LECTURE 24

- **Reference:** Section 9.3
- Course Evaluations (until 12/16)
  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 θ that "makes data most likely"

$$\max_\theta p_X(x;\theta)$$

  - Compare to Bayesian MAP estimation:

$$\max_\theta p_{\Theta|X}(\theta \mid x) \quad \text{or} \quad \max_\theta \frac{p_{X|\Theta}(x|\theta)p_\Theta(\theta)}{p_Y(y)}$$

- **Sample mean estimate of** $\theta = \mathbf{E}[X]$

$$\hat{\Theta}_n = (X_1 + \cdots + X_n)/n$$

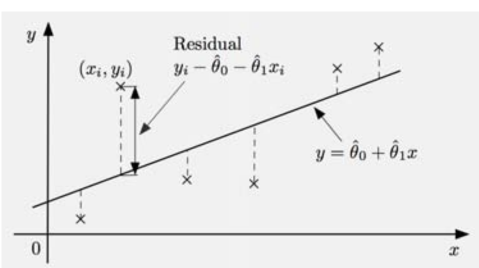
- $1-\alpha$ **confidence interval**

$$\mathbf{P}(\hat{\Theta}_n^- \le \theta \le \hat{\Theta}_n^+) \ge 1-\alpha, \quad \forall\ \theta$$

- confidence interval for sample mean
  - let z be s.t. $\Phi(z) = 1-\alpha/2$

$$\mathbf{P}\left(\hat{\Theta}_n - \frac{z\sigma}{\sqrt{n}} \le \theta \le \hat{\Theta}_n + \frac{z\sigma}{\sqrt{n}}\right) \approx 1-\alpha$$

Regression



- Data: $(x_1,y_1), (x_2,y_2), \ldots, (x_n,y_n)$
- Model: $y \approx \theta_0 + \theta_1 x$

$$\min_{\theta_0,\theta_1} \sum_{i=1}^n (y_i - \theta_0 - \theta_1 x_i)^2 \qquad (*)$$

- One interpretation:
  $Y_i = \theta_0 + \theta_1 x_i + W_i, \quad W_i \sim N(0,\sigma^2)$, 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):

$$\overline{x} = \frac{x_1 + \cdots + x_n}{n}, \qquad \overline{y} = \frac{y_1 + \cdots + y_n}{n}$$

$$\hat\theta_1 = \frac{\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^n (x_i - \overline{x})^2}$$

$$\hat\theta_0 = \overline{y} - \hat\theta_1 \overline{x}$$

- **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

$$\theta_1 = \frac{\text{cov}(X,Y)}{\text{var}(X)} = \frac{\mathbf{E}\big[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])\big]}{\mathbf{E}\big[(X - \mathbf{E}[X])^2\big]}$$

  - Solution formula for $\hat\theta_1$ 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:

$$\min_{\theta} \sum_{i=1}^{n} (y_i - \theta_0 - \theta_1 h_1(x_i))^2$$

The world of regression (ctd.)

- **In practice,** one also reports
  - Confidence intervals for the θ_i
  - "Standard error" (estimate of σ)
  - R^2, a measure of "explanatory power"

- **Some common concerns**
  - Heteroskedasticity
  - Multicollinearity
  - Sometimes misused to conclude causal relations
  - etc.

Binary hypothesis testing

- Binary θ; 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

$$\alpha(R) = \mathbf{P}(X \in R\,; H_0)$$

  - **Type II** (**false acceptance**, missed detection):
    H_0 false, but accepted

$$\beta(R) = \mathbf{P}(X \notin R\,; H_1)$$

Likelihood ratio test (LRT)

- Bayesian case (MAP rule): choose H_1 if:
  $\mathbf{P}(H_1 \mid X = x) > \mathbf{P}(H_0 \mid X = x)$
  or

$$\frac{\mathbf{P}(X = x \mid H_1)\mathbf{P}(H_1)}{\mathbf{P}(X = x)} > \frac{\mathbf{P}(X = x \mid H_0)\mathbf{P}(H_0)}{\mathbf{P}(X = x)}$$

  or

$$\frac{\mathbf{P}(X = x \mid H_1)}{\mathbf{P}(X = x \mid H_0)} > \frac{\mathbf{P}(H_0)}{\mathbf{P}(H_1)}$$

  (likelihood ratio test)

- Nonbayesian version: choose H_1 if

$$\frac{\mathbf{P}(X = x; H_1)}{\mathbf{P}(X = x; H_0)} > \xi \quad \text{(discrete case)}$$

$$\frac{f_X(x; H_1)}{f_X(x; H_0)} > \xi \quad \text{(continuous case)}$$

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