CLARK UNIVERSITY 1887

A short introduction to Bayesian statistics, part V
Math 217 Probability and Statistics

Prof. D. Joyce, Fall 2014

Normal distributions with unknown variances. Let's now consider normal distributions with both unknown mean μ and unknown variance σ^2. We'll discover a way to find a natural family of conjugate priors $f(\mu, \sigma^2)$, one that works in general.

Before we get started though, let's use the *precision* θ, which is defined to be the reciprocal of the variance, $\theta = 1/\sigma^2$, as our second parameter for the family of normal distributions. It turns out that will simplify the mathematics considerably. Then we're looking for a family of density functions $f(\mu, \theta)$ to use for our prior and posterior densities.

Let $\mathbf{x} = (x_1, \ldots, x_n)$ be a random sample. Then, as always, the posterior density is determined from the prior by the proportion

$$f(\mu, \theta \,|\, \mathbf{x}) \propto f(\mathbf{x} \,|\, \mu, \theta)\, f(\mu, \theta)$$

Now, the prior is updated to the posterior by a factor proportional to $f(\mathbf{x} \,|\, \mu, \theta)$. In the following computations, anywhere σ^2 appears, it's just $1/\theta$.

$$\begin{aligned}
& f(\mathbf{x} \,|\, \mu, \theta) \\
= & \prod_{i=1}^n f(x_i \,|\, \mu, \theta) \\
= & \prod \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right) \\
= & \prod \sqrt{\tfrac{\theta}{2\pi}} \exp\left(-\tfrac{\theta}{2}(x_i-\mu)^2\right) \\
= & \left(\tfrac{\theta}{2\pi}\right)^{n/2} \exp\left(-\tfrac{\theta}{2}\sum(x_i-\mu)^2\right) \\
= & \left(\tfrac{\theta}{2\pi}\right)^{n/2} \exp\left(-\tfrac{\theta}{2}(\sum x_i^2 - 2\mu\sum x_i + \mu^2 n)\right)
\end{aligned}$$

Note that this factor depends only on three statistics from the sample, namely n, $\sum x_i$, and $\sum x_i^2$. A set of statistics of this sort are called *sufficient statistics*. That suggests that we should be able to find a 3-parameter family of conjugate priors for the joint distribution of (μ, θ). Let's let $\alpha = n$, $\beta = \sum x_i$, and $\gamma = \sum x_i^2$. (Note that since $n\sum x_i^2 - (\sum x_i)^2 \geq 0$, therefore $\alpha\gamma - \beta^2 \geq 0$.) Then we can write

$$f(\mu, \theta \,|\, \mathbf{x}) \propto \theta^{\alpha/2} \exp\left(-\tfrac{\theta}{2}(\gamma - 2\mu\beta + \mu^2\alpha)\right).$$

The symbol $[\mu, \theta; \alpha, \beta, \gamma]$. Let's introduce the notation

$$[\mu, \theta; \alpha, \beta, \gamma]$$

for the expression

$$\theta^{\alpha/2} \exp\left(-\tfrac{\theta}{2}(\gamma - 2\mu\beta + \mu^2\alpha)\right).$$

(We'll always assume that $\alpha\gamma - \beta^2 > 0$ when we use this expression.) That gives us the proportion

$$f(\mu, \theta \,|\, \mathbf{x}) \propto [\mu, \theta; n, \sum x_i, \sum x_i^2]\, f(\mu, \theta).$$

The symbol $[\mu, \theta; \alpha, \beta, \gamma]$ has a nice additive property (easily proved from its definition)

$$\begin{aligned}
& [\mu, \theta; \alpha_1, \beta_1, \gamma_1]\,[\mu, \theta; \alpha_2, \beta_2, \gamma_2] \\
= & [\mu, \theta; \alpha_1 + \alpha_2, \beta_1 + \beta_2, \gamma_1 + \gamma_2]
\end{aligned}$$

which corresponds to the statement that if you gather some data and update the distribution for (μ, θ), then you gather more data and update it again, you get the same thing as if you took all the data and just updated once.

The natural thing to do is just to take our prior to be proportional to this symbol, that is,

$$f(\mu, \theta) \propto [\mu, \theta; \alpha, \beta, \gamma]$$

for some particular triple of values α, β, and γ, but that requires the double integral

$$\int_0^\infty \int_{-\infty}^\infty [\mu, \theta; \alpha, \beta, \gamma]\, d\mu\, d\theta$$

to be finite (that finite number being the constant of proportionality). It will be so, at least when

$$\alpha\gamma - \beta^2 > 0 \quad \text{and} \quad \alpha > -\tfrac{1}{2},$$

as we'll see when we analyze this function.

