

# Confidence & Credible Regions

STA 721: Lecture 15

Merlise Clyde (clyde@duke.edu)

Duke University



# Outline

- Confidence Intervals from Test Statistics
- Pivotal Quantities
- Confidence intervals for parameters
- Prediction Intervals
- Bayesian Credible Regions and Intervals

## Readings:

- Christensen Appendix C, Chapter 3



# Goals

For the regression model  $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$  we usually want to do more than just testing that  $\boldsymbol{\beta}$  is zero

- what is a plausible range for  $\beta_j$ ?
- what is a plausible set of values for  $\beta_j$  and  $\beta_k$ ?
- what is a plausible range of values for  $\mathbf{x}^T \boldsymbol{\beta}$  for a particular  $\mathbf{x}$ ?
- what is a plausible range of values for  $\mathbf{Y}_{n+1}$  for a given value of  $\mathbf{x}_{n+1}$ ?

Look at confidence intervals, confidence regions, prediction regions and Bayesian credible regions/intervals



# Confidence Sets

For a random variable  $\mathbf{Y} \sim \mathbf{P} \in \{P_{\theta} : \theta \in \Theta\}$

## ▼ Definition: Confidence Region

A set valued function  $C$  is a  $(1 - \alpha) \times 100\%$  confidence region for  $\theta$  if

$$P_{\theta}(\{\theta \in C(\mathbf{Y})\}) = 1 - \alpha \forall \theta \in \Theta$$

- In this case we say  $C(Y)$  is a  $1 - \alpha$  confidence region for the parameter  $\theta$
- there is some true value of  $\theta$ , and the confidence region will cover it with probability  $1 - \alpha$  no matter what it is.
- the randomness is due to  $\mathbf{Y}$  and  $C(\mathbf{Y})$
- once we observe  $\mathbf{Y}$  everything is fixed, so region may not include the true  $\theta$



# Hypothesis Tests and Rejection/Acceptance Regions

Recall for a level  $\alpha$  test of a point null hypothesis

- we reject  $H$  with probability  $\alpha$  when  $H$  is true
- for each test we can construct:
  - a rejection region  $R(\boldsymbol{\theta}) \subset \mathcal{Y}$ , the  $Y$  values for which we reject  $H$
  - an acceptance region  $A(\boldsymbol{\theta}) \subset \mathcal{Y}$ , the  $Y$  values for which we accept  $H$
- these sets are complements of each other (for non-randomized tests)

$$\Pr(\mathbf{Y} \in A(\boldsymbol{\theta}) \mid \boldsymbol{\theta}) = 1 - \alpha$$



# Duality of Hypothesis-Testing/Confidence Regions

Suppose we have a level  $\alpha$  test for every possible value of  $\theta$

- for each  $\theta \in \Theta$ , let  $A(\theta)$  be the acceptance region of the test  $\mathbf{Y} \sim P_\theta$
- then  $P(\mathbf{Y} \in A(\theta) \mid \theta) = 1 - \alpha$  for each  $\theta \in \Theta$
- This collection of hypothesis tests can be “inverted” to construct a confidence region for  $\theta$ , as follows:
- define  $C(\mathbf{Y}) = \{\theta \in \Theta : \mathbf{Y} \in A(\theta)\}$
- this is the set of  $\theta$  values that are not rejected when  $\mathbf{Y} = \mathbf{y}$  is observed
- then  $C$  is a  $1 - \alpha$  confidence region for  $\theta$



# Confidence Intervals for Regression Parameters

For the linear model  $\mathbf{Y} \sim \mathbf{N}(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$ , confidence intervals for  $\beta_j$  can be constructed from inverting the appropriate  $t$ -test.



# Acceptance Region & Confidence Interval





# Confidence Intervals for Linear Functions

For a linear function of the parameters  $\lambda = \mathbf{a}^T \boldsymbol{\beta}$  we can construct a confidence interval by inverting the appropriate  $t$ -test

- most important example  $\mathbf{a}^T \boldsymbol{\beta} = \mathbf{x}^T \boldsymbol{\beta} = \mathbf{E}[\mathbf{Y} \mid \mathbf{x}]$
- suppose you are testing  $H : \mathbf{a}^T \boldsymbol{\beta} = m$
- If  $H$  is true,  $\mathbf{a}^T \boldsymbol{\beta} - m \sim \mathbf{N}(0, \sigma^2 v)$  where  $v = \mathbf{a}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{a}$
- $s^2 \sim \sigma^2 \chi_{n-p}^2 / (n - p)$  independent of  $\mathbf{a}^T \hat{\boldsymbol{\beta}}$
- then  $t = \frac{\mathbf{a}^T \hat{\boldsymbol{\beta}} - m}{s \sqrt{v}} \sim t_{n-p}$
- a  $1 - \alpha$  confidence interval for  $\mathbf{a}^T \boldsymbol{\beta}$  is

$$\mathbf{a}^T \hat{\boldsymbol{\beta}} \pm s \sqrt{v} t_{n-p, 1-\alpha/2}$$



# Prediction Regions and Intervals

Related to CI for  $E[Y \mid \mathbf{x}] = \mathbf{x}^T \boldsymbol{\beta}$ , we may wish to construct a prediction interval for a new observation  $Y^*$  at  $\mathbf{x}_*$

