SDS 383D, Exercises 2: Bayes and the Gaussian Linear Model

Jan-Michael Cabrera

February 22, 2019

A simple Gaussian location model

(A) Show that the marginal prior, $p(\theta)$ takes on the follwing form:

$$p(\theta) \propto \left(1 + \frac{1}{\nu}\frac{(x-m)^2}{s^2}\right)^{-\frac{\nu+1}{2}}$$

The joint prior for θ and ω can be expressed as

$$p(\theta,\omega) \propto \omega^{(d+1)/2-1} \exp\left(-\omega\frac{k(\theta-\mu)^2}{2}\right)\exp\left(-\omega\frac{\eta}{2}\right)$$

This joint normal-gamma prior can be marginalized over ω to obtain a distribution for θ. Combining terms in the exponents and letting $a = \frac{d+1}{2}$ and $b = \frac{k(\theta-\mu)^2}{2} + \frac{\eta}{2}$, we get

$$\begin{aligned}
p(\theta) &\propto \int p(\theta,\omega)d\omega \\
&\propto \int \omega^{a-1}\exp(-\omega b)d\omega
\end{aligned}$$

We immediately realize that the term to the right of the integral is the unnormalized form of a gamma distribution. We add these normalizing components to ensure the intergral integrates to one and reduce the result,

$$\begin{aligned}
p(\theta) &\propto \frac{\Gamma(a)}{b^a}\int \frac{b^a}{\Gamma(a)}\omega^{a-1}\exp(-\omega b)d\omega \\
&\propto \Gamma\left(\frac{d+1}{2}\right)\left[\frac{\eta}{2} + \frac{k(\theta-\mu)^2}{2}\right]^{-\frac{d+1}{2}} \\
&\propto \left[\frac{\eta}{2}\left(1 + \frac{kd(\theta-\mu)^2}{2d\eta/2}\right)\right]^{-\frac{d+1}{2}} \\
&\propto \left(1 + \frac{1}{d}\frac{(\theta-\mu)^2}{(d\eta/k)}\right)^{-\frac{d+1}{2}}.
\end{aligned}$$

The result can be recognized as having the form of the first equation above with the following variable relations.

$$
\begin{aligned}
\nu &= d \\
x &= \theta \\
s^2 &= \frac{d\eta}{k} \\
m &= \mu
\end{aligned}
$$

(B) Show that $p(\theta, \omega|\mathbf{y})$ has the form

$$p(\theta, \omega|\mathbf{y}) \propto \omega^{(d^*+1)/2-1}\exp\left[-\omega\frac{k^*(\theta-\mu^*)^2}{2}\right]\exp\left[-\omega\frac{\eta^*}{2}\right]$$

The sampling distribution is given as the following, with the term in the exponent reduced to sufficient statistics for a normal distribution,

$$p(\mathbf{y}|\theta, \omega) \propto \omega^{n/2}\exp\left[-\omega\left(\frac{S_y + n(\bar{y}-\theta)^2}{2}\right)\right]$$

We multiply our normal-gamma prior with the normal sampling distribution and collect similar terms in the exponents.

$$
\begin{aligned}
p(\theta, \omega|\mathbf{y}) &\propto \omega^{(d+1)/2-1}\exp\left[-\omega\frac{(\theta-\mu)^2}{2}\right]\exp\left[-\omega\frac{\eta}{2}\right]\omega^{n/2}\exp\left[-\omega\left(\frac{S_y+n(\bar{y}-\theta)^2}{2}\right)\right] \\
&\propto \omega^{(d+n+1)/2-1}\exp\left[-\frac{\omega}{2}\left(k(\theta-\mu)^2+\eta+S_y+n(\bar{y}-\theta)^2\right)\right]
\end{aligned}
$$

We expand the term in the $e^{(\cdots)}$ term and collect terms together,

$$
\begin{aligned}
k(\theta-\mu)^2+\eta+S_y+n(\bar{y}-\theta)^2 &= k\theta^2 - 2k\mu\theta + k\mu^2 + \eta + S_y + n\bar{y}^2 - 2n\theta\bar{y} + n\theta^2 \\
&= (k+n)\theta^2 + (-2k\mu - 2n\bar{y})\theta + k\mu^2 + \eta + S_y + n\bar{y}^2.
\end{aligned}
$$

