TESTS, MODEL CRITICISM

1. Old business

- Lancaster's definition of a natural conjugate prior: It implies that any prior pdf that has the form $\pi_0(\beta)\ell(Y^*, \theta)$, where $\ell(Y^*, \theta)$ is what we have called a conjugate prior, is a "Lancaster-conujugate" prior, regardless of what π_0 is. This is not a standard definition, and I think it is unintentionally broad. Lancaster's examples, as far as I've looked at them, all fit our lecture definition.
- Gelman, Carlin, Stern and Rubin label Lancaster's "natural conjugate" priors just plain "conjugate" priors, and use "natural conjugate" to refer to priors that are in the same family as the likelihood function. Probably this is the best usage. So our definition of "conjugate" in lecture should instead be labeled "natural conjugate", and Lancaster should revise his book to make the distinction.

2. "Significant" and "insignificant" results

- What we say here applies about equally to confidence sets and to minimum-size posterior probability sets.
- There is a big difference between a result that posterior probability is concentrated in a small (from the point of view of the substance of the problem) region around $\beta = 0$ and the result that the sample data is so uninformative that the posterior probability is spread widely, with a 95% HPD region therefore including $\beta = 0$.
- The former says we are quite sure that β is substantively small. The latter says β could be very big, indeed from looking at the data alone seems more likely to be big in absolute value than small in absolute value.
- Yet it is not uncommon to see one regression study, which found an "insignificant" effect of a variable X, cited as contradicting another study which found a "significant" effect, without any attention to what the probability intervals were and the degree to which they overlap.

3. A general method for constructing confidence regions

- For each $\theta \in \Theta$, choose a test statistic $T(Y, \theta)$, where Y is the observable data.
- For each θ choose a **rejection region** $R(\theta) \subset \Theta$ such that $P[T(Y, \theta) \in R(\theta) \mid \theta] = \alpha$.
- Define $S(Y) = \{\theta \mid T(Y, \theta) \notin R(\theta)\}$.
- Then $S(Y)$ is a $100(1 - \alpha)\%$ confidence region for θ.

4. Remarks on the general method

- Regions constructed this way are exact confidence regions.
- When Y is continuously distributed, it is always possible to find $T(Y,\theta)$'s and $R(\theta)$'s to implement this idea.
- It is generally not possible to use this idea to produce confidence regions for individual elements of the θ vector or linear combinations of them.
- There are obviously many ways to choose the T and R functions, and thus many confidence regions, for a given α value.

5. Pivotal quantities

- Sometimes we can find a collection of functions $T^*(Y,\theta)$ with the property that the distribution of $\{T^*(Y,\theta) \mid \theta\}$ does not depend on θ. In this case we call T^* a **pivotal quantity** or **pivot** for θ.
- Constructing confidence regions based on pivots is particularly easy: One defines a single region R^* such that $P[T^*(Y,\theta) \in R] = \alpha$, and this rejection region defines the test used to construct the confidence region for all values of θ. Most confidence regions actually used in practice are based on pivots.
- Certain kinds of pivots also make confidence regions based on them correspond to posterior probability regions with probabilities matching the confidence levels under certain flat priors.
- A leading special case is the SNLM, in which
$$\frac{\hat{u}'\hat{u}}{\sigma^2} \text{ and } \frac{(\beta-\hat{\beta})'X'X(\beta-\hat{\beta})}{\hat{u}'\hat{u}}$$
are a pair of jointly pivotal quantities and confidence regions based on the distribution of these statistics conditional on β, σ turn out to coincide with posterior probability regions with probabilities corresponding to $1-\alpha$ if the $(1/\sigma)d\sigma\, d\beta$ flat prior is used.
- Other cases: scale parameters in general, like α in the Gamma or σ in the t distribution, allow this kind of construction, as do location parameters in general, like μ in the t distribution.
- Most confidence regions in actual use are not only based on pivots, they are based on location-scale pivots that allow this flat-prior Bayesian interpretation.

6. Tests

- Each $T(Y,\theta)$, $R(\theta)$ pair in our construction of a confidence region makes up what is known as a **statistical test** of the **null hypothesis** that θ is the true value of the parameter. The parameter α is known as the **significance level** or **size** of the test.
- So an exact confidence region can always be interpreted as a collection of statistical tests with significance level α.

- More generally, we can consider tests of hypotheses that do not consist of a single point in Θ. For such compound hypotheses, DeGroot and Schervish distinguish level, or significance level, and size. The null hypothesis is some set $\Omega_0 \in \Theta$ and the test still takes the form of a statistic $T(Y)$ and rejection region R. Only now it is possible that $P[T(Y) \in R \mid \theta]$ differs across θ's in Ω_0. The standard definitions now say that the test has significance level α if $P[T(Y) \in R \mid \theta] \leq \alpha$ for all $\theta \in \Omega_0$, and that it has size α if it has level γ for all $\gamma \geq \alpha$.

7. POWER

- The **power function** of a test is $P[T(Y) \in R \mid \theta]$ considered as a function of θ over $\Theta - \Omega_0$. (It could be extended to range over Ω_0 also.)
- We would like a test to have a small size and have a large power function.
- We would like a test to be **unbiased**, meaning that the infimum of $P[T(Y) \in R \mid \theta]$ over $\Theta - \Omega_0$ is no smaller than the supremum over Ω_0 of the same thing.