LECTURE 24

- Reference: Section 9.3
- Course Evaluations (until 12/16)  
  [link](http://web.mit.edu/subjectevaluation)

Outline

- Review
  - Maximum likelihood estimation
  - Confidence intervals
- Linear regression
- Binary hypothesis testing
  - Types of error
  - Likelihood ratio test (LRT)

Review

- Maximum likelihood estimation
  - Have model with unknown parameters:  
    \[ X \sim p_X(x; \theta) \]
  - Pick \( \theta \) that “makes data most likely”  
    \[ \max_{\theta} p_X(x; \theta) \]
  - Compare to Bayesian MAP estimation:  
    \[ \max_{\theta} p_{\theta|X}(\theta | x) \text{ or } \max_{\theta} \frac{p_X(x|\theta)p_\theta(\theta)}{p_Y(y)} \]
- Sample mean estimate of \( \theta = E[X] \)  
  \[ \hat{\Theta}_n = (X_1 + \cdots + X_n)/n \]
- 1 - \( \alpha \) confidence interval  
  \[ P(\hat{\Theta}_n^- \leq \theta \leq \hat{\Theta}_n^+) \geq 1 - \alpha, \quad \forall \theta \]
  - confidence interval for sample mean
    - let \( z \) be s.t. \( \Phi(z) = 1 - \alpha/2 \)
    \[ P(\hat{\Theta}_n - \frac{z\sigma}{\sqrt{n}} \leq \theta \leq \hat{\Theta}_n + \frac{z\sigma}{\sqrt{n}}) \approx 1 - \alpha \]

Regression

- Data: \((x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\)
- Model: \( y \approx \theta_0 + \theta_1 x \)
  \[ \min_{\theta_0, \theta_1} \sum_{i=1}^{n} (y_i - \theta_0 - \theta_1 x_i)^2 \]  
  \((*)\)
- One interpretation:  
  \[ Y_i = \theta_0 + \theta_1 x_i + W_i, \quad W_i \sim N(0, \sigma^2), \text{ i.i.d.} \]
  - Likelihood function \( f_{X,Y|\theta}(x,y; \theta) \) is:  
    \[ c \cdot \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \theta_0 - \theta_1 x_i)^2 \right\} \]
  - Take logs, same as \((*)\)
  - Least sq. \( \leftrightarrow \) pretend \( W_i \) i.i.d. normal

Linear regression

- Model \( y \approx \theta_0 + \theta_1 x \)  
  \[ \min_{\theta_0, \theta_1} \sum_{i=1}^{n} (y_i - \theta_0 - \theta_1 x_i)^2 \]  
  \((*)\)
- Solution (set derivatives to zero):  
  \[ x = \frac{x_1 + \cdots + x_n}{n}, \quad y = \frac{y_1 + \cdots + y_n}{n} \]
  \[ \bar{\theta}_1 = \frac{\sum_{i=1}^{n} (x_i - x)(y_i - y)}{\sum_{i=1}^{n} (x_i - x)^2} \]
  \[ \bar{\theta}_0 = y - \bar{\theta}_1 x \]
- Interpretation of the form of the solution
  - Assume a model \( Y = \theta_0 + \theta_1 X + W \)
    \( W \) independent of \( X \), with zero mean
  - Check that  
    \[ \theta_1 = \frac{\text{cov}(X,Y)}{\text{var}(X)} = \frac{E[(X - E[X])(Y - E[Y])]}{E[(X - E[X])^2]} \]
  - Solution formula for \( \bar{\theta}_1 \) uses natural estimates of the variance and covariance
The world of linear regression

- **Multiple linear regression:**
  - **data:** $(x_i, x'_i, x''_i, y_i), i = 1, \ldots, n$
  - **model:** $y \approx \theta_0 + \theta x + \theta'x' + \theta''x''$
  - **formulation:**
    $$\min_{\theta, \theta', \theta''} \sum_{i=1}^{n} (y_i - \theta_0 - \theta x_i - \theta'x'_i - \theta''x''_i)^2$$

- **Choosing the right variables**
  - **model:** $y \approx \theta_0 + \theta_1 h(x)$
    e.g., $y \approx \theta_0 + \theta_1 x^2$
  - **work with data points** $(y_i, h(x))$
  - **formulation:**
    $$\min_{\theta_0, \theta_1} \sum_{i=1}^{n} (y_i - \theta_0 - \theta_1 h(x_i))^2$$

The world of regression (ctd.)

- **In practice,** one also reports
  - Confidence intervals for the $\theta_i$
  - “Standard error” (estimate of $\sigma$)
  - $R^2$, a measure of “explanatory power”

Some common concerns

- Heteroskedasticity
- Multicollinearity
- Sometimes misused to conclude causal relations
- etc.

Binary hypothesis testing

- Binary $\theta$; new terminology:
  - **null hypothesis** $H_0$: $X \sim p_X(x; H_0)$ [or $f_X(x; H_0)$]
  - **alternative hypothesis** $H_1$: $X \sim p_X(x; H_1)$ [or $f_X(x; H_1)$]

- Partition the space of possible data vectors
  **Rejection region** $R$: reject $H_0$ iff data $\in R$

- Types of errors:
  - **Type I (false rejection, false alarm):** $H_0$ true, but rejected
    $$\alpha(R) = P(X \in R; H_0)$$
  - **Type II (false acceptance, missed detection):** $H_0$ false, but accepted
    $$\beta(R) = P(X \notin R; H_1)$$

Likelihood ratio test (LRT)

- Bayesian case (MAP rule): choose $H_1$ if:
  $$\frac{P(X = x | H_1)P(H_1)}{P(X = x | H_0)P(H_0)} > \frac{P(X = x | H_0)}{P(X = x)}$$
  or
  $$\frac{P(X = x | H_1)}{P(X = x | H_0)} > \frac{P(H_0)}{P(H_1)}$$
  (likelihood ratio test)

- Nonbayesian version: choose $H_1$ if
  $$\frac{f_X(x; H_1)}{f_X(x; H_0)} > \xi$$ (discrete case)
  $$\frac{f_X(x; H_1)}{f_X(x; H_0)} > \xi$$ (continuous case)

- threshold $\xi$ trades off the two types of error
  - choose $\xi$ so that $P(\text{reject } H_0; H_1) = \alpha$
    (e.g., $\alpha = 0.05$)
Fall 2010

For information about citing these materials or our Terms of Use, visit: http://ocw.mit.edu/terms.