STANDARD NORMAL LINEAR REGRESSION

1. Bayesian Mechanics for the Standard Normal Linear [Regression] Model: SNLM

$$
\begin{aligned}
&Y = X\beta + \varepsilon \\
&p(\underset{T\times 1}{Y} \mid \underset{T\times k}{X}, \beta, \sigma^2) = \varphi(Y - X\beta; \sigma^2 I) \\
&\qquad = (2\pi)^{-T/2}\sigma^{-T} \exp\left(-\frac{(Y - X\beta)'(Y - X\beta)}{2\sigma^2}\right)
\end{aligned}
$$

- A normal marginal (i.e. a normal prior) for $\beta \mid \left\{X, \sigma^2\right\}$ will be convenient.
- It will make $\beta \mid \left\{Y, X, \sigma^2\right\}$ itself normal.

2. Conjugate prior

- *Really* convenient prior: $\pi(\beta, \sigma^2 \mid X) \propto \phi(Y^* \mid X^*, \beta, \sigma^2)$.
- This makes $\pi \cdot p$ have the same form as

$$
p\left(\begin{bmatrix} Y \\ Y^* \end{bmatrix} \middle| \begin{bmatrix} X \\ X^* \end{bmatrix}, \beta, \sigma^2\right) ,
$$

  i.e., the joint pdf of data and parameters has the same form as the likelihood for a data set expanded by the "dummy observations" (Y^*, X^*).
- Convenient both mechanically and intuitively. Prior formulated as if one has "observations" based on pre-sample knowledge. A prior like this, that has the same form as a likelihood function from a sample using the same model, is known as a **conjugate** prior.

3. Marginal for σ^2, assuming dummy observations in Y, X

$$u(\beta) = Y - X\beta$$
$$\hat{\beta} = (X'X)^{-1}X'Y, \text{ the \textbf{OLS estimator}}$$
$$\hat{u} = u(\hat{\beta})$$
$$v = 1/\sigma^2$$
$$s^2 = \hat{u}'\hat{u}$$

4. Posterior

$$\sigma^{-T} \exp\left(-\frac{\hat{u}'\hat{u} + (\beta - \hat{\beta})'X'X(\beta - \hat{\beta})}{2\sigma^2}\right) d\beta \, d\sigma$$
$$= \sigma^{-T} \exp\left(-\frac{\hat{u}'\hat{u}}{2\sigma^2}\right) \left(\sigma^k \left|X'X\right|^{-\frac{1}{2}}\right)$$
$$\sigma^{-k} \left|X'X\right|^{\frac{1}{2}} \exp\left(-\frac{(\beta - \hat{\beta})'X'X(\beta - \hat{\beta})}{2\sigma^2}\right) d\beta \, d\sigma$$
$$\propto \sigma^{-T+k} \left|X'X\right|^{-\frac{1}{2}} \exp\left(-\frac{\hat{u}'\hat{u}}{2\sigma^2}\right) \varphi\left(\beta - \hat{\beta}; \sigma^2 (X'X)^{-1}\right) d\beta \, d\sigma$$
$$\propto v^{(T-k)/2} \exp\left(\frac{\hat{u}'\hat{u}}{2} v\right) \varphi\left(\beta - \hat{\beta}; \sigma^2 (X'X)^{-1}\right) d\beta \, \frac{dv}{v^{3/2}} \, .$$

5. Integrating to get the marginal on $v = 1/\sigma^2$

Integrating this expression w.r.t. β and setting $\alpha = \hat{u}'\hat{u}/2$ gives us an expression proportional to

$$v^{(T-k-3)/2} \exp\left(-\frac{\hat{u}'\hat{u}}{2} v\right) dv \propto \alpha^{(T-k-1)/2} v^{(T-k-3)/2} e^{-\alpha v} dv \, ,$$

which is a standard $\Gamma((T-k-1)/2, \alpha)$ pdf. Because it is $v = 1/\sigma^2$ that has the gamma distribution, we say that σ^2 itself has an **inverse-gamma** distribution.