The result from expanding the terms can be reduced to the following form to recover the normal part of the posterior for θ,

$$ax^2 + bx + c = a(x-h)^2 + l, \quad h = -\frac{b}{2a}, \quad l = c - ah^2$$

We carry out the process of completing the square the θ term and reduce other parts of the expression,

$$
\begin{aligned}
&(k+n)\theta^2 + (-2k\mu - 2n\bar{y})\theta + k\mu^2 + \eta + S_y + n\bar{y}^2 = \\
&(k+n)\left(\theta - \frac{k\mu+n\bar{y}}{k+n}\right)^2 + k\mu^2 + \eta + S_y + n\bar{y}^2 - (k+n)\left(\frac{k\mu+n\bar{y}}{k+n}\right)^2 = \\
&(k+n)\left(\theta - \frac{k\mu+n\bar{y}}{k+n}\right)^2 + \frac{nk(\mu-\bar{y})^2}{k+n} + \eta + S_y.
\end{aligned}
$$

The result is plugged back into the exponential term, and the terms are collected to obtain the form of a normal-gamma posterior.

$$
\begin{aligned}
p(\mathbf{y}|\theta,\omega) &\propto \omega^{(d+n+1)/2-1}\text{exp}\left[-\frac{\omega}{2}\left((k+n)\left(\theta-\frac{k\mu+n\bar{y}}{k+n}\right)^2+\frac{nk(\mu-\bar{y})^2}{k+n}+\eta+S_y\right)\right] \\
&\propto \omega^{(d+n+1)/2-1}\text{exp}\left[-\frac{\omega}{2}\left((k+n)\left(\theta-\frac{k\mu+n\bar{y}}{k+n}\right)^2\right)\right]\text{exp}\left[-\frac{\omega}{2}\left(\frac{nk(\mu-\bar{y})^2}{k+n}+\eta+S_y\right)\right]
\end{aligned}
$$

The normal-gamma posterior has the form of the normal-gamma prior with the following variable relations.

$$
\begin{aligned}
d^* &= d+n \\
k^* &= k+n \\
\mu^* &= \frac{k\mu+n\bar{y}}{k+n} \\
\eta^* &= \frac{nk(\mu-\bar{y})^2}{k+n}+\eta+S_y
\end{aligned}
$$

(C) From the joint posterior, what is the conditional posterior distribution $p(\theta|\mathbf{y},\omega)$?

$$
\begin{aligned}
p(\theta|\mathbf{y},\omega) &= \int_0^\infty p(\theta,\omega|\mathbf{y})d\omega \\
&\propto \text{exp}\left[-\frac{\omega}{2}\left((k+n)\left(\theta-\frac{k\mu+n\bar{y}}{k+n}\right)^2\right)\right]\int_0^\infty \omega^{(d+n+1)/2-1}\text{exp}\left[-\frac{\omega}{2}\left(\frac{nk(\mu-\bar{y})^2}{k+n}+\eta+S_y\right)\right]d\omega
\end{aligned}
$$

The conditional posterior distribution is simply the joint posterior marginalized over ω. We also can recognize that the normal distribution can be completely characterized by its mean and variance. Therefore,

$$
(\theta|\mathbf{y},\omega) \sim \text{N}\left(\frac{k\mu+n\bar{y}}{k+n},(\omega(k+n))^{-1}\right)
$$

(D) From the joint posterior, what is the marginal posterior distribution $p(\omega|\mathbf{y})$?

$$
\begin{aligned}
p(\omega|\mathbf{y}) &= \int_0^\infty p(\theta,\omega|\mathbf{y})d\theta \\
&\propto \omega^{(d+n)/2-1}\text{exp}\left[-\frac{\omega}{2}\left(\frac{nk(\mu-\bar{y})^2}{k+n}+\eta+S_y\right)\right]\int_{-\infty}^\infty \omega^{1/2}\text{exp}\left[-\frac{\omega}{2}\left((k+n)\left(\theta-\frac{k\mu+n\bar{y}}{k+n}\right)^2\right)\right]d\theta
\end{aligned}
$$

The marginal posterior for ω is found in a similar fashion as above, noting that part of the normal distribution to be marginalized depends on ω. Including this term in the integral ensures that it integrates to a constant not dependent on ω. The distribution then has the following form,

$$
\begin{aligned}
\omega|\mathbf{y} &\sim \text{Gamma}\left(\frac{d^*}{2},\frac{\eta^*}{2}\right) \\
d^* &= d+n \\
\eta^* &= \frac{nk(\mu-\bar{y})^2}{k+n}+\eta+S_y
\end{aligned}
$$

