Online Review Course of
Undergraduate Probability and Statistics

# Review Lecture 16
# Linear Regression, part 1

Chris A. Mack
*Adjunct Associate Professor*

Course Website: www.lithoguru.com/scientist/statistics/review.html

 



## A Linear Model

- Consider two random variables, $X$ and $Y$. What does it mean to say they are linearly related?
- Approach 1: $Y = aX + b$ ($a$ and $b$ are scalars)
  – This model assumes that $Y$ is completely determined by knowing $X$
- Approach 2: $E[Y|X] = aX + b$
  – The mean value of $Y$ is controlled by $X$
  – There is some variation in $Y$ that is not controlled by $X$
  – Note: $\mathrm{E}[Y|X]$ is called the conditional expectation
  – Law of Iterated Expectations: $\mathrm{E}\big[\mathrm{E}[Y|X]\big] = E[Y]$

 



## A Linear Model

- Our model: $E[Y|X] = aX + b$
- Take the expectation of both sides, leading to
  – $E[Y] = aE[X] + b,$ or $b = E[Y] - aE[X]$
- Find the covariance of X and Y, leading to
  – $cov(X,Y) = a\,var(X),$ or

$$a = \frac{cov(X,Y)}{var(X)}$$

 



## A Linear Model

- Our model: $E[Y|X] = aX + b$

$$b = E[Y] - aE[X] \qquad a = \frac{cov(X,Y)}{var(X)}$$

- We need estimators for:
  – $E[X], E[Y], cov(X,Y)$, and $var(X)$
- Example: use the estimators that we have already used
  – Are there other estimators?
  – Which estimators are the “best”

 



## Estimators

- Let $\theta$ be the parameter and $\hat{\theta}$ be the estimator for that parameter
- What properties do we want our estimator to have?
  – Unbiased: $\mathrm{E}[\hat{\theta}] = \theta$
  – Minimum variance: make $var[\hat{\theta}]$ as small as possible
  – Robustness: a few bad data points shouldn't completely ruin our estimate
- There are many different estimators for each parameter, with trade-offs of bias, variance, and robustness
  – Mean vs. median vs. trimmed mean as estimator for $\mathrm{E}[\mathrm{X}]$

 



## Maximum Likelihood Estimators

- Suppose we run an experiment and get a set of data $x_1, x_2, \ldots, x_n$.
- What value of the parameter maximizes the likelihood of getting that exact data set?
  – $\mathbb{P}(X_1 = x_1|\theta), \mathbb{P}(X_2 = x_2|\theta)$, etc.
  – If each measurement is independent, the likelihood function is the product of all the probabilities
  – To evaluate the likelihood function (and thus maximize it) requires a model for the probabilities

 

## Maximum Likelihood Estimator (MLE)

- Example: suppose iid $X_i \sim N(\mu, \sigma^2)$
- Let's find the MLE for the mean, μ

$$\mathbb{P}(X_i = x_i|\mu) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x_i-\mu)^2/2\sigma^2}$$

Likelihood:

$$\prod_{i=1}^{n} \mathbb{P}(X_i = x_i|\mu) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\Sigma(x_i-\mu)^2/2\sigma^2}$$

- Now find the maximum: take derivative wrt μ, set equal to 0 (hint: take the log first, then the derivative)
  - Result is the familiar formula for the sample mean





## Maximum Likelihood Estimator (MLE)

- Some MLE estimators, assuming iid normal distribution

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$\widehat{cov(x,y)} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

- The MLE estimators for variance and covariance are biased; we often modify them to remove bias
- These estimators are not robust





## MLE Linear Model

- Our model: $E[Y|X] = aX + b$

$$a = \frac{cov(X,Y)}{var(X)} \qquad b = E[Y] - aE[X]$$

- Our MLE estimators:

$$a = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad b = \bar{y} - a\bar{x}$$





## Least Square Error (LSE)

- Another approach: minimize the square of the difference between the model and the data

Model: $y_i = ax_i + b + \epsilon_i$ (residual: $\epsilon_i$)

- This model assumes all uncertainty comes from y (there is no uncertainty in x)
- Now, find the values of $a$ and $b$ that minimize the sum of the squares of the residuals (SSE)

$$SSE = \sum_{i=1}^{n}\epsilon_i^2 = \sum_{i=1}^{n}(y_i - ax_i - b)^2$$





## Least Square Error (LSE)

- Properties of the LSE solution
  - $\sum_{i=1}^{n}\epsilon_i = 0$
  - Unbiased estimator for $var(\epsilon)$ is SSE/(n-2)
- The LSE estimates of the slope and intercept are exactly the same as the MLE estimates when the data are iid normally distributed
  - This is our motivation for least-squares fitting





## Review #16: What have we learned?

- What is our model for a linear regression?
- What four statistical parameters must we estimate in order to find the slope and intercept of our line?
- What is a maximum likelihood estimator?
- What is a least squares estimator?
- When do MLE and LSE give the same answers?