- a  $1 - \alpha$  prediction interval for  $Y^*$  is a set valued function of  $\mathbf{Y}$ ,  $C(\mathbf{Y})$  such that

$$\Pr(\mathbf{Y}^* \in C(\mathbf{Y}) \mid \boldsymbol{\beta}, \sigma^2) = 1 - \alpha$$

where the distribution is computed using the distribution of  $\mathbf{Y}^*$

- this use the idea of a *pivotal quantity*: a function of the data and the parameters that has a known distribution that does not depend on any unknown parameters.
- for prediction,  $Y^* = \mathbf{x}_*^T \boldsymbol{\beta} + \boldsymbol{\epsilon}^*$  where  $\boldsymbol{\epsilon}^* \sim N(0, \sigma^2)$  independent of  $\boldsymbol{\epsilon}$

$$E[Y^* - \mathbf{x}_*^T \hat{\boldsymbol{\beta}}] = \mathbf{x}_*^T \boldsymbol{\beta} - \mathbf{x}_*^T \hat{\boldsymbol{\beta}} = 0$$

$$\text{Var}(Y^* - \mathbf{x}_*^T \hat{\boldsymbol{\beta}}) = \text{Var}(\boldsymbol{\epsilon}^*) + \text{Var}(\mathbf{x}_*^T \hat{\boldsymbol{\beta}}) = \sigma^2 + \sigma^2 \mathbf{x}_*^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_*$$

$$Y^* - \mathbf{x}_*^T \hat{\boldsymbol{\beta}} \sim N(0, \sigma^2 (1 + \mathbf{x}_*^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_*))$$



# Pivotal Quantity and Prediction Intervals

Since  $\hat{\beta}$  and  $s^2$  are independent, we can construct a pivotal quantity for  $Y^* - \mathbf{x}_*^T \hat{\beta}$ :

$$\frac{Y^* - \mathbf{x}_*^T \hat{\beta}}{s \sqrt{1 + \mathbf{x}_*^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_*}} \sim t_{n-p}$$

- therefore

$$\Pr \left( \frac{|Y^* - \mathbf{x}_*^T \hat{\beta}|}{s \sqrt{1 + \mathbf{x}_*^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_*}} < t_{n-p, 1-\alpha/2} \right) = 1 - \alpha$$

- Rearranging gives a  $1 - \alpha$  prediction interval for  $Y^*$ :

$$\mathbf{x}_*^T \hat{\beta} \pm s \sqrt{1 + \mathbf{x}_*^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_*} t_{n-p, 1-\alpha/2}$$



# Joint Confidence Regions for $\beta$

- we can construct a joint confidence region for  $\beta$  based on inverting a test  $H : \beta = \beta_0$ . Recall:

$$\begin{aligned}\hat{\beta} - \beta &\sim \mathbf{N}(0, \sigma^2(\mathbf{X}^T \mathbf{X})^{-1}) \\ (\mathbf{X}^T \mathbf{X})^{-1/2}(\hat{\beta} - \beta) &\sim \mathbf{N}(0, \sigma^2 \mathbf{I}) \\ (\hat{\beta} - \beta)^T (\mathbf{X}^T \mathbf{X})^{-1} (\hat{\beta} - \beta) &\sim \sigma^2 \chi_p^2\end{aligned}$$



# Bayesian Credible Regions

- In a Bayesian setting, we have a posterior distribution for  $\beta$  given the data  $\mathbf{Y}$
- a set  $C \in \mathbb{R}^p$  is a  $1 - \alpha$  posterior credible region (sometimes called a Bayesian confidence region) if  $\Pr(\beta \in C \mid \mathbf{Y}) = 1 - \alpha$
- lots of sets have this property, but we usually want the most probable values of  $\beta$  given the data
- this motivates looking at the highest posterior density (HPD) region which is a  $1 - \alpha$  credible set  $C$  such that the values in  $C$  have higher posterior density than those outside of  $C$
- the HPD region is the smallest region that contains  $1 - \alpha$  of the posterior probability



# Bayesian Credible Regions

- For a normal prior and normal likelihood, the posterior for  $\beta$  conditional on  $\sigma^2$  is normal with say posterior mean  $\mathbf{b}_n$  and posterior precision  $\mathbf{\Phi}_n$
- the posterior density as a function of  $\beta$  for a fixed  $\sigma^2$  is

$$p(\beta \mid \mathbf{Y}) \propto \exp \left\{ -(\beta - \mathbf{b}_n)^T \mathbf{\Phi}_n (\beta - \mathbf{b}_n) / 2 \right\}$$

