Sampling Distributions of MLEs

Merlise Clyde

STA721 Linear Models

Duke University

September 7, 2017

Outline

Topics

- Student t Distributions
- Chi-squared Distributions

Readings: Christensen Appendix C, Chapter 2

Distribution of $\boldsymbol{\beta}$

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

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

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

Unknown σ^2

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

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

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

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

Central Student t Distribution

Definition

Let $Z \sim \mathsf{N}(0, 1)$ and $S \sim \chi^2_p$ 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 $Z \sim \mathrm{N}(0,1)$ then $Z^2 \sim \chi^2_1$ (A Chi-squared distribution with one degree of freedom)

- Density

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

- Characteristic Function

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

Chi-Squared Distribution with p Degrees of Freedom

If $Z_j \stackrel{\text{iid}}{\sim} \mathsf{N}(0,1)$ $j = 1, \ldots p$ then $X \equiv \mathbf{Z}^T\mathbf{Z} = \sum_j^p Z_j^2 \sim \chi^2_p$

Characteristic Function

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

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

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

Spectral Decompostion of Projection Matrices

If $\mathbf{P}$ is an orthogonal projection matrix, then its eigenvalues λ_i are either zero or one with $\text{tr}(\mathbf{P}) = \sum_i(\lambda_i) = r(\mathbf{P})$

- $\mathbf{P} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$
- $\mathbf{P} = \mathbf{P}^2 \Rightarrow \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T = \mathbf{U}\boldsymbol{\Lambda}^2\mathbf{U}^T$
- $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}^2$ is true only for $\lambda_i = 1$ or $\lambda_i = 0$
- Since $r(\mathbf{P})$ is the number of non-zero eigenvalues, $r(\mathbf{P}) = \sum \lambda_i = \text{tr}(\mathbf{P})$

$$\mathbf{P} = [\mathbf{U}_P \mathbf{U}_{P^\perp}] \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_{n-r} \end{bmatrix} \begin{bmatrix} \mathbf{U}_P^T \\ \mathbf{U}_{P^\perp}^T \end{bmatrix} = \mathbf{U}_P \mathbf{U}_P^T$$

$$\mathbf{P} = \sum_{i=1}^{r} \mathbf{u}_i \mathbf{u}_i^T$$

sum of r rank 1 projections.

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* $C(\mathbf{X})^\perp$, $(\mathbf{Y}^T\mathbf{Q}\mathbf{Y})/\sigma^2 \sim \chi^2_k$

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)

$$\begin{aligned}
\mathbf{Y}^T\mathbf{Q}\mathbf{Y} &= \mathbf{Y}^T\mathbf{U}_k\mathbf{U}_k^T\mathbf{Y} \\
\mathbf{Z} &= \mathbf{U}_k^T\mathbf{Y}/\sigma \sim \mathrm{N}(\mathbf{U}_k^T\boldsymbol{\mu}, \mathbf{U}_k^T\mathbf{U}_k) \\
\mathbf{Z} &\sim \mathrm{N}(\mathbf{0}, \mathbf{I}_k) \\
\mathbf{Z}^T\mathbf{Z} &\sim \chi^2_k
\end{aligned}$$

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

$\square$

Residual Sum of Squares Error

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

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

Estimated Coefficients and Residuals are Independent

If $\mathbf{Y} \sim \mathrm{N}(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n)$
Then $\mathrm{Cov}(\mathbf{e}, \hat{\boldsymbol{\beta}}) = \mathbf{0}$ which implies independence

Putting it all together

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

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

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

- Standard Error $\mathrm{SE}(\hat{\beta}_j) = \hat{\sigma}\sqrt{[(\mathbf{X}^T\mathbf{X})^{-1}]_{jj}}$

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

Inference

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

Pivotal Quantities and CI

Other Quantities

Linear Combinations: $\lambda^T \hat{\beta}$

Fitted Values

Unknown Mean: $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$

Prostate Example

```
suppressPackageStartupMessages(library(lasso2))
library(knitr)
data(Prostate)
prostate.lm = lm(lpsa ~ ., data=Prostate)
ci=confint(prostate.lm); kable(ci, digits=2)
```

	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

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

- 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 $\mathbf{Y}$
- if`cavol` increases by 10% then we expect the median `PSA` to increase by $1.10^{(CI)} = (1.04, 1.075)$ or by 4.0 to 7.5 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

Next

Next Class: Predictive Distributions