Probability Theory Lecture 11

Padmini Mukkamala

Budapest University of Technology and Economics

January 21, 2021

Overview

Recap: Distance from mean - 2

Steiner Equality

For any random variable X, $E((X-c)^2)$ is minimized when $c=\mu$, where $\mu = E(X)$.

Proof:
$E((X-c)^2) = E(((X-\mu)+(\mu-c))^2) = Var(X) + (\mu - c)^2$.

Introduction

Consider a pair of random variables X, Y. We want to find a linear equation in X that would best approximate Y. More precisely, we want to find two constants α, β such that Y can be approximated by the line equation $\beta X + \alpha$, which we will also call our estimate for Y.

One way to figure out the constants α and β would be to minimize the error of $Y - (\beta X + \alpha)$ (the difference between the actual value and our estimate), by minimizing $E((Y - (\beta X + \alpha))^2)$. This is just **one heuristic**. Different heuristics will give different equations.

Method of Least Squares

When we chose α, β to minimize $E((Y - (\beta X + \alpha))^2)$ then it is said that we are using the **Method of least squares**.

$$E((Y - (\beta X + \alpha))^2)$$
$$= E(Y^2 + \beta^2 X^2 + \alpha^2 + 2\alpha\beta X - 2\beta XY - 2\alpha Y)$$

This considered as a function of α and β, is minimized when the partial derivatives with respect to α and β evaluate to 0.

Method of Least Squares cont.

Taking the partial derivatve with respect to α, we get,

$$E(2\alpha + 2\beta X - 2Y) = 0$$

Using Linearity, this is

$$2\alpha + 2\beta E(X) - 2E(Y) = 0$$

or,

$$\alpha = E(Y) - \beta E(X)$$

We will call this **Equation 1**.

Method of Least Squares cont.

Taking the partial derivatve with respect to β, we get,

$$E(2\beta X^2 + 2\alpha X - 2XY) = 0$$

Using the value of α from Equation 1, we get,

$$E(2\beta X^2 + 2(\mu_Y - \beta\mu_X)X - 2XY) = 0$$

$$\implies E(2\beta X^2 + 2\mu_Y X - \beta\mu_X X - 2XY) = 0$$

$$\implies 2\beta E(X^2) + 2\mu_Y E(X) - \beta\mu_X E(X) - 2E(XY) = 0$$

$$\implies \beta(E(X^2) - \mu_X^2) = E(XY) - \mu_X\mu_Y = Cov(X, Y)$$

$$\implies \beta = \frac{Cov(X, Y)}{\sigma_X^2}$$

Method of Least Squares cont.

From here, we can derive that,

$$\alpha = E(Y) - \frac{Cov(X, Y)}{\sigma_X^2} E(X)$$

Method of Least Squares cont.

Summary

The **Linear Regression** of Y in terms of X is the line equation $\beta X + \alpha$ where the constants α and β are given by:

$$\beta = \frac{Cov(X, Y)}{\sigma_X^2}$$

$$\alpha = E(Y) - \frac{Cov(X, Y)}{\sigma_X^2} E(X)$$

Estimating the Error

Recall that the error of our Linear Regression is given by $E((Y-(\beta X+\alpha))^2)$. We note that since $E(Y-(\beta X+\alpha))=0$, so then we have,

$$E((Y-(\beta X+\alpha))^2) = Var(Y-(\beta X+\alpha))$$

$$= Var(Y) + \beta^2 Var(X) - 2\beta Cov(X, Y)$$

Estimating the Error

Plugging in the value of β, we get,

$$= Var(Y) + \frac{(Cov(X,Y))^2}{(Var(X))^2} Var(X) - 2\frac{Cov(X,Y)}{Var(X)} Cov(X,Y)$$

$$= Var(Y) - \frac{(Cov(X,Y))^2}{Var(X)} = Var(Y)(1 - \frac{(Cov(X,Y))^2}{Var(X)Var(Y)})$$

$$= \sigma_Y^2(1 - (\rho(X,Y))^2)$$

where $\rho(X,Y)$ is the correlation coefficient.