6. Marginal on β

Rewrite the likelihood:

$$v^{(T-3)/2} \exp\left(-\frac{1}{2} u(\beta)' u(\beta) v\right) dv \, d\beta \, .$$

As a function of v, this is proportional to a standard $\Gamma\big((T-1)/2, u(\beta)'u(\beta)/2\big)$ pdf, but here there is a missing normalization factor that depends on β. When we integrate with respect to v, therefore, we arrive at

$$\left(\frac{u(\beta)'u(\beta)}{2}\right)^{-(T-1)/2} d\beta \quad \propto \quad \left(1 + \frac{(\beta-\hat{\beta})'X'X(\beta-\hat{\beta})}{s^2}\right)^{-(T-1)/2} d\beta\,.$$

7. Remarks on the marginal pdf for β

$$\propto \text{multivariate } t_n\big(0, (s^2/n)(X'X)^{-1}\big)$$
$$(n = T-k-1: \text{ the } \textbf{degrees of freedom})$$
$$\beta_i \sim t_n(\hat{\beta}, s_\beta^2),\; s_\beta^2 = (s^2/n)(X'X)_{ii}^{-1}$$

General form of the $t_n(\mu;\Sigma)$ pdf, with argument ξ:

$$\frac{\kappa\,|\Sigma|^{-1/2}}{\big(n + \xi'\Sigma^{-1}\xi\big)^{(n+k)/2}}$$

where the normalizing constant κ is given by

$$\frac{\Gamma((n+k)/2)}{\pi^{n/2}\Gamma(n/2)}$$

8. Choosing a prior

The most common framework for Bayesian analysis of this model asserts a prior that is flat in β and $\log\sigma$ or $\log\sigma^2$, i.e. $d\sigma/\sigma$ or $d\sigma^2/\sigma^2$. However, there are arguments in favor of other improper priors as a starting point, most prominently for using $d\sigma/\sigma^{k+1}$. This latter is the prior to which the reasoning behind **Jeffreys priors** leads. Jeffreys himself, though, favored the $d\sigma/\sigma$ prior for this model. You are not expected for this course to learn how Jeffreys priors are derived. We will assume the prior has the form $d\sigma/\sigma^p$, then discuss how the results depend on p.

9

The prior arising from the multivariate Jeffreys analysis, $p = k+1$, therefore gives a $\Gamma(T/2, \alpha)$ pdf for v, regardless of k. The prior more usually called a Jeffreys prior, $d\sigma/\sigma$, produces a $\Gamma((T-k)/2, \alpha)$ distribution for v. The number $T-k$ is what is in this model called the **degrees of freedom**. Note that unless there are positive degrees of freedom, the $X'X$ matrix will not be invertible, the prior times the likelihood will therefore not be integrable in β, and the derivation we have just given does not go through. Because it is $v = 1/\sigma^2$ that has the Γ distribution, we say that σ^2 itself has an **inverse-gamma** distribution. Since a $\Gamma(n/2, 1)$ variable, multiplied by 2, is a $\chi^2(n)$ random variable, some prefer to say that $\hat{u}'\hat{u}/\sigma^2$ has a $\chi^2(T-k)$ distribution, and thus that σ^2 has an inverse-chi-squared distribution.

10. Using the multivariate t

- Individual elements of a vector that has a multivariate $t_n(\mu, \Sigma)$ distribution have a univariate $t_n(\mu_i, \Sigma_{ii})$ distribution.
- Therefore
$$P[\beta_i < a] = P\left[\tau < \frac{a - \hat{\beta}_i}{\sqrt{\frac{s^2}{T-k}(X'X)^{-1}_{ii}}}\right] ,$$
  where τ is a $t_{T-k}(0, 1)$ variate. This probability can be looked up in a table in a reference book or evaluated with a call to a standard function in many programs. (In R, $P[\tau < a]$ is `pt(a,T-k)`.)
- More generally, if c is a $q \times k$ matrix of constants and $\tau \sim t_n(\mu, \Sigma)$, then $c\tau \sim t_n(c\mu, c\Sigma c')$. In particular, any linear combination of β's has a univariate t distribution, so probabilities can be calculated for the linear combination just as for the single-coefficient case already discussed.