

# Statistical Modeling 2

## Exercise 1

January 24, 2017

### 1 Bayesian inference in simple conjugate families

#### A

$$\begin{aligned}
 p(w \mid x_1, \dots, x_N) &\propto p(x_1, \dots, x_N \mid w)p(w) && \text{(Bayes rule)} \\
 &\propto \prod_{i=1}^N p(x_i \mid w) w^{a-1} (1-w)^{b-1} && \text{(independence)} \\
 &\propto w^s (1-w)^{N-s} w^{a-1} (1-w)^{b-1} && \text{(let } s = \sum_{i=1}^N x_i) \\
 &= w^{s+a-1} (1-w)^{N-s+b-1} \\
 &\propto \text{Beta}(s+a, N-s+b)
 \end{aligned}$$

#### B

Let  $f(x_1, x_2) = (y_1, y_2) = (x_1/(x_1 + x_2), x_1 + x_2)$ , we have:

$$f^{-1}(y_1, y_2) = (x_1, x_2) = (y_1 y_2, y_2 - y_1 y_2)$$

We then calculate the Jacobian of  $f^{-1}$ :

$$\begin{aligned}
 \partial x_1 / \partial y_1 &= y_2 \\
 \partial x_1 / \partial y_2 &= y_1 \\
 \partial x_2 / \partial y_1 &= -y_2 \\
 \partial x_2 / \partial y_2 &= 1 - y_1
 \end{aligned}$$

Therefore,

$$\begin{aligned}
 |J(f^{-1})| &= \begin{vmatrix} y_2 & y_1 \\ -y_2 & 1 - y_1 \end{vmatrix} \\
 &= y_2(1 - y_1) + y_1 y_2 \\
 &= y_2
 \end{aligned}$$

Let  $p_X$  be the joint density of  $(x_1, x_2)$ . We have the joint density of  $y_1$  and  $y_2$ :

$$\begin{aligned}
p(y_1, y_2) &= p_X(f^{-1}(y_1, y_2)) |J(f^{-1}(y_1, y_2))| \\
&= \text{Ga}(y_1 y_2; a_1, 1) \text{Ga}(y_2 - y_1 y_2; a_2, 1) y_2 \\
&= \frac{(y_1 y_2)^{a_1-1} \exp(-y_1 y_2)}{\Gamma(a_1)} \frac{((1 - y_1) y_2)^{a_2-1} \exp(y_1 y_2 - y_2)}{\Gamma(a_2)} y_2 \\
&= \frac{y_1^{a_1-1} y_2^{a_1+a_2-1} (1 - y_1)^{a_2-1} \exp(-y_2)}{\Gamma(a_1) \Gamma(a_2)}
\end{aligned}$$

The marginals are:

$$\begin{aligned}
p(y_1) &= \int_0^\infty p(y_1, y_2) dy_2 \\
&= \frac{y_1^{a_1-1} (1 - y_1)^{a_2-1}}{\Gamma(a_1) \Gamma(a_2)} \int_0^\infty y_2^{a_1+a_2-1} \exp(-y_2) dy_2 \\
&= \frac{y_1^{a_1-1} (1 - y_1)^{a_2-1} \Gamma(a_1 + a_2)}{\Gamma(a_1) \Gamma(a_2)}
\end{aligned}$$

and

$$\begin{aligned}
p(y_2) &= \int_0^\infty p(y_1, y_2) dy_1 \\
&= \frac{y_2^{a_1+a_2-1} \exp(-y_2)}{\Gamma(a_1) \Gamma(a_2)} \int_0^\infty y_1^{a_1-1} (1 - y_1)^{a_2-1} dy_1 \\
&= \frac{y_2^{a_1+a_2-1} \exp(-y_2)}{\Gamma(a_1) \Gamma(a_2)} \text{Beta}(a_1, a_2)
\end{aligned}$$

We can simulate a Beta random variable by taking two Gamma random variable as  $x_1$  and  $x_2$  and evaluate  $y_1$ .

