SDS 383D: Homework 1

Giorgio Paulon

January 22, 2017

Problem 1. Bayesian inference in simple conjugate families

We start with a few of the simplest building blocks for complex multivariate statistical models: the beta/binomial, normal, and inverse- gamma conjugate families.

(A) *Suppose that we take independent observations $X_1, \dots, X_N$ from a Bernoulli sampling model with unknown probability w. That is, the X_i are the results of flipping a coin with unknown bias. Suppose that w is given a Beta(a, b) prior distribution:*

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

where $\Gamma(\cdot)$ denotes the Gamma function. Derive the posterior distribution $p(w|x_1, \dots, x_N)$.

The following model

$$\begin{aligned} &X_1, \dots, X_n | w \overset{\text{iid}}{\sim} Bernoulli(w) \\ &w \sim Beta(a, b) \end{aligned}$$

leads to the following posterior distribution:

$$\begin{aligned} p(w|\boldsymbol{x}) &\propto p(\boldsymbol{X}|w)p(w) \\ &= \prod_{i=1}^{N} p(X_i|w)p(w) \\ &\propto \prod_{i=1}^{N} \left\{ w^{x_i} (1-w)^{1-x_i} \right\} w^{a-1} (1-w)^{b-1} \mathcal{I}_{[0,1]}(w) \\ &= w^{\sum_{i=1}^{N} x_i} (1-w)^{n - \sum_{i=1}^{N} x_i} w^{a-1} (1-w)^{b-1} \mathcal{I}_{[0,1]}(w) \\ &= w^{a + \sum_{i=1}^{N} x_i - 1} (1-w)^{b + (n - \sum_{i=1}^{N} x_i) - 1} \mathcal{I}_{[0,1]}(w). \end{aligned}$$

Therefore,

$$p(w|\boldsymbol{x}) = \frac{\Gamma(a+b+n)}{\Gamma(a + \sum_{i=1}^{N} x_i)\Gamma(b + (n - \sum_{i=1}^{N} x_i))} w^{a + \sum_{i=1}^{N} x_i - 1} (1-w)^{b + (n - \sum_{i=1}^{N} x_i) - 1} \mathcal{I}_{[0,1]}(w)$$

that is,

$$W|\boldsymbol{x} \sim Beta\left(a + \sum_{i=1}^{N} x_i, b + (n - \sum_{i=1}^{N} x_i) \right).$$

This is the so-called **Bernoulli-Beta model**.

(B) *The probability density function (PDF) of a gamma random variable, $X \sim Ga(a, b)$, is*

$$p(x) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}.$$

Suppose that $X_1 \sim Ga(a_1, 1)$ and $X_2 \sim Ga(a_2, 1)$. Define two new random variables $Y_1 = X_1/(X_1 + X_2)$ and $Y_2 = X_1 + X_2$. Find the joint density for (Y_1, Y_2) using a direct PDF transformation (and its Jacobian). Use this to characterize the marginals $p(y_1)$ and $p(y_2)$, and propose *a method that exploits this result to simulate beta random variables, assuming you have a source of gamma random variables.*

Let $X_1 \sim Ga(a_1, 1)$, $X_2 \sim Ga(a_2, 1)$ and let us define $(Y_1, Y_2) = g(X_1, X_2) = \left(\frac{X_1}{X_1+X_2}, X_1 + X_2\right)$. The joint density of (Y_1, Y_2) can be found via pdf transformation, i.e.

$$\begin{cases} y_1 = g_1(x_1, x_2) = \frac{x_1}{x_1+x_2} \\ y_2 = g_2(x_1, x_2) = x_1 + x_2 \end{cases} \Rightarrow \begin{cases} x_1 = g_1^{-1}(y_1, y_2) = y_1 y_2 \\ x_2 = g_2^{-1}(y_1, y_2) = y_2(1 - y_1). \end{cases} \tag{1}$$

The inverse transformation has a unique solution and therefore the mapping is one-to-one. Moreover, the domain $\mathcal{X}_1 \times \mathcal{X}_2 = [0, \infty)^2$ is mapped to $\mathcal{Y}_1 \times \mathcal{Y}_2 = [0, 1] \times [0, \infty)$.

