LECTURE 14

- **Readings:** Sections 4.5, 4.6

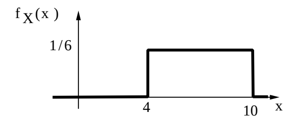
Outline

- Least squares prediction
  - Conditional variance
  - Linear prediction

Review

- $\mathbf{E}[X \mid Y]$ is a random variable whose value is $\mathbf{E}[X \mid Y = y]$ when $Y = y$
  - It is a function of Y
  - $\mathbf{E}[\mathbf{E}[X \mid Y]] = \mathbf{E}[X]$

Prediction in the absence of information

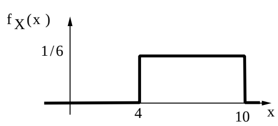


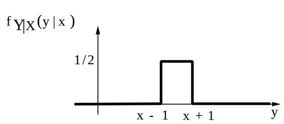
- prediction c

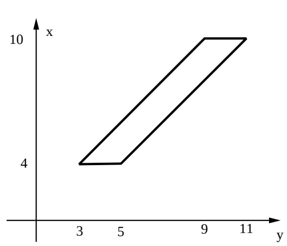
$$\text{minimize } \mathbf{E}\left[(X - c)^2\right]$$

- $c = \mathbf{E}[X]$
- Optimal mean squared error:

$$\mathbf{E}\left[(X - \mathbf{E}[X])^2\right] = \text{Var}(X)$$



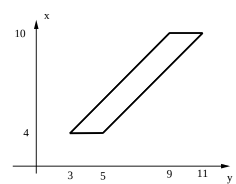


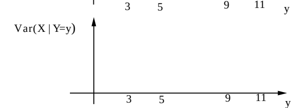


Conditional variance

- $\text{Var}(X \mid Y = y)$: variance of the conditional distribution of X

$$\mathbf{E}[(X - E[X \mid Y])^2 \mid Y = y]$$





Predicting X based on Y

- Two r.v.'s X, Y
- we observe that $Y = y$
  - new universe: condition on $Y = y$
- $\mathbf{E}\left[(X - c)^2 \mid Y = y\right]$ is minimized by $c =$
- View predictor as a function $g(y)$
- $\mathbf{E}[X \mid Y]$ minimizes

  $$\mathbf{E}[(X - g(Y))^2]$$

  over all predictors $g(\cdot)$

Prediction given several measurements

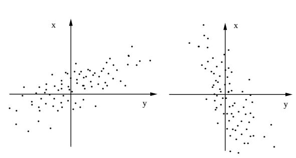
- Unknown r.v. X
- Observe values of r.v.'s $Y_1, \ldots, Y_n$
- Best prediction: $\mathbf{E}[X \mid Y_1, \ldots, Y_n]$
- Can be hard to compute/implement
  - need model $f_{X,Y_1,\ldots,Y_n}$
  - even with model, computations are hard

Linear prediction

- Form a predictor (of X) of the form $aY + b$
- Minimize $\mathbf{E}\left[(X - aY - b)^2\right]$
- Best predictor:

  $$\mathbf{E}[X] + \frac{\text{Cov}(X,Y)}{\text{var}(Y)}(Y - \mathbf{E}[Y])$$

  $$\text{Cov}(X,Y) = \mathbf{E}\left[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])\right]$$



Covariance and correlation

- Covariance:

  $$\text{Cov}(X,Y) = \mathbf{E}\left[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])\right]$$

- Correlation (dimensionless version of covariance)

  $$\rho = \mathbf{E}\left[\frac{(X - \mathbf{E}[X])}{\sigma_X} \cdot \frac{(Y - \mathbf{E}[Y])}{\sigma_Y}\right]$$

- $-1 \le \rho \le 1$
- Independence implies zero covariance (converse is not true)