mathcal{N}(\mu, \Sigma)$. Define the transformation $f : z \mapsto x$, $x = f(z) = Lz + \mu$, and its inverse transformation, $f^{-1} = g : x \mapsto z$, $z = g(x) = L^{-1}(x - \mu)$, where z follows the standard multivariate distribution. The PDF of x is

$$f_x(x) = f_z(g(x)) \cdot |J(y)|, \tag{98}$$

where $J(y)$ is the Jacobian determinant of the transformation g, which in this case is just $\det\left(L^{-1/2}\right)$,

$$\begin{aligned}
f_x(x) &= (2\pi)^{-p/2}\exp\left(-\frac{1}{2}\left[L^{-1}(x-\mu)\right]^T\left[L^{-1}(x-\mu)\right]\right)\left|\det\left(L^{-1}\right)\right| && (99) \\
&= (2\pi)^{-p/2}\exp\left(-\frac{1}{2}(x-\mu)^T\left(L^{-1}\right)^T L^{-1}(x-\mu)\right)[\det(\Sigma)]^{-1/2} && (100) \\
&= (2\pi)^{-p/2}[\det(\Sigma)]^{-1/2}\exp\left(-\frac{1}{2}(x-\mu)^T\left(L^T\right)^{-1}L^{-1}(x-\mu)\right) && (101) \\
&= (2\pi)^{-p/2}[\det(\Sigma)]^{-1/2}\exp\left(-\frac{1}{2}(x-\mu)^T\left(LL^T\right)^{-1}(x-\mu)\right) && (102) \\
&= (2\pi)^{-p/2}[\det(\Sigma)]^{-1/2}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right) && (103)
\end{aligned}$$

(G) Let $x_1 \sim \mathcal{N}(\mu_1, \Sigma_1)$ independent of $x_2 \sim \mathcal{N}(\mu_2, \Sigma_2)$, and define $y = Ax_1 + Bx_2$. The MGFs of x_1 and x_2 are, respectively,

$$\begin{aligned}
M_{x_1}(s) &= E\left(\exp\left[s^T x_1\right]\right) = \exp\left(s^T\mu_1 + s^T\Sigma_1 s/2\right) && (104) \\
M_{x_2}(s) &= E\left(\exp\left[s^T x_2\right]\right) = \exp\left(s^T\mu_2 + s^T\Sigma_2 s/2\right). && (105)
\end{aligned}$$

We will characterize y by its MGF,

$$\begin{aligned}
M_y(t) &= E\left(\exp\left[t^T y\right]\right) && (106)\\
&= E\left(\exp\left[t^T\left(Ax_1 + Bx_2\right)\right]\right) && (107)\\
&= E\left(\exp\left[t^T Ax_1 + t^T Bx_2\right]\right) && (108)\\
&= E\left(\exp\left[t^T Ax_1\right]\exp\left[t^T Bx_2\right]\right) && (109)\\
&= E\left(\exp\left[t^T Ax_1\right]\right)E\left(\exp\left[t^T Bx_2\right]\right) \Leftarrow x_1 \perp\!\!\!\perp x_2 && (110)\\
&= M_{x_1}(A^T t)M_{x_2}(B^T t) && (111)\\
&= \exp\left(t^T A\mu_1 + t^T A\Sigma_1 A^T t/2\right)\exp\left(t^T B\mu_2 + t^T B\Sigma_2 B^T t/2\right) && (112)\\
&= \exp\left(t^T A\mu_1 + t^T A\Sigma_1 A^T t/2 + t^T B\mu_2 + t^T B\Sigma_2 B^T t/2\right) && (113)\\
&= \exp\left(t^T(A\mu_1 + B\mu_2) + t^T(A\Sigma_1 A^T + B\Sigma_2 B^T)t/2\right). && (114)
\end{aligned}$$

Therefore, $y \sim \mathcal{N}(A\mu_1 + B\mu_2, A\Sigma_1 A^T + B\Sigma_2 B^T)$.

R code is shown on the following page.

```
##########################################################
######### Created by Spencer Woody on 22 Jan 2017 #########
##########################################################
```