Let's start with the inner integral which is proportional to the marginal density function for θ.

$$\begin{aligned} f(\theta) &\propto \int_{-\infty}^{\infty} [\mu, \theta; \alpha, \beta, \gamma]\, d\mu \\ &= \int \theta^{\alpha/2} \exp\left(-\tfrac{\theta}{2}(\gamma - 2\mu\beta + \mu^2\alpha)\right) d\mu \\ &= \theta^{\alpha/2} \int \exp\left(-\tfrac{\theta}{2}(\gamma - 2\mu\beta + \mu^2\alpha)\right) d\mu \end{aligned}$$

First, we'll complete the square of the expression $\gamma - 2\mu\beta + \mu^2\alpha$ in the exponent.

$$\begin{aligned} \gamma - 2\mu\beta + \mu^2\alpha &= \alpha\left(\mu^2 - \frac{2\beta}{\alpha}\mu + \frac{\gamma}{\alpha}\right) \\ &= \alpha\left(\left(\mu - \frac{\beta}{\alpha}\right)^2 + \frac{\alpha\gamma - \beta^2}{\alpha^2}\right) \end{aligned}$$

Our integral now looks like

$$\theta^{\alpha/2} \int \exp\left(-\tfrac{\theta\alpha}{2}\left((\mu - \tfrac{\beta}{\alpha})^2 + \tfrac{\alpha\gamma-\beta^2}{\alpha^2}\right)\right) d\mu$$

which we can rewrite as

$$\theta^{\alpha/2} \exp\left(-\tfrac{\theta(\alpha\gamma-\beta^2)}{2\alpha}\right) \int \exp\left(-\tfrac{\theta\alpha}{2}(\mu - \tfrac{\beta}{\alpha})^2\right) d\mu$$

and that looks something like the density function for a normal distribution. In fact, by adjusting the constants, we can make it the density for $\textsc{Normal}(\frac{\beta}{\alpha}, \frac{\sigma^2}{\alpha})$ as follows. Multiply the constant $\sqrt{\frac{2\pi}{\theta\alpha}} = \sqrt{2\pi\sigma^2/\alpha}$ outside the integral, and divide by it inside the integral. Then the integral part of the expression is

$$\int \frac{1}{\sqrt{2\pi\sigma^2/\alpha}} \exp\left(-\tfrac{\alpha}{2\sigma^2}(\mu - \tfrac{\beta}{\alpha})^2\right) d\mu$$

which is the integral of the density for a $\textsc{Normal}(\frac{\beta}{\alpha}, \frac{\sigma^2}{\alpha})$ distribution, and therefore equals 1. Therefore, what remains is the constant outside the integral

$$\theta^{\alpha/2} \exp\left(-\tfrac{\theta(\alpha\gamma-\beta^2)}{2\alpha}\right) \sqrt{\frac{2\pi}{\theta\alpha}}$$

Thus, we've found the inner integral. We can simplify it a bit and find that it's proportional to to

$$\theta^{(\alpha-1)/2} \exp\left(-\tfrac{\theta(\alpha\gamma-\beta^2)}{2\alpha}\right)$$

That expression is proportional to the marginal density for θ, but it's also proportional to a gamma distribution, namely $\textsc{Gamma}(\frac{\alpha\gamma-\beta^2}{2\alpha}, \frac{\alpha+1}{2})$, since a generic gamma distribution $\textsc{Gamma}(\lambda, r)$ has a density

$$f(x) = \frac{\lambda^r x^{r-1} e^{-\lambda x}}{\Gamma(r)}$$

so the $\textsc{Gamma}(\frac{\alpha\gamma-\beta^2}{2\alpha}, \frac{\alpha+1}{2})$ distribution has a density proportional to

$$f(x) = x^{(\alpha-1)/2} \exp\left(-\frac{\alpha\gamma - \beta^2}{2\alpha} x\right).$$

(Note that in order for this to be a valid gamma distribution, $\alpha\gamma - \beta^2$ has to be positive, but we're assuming that. We'll also have to assume that $\alpha > -1$.)

Summary. In summary, we have have found a family of conjugate priors for the family of normal distributions with unknown mean μ and precision θ. It is a three-parameter family parametrized by $[\alpha, \beta, \gamma]$. We found that if we take the prior distribution on the pair (μ, θ) to have a density proportional to $[\mu, \theta; \alpha_0, \beta_0, \gamma_0]$, then after observing the sample $\mathbf{x}$, the posterior distribution will be proportional to $[\mu, \theta; \alpha_0 + n, \beta_0 + \sum x_i, \gamma_0 + \sum x_i^2]$. As designed, α keeps track of the number of observations, β their sum, and γ the sum of their squares.

In the process of the computation, we discovered that the marginal distribution for θ is the gamma distribution $\textsc{Gamma}(\frac{\alpha\gamma-\beta^2}{2\alpha}, \frac{\alpha+1}{2})$. Before that, we discovered the conditional distribution for $\mu|\theta$ was a normal distribution $\textsc{Normal}(\frac{\beta}{\alpha}, \frac{1}{\theta\alpha})$, where $\sigma^2 = 1/\theta$, of course. Thus, the joint density $f(\mu, \theta)$ is the product $f(\mu|\theta) f(\theta)$ of a normal distribution and a gamma distribution.

