LEC
cTUE 24

• Reference: Section 9.3

• Course VI Underground Guide Evaluations
https://sixweb.mit.edu/student/evaluate/6.041-f2010
https://sixweb.mit.edu/student/evaluate/6.431-f2010

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
  \[
  P(\hat{\Theta}_n^- \leq \theta \leq \hat{\Theta}_n^+) \geq 1 - \alpha, \forall \theta
  \]
• confidence interval for sample mean
  – let \( z \) be s.t. \( \Phi(z) = 1 - \alpha/2 \)
    \[
    P\left( \hat{\Theta}_n - \frac{z\sigma}{\sqrt{n}} \leq \theta \leq \hat{\Theta}_n + \frac{z\sigma}{\sqrt{n}} \right) \approx 1 - \alpha
    \]

Regression

• Data: \((x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\)
• Model: \( y \approx \theta_0 + \theta_1 x \)
  \[
  \min_{\theta_0, \theta_1} \sum_{i=1}^{n} (y_i - \theta_0 - \theta_1 x_i)^2 \quad (*)
  \]
• 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):
  \[
  \bar{x} = \frac{x_1 + \cdots + x_n}{n}, \quad \bar{y} = \frac{y_1 + \cdots + y_n}{n}
  \]
  \[
  \hat{\theta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}
  \]
  \[
  \hat{\theta}_0 = \bar{y} - \hat{\theta}_1 \bar{x}
  \]
• Interpretation of the form of the solution
  – Assume a model \( Y = \theta_0 + \theta_1 X + W \)
    \( W \) independent of \( X \) and \( Y \), 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 \( \hat{\theta}_1 \) is a natural estimate of the 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:**
  \[\min_{\theta} \sum_{i=1}^n (y_i - \theta_0 - \theta_1 h(x_i))^2\]

-- 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:
  \[
P(H_1 | X = x) > P(H_0 | X = x)
  \]
  or
  \[
  P(X = x | H_1)P(H_1) > P(X = x | H_0)P(H_0)
  \]
  or
  \[
  \frac{P(X = x | H_1)}{P(X = x | H_0)} > \frac{P(H_1)}{P(H_0)}
  \]
  (likelihood ratio test)

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

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