Explanatory Notes 1 for Bayesian Inference

Why is normal is conjugate for normal?

Takis Konstantopoulos

Spring 2009

In the notes you have there is a statement (SZ, p.9) that if we are given a normal distribution with mean which is a priori normally distributed then it is also a posteriori normally distributed. In the notes you are asked to perform some tedious algebraic manipulations in order to find out the exact posterior distribution.

The problem with this approach is that, even if do the algebra correctly, you will not understand why the posterior is again normal and where its parameters come from.

I always maintain that it is impossible to learn something unless you understand it. So, to actually understand the result, we shall proceed in small, easy-to-digest, steps.

Recall that the density of a standard normal random variable X is

$$\varphi(x) = (2\pi)^{-1/2} \exp(-x^2/2).$$

We define $\mathcal{N}(\mu, \sigma^2)$ to be the law of $\sigma X + \mu$. The density of $\mathcal{N}(\mu, \sigma^2)$ is

$$\varphi_{\mu,\sigma^2}(x) = \frac{1}{\sigma}\varphi\Big(\frac{x-\mu}{\sigma}\Big).$$

Definition 1. *Recall that the random variables $(X_1, \ldots, X_d)$ are said to be multivariate normal (or jointly normal, or jointly Gaussian) if any linear combination of them is normal.*

Exercise 1. *Find the joint density (if it exists) of d jointly normal random variables and state a necessary and sufficient condition under which this density exists.*

Hint: Let R be the matrix with entries the covariances between the variables. A necessary and sufficient condition for the existence of density is that R be non-singular matrix (its determinant must be nonzero). The joint density is found by considering Definition 1.

Exercise 2. *Show that if (X, Y) are jointly normal and uncorrelated then they are independent.*

Lemma 1. *If (U, V) are jointly normal with $EU = EV = 0$, $EV^2 \neq 0$ then* *(i)* $E(U|V)$ *is a linear function of* V*; that is,*

$$\boxed{E(U|V) = \alpha V, \text{ where } \alpha = \frac{EUV}{EV^2}.}$$

(ii) $\text{var}(U|V)$ *is not a random variable; in fact,*

$$\boxed{\text{var}(U|V) = \text{var}(U) - \text{var}(E(U|V))}.$$

(iii) the law of U *conditional on* V *is* $\mathcal{N}\big(E(U|V), \text{var}(U|V)\big)$.

Proof. Assume that we have shown that $E(U|V) = \alpha V$ for some real number α. We can find this α by recalling that, by definition of conditional expectation, $U - \alpha V$ and any (linear or nonlinear) function of V must be uncorrelated. In particular, $U - \alpha V$ and V are uncorrelated:

$$E[(U - \alpha V)V] = 0. \tag{1}$$

This gives $\alpha = EUV/EV^2$. To show that $E(U|V) = \alpha V$ with α as above, observe that (1) and Exercise 2 imply the very important fact that

$$U - \alpha V \text{ and } V \text{ are independent.} \tag{2}$$

Hence

$$E(U - \alpha V|V) = E(U - \alpha V) = 0.$$

But $U = (U - \alpha V) + \alpha V$ and so $E(U|V) = E(U - \alpha V|V) + E(\alpha V|V) = 0 + \alpha E(V|V) = \alpha V$, as claimed. So (i) holds. To show (ii), recall that

$$\text{var}(U|V) = E\big((U - E(U|V))^2|V\big),$$

and use again the important fact (2) to get rid of the conditioning with respect to V:

$$\text{var}(U|V) = E\big((U - E(U|V))^2\big).$$

But $U = (U - E(U|V)) + E(U|V)$, and, by independence,

$$\text{var}(U) = \text{var}(U - E(U|V)) + \text{var}(E(U|V))$$

and so (ii) holds. The last part (iii) follows again from the important observation (2). $\square$

Exercise 3. *Suppose that EU, EV are not necessarily zero. Then argue that*

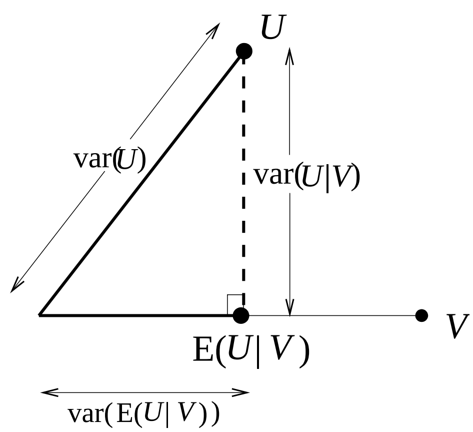
$$E(U|V) = E(U|V - EV) = E(U - EU|V - EV) + EU,$$

where the first equality follows from the fact that V and $V - EV$ convey the same information, while the second equality is trivial. Thus reduce the problem to the previous case, therefore justifying the formula

$$E(U|V) = \alpha(V - EV) + EU,$$

where $\alpha = E[(U - EU)(V - EV)]/\operatorname{var}(V)$, as long as $\operatorname{var}(V) \neq 0$. What happens when $\operatorname{var}(V) = 0$?