The Jacobian of the transformation is

$$J = \begin{pmatrix} \dfrac{\partial g_1^{-1}(y_1, y_2)}{\partial y_1} & \dfrac{\partial g_1^{-1}(y_1, y_2)}{\partial y_2} \\ \dfrac{\partial g_2^{-1}(y_1, y_2)}{\partial y_1} & \dfrac{\partial g_2^{-1}(y_1, y_2)}{\partial y_2} \end{pmatrix} = \begin{pmatrix} y_2 & y_1 \\ -y_2 & 1 - y_1 \end{pmatrix}$$

and the absolute value of its determinant is $|J| = |y_2 - y_1 y_2 + y_1 y_2| = y_2$. The joint pdf of (X_1, X_2) is, since $X_1 \perp X_2$,

$$f_{(X_1,X_2)}(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2) = \frac{1}{\Gamma(a_1)\Gamma(a_2)} x_1^{a_1-1} e^{-x_1} x_2^{a_2-1} e^{-x_2} \mathcal{I}_{[0,\infty)}(x_1) \mathcal{I}_{[0,\infty)}(x_2).$$

Therefore

$$\begin{aligned} f_{(Y_1,Y_2)}(y_1, y_2) &= f_{(X_1,X_2)}(g_1^{-1}(y_1, y_2), g_2^{-1}(y_1, y_2))|J| \\ &= \frac{1}{\Gamma(a_1)\Gamma(a_2)} y_1^{a_1-1} y_2^{a_1-1} e^{-y_1 y_2} y_2^{a_2-1} (1 - y_1)^{a_2-1} e^{-y_2(1-y_1)} \mathcal{I}_{[0,1]}(y_1) \mathcal{I}_{[0,\infty)}(y_2) y_2 \\ &= \frac{\Gamma(a_1 + a_2)}{\Gamma(a_1)\Gamma(a_2)} y_1^{a_1-1} (1 - y_1)^{a_2-1} \mathcal{I}_{[0,1]}(y_1) \frac{1}{\Gamma(a_1 + a_2)} \times y_2^{a_1+a_2-1} e^{-y_2} \mathcal{I}_{[0,\infty)}(y_2). \end{aligned}$$

The joint density factors, and therefore we recognize that

$$f_{(Y_1,Y_2)}(y_1, y_2) = Beta(a_1, a_2) \times Gamma(a_1 + a_2, 1),$$

that is,

$$\begin{aligned} &Y_1 \sim Beta(a_1, a_2) \\ &Y_2 \sim Gamma(a_1 + a_2, 1) \\ &Y_1 \perp Y_2 \end{aligned}$$

Therefore, in order to sample $Y_1 \sim Beta(a_1, a_2)$ we can sample $X_1 \sim Gamma(a_1, 1)$, then $X_2 \sim Gamma(a_2, 1)$ independently. The resulting $Y_1 = \frac{X_1}{X_1+X_2}$ is a draw from $Beta(a_1, a_2)$.

(C) *Suppose that we take independent observations $X_1, \ldots, X_N$ from a normal sampling model with unknown mean θ and known variance σ^2: $X_i \overset{\text{iid}}{\sim} N(\theta, \sigma^2)$. Suppose that θ is given a normal prior distribution with mean m and variance v. Derive the posterior distribution $p(\theta|x_1, \ldots, x_N)$.*

The following model

$$
\begin{aligned}
&X_1, \dots, X_n | \theta \overset{\text{iid}}{\sim} N(\theta, \sigma^2) \\
&\theta \sim N(m, v)
\end{aligned}
$$

leads to the following posterior distribution:

