

# Sampling Distributions

Merlise Clyde

STA721 Linear Models

Duke University

September 3, 2015

# Outline

## Topics

- Normal Theory
- Chi-squared Distributions
- Student  $t$  Distributions

Readings: Christensen Apendix C, Chapter 1-2

# Prostate Example

```
> library(lasso2); data(Prostate)      # n = 97, 9 variables
> summary(lm(lpsa ~ ., data=Prostate))
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	0.669399	1.296381	0.516	0.60690	
lcavol	0.587023	0.087920	6.677	2.11e-09	***
lweight	0.454461	0.170012	2.673	0.00896	**
age	-0.019637	0.011173	-1.758	0.08229	.
lbph	0.107054	0.058449	1.832	0.07040	.
svi	0.766156	0.244309	3.136	0.00233	**
lcp	-0.105474	0.091013	-1.159	0.24964	
gleason	0.045136	0.157464	0.287	0.77506	
pgg45	0.004525	0.004421	1.024	0.30885	

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7084 on 88 degrees of freedom

Multiple R-squared: 0.6548, Adjusted R-squared: 0.6234

# Summary of Distributions

Models: Full  $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$

Assume  $\mathbf{X}$  is full rank with the first column of ones  $\mathbf{1}_n$  and  $p$  additional predictors  $r(\mathbf{X}) = p + 1$

$$\hat{\boldsymbol{\beta}} \mid \sigma^2 \sim \mathbf{N}(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^T \mathbf{X})^{-1})$$

$$\frac{\text{RSS}}{\sigma^2} \sim \chi^2_{n-r(\mathbf{X})}$$

$$\frac{\hat{\beta}_j - \beta_j}{\text{SE}(\hat{\beta}_j)} \sim t_{n-r(\mathbf{X})}$$

where  $\text{SE}(\hat{\beta}_j)$  is the square root of the  $j$ th diagonal element of  $\hat{\sigma}^2(\mathbf{X}^T \mathbf{X})^{-1}$  and  $\hat{\sigma}^2$  is the unbiased estimate of  $\sigma^2$

# Sampling Distribution of $\beta$

If  $\mathbf{Y} \sim N(\mathbf{X}\beta, \sigma^2 \mathbf{I}_n)$

Then  $\hat{\beta} \sim N(\beta, \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1})$

Unknown  $\sigma^2$ 

$$\hat{\beta}_j \mid \beta_j, \sigma^2 \sim N(\beta_j, \sigma^2 [(\mathbf{X}^T \mathbf{X})^{-1}]_{jj})$$

What happens if we substitute  $\hat{\sigma}^2 = \mathbf{e}^T \mathbf{e} / (n - r(\mathbf{X}))$  in the above?

$$\frac{(\hat{\beta}_j - \beta_j) / \sigma \sqrt{[(\mathbf{X}^T \mathbf{X})^{-1}]_{jj}}}{\sqrt{\mathbf{e}^T \mathbf{e} / (\sigma^2 (n - r(\mathbf{X})))}} \stackrel{D}{=} \frac{N(0, 1)}{\sqrt{\chi_{n-r(\mathbf{X})}^2 / (n - r(\mathbf{X}))}} \sim t(n - r(\mathbf{X}), 0, 1)$$

Need to show that  $\mathbf{e}^T \mathbf{e} / \sigma^2$  has a  $\chi^2$  distribution and is independent of the numerator!

# Central Student $t$ Distribution

## Definition

Let  $Z \sim N(0, 1)$  and  $S \sim \chi_p^2$  with  $Z$  and  $S$  independent, then

$$W = \frac{Z}{\sqrt{S/p}}$$

has a (central) Student  $t$  distribution with  $p$  degrees of freedom

See Casella & Berger or DeGroot & Schervish for derivation - nice change of variables and marginalization problem!

# Chi-Squared Distribution

## Definition

If  $Z \sim N(0, 1)$  then  $Z^2 \sim \chi_1^2$  (A Chi-squared distribution with one degree of freedom)

- Density

$$f(x) = \frac{1}{\Gamma(1/2)} (1/2)^{-1/2} x^{1/2-1} e^{-x/2} \quad x > 0$$

- Characteristic Function

$$E[e^{itZ^2}] = \varphi(t) = (1 - 2it)^{-1/2}$$



# Chi-Squared Distribution with $p$ Degrees of Freedom

If  $Z_j \stackrel{\text{iid}}{\sim} N(0, 1)$   $j = 1, \dots, p$  then  $X \equiv \mathbf{Z}^T \mathbf{Z} = \sum_{j=1}^p Z_j^2 \sim \chi_p^2$

## Characteristic Function

$$\begin{aligned}
 \varphi_X(t) &= E[e^{it \sum_{j=1}^p Z_j^2}] \\
 &= \prod_{j=1}^p E[e^{it Z_j^2}] \\
 &= \prod_{j=1}^p (1 - 2it)^{-1/2} \\
 &= (1 - 2it)^{-p/2}
 \end{aligned}$$

A Gamma distribution with shape  $p/2$  and rate  $1/2$ ,  $G(p/2, 1/2)$