The means and variances of μ, θ, and σ^2 for $[\alpha, \beta, \gamma]$. Suppose that we have one of the $[\alpha, \beta, \gamma]$ joint distribution for (μ, θ). Since the marginal distribution for the precision θ is the gamma distribution

$$\text{Gamma}(\lambda, r) = \text{Gamma}(\tfrac{\alpha\gamma - \beta^2}{2\alpha}, \alpha + \tfrac{1}{2}),$$

therefore the mean and variance of θ are

$$\begin{aligned} E(\theta) &= \frac{r}{\lambda} = \frac{\alpha(\alpha+1)}{\alpha\gamma - \beta^2} \\ \text{Var}(\theta) &= \frac{r}{\lambda^2} = \frac{2\alpha^2(\alpha+1)}{(\alpha\gamma - \beta^2)^2} \end{aligned}$$

You can also find the mean and variance of σ^2 by direct computations of integrals, and they turn out to be

$$\begin{aligned} E(\sigma^2) &= \frac{\lambda}{r-1} = \frac{\alpha\gamma - \beta^2}{\alpha(\alpha-1)} \\ \text{Var}(\sigma^2) &= \frac{\lambda^2}{(r-1)^2(r-2)} \\ &= \frac{2(\alpha\gamma - \beta^2)^2}{\alpha^2(\alpha-1)(\alpha-3)} \end{aligned}$$

Since the conditional distribution for $\mu|\theta$ is $\text{Normal}(\frac{\beta}{\alpha}, \frac{1}{\theta\alpha})$, therefore the conditional expectation is $E(\mu|\theta) = \beta/\alpha$, but that doesn't depend on θ, so $E(\mu) = \beta/\alpha$. The conditional variance is $\text{Var}(\mu|\theta) = 1/(\theta\alpha)$, so the variance is

$$\text{Var}(\mu) = E(\tfrac{1}{\theta\alpha}) = \tfrac{1}{\alpha} E(\tfrac{1}{\theta}) = \tfrac{1}{\alpha} E(\sigma^2) = \frac{\alpha\gamma - \beta^2}{\alpha^2(\alpha-1)}.$$

Incidentally, the marginal distribution of μ (not the conditional $\mu|\theta$) is a scaled T-distribution.

In summary, the means and variances of μ and σ^2 in terms of $[\alpha, \beta, \gamma]$ are

$$\begin{aligned} E(\mu) &= \frac{\beta}{\alpha} \\ \text{Var}(\mu) &= \frac{\alpha\gamma - \beta^2}{\alpha^2(\alpha-1)} \\ E(\sigma^2) &= \frac{\alpha\gamma - \beta^2}{\alpha(\alpha-1)} \\ \text{Var}(\sigma^2) &= \frac{2(\alpha\gamma - \beta^2)^2}{\alpha^2(\alpha-1)(\alpha-3)} \end{aligned}$$

Note that $\text{Var}(\mu)$ and $E(\sigma^2)$ differ only by a factor of α.

The two expectation equations give Baysian point estimators $\hat{\mu} = E(\mu) = \frac{\beta}{\alpha}$ and $\hat{\sigma}^2 = E(\sigma^2) = \frac{\alpha\gamma - \beta^2}{\alpha(\alpha-1)}$ for the unknown parameters μ and σ^2.

Selecting the prior distribution. A know-nothing prior has $[\alpha_0, \beta_0, \gamma_0] = [0, 0, 0]$. As usual, this is not an actual distribution, but after observing $\mathbf{x}$, it gives the posterior distribution $[n, \sum x_i, \sum x_i^2]$. Indeed, this won't be an actual distribution until $n > 0$, and it won't be until $n > 2$ that $\text{Var}(\sigma^2)$ is finite. With this know-nothing prior, the estimators $\hat{\mu}$ and $\hat{\sigma}^2$ are

$$\begin{aligned} \hat{\mu} &= \frac{\beta}{\alpha} = \frac{\sum x_i}{n} = \overline{x} \\ \hat{\sigma}^2 &= \frac{\alpha\gamma - \beta^2}{\alpha(\alpha-1)} = \frac{n\sum x_i^2 - (\sum x_i)^2}{n(n-1)} = s^2 \end{aligned}$$

where s^2 is the sample variance $\frac{1}{n-1}\sum(x_i - \overline{x})^2$.

You may, however, have a preconception about the distributions of μ and σ^2, for instance, you might think that μ has a particular mean and variance, and you have an idea that σ^2 has some particular mean. By the equations above, you can determine from these three values what $[\alpha, \beta, \gamma]$ should be for the prior distribution that your preconceptions suggest.

A four-parameter family of conjugate priors is sometimes used to accommodate four preconceived values, but three parameters are enough to get the job done.

Math 217 Home Page at
`http://math.clarku.edu/~djoyce/ma217/`