$$
\begin{aligned}
p(\theta|\boldsymbol{x}) &\propto p(\boldsymbol{X}|\theta)p(\theta) \\
&\propto \prod_{i=1}^{N} \left\{ \left(\frac{1}{2\pi\sigma^2}\right)^{1/2} e^{-\frac{1}{2\sigma^2}(x_i-\theta)^2} \right\} \left(\frac{1}{2\pi v}\right)^{1/2} e^{-\frac{1}{2v}(\theta-m)^2} \\
&\propto e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-\theta)^2} e^{-\frac{1}{2v}(\theta-m)^2} \\
&= e^{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{N} x_i^2 + n\theta^2 - 2\theta\sum_{i=1}^{N} x_i\right)} e^{-\frac{1}{2v}(\theta^2+m^2-2m\theta)} \\
&\propto e^{-\frac{n}{2\sigma^2}\left(\theta^2-2\overline{x}\theta\right)} e^{-\frac{1}{2v}(\theta^2-2m\theta)} \\
&= e^{-\frac{n}{2\sigma^2 v}\left(v\theta^2-2\overline{x}v\theta+\frac{\sigma^2}{n}\theta^2-2\frac{\sigma^2}{n}m\theta\right)} \\
&= e^{-\frac{n}{2\sigma^2 v}\left(v\theta^2-2\overline{x}v\theta+\frac{\sigma^2}{n}\theta^2-2\frac{\sigma^2}{n}m\theta\right)} \\
&= e^{-\frac{n}{2\sigma^2 v}\left[(v+\frac{\sigma^2}{n})\theta^2-2(\overline{x}v+\frac{\sigma^2}{n}m)\theta\right]} \\
&= e^{-\frac{nv+\sigma^2}{2\sigma^2 v}\left(\theta^2-2\frac{\overline{x}nv+\sigma^2 m}{nv+\sigma^2}\theta\right)} \\
&\propto e^{-\frac{nv+\sigma^2}{2\sigma^2 v}\left(\theta-\frac{\overline{x}nv+\sigma^2 m}{nv+\sigma^2}\theta\right)^2} \\
&\propto N\left(\frac{v}{v+\sigma^2/n}\overline{x} + \frac{\sigma^2/n}{v+\sigma^2/n}m; \left(\frac{n}{\sigma^2}+\frac{1}{v}\right)^{-1}\right)
\end{aligned}
$$

This is the so-called **Normal-Normal model for unknown mean and known variance**.

(D) *Suppose that we take independent observations $X_1, \dots, X_N$ from a normal sampling model with known mean θ but unknown variance σ^2 (this seems even more artificial than the last, but is conceptually important). To make this easier, we will re-express things in terms of the precision, or inverse variance $\omega = 1/\sigma^2$:*

$$
p(x_i|\theta, \omega) = \left(\frac{\omega}{2\pi}\right)^{1/2} \exp\left\{-\frac{\omega}{2}(x_i-\theta)^2\right\}.
$$

Suppose that ω has a gamma prior with parameters a and b, implying that σ^2 has what is called an inverse-gamma prior. Derive the posterior distribution $p(\omega|x_1, \dots, x_N)$. Re-express this as a posterior for σ^2, the variance.

The following model

$$
\begin{aligned}
&X_1, \dots, X_n | \sigma^2 \overset{\text{iid}}{\sim} N(\theta, \sigma^2) \\
&\sigma^2 \sim IG(a, b)
\end{aligned}
$$

can be rewritten in terms of the precision parameter $\omega = 1/\sigma^2$ as

$$
\begin{aligned}
&X_1, \dots, X_n | \omega \overset{\text{iid}}{\sim} N(\theta, \omega) \\
&\omega \sim Gamma(a, b).
\end{aligned}
$$

This model leads to the following posterior distribution:

$$
\begin{aligned}
p(\omega|\boldsymbol{x}) &\propto p(\boldsymbol{X}|\omega)p(\omega) \\
&\propto \prod_{i=1}^{N}\left\{\left(\frac{\omega}{2\pi}\right)^{1/2} e^{-\frac{\omega}{2}(x_i-\theta)^2}\right\}\frac{b^a}{\Gamma(a)}\omega^{a-1}e^{-b\omega}\mathcal{I}_{[0,\infty)}(\omega) \\
&\propto \omega^{N/2}e^{-\frac{\omega}{2}\sum_{i=1}^{N}(x_i-\theta)^2}\omega^{a-1}e^{-b\omega}\mathcal{I}_{[0,\infty)}(\omega) \\
&= \omega^{a+N/2-1}e^{-\omega(b+\frac{1}{2}\sum_{i=1}^{N}(x_i-\theta)^2)}\mathcal{I}_{[0,\infty)}(\omega) \\
&\propto Gamma\left(a+N/2; b+\frac{1}{2}\sum_{i=1}^{N}(x_i-\theta)^2)\right).
\end{aligned}
$$

and, subsequently, Therefore,

$$
\sigma^2|\boldsymbol{x} \sim IG\left(a+N/2; b+\frac{1}{2}\sum_{i=1}^{N}(x_i-\theta)^2)\right).
$$