Interpretation. *Variances measure (squared) lengths of vectors. Correlation corresponds to inner product. Zero correlation means orthogonality. In the Gaussian World, $E(U|V)$ corresponds to the projection of U onto the space defined by V. The Pythagorean theorem holds: the square of the hypotenuse of a right triangle equals the sum of the squares of the other two sides.*



We can now show the main result. The statement on p.9 of the SZ notes is that

> ...we have n independent observations $\boldsymbol{y} = (y_1, \ldots, y_n)$ from a normal distribution $\mathcal{N}(\mu, \sigma^2)$ with known variance σ^2 and unknown mean μ to be estimated...

It is further written that:

> ...[s]uppose now that we choose a prior distribution for μ a $\mathcal{N}(\mu_0, \sigma_0^2)$ distribution...

These statements are informal wordings of the following assumption:

Assumption 1. *Let μ be a random variable whose law is $\mathcal{N}(\mu_0, \sigma_0^2)$ and let $(Y_1, \ldots, Y_n)$ be such that, conditional on μ, they are i.i.d. $\mathcal{N}(\mu, \sigma^2)$ each.*

Lemma 2. *Under Assumption 1, the distribution of μ conditional on $(Y_1, \ldots, Y_n)$ is again normal with*

mean equal to	$\mu_0 + \dfrac{\sigma_0^2}{\frac{1}{n}\sigma^2 + \sigma_0^2} (\overline{Y} - \mu_0)$
and variance equal to	$\dfrac{\frac{1}{n}\sigma^2\sigma_0^2}{\frac{1}{n}\sigma^2 + \sigma_0^2},$

where $\overline{Y} = \frac{1}{n}\sum_{k=1}^n Y_k$.

Proof. I shall do the proof for the case where $\mu_0 = 0$ and $n = 1$ and leave the rest for an (easy) exercise. The assumption says that the law of μ is $\mathcal{N}(0, \sigma_0^2)$ and that the law of Y, conditional on μ is $\mathcal{N}(\mu, \sigma^2)$. We wish to compute the law of μ conditional on Y. By (iii) of Lemma 1, we have that this conditional law is $\mathcal{N}\big(E(\mu|Y), \mathrm{var}(\mu|Y)\big)$, so all we have to do is compute $E(\mu|Y)$ and $\mathrm{var}(\mu|Y)$. From (i) of Lemma 1 we have

$$E(\mu|Y) = \alpha Y, \text{ where } \alpha = \frac{E\mu Y}{EY^2}.$$

But

$$Y = (Y - \mu) + \mu \tag{3}$$

and the terms on the right are independent zero-mean normal random variables. So

$$E\mu Y = E[\mu(Y-\mu)] + E\mu^2 = E(\mu)E(Y-\mu) + \sigma_0^2 = \sigma_0^2$$
$$EY^2 = E(Y-\mu)^2 + E\mu^2.$$

But

$$E(Y-\mu)^2 = E(E(Y-\mu)^2|\mu) = E\,\mathrm{var}(Y|\mu) = \mathrm{var}(Y|\mu) = \sigma^2,$$

by assumption. Hence

$$EY^2 = \sigma^2 + \sigma_0^2.$$

Next, by (ii) of Lemma 1,

$$\begin{aligned}
\operatorname{var}(\mu|Y) &= \operatorname{var}(\mu) - \operatorname{var}(E(\mu|Y)) \\
&= \sigma_0^2 - \operatorname{var}(\alpha Y) \\
&= \sigma_0^2 - \alpha^2 \operatorname{var}(Y) \\
&= \sigma_0^2 - \left(\frac{\sigma_0^2}{\sigma^2 + \sigma_0^2}\right)^2 \sigma^2 \\
&= \frac{\sigma^2 \sigma_0^2}{\sigma^2 + \sigma_0^2}.
\end{aligned}$$

□

Exercise 4. *Complete the proof of Lemma 2 for general n and μ_0.*

Hint: Observe that

$$E(\mu|Y_1, \ldots, Y_n) = E(\mu|Y_1 + \cdots + Y_n) = E(\mu|\overline{Y}).$$

Then replace (3) above by

$$(\overline{Y} - \mu_0) = (\overline{Y} - \mu) + (\mu - \mu_0),$$

observing, again, that the two terms on the right are independent, zero-mean normal random variables.

Exercise 5. *Let us now have some fun, as follows: Suppose that we have a random variable Y which, conditionally on μ, has law $\mathcal{N}(\mu, \sigma^2)$. Suppose that μ, conditionally on μ_0, has law $\mathcal{N}(\mu_0, \sigma_0^2)$. Suppose that μ_0, conditionally on μ_1, has law $\mathcal{N}(\mu_1, \sigma_1^2)$. Suppose that μ_1, conditionally on μ_2, has law $\mathcal{N}(\mu_2, \sigma_2^2)$. And so on. Can we take this ad infinitum? Under what conditions? What about the limit?*