$$f(x) = \frac{1}{\Gamma(p/2)} (1/2)^{-p/2} x^{p/2-1} e^{-x/2} \quad x > 0$$

# Quadratic Forms

## Theorem

Let  $\mathbf{Y} \sim N(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_n)$  with  $\boldsymbol{\mu} \in C(\mathbf{X})$  then if  $\mathbf{Q}$  is a rank  $k$  orthogonal projection on to  $C(\mathbf{X})^\perp$ ,  $(\mathbf{Y}^T \mathbf{Q} \mathbf{Y}) / \sigma^2 \sim \chi_k^2$

## Proof.

For an orthogonal projection  $\mathbf{Q} = \mathbf{U} \boldsymbol{\Lambda} \mathbf{U}^T = \mathbf{U}_k \mathbf{U}_k^T$  where  $C(\mathbf{Q}) = C(\mathbf{U}_k)$  and  $\mathbf{U}_k^T \mathbf{U}_k = \mathbf{I}_k$  (Spectral Theorem)

$$\mathbf{Y}^T \mathbf{Q} \mathbf{Y} = \mathbf{Y}^T \mathbf{U}_k \mathbf{U}_k^T \mathbf{Y}$$

$$\mathbf{Z} = \mathbf{U}_k^T \mathbf{Y} / \sigma \sim N(\mathbf{U}_k^T \boldsymbol{\mu}, \mathbf{U}_k^T \mathbf{U}_k)$$

$$\mathbf{Z} \sim N(\mathbf{0}, \mathbf{I}_k)$$

$$\mathbf{Z}^T \mathbf{Z} \sim \chi_k^2$$

Since  $\mathbf{U}^T \mathbf{Y} / \sigma \stackrel{D}{=} \mathbf{Z}$ ,  $\frac{\mathbf{Y}^T \mathbf{Q} \mathbf{Y}}{\sigma^2} \sim \chi_k^2$



# Residual Sum of Squares Example

## Sum of Squares Error (SSE)

Let  $\mathbf{Y} \sim N(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_n)$  with  $\boldsymbol{\mu} \in C(\mathbf{X})$ .

Because  $\boldsymbol{\mu} \in C(\mathbf{X})$ ,  $\mathbf{I} - \mathbf{P}_\mathbf{X}$  is a projection on  $C(\mathbf{X})^\perp$

$$\frac{\mathbf{e}^T \mathbf{e}}{\sigma^2} = \mathbf{Y}^T \frac{(\mathbf{I}_n - \mathbf{P}_\mathbf{X})^2}{\sigma} \mathbf{Y} \sim \chi_{n-r(\mathbf{X})}^2$$

# Estimated Coefficients and Residuals are Independent

If  $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n)$

Then  $\text{Cov}(\hat{\boldsymbol{\beta}}, \mathbf{e}) = \mathbf{0}$  which implies independence

Functions of independent random variables are independent (show characteristic functions or densities factor)

# Putting it all together

$$\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^T \mathbf{X})^{-1})$$

- $(\hat{\beta}_j - \beta_j)/\sigma[(\mathbf{X}^T \mathbf{X})^{-1}]_{jj} \sim N(0, 1)$
- $\mathbf{e}^T \mathbf{e} / \sigma^2 \sim \chi^2_{n-r(\mathbf{X})}$
- $\hat{\boldsymbol{\beta}}$  and  $\mathbf{e}$  are independent

$$\frac{(\hat{\beta}_j - \beta_j)/\sigma[(\mathbf{X}^T \mathbf{X})^{-1}]_{jj}}{\sqrt{\mathbf{e}^T \mathbf{e} / (\sigma^2(n - r(\mathbf{X})))}} \sim t(n - r(\mathbf{X}), 0, 1)$$

# Inference

- 95% Confidence interval:  $\hat{\beta}_j \pm t_{\alpha/2} \text{SE}(\hat{\beta}_j)$  use `qt(a, df)` for  $t_a$  quantile
- derive from pivotal quantity  $t = (\hat{\beta}_j - \beta_j) / \text{SE}(\hat{\beta}_j)$  where  $P(t \in (t_{\alpha/2}, t_{1-\alpha/2})) = 1 - \alpha$

# Prostate Example

`xtable(confint(prostate.lm))` from `library(MASS)` and `library(xtable)`

	2.5 %	97.5 %
(Intercept)	-1.91	3.25
lcavol	0.41	0.76
lweight	0.12	0.79
age	-0.04	0.00
lbph	-0.01	0.22
svi	0.28	1.25
lcp	-0.29	0.08
gleason	-0.27	0.36
pgg45	-0.00	0.01

# interpretation

- For a “1” unit increase in  $\mathbf{X}_j$ , expect  $\mathbf{Y}$  to increase by  $\hat{\beta}_j \pm t_{\alpha/2} \text{SE}(\hat{\beta}_j)$
- for log transforms

$$\mathbf{Y} = \exp(\mathbf{X}\beta + \epsilon) = \prod \exp(\mathbf{X}_j\beta_j) \exp(\epsilon)$$

- if  $\mathbf{X} = \log(\mathbf{W}_j)$  then look at 2-fold or % increases in  $\mathbf{W}$  to look at multiplicative increase in median of  $\mathbf{Y}$
- ifcavol increases by 10% then we expect PSA to increase by  $1.10^{(CI)} = (1.0398\%, 1.0751\%)$  or by 3.98 to 7.51 percent

For a 10% increase in cancer volume, we are 95% confident that the PSA levels will increase by approximately 4 to 7.5 percent.



# Derivation