In terms of the pdf of σ^2, let us find the generic pdf of $IG(a,b)$. We know that, if $X \sim Gamma(a,b)$, then $Y = 1/X \sim IG(a,b)$. The transformation $g(x) = 1/x$ is monotone and its inverse is $g^{-1}(y) = 1/y$. The derivative of the inverse transformation is $\frac{\partial}{\partial y}g^{-1}(y) = -\frac{1}{y^2}$. Thus,

$$
\begin{aligned}
f_Y(y) &= f_X(1/y)\left|\frac{\partial}{\partial y}g^{-1}(y)\right| \\
&= \frac{b^a}{\Gamma(a)}\left(\frac{1}{y}\right)^{a-1}e^{-\frac{b}{y}}\frac{1}{y^2}\mathcal{I}_{[0,\infty)}(y) \\
&= \frac{b^a}{\Gamma(a)}y^{-a-1}e^{-\frac{b}{y}}\mathcal{I}_{[0,\infty)}(y).
\end{aligned}
$$

Therefore, if $\sigma^2|\boldsymbol{x} \sim IG\left(a+N/2; b+\frac{1}{2}\sum_{i=1}^{N}(x_i-\theta)^2\right)$, the pdf is

$$
f_{\sigma^2|\boldsymbol{x}}(\sigma^2|\boldsymbol{x}) = \frac{(b+\frac{1}{2}\sum_{i=1}^{N}(x_i-\theta)^2)^{a+N/2}}{\Gamma(a+N/2)}\sigma^{2(-a-N/2-1)}e^{-\frac{b+\frac{1}{2}\sum_{i=1}^{N}(x_i-\theta)^2}{\sigma^2}}\mathcal{I}_{[0,\infty)}(\sigma^2).
$$

This is the so-called **Normal-inverse gamma model for known mean and unknown variance**.

(E) *Suppose that, as above, we take independent observations $X_1, \ldots, X_N$ from a normal sampling model with unknown, common mean θ. This time, however, each observation has its own idiosyncratic (but known) variance: $X_i \overset{\text{ind}}{\sim} N(\theta, \sigma_i^2)$. Suppose that θ is given a normal prior distribution with mean m and variance v. Derive the posterior distribution $p(\theta|x_1, \ldots, x_N)$. Express the posterior mean in a form that is clearly interpretable as a weighted average of the observations and the prior mean.*

The following model

$$
\begin{aligned}
&X_1, \ldots, X_n|\theta \overset{\text{iid}}{\sim} N(\theta, \sigma_i^2) \\
&\theta \sim N(m, v)
\end{aligned}
$$

leads to the following posterior distribution:

$$\begin{aligned}
p(\theta|\boldsymbol{x}) &\propto p(\boldsymbol{X}|\theta)p(\theta) \\
&\propto \prod_{i=1}^{N} \left\{ \left(\frac{1}{2\pi\sigma_i^2}\right)^{1/2} e^{-\frac{1}{2\sigma_i^2}(x_i-\theta)^2} \right\} \left(\frac{1}{2\pi v}\right)^{1/2} e^{-\frac{1}{2v}(\theta-m)^2} \\
&\propto e^{-\frac{1}{2}\sum_{i=1}^{N}\frac{(x_i-\theta)^2}{\sigma_i^2}} e^{-\frac{1}{2v}(\theta-m)^2}.
\end{aligned}$$

The exponents can be rewritten as

$$\begin{aligned}
& -\frac{1}{2}\left[\sum_{i=1}^{N}\left(\frac{x_i-\theta}{\sigma_i}\right)^2 + \frac{(\theta-m)^2}{v}\right] \\
&= -\frac{1}{2}\left[\sum_{i=1}^{N}\left(\frac{x_i^2}{\sigma_i^2} + \frac{\theta^2}{\sigma_i^2} - 2\frac{x_i\theta}{\sigma_i^2}\right) + \frac{\theta^2+m^2-2m\theta}{v}\right] \\
&\propto -\frac{1}{2}\left[\theta^2\left(\frac{1}{v}+\sum_{i=1}^{N}\frac{1}{\sigma_i^2}\right) - 2\theta\left(\frac{m}{v}+\sum_{i=1}^{N}\frac{x_i}{\sigma_i^2}\right)\right].
\end{aligned}$$

Therefore, we get

$$\theta|\boldsymbol{x} \sim N\left(\frac{\frac{m}{v}+\sum_{i=1}^{N}\frac{x_i}{\sigma_i^2}}{\frac{1}{v}+\sum_{i=1}^{N}\frac{1}{\sigma_i^2}}; \left(\frac{1}{v}+\sum_{i=1}^{N}\frac{1}{\sigma_i^2}\right)^{-1}\right).$$

This is the so-called **Normal-Normal model for unknown mean and known idiosyncratic variances**.

