Lecture 23: Likelihood ratio tests

Ciaran Evans

Course logistics

- HW 6 due today, HW 7 on course website
- Exam 1 done (woo!)
- Exam 2 plan: released on April 4, due April 11
  - Focus on convergence, hypothesis testing, maybe confidence intervals
  - Would cover HW 5 – HW 8

Last time: binary classification and classification error

Consider data (X, Y) with $X \in \mathbb{R}^d$ and $Y \in \{0, 1\}$. Fit a model to estimate

$$p(x) = P(Y = 1 | X = x)$$

Our binary predictions are

$$\widehat{Y} = \begin{cases} 1 & p(x) \geq h \\ 0 & p(x) < h \end{cases}$$

The **classification error** is given by $P(\widehat{Y} \neq Y)$.

Result: For any binary classifier, $h = 0.5$ minimizes classification error.

Changing the threshold

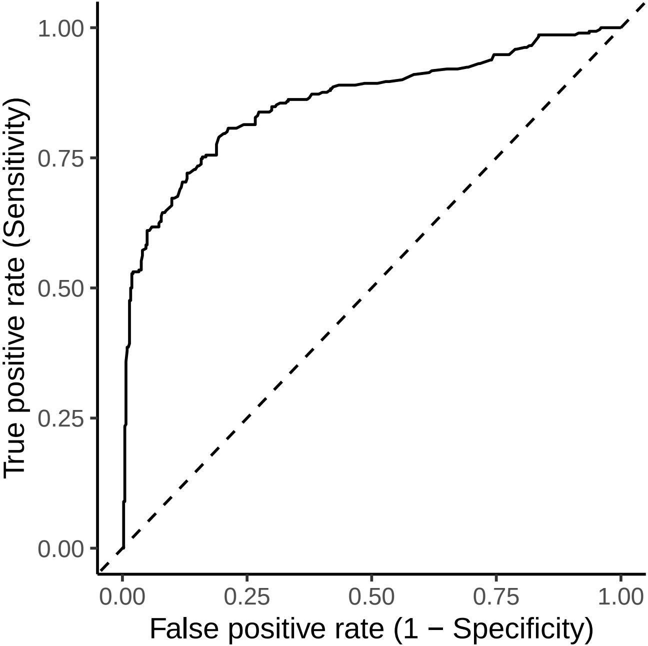
Threshold of 0.7:

		Actual	
		$Y = 0$	$Y = 1$
Predicted	$\widehat{Y} = 0$	412	136
	$\widehat{Y} = 1$	12	154

Threshold of 0.3:

		Actual	
		$Y = 0$	$Y = 1$
Predicted	$\widehat{Y} = 0$	309	49
	$\widehat{Y} = 1$	115	241

ROC curve: consider all thresholds



Binary classification vs. hypothesis testing

- Both binary classification and hypothesis testing involve deciding between two options
- Error metrics for both involve looking at correct decisions, false positives (type I errors), false negatives (type II errors)

Question: How do binary classification and hypothesis testing *differ*?

Binary classification vs. hypothesis testing

Binary classification:

- Can use training data to estimate performance and so choose a threshold
- Thresholds are chosen to maximize some combination of sensitivity and specificity

Hypothesis testing:

- Conceptually a two-step approach: control type I error, then hope to have good power (i.e., don't consider tests which have high type I error)
- Only see one test result; don't get to estimate type I error or power from a single test
- Want theoretical guarantees that (if assumptions are met) type I error can be controlled at desired level

Binary classification vs. hypothesis testing

- Usual approach to binary classification: maximize some combination of sensitivity and specificity
- Neyman-Pearson classification[1]: control probability of false positives (1 - specificity) at desired level, then try to maximize sensitivity

Question: Why might you choose one of these approaches over the other?

[1] Scott, C., & Nowak, R. (2005). A Neyman-Pearson approach to statistical learning. *IEEE Transactions on Information Theory*, 51(11), 3806-3819.

Previously: Neyman-Pearson test

Example: Let $X_1, ..., X_n \overset{iid}{\sim} \textit{Exponential}(\theta)$, with pdf $f(x|\theta) = \theta e^{-\theta x}$. We want to test

$$H_0 : \theta = \theta_0 \qquad H_A : \theta = \theta_1,$$

where $\theta_1 < \theta_0$. The Neyman-Pearson test rejects when

$$\frac{L(\theta_1|\mathbf{X})}{L(\theta_0|\mathbf{X})} > k.$$

Question: What should I do if I want to test the hypotheses

$$H_0 : \theta = \theta_0 \qquad H_A : \theta \neq \theta_0$$

Likelihood ratio test

Let $X_1, ..., X_n$ be a sample from a distribution with parameter $\theta \in \mathbb{R}^d$. We wish to test $H_0 : \theta \in \Theta_0$ vs. $H_A : \theta \in \Theta_1$.

The **likelihood ratio test** (LRT) rejects H_0 when

$$\frac{\sup\limits_{\theta \in \Theta_1} L(\theta|\mathbf{X})}{\sup\limits_{\theta \in \Theta_0} L(\theta|\mathbf{X})} > k,$$

where k is chosen such that $\sup\limits_{\theta \in \Theta_0} \beta_{LR}(\theta) \leq \alpha$.

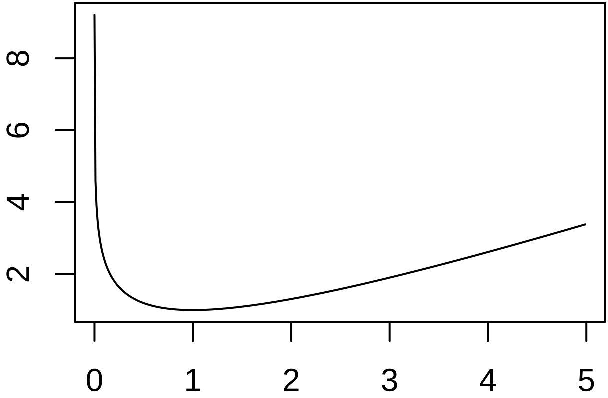
Example

Let $X_1, ..., X_n \stackrel{iid}{\sim} Exponential(\theta)$, with pdf $f(x|\theta) = \theta e^{-\theta x}$. We want to test

$$H_0 : \theta = \theta_0 \qquad H_A : \theta \neq \theta_0$$

Example

Plot of $\theta_0\overline{X} - \log(\theta_0\overline{X})$:



Example: linear regression with normal data

Suppose we observe $(X_1, Y_1), ..., (X_n, Y_n)$, where $Y_i = \beta^T X_i + \varepsilon_i$ and $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$. Partition $\beta = (\beta_{(1)}, \beta_{(2)})^T$. We wish to test $H_0 : \beta_{(2)} = 0$ vs. $H_A : \beta_{(2)} \neq 0$.