Example

We want to see if the number of hours a student studies per week has a linear relation with their grade in the course. Let X denote the number of hours a student studied on an average every week, and let Y denote their grade on the course (for simplicity assumed to be a real number in the interval $(0, 5)$.

From past data it was observed that students study an average of 1 hour per week per course, with $\sigma_X^2 = 1$. It was also absorved that the average grade of the students in the course was 3 with $\sigma_Y^2 = 2$. If $Cov(X, Y) = 1$, what is the best linear approximation of Y in terms of X? Use the linear regression to predict the grade of the student if he studies 3 hours every week. What is the error of your prediction? If instead σ_Y^2 was also equal to 1, what can you say about your linear regression line?

2D Normal Distribution

Recall that if (X, Y) is a random variable vector with 2D normal distribution, then we stated (without proof) that the regression is the same as Linear Regression. So,

Regression Function

$E(Y|X) = \beta X + \alpha$, where the constants α and β are given by:

$$\beta = \frac{Cov(X, Y)}{\sigma_X^2}$$

$$\alpha = E(Y) - \frac{Cov(X, Y)}{\sigma_X^2} E(X)$$

Joint Conditional Mass Function

Discrete case: Joint Conditional PMF

For two random variables X, Y, the joint conditional probability mass function of X conditioned on a specific value of $Y = y$ where $p_Y(y) \neq 0$, is given by,

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$$

We must check that this function $p_{X|Y}(\cdot|y)$ is a proper probability mass function for the variable X. We can do so by checking,

$$\sum_x p_{X|Y}(x|y) = 1$$

Further properties

We notice that if X and Y are independent random variables, then,

$$p_{X|Y}(x|y) = p_X(x)$$

Joint Conditional CDF

The joint conditional cumulative distribution function, then, can be computed as follows:

$$F_{X|Y}(a|y) = \sum_{x \leq a} p_{X|Y}(x|y)$$

Example - 1

This is example 4a from the book: Consider the following joint pmf table.

$X \backslash Y$	0	1	$p_X(x)$
0	0.4	0.2	0.6
1	0.1	0.3	0.4
$p_Y(y)$	0.5	0.5	1

Table: Joint pmf of X, Y

What is the joint conditional pmf $p_{X|Y}(x|1)$?

Example - 2

Let X the sum of two independent dice rolls, and let Y denote the outcome of the first dice roll. What is the conditional pmf of $p_{Y|X}(y|5)$ and $p_{Y|X}(y|7)$?
What is $p_{Y|X}(y|x)$, where now we try to find the joint pmf as a function of x?

Example - 2 cont.

Y	$p_{Y\|X}(y\|5)$
1	$\frac{1}{4}$
2	$\frac{1}{4}$
3	$\frac{1}{4}$
4	$\frac{1}{4}$
5	0
6	0

Y	$p_{Y\|X}(y\|7)$
1	$\frac{1}{6}$
2	$\frac{1}{6}$
3	$\frac{1}{6}$
4	$\frac{1}{6}$
5	$\frac{1}{6}$
6	$\frac{1}{6}$

Table: Conditional pmf of $p_{Y|X}(y|5)$ and $p_{Y|X}(y|7)$

Example - 2 cont.

We can check that the joint pmf, as a function of x can be denoted as: For $2 \leq x \leq 7$,

$$p_{Y|X}(y|x) = \begin{cases} \frac{1}{x-1} & 1 \leq y \leq x-1 \\ 0 & \text{otherwise} \end{cases}$$

and for $7 < x \leq 12$,

$$p_{Y|X}(y|x) = \begin{cases} \frac{1}{13-x} & x-6 \leq y \leq 6 \\ 0 & \text{otherwise} \end{cases}$$

Example - 3

Let a fair dice be rolled n times. Let $X \sim Bin(n, \frac{1}{6})$ denote the number of times 4 appears in these n dice rolls. Also, let $Y \sim Bin(n, \frac{1}{3})$ be the number of times a perfect square $(1, 4)$ appears in these n dice rolls. What is the distribution of $p_{X|Y}(x|y)$?

