Outline

The Unknown σ Model

The model is

$$Y_i = \mu + \epsilon_i, \;\; \epsilon_i \sim N(0, \sigma^2), \; i = 1, 2, \ldots, n.$$

We will assume we know μ and want inference for σ.

Given we know μ, we act as if we observe the errors

$$Y_i - \mu = \epsilon_i \sim N(0, \sigma^2), i = 1, 2, \ldots, n.$$

The χ^2 Distribution

Recall the χ^2 distribution.

If $Z_i \sim N(0,1), i = 1,2,\ldots,\nu$, then

$$X = \sum_{i=1}^{\nu} Z_i^2 \sim \chi_\nu^2.$$

$E(X) = \nu.$

The pdf of X is

$$f(x) = \frac{\left(\frac{1}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)}\, x^{\frac{\nu}{2}-1}\, e^{-\frac{x}{2}}.$$

Conjugate Prior for Normal σ

A prior for a standard deviation (or, equivalently, the variance) we will use a lot is

$$\sigma^2 \sim \frac{\nu \, \lambda}{\chi^2_\nu}.$$

This is called an inverse-χ^2 distribution for obvious reasons.

Lot's of people work with σ^2 as the variable but I prefer to work with σ. I don't like having a variable "x^2" and σ is actually the more interpretable quantitity.

I need the pdf of σ.

$$\sigma^2 \sim \frac{\nu\,\lambda}{X},\; X \sim \chi^2_\nu.$$

$$x = \frac{\nu\,\lambda}{\sigma^2}, \qquad |\frac{dx}{d\sigma}| = \frac{2\nu\lambda}{\sigma^3}.$$

$$p(\sigma) = p_X(x(\sigma)) \;\times\; |\frac{dx}{d\sigma}| =$$

$$p(\sigma) = \frac{(\frac{1}{2})^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})}\;\left(\frac{\nu\,\lambda}{\sigma^2}\right)^{\frac{\nu}{2}-1}\; e^{-\frac{\nu\lambda}{2\sigma^2}} \;\times\; \frac{2\nu\lambda}{\sigma^3}$$

$$= \frac{(\nu\lambda)^{\frac{\nu}{2}}}{2^{\frac{\nu}{2}-1}\,\Gamma(\frac{\nu}{2})}\;\sigma^{-(\nu+1)}\; e^{-\frac{\nu\lambda}{2\sigma^2}}.$$

The Likelihood

$$p(\epsilon_1, \epsilon_2, \ldots, \epsilon_n \mid \sigma) = \Pi_{i=1}^{n} (2\pi)^{-1/2} \, \sigma^{-1} \, e^{-\frac{\epsilon_i^2}{2\sigma^2}}.$$

Let $S = \sum_{i=1}^{n} \epsilon_i^2$.

$$L(\sigma) \propto \sigma^{-n} \, e^{-\frac{S}{2\sigma^2}}.$$

The Posterior

$$p(\sigma \mid \epsilon_1, \epsilon_2, \ldots, \epsilon_n) = p(\sigma \mid S)$$

$$\propto L(\sigma) \; \times \; p(\sigma)$$

$$\propto \sigma^{-n} \, e^{-\frac{S}{2\sigma^2}} \; \times \; \sigma^{-(\nu+1)} \; e^{-\frac{\nu\lambda}{2\sigma^2}}$$

$$= \sigma^{-(\nu+n+1)} \; e^{-\frac{\nu\lambda+S}{2\sigma^2}}$$

Let

$$\nu' = \nu + n, \ \ \nu' \lambda' = \nu\lambda + S.$$

then,

$$\sigma^2 \mid \text{data} \sim \frac{\nu' \lambda'}{\chi^2_{\nu'}}.$$

So, the prior is indeed conjugate with remarkably simple updates for the parameters ν and λ.

Ball-parking the Prior

Since $E(\chi^2_\nu) = \nu$, for large ν we have

$$\sigma^2 \sim \frac{\nu\,\lambda}{\chi^2_\nu} \approx \lambda.$$

Large ν means a "tight" prior.

So, you can roughly pick the prior by choosing $\sqrt{\lambda}$ to be your "guess" at σ and ν "small enough" to capture your uncertainty.

Prediction

To predict the next ϵ, we need to compute

$$
\begin{aligned}
& p(\epsilon \mid \epsilon_1, \epsilon_2, \ldots, \epsilon_n) \\
& \quad = \int p(\epsilon, \sigma \mid \epsilon_1, \epsilon_2, \ldots, \epsilon_n) \, d\sigma \\
& \quad = \int p(\epsilon \mid \sigma, \epsilon_1, \epsilon_2, \ldots, \epsilon_n) \, p(\sigma \mid \epsilon_1, \epsilon_2, \ldots, \epsilon_n) \, d\sigma \\
& \quad = \int p(\epsilon \mid \sigma) \, p(\sigma \mid S) \, d\sigma.
\end{aligned}
$$

Let's just use ν and λ to denote the distribution of σ since we are using the conjugate prior.

$$
\begin{aligned}
&\propto \int \left(\frac{1}{\sigma}\, e^{-\frac{1}{\sigma^2}\epsilon^2}\right)\left(\sigma^{-(\nu+1)}\, e^{-\frac{\nu\lambda}{2\sigma^2}}\right)\, d\sigma \\
&\propto \int \sigma^{-(\nu+1+1)}\, e^{-\frac{1}{2\sigma^2}(\epsilon^2+\nu\lambda)}\, d\sigma \\
&\propto (\nu\lambda+\epsilon^2)^{-\frac{\nu+1}{2}} \\
&\propto (1+\frac{1}{\nu}(\frac{\epsilon}{\sqrt{\lambda}})^2)^{-\frac{\nu+1}{2}}
\end{aligned}
$$

Now note:

(i)

If X has a t distribution with ν degrees of freedom then

$$X \sim t_\nu \Rightarrow f(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} (1 + \frac{x^2}{\nu})^{-\frac{\nu+1}{2}}$$

(ii)

If $Y = c\,X$ then

$$f_Y(y) = \frac{1}{c}\, f_X(\frac{y}{c}).$$

Hence

$$\epsilon/\sqrt{\lambda} \sim t_\nu, \text{or, } \epsilon = \sqrt{\lambda}\, t_\nu.$$

An alternative derivation:

$$\epsilon = \sigma Z,\ Z \sim N(0,1),\ \sigma^2 = \frac{\nu\lambda}{\chi^2_\nu}.$$

so,

$$\epsilon = \sqrt{\lambda}\,\frac{Z}{\sqrt{\chi^2_\nu/\nu}} \sim \sqrt{\lambda}\,t_\nu.$$

An alternative computation:

Suppose we have $p(\theta)$ that we can draw from and $p(y \mid \theta)$ that we can draw from.

We can always draw from the marginal distribution of Y by drawing (θ, y) from the joint and then discarding θ.

We can draw from the joint by drawing $\theta \sim p(\theta)$ and then $y \sim p(y \mid \theta)$.

In this case we draw $\sigma = \sqrt{\frac{\nu\lambda}{\chi^2_\nu}}$ and then $\epsilon = \sigma Z$, $Z \sim N(0, 1)$.

You might notice your are drawing a t, or, you might not!