(E) Show that the marginal posterior $p(\theta|\mathbf{y})$ takes the form of a centered, scaled t distribution and express the parameters in terms of the four parameters of the normal-gamma posterior for (θ, ω).

$$p(\theta) \propto \left(1 + \frac{1}{d}\frac{(\theta-\mu)^2}{(d\eta/k)}\right)^{-\frac{d+1}{2}}$$

This simply reqiures plugging in the respective starred values into the result above producing,

$$p(\theta|\mathbf{y}) \propto \left(1 + \frac{1}{d^*}\frac{(\theta-\mu^*)^2}{(d^*\eta^*/k^*)}\right)^{-\frac{d^*+1}{2}}$$

$$\begin{aligned}
d^* &= d + n \\
k^* &= k + n \\
\mu^* &= \frac{k\mu + n\bar{y}}{k+n} \\
\eta^* &= \frac{nk(\mu - \bar{y})^2}{k+n} + \eta + S_y
\end{aligned}$$

(F) True or false: in the limit as the prior parameters k, d, and η approach zero, the priors $p(\theta)$ and $p(\omega)$ are valid probability distributions.

To check this result we simply check to see if the resultant distribution integrates to one or a constant in the limit. The prior for θ is,

$$p(\theta) \propto \left[1 + \frac{1}{d}\frac{(\theta-\mu)^2}{\eta/kd}\right]^{-\frac{d+1}{2}}$$

$$\lim_{k,d,\eta\to 0}\left[1 + \frac{1}{d}\frac{(\theta-\mu)^2}{\eta/kd}\right]^{-\frac{d+1}{2}} = \lim_{k,d,\eta\to 0}\left[1 + \frac{k(\theta-\mu)^2}{\eta}\right]^{-\frac{d+1}{2}}$$

$$\left[1 + \frac{0(\theta-\mu)^2}{0}\right]^{-\frac{1}{2}}$$

In the limit as the parameters approach zero, we have an undefined function that cannot be integrated. For ω we have,

$$p(\omega) \propto \omega^{d/2-1}\exp\left(-\omega\frac{\eta}{2}\right)$$

Taking the limit for this expression,

$$\lim_{d,\eta\to 0}\omega^{d/2-1}\exp\left(-\omega\frac{\eta}{2}\right) = \omega^{-1}$$

The resultant expression has a singularity at $\omega = 0$ and will therefore integrate to infinity.

False

(G) True or false: in the limit as the prior parameters k, d, and η approach zero, the posteriors $p(\theta|\mathbf{y})$ and $p(\omega|\mathbf{y})$ are valid probability distributions.

We first take the limit in the starred expressions,

$$
\begin{aligned}
d^* &= n + d \quad \xrightarrow[d \to 0]{} \quad d^* = n \\
k^* &= k + n \quad \xrightarrow[k \to 0]{} \quad k^* = n \\
\mu^* &= \frac{k\mu + n\bar{y}}{k + n} \quad \xrightarrow[k \to 0]{} \quad \mu^* = \bar{y} \\
\eta^* &= \frac{nk(\mu - \bar{y})^2}{k + n} + \eta + S_y \quad \xrightarrow[k, \eta \to 0]{} \quad \eta^* = S_y
\end{aligned}
$$

Plugging in these values into our posteriors we get,

$$p(\theta|\mathbf{y}) \propto \left[1 + \frac{1}{n}\frac{(\theta - \bar{y})^2}{S_y/n^2}\right]^{-\frac{n+1}{2}}$$

$$p(\omega|\mathbf{y}) \propto \omega^{n/2-1}\exp\left(-\omega\frac{S_y}{2}\right)$$

Both are kernels for valid probability distributions

True

(H) True or false: In the limit as the prior parameters k, d, and η approach zero, the Bayesian credible interval for θ becomes identical to the classical (frequentist) confidence interval for θ at the same confidence level.