## C

The posterior is:

$$\begin{aligned}
p(\theta|x_1, \dots, x_N) &\propto p(x_1, \dots, x_N|\theta)p(\theta) \\
&= \prod_{i=1}^N N(x_i; \theta, \sigma^2) N(\theta; m, v) \\
&\propto \prod_{i=1}^N \exp\left(-\frac{(x_i - \theta)^2}{2\sigma^2}\right) \exp\left(-\frac{(\theta - m)^2}{2v}\right) \\
&= \prod_{i=1}^N \exp\left(-\frac{x_i^2 - 2x_i\theta + \theta^2}{2\sigma^2}\right) \exp\left(-\frac{\theta^2 - 2\theta m + m^2}{2v}\right) \\
&= \exp\left(\frac{-\sum_i x_i^2 + 2\sum_i x_i\theta - N\theta^2}{2\sigma^2}\right) \exp\left(\frac{-\theta^2 + 2\theta m - m^2}{2v}\right) \\
&= \exp\left(-\theta^2\left(\frac{N}{2\sigma^2} + \frac{1}{2v}\right) + \theta\left(\frac{\sum_i x_i}{\sigma^2} + \frac{m}{v}\right) - \frac{\sum_i x_i^2}{2\sigma^2} - \frac{m^2}{2v}\right)
\end{aligned}$$

We then complete the square by setting the posterior to:

$$\begin{aligned}
&= \exp\left[-a(\theta^2 - 2b\theta + b^2)\right] \\
&= \exp\left[-a(\theta - b)^2\right] \\
&= \exp\left[-\frac{(\theta - b)^2}{2(1/(2a))}\right]
\end{aligned}$$

We calculate  $a, b$  by matching coefficients:

$$\begin{aligned}
a &= \frac{N}{2\sigma^2} + \frac{1}{2v} = \frac{Nv + \sigma^2}{2\sigma^2 v} \\
2ab &= \frac{\sum_i x_i}{\sigma^2} + \frac{m}{v} \\
\implies b &= \frac{v \sum_i x_i + m\sigma^2}{v\sigma^2} \frac{\sigma^2 v}{Nv + \sigma^2} \\
&= \frac{v \sum_i x_i + m\sigma^2}{Nv + \sigma^2}
\end{aligned}$$

The posterior is then:

$$\begin{aligned}
&N(b, 1/(2a)) \\
&= N\left(\frac{v \sum_i x_i + m\sigma^2}{Nv + \sigma^2}, \frac{\sigma^2 v}{Nv + \sigma^2}\right)
\end{aligned}$$

## D

$$\begin{aligned}
p(\omega \mid x_1, \dots, x_N) &\propto \prod_{i=1}^N p(x_i \mid \theta, \omega) p(\omega) \\
&\propto \prod_{i=1}^N \omega^{1/2} \exp \left[ -\frac{\omega}{2} (x_i - \theta)^2 \right] \frac{b^a}{\Gamma(a)} \omega^{a-1} \exp(-b\omega) \\
&\propto \omega^{N/2+a-1} \exp \left[ -\omega \left( b + \frac{\sum_i (x_i - \theta)^2}{2} \right) \right] \\
&\propto \text{Ga} \left( a + \frac{N}{2}, b + \frac{\sum_i (x_i - \theta)^2}{2} \right)
\end{aligned}$$

We have the posterior of the variance:

$$p(\sigma^2 \mid x_1, \dots, x_N) = \text{IG} \left( a + \frac{N}{2}, b + \frac{\sum_i (x_i - \theta)^2}{2} \right)$$

## E

The posterior is:

$$\begin{aligned}
p(\theta \mid x_1, \dots, x_N) &\propto p(x_1, \dots, x_N \mid \theta) p(\theta) \\
&= \prod_{i=1}^N \text{N}(x_i; \theta, \sigma_i^2) \text{N}(\theta; m, v) \\
&\propto \prod_{i=1}^N \exp \left( -\frac{(x_i - \theta)^2}{2\sigma_i^2} \right) \exp \left( -\frac{(\theta - m)^2}{2v} \right) \\
&= \exp \left( -\sum_{i=1}^n \frac{(x_i - \theta)^2}{2\sigma_i^2} - \frac{(\theta - m)^2}{2v} \right) \\
&= \exp \left[ -\frac{1}{2} \left( \sum_{i=1}^n \frac{(x_i - \theta)^2}{\sigma_i^2} + \frac{(\theta - m)^2}{v} \right) \right] \\
&= \exp \left[ -\frac{1}{2} \left( \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} + \sum_{i=1}^n \frac{-2x_i\theta}{\sigma_i^2} + \sum_{i=1}^n \frac{\theta^2}{\sigma_i^2} + \frac{\theta^2 - 2\theta m + m^2}{v} \right) \right] \\
&= \exp \left\{ -\frac{1}{2} \left[ \theta^2 \left( \sum_{i=1}^n \frac{1}{\sigma_i^2} + \frac{1}{v} \right) - 2\theta \left( \sum_{i=1}^n \frac{x_i}{\sigma_i^2} + \frac{m}{v} \right) + \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} + \frac{m^2}{v} \right] \right\} \\
\text{set } &= \exp \left\{ -\frac{1}{2} [a(\theta^2 - 2\theta b + b^2)] \right\} \\
&= \exp \left\{ -\frac{1}{2} \left[ \frac{(\theta - b)^2}{1/a} \right] \right\}
\end{aligned}$$

Matching the coefficients, we have:

$$a = \sum_{i=1}^n \frac{1}{\sigma_i^2} + \frac{1}{v}$$

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

The posterior is:

$$N(b, 1/a)$$

## F

$$\begin{aligned} p(x) &= \int_0^\infty p(x | \sigma^2) p(\sigma^2) d\sigma^2 \\ &= \int_0^\infty p(x | \omega) p(\omega) d\omega \\ &= \int_0^\infty \left( \frac{\omega}{2\pi} \right)^{1/2} \exp\left(-\frac{\omega}{2} x^2\right) \frac{b^a}{\Gamma(a)} \omega^{a-1} \exp(-b\omega) d\omega \\ &= \frac{b^a}{(2\pi)^{1/2} \Gamma(a)} \int_0^\infty \omega^{1/2+a-1} \exp\left(-\omega \left(\frac{x^2}{2} + b\right)\right) d\omega \\ &= \frac{b^a}{(2\pi)^{1/2} \Gamma(a)} \frac{\Gamma(a+1/2)}{(b + x^2/2)^{a+1/2}} \quad (\text{Gamma integral}) \\ &= \frac{\Gamma(a+1/2)}{(2\pi b)^{1/2} \Gamma(a) (1 + \frac{x^2}{2b})^{a+1/2}} \end{aligned}$$

Let  $\nu = 2a$  and  $\lambda = a/b$ , we have:

$$p(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\nu/2)} \left( \frac{\lambda}{\pi\nu} \right)^{1/2} \left( 1 + \frac{\lambda x^2}{2} \right)^{-\frac{\nu+1}{2}}$$

This is the Student t distribution with  $\nu$  degree of freedom and ‘precision’  $\lambda$ .

## 2 The multivariate normal distribution

### A

$$\begin{aligned} \text{cov}(x) &= E\{(x - \mu)(x - \mu)^T\} \\ &= E\{xx^T - x\mu^T - \mu x^T + \mu\mu^T\} \\ &= E(xx^T) - E(x)\mu^T - \mu E(x)^T + \mu\mu^T \\ &= E(xx^T) - \mu\mu^T \end{aligned}$$

We have:

$$E(Ax + b) = AE(x) + b = A\mu + b$$

then

$$\begin{aligned} \text{cov}(Ax + b) &= E\{[(Ax + b) - (A\mu + b)][(Ax + b) - (A\mu + b)]^T\} \\ &= E\{(Ax - A\mu)(Ax - A\mu)^T\} \\ &= E\{A(x - \mu)(x - \mu)^T A^T\} \\ &= AE\{(x - \mu)(x - \mu)^T\}A^T \\ &= A\text{cov}(x)A^T \end{aligned}$$

## B

$$\begin{aligned} p(z) &= \prod_{i=1}^p p(z_i) \\ &= \frac{1}{(\sqrt{2\pi})^p} \exp\left(-\sum_{i=1}^p \frac{z_i^2}{2}\right) \\ &= \frac{1}{(\sqrt{2\pi})^p} \exp\left(-\frac{z^T z}{2}\right) \end{aligned}$$