(F) *Suppose that $(X|\sigma^2) \sim N(0,\sigma^2)$, and that $1/\sigma^2$ has a $Gamma(a,b)$ prior, defined as above. Show that the marginal distribution of X is Student's t. This is why the t distribution is often referred to as a scale mixture of normals.*

Problem 2. The multivariate normal distribution

Basics

We all know the univariate normal distribution, whose long history began with de Moivre's 18th-century work on approximating the (analytically inconvenient) binomial distribution. This led to the probability density function

$$p(x) = \frac{1}{\sqrt{2\pi v}} \exp\left\{-\frac{(x-m)^2}{2v}\right\}$$

for the normal random variable with mean m and variance v, written $X \sim N(m,v)$.

Here's an alternative characterization of the univariate normal distribution in terms of moment-generating functions: a random variable X has a normal distribution if and only if $E\{\exp(tx)\} = \exp(mt + vt^2/2)$ for some real m and positive real v. Remember that $E(\cdot)$ denotes the expected value of its argument under the given probability distribution. We will generalize this definition to the multivariate normal.

(A) *First, some simple moment identities. The covariance matrix cov*$(\boldsymbol{X})$ *of a vector-valued random variable* $\boldsymbol{X}$ *is defined as the matrix whose* (i, j) *entry is the covariance between* X_i *and* X_j*. In matrix notation, cov*$(\boldsymbol{X}) = E\{(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^T\}$*, where* $\boldsymbol{\mu}$ *is the mean vector whose* i*th component is* $E(X_i)$*. Prove the following: (1) cov*$(\boldsymbol{X}) = E(\boldsymbol{X}\boldsymbol{X}^T) - \boldsymbol{\mu}\boldsymbol{\mu}^T$*; and (2) cov*$(A\boldsymbol{X} + \boldsymbol{b}) = A$*cov*$(\boldsymbol{X})A^T$ *for matrix* A *and vector* $\boldsymbol{b}$*.*

(B) *Consider the random vector* $\boldsymbol{Z} = (Z_1, \ldots, Z_p)^T$*, with each entry having an independent standard normal distribution (that is, mean* 0 *and variance* 1*). Derive the probability density function (PDF) and moment-generating function (MGF) of* $\boldsymbol{Z}$*, expressed in vector notation. We say that* $\boldsymbol{Z}$ *has a standard multivariate normal distribution.*

(C) *A vector-valued random variable* $\boldsymbol{X} = (X_1, \ldots, X_p)^T$ *has a multivariate normal distribution if and only if every linear combination of its components is univariate normal. That is, for all vectors* $\boldsymbol{a}$ *not identically zero, the scalar quantity* $Z = \boldsymbol{a}^T\boldsymbol{X}$ *is normally distributed. From this definition, prove that* $\boldsymbol{X}$ *is multivariate normal, written* $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \Sigma)$*, if and only if its moment-generating function is of the form* $E(\exp\{\boldsymbol{t}^T\boldsymbol{X}\}) = \exp(\boldsymbol{t}^T\boldsymbol{\mu} + \boldsymbol{t}^T\Sigma\boldsymbol{t}/2)$*. Hint: what are the mean, variance, and moment-generating function of* Z*, expressed in terms of moments of* $\boldsymbol{X}$*?*

(D) *Another basic theorem is that a random vector is multivariate normal if and only if it is an affine transformation of independent univariate normals. You will first prove the "if" statement. Let* $\boldsymbol{Z}$ *have a standard multivariate normal distribution, and define the random vector* $\boldsymbol{X} = L\boldsymbol{Z} + \boldsymbol{\mu}$ *for some* $p \times p$ *matrix* L *of full column rank. Prove that* $\boldsymbol{X}$ *is multivariate normal. In addition, use the moment identities you proved above to compute the expected value and covariance matrix of* $\boldsymbol{X}$*.*

(E) *Now for the "only if", Suppose that* $\boldsymbol{X}$ *has a multivariate normal distribution. Prove that* $\boldsymbol{X}$ *can be written as an affine transformation of standard normal random variables. (Note: a good way to prove that something can be done is to do it!) Use this insight to propose an algorithm for simulating multivariate normal random variables with a specified mean and covariance matrix.*

Appendix A

R code