$$\theta \in m \pm t^* \cdot s$$

$$
\begin{aligned}
m = \mu^* \to \bar{y} &\quad \text{for } k, d, \eta \to 0 \\
s^{*2} = \frac{S_y}{n^2} \to s &= \frac{1}{n}\left[\sum_{i=1}^{n}(y_i - \bar{y})^2\right]^{1/2}
\end{aligned}
$$

$$\theta \in \bar{y} \pm t^* \cdot \frac{1}{n}\left[\sum_{i=1}^{n}(y_i - \bar{y})^2\right]^{1/2}$$

True

The Conjugate Gaussian Linear Model

(A) Derive the conditional posterior $p(\beta|\mathbf{y}, \omega)$

$$\beta|\omega \sim \mathrm{N}(m, (\omega K)^{-1})$$
$$\omega \sim \mathrm{Gamma}\left(\frac{d}{2}, \frac{\eta}{2}\right)$$
$$\mathbf{y}|\beta, \omega \sim \mathrm{N}(X\beta, (\omega\Lambda)^{-1})$$

First we find the joint posterior for β and ω given the multivariate sampling distribution.

$$\begin{aligned}
p(\beta, \omega|\mathbf{y}) &\propto p(\omega)p(\beta|\omega)p(\mathbf{y}|\beta, \omega) \\
&\propto \omega^{d/2-1}\exp\left(-\omega\frac{\eta}{2}\right)\omega^{p/2}\exp\left[-\frac{\omega}{2}(\beta-m)^T K(\beta-m)\right]\omega^{n/2}\exp\left[-\frac{\omega}{2}(y-X\beta)^T\Lambda(y-X\beta)\right] \\
&\propto \omega^{\frac{d+p+n}{2}-1}\exp\left[-\frac{\omega}{2}\left((\beta-m)^T K(\beta-m)+(y-X\beta)^T\Lambda(y-X\beta)+\eta\right)\right]
\end{aligned}$$

After combining exponents, we distribute the terms and collect the β terms together,

$$\begin{aligned}
&(\beta-m)^T K(\beta-m)+(y-X\beta)^T\Lambda(y-X\beta)+\eta \\
&= \beta^T K\beta - \beta^T Km - m^T K\beta + m^T Km + y^T\Lambda y - y^T\Lambda X\beta - \beta^T X^T\Lambda y + \beta^T X^T\Lambda X\beta + \eta \\
&= \beta^T K\beta - 2\beta^T Km + m^T Km + y^T\Lambda y - 2\beta^T X^T\Lambda y + \beta^T X^T\Lambda X\beta + \eta \\
&= \beta^T(K + X^T\Lambda X)\beta - 2\beta^T(Km + X^T\Lambda y) + m^T Km + y^T\Lambda y + \eta
\end{aligned}$$

As before, we can realize we need to get the expression in the proper form that characterizes the multivariate normal distribution. We do this by completing the square. The equation has the from $Q^T AQ + Q^T b + c = (Q-h)^T A(Q-h) + k$, with $h = -\frac{1}{2}A^{-1}b$ and $k = c - \frac{1}{4}b^T A^{-1}b$.

Let $A = (K + X^T\Lambda X) = K^*$, $b = -2(Km + X^T\Lambda y)$, and $c = m^T Km + y^T\Lambda y + \eta$. The term h reduces to the following,

$$\begin{aligned}
h &= -\frac{1}{2}(K + X^T\Lambda X)^{-1}\cdot(-2(Km + X^T\Lambda y)) \\
&= (K + X^T\Lambda X)^{-1}(Km + X^T\Lambda y) = m^*
\end{aligned}$$

Plugging these terms back into the equation we get,

$$\begin{aligned}
&\beta^T(K + X^T\Lambda X)\beta - 2\beta^T(Km + X^T\Lambda y) + m^T Km + y^T\Lambda y + \eta \\
&= (\beta - m^*)^T K^*(\beta - m) + m^T Km + y^T\Lambda y + \eta - (Km + X^T\Lambda y)^T {K^*}^{-1}(Km + X^T\Lambda y)
\end{aligned}$$

These are then reintroduced into the posterior,

$$\begin{aligned}
p(\beta, \omega|\mathbf{y}) &\propto \omega^{\frac{d+p+n}{2}-1}\exp\left[-\frac{\omega}{2}\left((\beta - m^*)^T K^*(\beta - m) + m^T Km + y^T\Lambda y + \eta - (Km + X^T\Lambda y)^T {K^*}^{-1}(Km + X^T\Lambda y)\right)\right] \\
&\propto \omega^{\frac{d^*}{2}-1}\exp\left[-\omega\frac{\eta^*}{2}\right]\omega^{\frac{p}{2}}\exp\left[-\frac{\omega}{2}(\beta - m^*)^T K^*(\beta - m^*)\right]
\end{aligned}$$