The MGF of  $z$  is:

$$\begin{aligned} E(\exp(t^T z)) &= E\left[\exp\left(\sum_{i=1}^p t_i z_i\right)\right] \\ &= E\left[\prod_{i=1}^p \exp(t_i z_i)\right] \\ &= \prod_{i=1}^p E[\exp(t_i z_i)] \\ &= \prod_{i=1}^p \exp(t_i^2/2) \\ &= \exp\left[\sum_{i=1}^p t_i^2/2\right] \\ &= \exp(t^T t/2) \end{aligned}$$

## C

We need to prove that for all vector  $a$  not identically zero, the scalar quantity  $z = a^T x$  is normally distributed if and only if

$$E[\exp(t^T x)] = \exp(t^T \mu + t^T \Sigma t/2)$$

**(only if)** We have that  $z = a^T x$  is normally distributed:

$$\text{MGF}_z(s) = E[\exp(sa^T x)] = \exp(ms + vs^2/2)$$

Consider:

**D**

We have  $z \sim N(0, I)$  and  $x = Lz + \mu$ .

The MGF of  $x$  is:

$$E(\exp(t^T x)) = E[\exp(t^T Lz + t^T \mu)]$$

The expectation is with respect to  $z$ ,  $t^T \mu$  is a constant, we then look at:

$$\begin{aligned} E[\exp(t^T Lz)] &= E \left[ \exp \left( \sum_{i=1}^p \sum_{j=1}^p t_i L_{ij} z_j \right) \right] \\ &= E \left[ \prod_{j=1}^p \exp \left( \sum_{i=1}^p t_i L_{ij} z_j \right) \right] \\ &= \prod_{j=1}^p E \left[ \exp \left( \sum_{i=1}^p t_i L_{ij} z_j \right) \right] \quad (\text{independence}) \\ &= \prod_{j=1}^p \text{MGF}_{z_j} \left( \sum_{i=1}^p t_i L_{ij} \right) \\ &= \prod_{j=1}^p \exp \left( \frac{1}{2} (t^T L_j)^2 \right) \\ &= \prod_{j=1}^p \exp \left( \frac{1}{2} t^T L_j L_j^T t \right) \\ &= \exp \left( \frac{1}{2} \sum_{j=1}^p t^T L_j L_j^T t \right) \\ &= \exp \left( \frac{1}{2} t^T L L^T t \right) \end{aligned}$$

Come back to the MGF of  $x$ :

$$E(\exp(t^T x)) = \exp \left( t^T \mu + \frac{1}{2} t^T L L^T t \right)$$

Therefore,  $x \sim N(\mu, L L^T)$ .

## E

We have that  $x$  has a multivariate normal distribution:  $x \sim N(\mu, \Sigma)$ . The covariance matrix  $\Sigma$  is symmetric positive definite and has a Cholesky decomposition:

$$\Sigma = LL^T$$

where  $L$  is a lower triangular matrix with positive diagonal entries and therefore invertible. Let

$$z = L^{-1}(x - \mu)$$

Consider the MGF of  $z$ :

$$\begin{aligned} E[\exp(t^T z)] &= E[\exp(t^T L^{-1}(x - \mu))] \\ &= E[\exp(t^T L^{-1}x) \cdot \exp(-t^T L^{-1}\mu)] \\ &= \text{MGF}_x(t^T L^{-1}) \cdot \exp(-t^T L^{-1}\mu) \\ &= \exp(t^T L^{-1}\mu + t^T L^{-1}\Sigma L^{-T}t/2) \cdot \exp(-t^T L^{-1}\mu) \\ &= \exp(t^T L^{-1}\Sigma L^{-T}t/2) \\ &= \exp(t^T L^{-1}(LL^T)L^{-T}t/2) \\ &= \exp(t^T t/2) \end{aligned}$$

We conclude that  $z$  has standard multivariate normal distribution and that  $x$  can be written as an affine transformation of standard normal distribution.