- so a highest posterior density region has the form

$$C = \{ \beta : (\beta - \mathbf{b}_n)^T \mathbf{\Phi}_n^{-1} (\beta - \mathbf{b}_n) < q \}$$

$$\beta - \beta_n \mid \sigma^2 \sim \mathbf{N}(0, \mathbf{\Phi}_n^{-1})$$

$$\mathbf{\Phi}_n^{1/2} (\beta - \beta_n) \mid \sigma^2 \sim \mathbf{N}(0, \mathbf{I})$$

$$(\beta - \beta_n)^T \mathbf{\Phi}_n (\beta - \beta_n) \mid \sigma^2 \sim \chi_p^2$$



# Bayesian HPD Regions For Unknown $\sigma^2$

- For unknown  $\sigma^2$  we need to integrate out  $\sigma^2$  to get the marginal posterior for  $\beta$
- for conjugate priors,  $\beta \mid \phi \sim \mathbf{N}(\mathbf{b}_0, (\phi \Phi_0)^{-1})$  and  $\phi \sim \mathbf{G}(a_0/2, b_0/2)$ , then

$$\begin{aligned}\beta \mid \phi, \mathbf{Y} &\sim \mathbf{N}(\mathbf{b}_n, (\phi \Phi_n)^{-1}) \\ \phi \mid \mathbf{Y} &\sim \mathbf{G}(a_n/2, b_n/2) \\ \beta \mid \mathbf{Y} &\sim \text{St}(a_n, \mathbf{b}_n, \hat{\sigma}^2 \Phi_n^{-1})\end{aligned}$$

where  $\text{St}(a_n, \mathbf{b}_n, \hat{\sigma}^2 \Phi_n^{-1})$  is a multivariate Student-t distribution with  $a_n$  degrees of freedom location  $\mathbf{b}_n$  and scale matrix  $\hat{\sigma}^2 \Phi_n^{-1}$  with  $\hat{\sigma}^2 = b_n/a_n$

- density of  $\beta$  is

$$p(\beta \mid \mathbf{Y}) \propto \left( 1 + \frac{(\beta - \mathbf{b}_n)^T \Phi_n (\beta - \mathbf{b}_n)}{a_n \hat{\sigma}^2} \right)^{-(a_n+p)/2}$$



# Reference Posterior Distribution

For the reference prior  $\pi(\boldsymbol{\beta}, \phi) \propto 1/\phi$  and the likelihood  $p(\mathbf{Y} \mid \boldsymbol{\beta})$ , the posterior is proportional to the likelihood times  $\phi^{-1}$

- (generalized) posterior distribution:

$$\begin{aligned}\boldsymbol{\beta} \mid \phi, \mathbf{Y} &\sim \mathbf{N}(\hat{\boldsymbol{\beta}}, (\phi \mathbf{X}^T \mathbf{X})^{-1}) \\ \phi \mid \mathbf{Y} &\sim \mathbf{G}((n - p)/2, \text{SSE}/2)\end{aligned}$$

if  $n > p$

- marginal posterior distribution for  $\boldsymbol{\beta}$  is multivariate Student-t with  $n - p$  degrees of freedom, location  $\hat{\boldsymbol{\beta}}$  and scale matrix  $\hat{\sigma}^2 \mathbf{X}^T \mathbf{X}^{-1}$





# Duality

- the posterior density  $p(\boldsymbol{\beta} | \mathbf{Y})$  is a monotonically decreasing function of  $Q(\boldsymbol{\beta}) \equiv (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T \mathbf{X}^T \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})$  so contours of  $p(\boldsymbol{\beta} | \mathbf{Y})$  are ellipsoidal in the parameter space of  $\boldsymbol{\beta}$
- the quantity  $Q(\boldsymbol{\beta}) / p\hat{\sigma}^2$  is distributed *a posteriori*

$$Q(\boldsymbol{\beta}) / p\hat{\sigma}^2 \sim F(p, n - p)$$

and the ellipsoidal contour of  $p(\boldsymbol{\beta} | \mathbf{Y})$  is defined as  $\frac{Q(\boldsymbol{\beta})}{p\hat{\sigma}^2} = F(p, n - p, \alpha)$ . (Box & Tiao 1973)

- then HPD regions for  $\boldsymbol{\beta}$  are the same as confidence regions for  $\boldsymbol{\beta}$  based on the  $F$ -distribution
- marginals of  $\beta_j$ ,  $\mathbf{x}^T \boldsymbol{\beta}$  and  $Y^*$  are also univariate Student-t with  $n - p$  degrees of freedom
- difference is in the interpretation of the regions i.e posterior probability that  $\boldsymbol{\beta}$  is in the given the data vs the probability *a priori* that the region covers the true  $\boldsymbol{\beta}$