We see that the form of the posterior is also multivariate normal-gamma with the following variables

$$
\begin{aligned}
d^* &= d + n \\
K^* &= (K + X^T \Lambda X) \\
\eta^* &= m^T K m + y^T \Lambda y + \eta - (Km + X^T \Lambda y)^T {K^*}^{-1} (Km + X^T \Lambda y) \\
m^* &= (K + X^T \Lambda X)^{-1} (Km + X^T \Lambda y)
\end{aligned}
$$

From this distribution, again since the multivariate normal is completely characterized by it's mean vector and covariance matrix, we get that the conditional posterior has the form,

$$
\beta | \omega, \mathbf{y} \sim \mathrm{N}\left(m^*, (\omega K^*)^{-1}\right).
$$

(B) Derive the marginal posterior $p(\omega|\mathbf{y})$

The marginal posterior for ω can be found by marginalizing the joint posterior over β,

$$
\begin{aligned}
p(\omega|\mathbf{y}) &\propto \int_{-\infty}^{\infty} p(\beta, \omega|\mathbf{y}) d\beta \\
&\propto \omega^{\frac{d^*}{2}-1} \exp\left[-\omega \frac{\eta^*}{2}\right] \int_{-\infty}^{\infty} \omega^{\frac{p}{2}} \exp\left[-\frac{\omega}{2}(\beta - m^*)^T K^* (\beta - m^*)\right] d\beta \\
&\propto \omega^{\frac{d^*}{2}-1} \exp\left[-\omega \frac{\eta^*}{2}\right].
\end{aligned}
$$

Note that again here we must be careful of the terms included within the integral when marginalizing. From this we find that the posterior also has the form of a gamma distribution,

$$
\omega | \mathbf{y} \sim \mathrm{Gamma}\left(\frac{d^*}{2}, \frac{\eta^*}{2}\right).
$$

This distribution is characterized by the following

$$
\begin{aligned}
d^* &= d + n \\
K^* &= (K + X^T \Lambda X) \\
\eta^* &= m^T K m + y^T \Lambda y + \eta - (Km + X^T \Lambda y)^T {K^*}^{-1} (Km + X^T \Lambda y)
\end{aligned}
$$

(C) Putting these together, derive the marginal posterior $p(\beta|\mathbf{y})$

The process here closely matches that of part A in the previous section. We marginalize the joint posterior over ω.

$$
\begin{aligned}
p(\beta|\mathbf{y}) &\propto \int_0^{\infty} p(\beta, \omega|\mathbf{y}) d\omega \\
&\propto \int_0^{\infty} \omega^{\frac{d^*}{2}-1} \exp\left[-\omega \frac{\eta^*}{2}\right] \omega^{\frac{p}{2}} \exp\left[-\frac{\omega}{2}(\beta - m^*)^T K^* (\beta - m^*)\right] d\omega \\
&\propto \int_0^{\infty} \omega^a \exp(-b\omega) d\omega
\end{aligned}
$$

Letting $a = \frac{d^*+p}{2}$ and $b = \frac{1}{2}\left(\eta^* + (\beta - m^*)^T K^*(\beta - m^*)\right)$ we see that the terms in the integral form an unnormalized gamma distribution. Normalizing the distribution we obtain,

$$
\begin{aligned}
p(\beta|\mathbf{y}) &\propto \frac{\Gamma(a)}{b^a} \int_0^\infty \frac{b^a}{\Gamma(a)} \omega^a \exp(-b\omega) d\omega \\
&\propto \Gamma(a) b^{-a} \\
&\propto \left[\eta^* + (\beta - m^*)^T K^*(\beta - m^*)\right]^{-\frac{d^*+p}{2}} \\
&\propto \left[1 + \frac{1}{d^*} \frac{(\beta - m^*)^T K^*(\beta - m^*)}{\eta^*/d^*}\right]^{-\frac{d^*+p}{2}} .
\end{aligned}
$$

This has the form of a multivariate t distribution.

(D) Bayesian linear model fit to data in "gdpgrowth.csv"