Example - 3

Let a fair dice be rolled n times. Let $X \sim Bin(n, \frac{1}{6})$ denote the number of times 4 appears in these n dice rolls. Also, let $Y \sim Bin(n, \frac{1}{3})$ be the number of times a perfect square $(1, 4)$ appears in these n dice rolls. What is the distribution of $p_{X|Y}(x|y)$? $Bin(y, \frac{1}{2})$.
What is the distribution of $p_{Y|X}(y|x)$?

Example - 3

Let a fair dice be rolled n times. Let $X \sim Bin(n, \frac{1}{6})$ denote the number of times 4 appears in these n dice rolls. Also, let $Y \sim Bin(n, \frac{1}{3})$ be the number of times a perfect square $(1, 4)$ appears in these n dice rolls. What is the distribution of $p_{X|Y}(x|y)$? $Bin(y, \frac{1}{2})$.

What is the distribution of $p_{Y|X}(y|x)$? $Bin(n-x, \frac{1}{5})$.

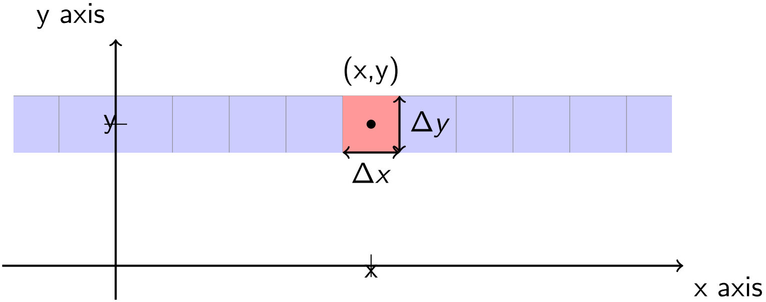
Continuous Case: Joint Conditional PDF

For two continuous random variables X, Y with joint pdf given by $f_{X,Y}(x, y)$, for the values of $Y = y$ where the density function $f_Y(y) \neq 0$, there the joint conditional probability density function is given by,

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

Proof: We look at a small rectangle of side lengths Δx and Δy in the neighborhood of the point (x, y). In the picture, we need the conditional probability of the small red rectangle, conditioned on the the fact that we are in the big blue rectangle. To get density, this will eventually be divided by Δx and we'll take the limit as $\Delta x \to 0$.

Continuous Case: Joint Conditional PDF



Continuous Case: Joint Conditional PDF

So the conditional probability of the pink rectangle, conditioning on the blue rectangle is: $P(X \in [x - \frac{\Delta x}{2}, x + \frac{\Delta x}{2}] | Y = y) = \lim_{\Delta y \to 0} \frac{f_{X,Y}(x,y)\Delta x \Delta y}{f_Y(y)\Delta y} = \frac{f_{X,Y}(x,y)\Delta x}{f_Y(y)}$

Then, $f_{X|Y}(x|y) = \lim_{\Delta x \to 0} \frac{P(X \in [x - \frac{\Delta x}{2}, x + \frac{\Delta x}{2}] | Y = y)}{\Delta x} = \lim_{\Delta x \to 0} \frac{1}{\Delta x} \frac{f_{X,Y}(x,y)\Delta x}{f_Y(y)} = \frac{f_{X,Y}(x,y)}{f_Y(y)}$

Continuous Case: Joint Conditional CDF

For the joint conditional cumulative distribution function, we have to note that y in $f_{X|Y}(x|y)$ is treated as a constant, and this is entirely a function of x, so,

$$F_{X|Y}(a|y) = \int_{-\infty}^{a} f_{X|Y}(x|y)dx$$

Examples

Let $f_{X,Y}(x,y) = \begin{cases} \frac{12}{5}y(2-x-y) & 0 < x < 1, 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$.

What is $f_{Y|X}(y|x)$?

Examples cont.

Let $f_{X,Y}(x,y) = \begin{cases} \frac{e^{-\frac{x}{y}} e^{-y}}{y} & 0 < x < \infty, 0 < y < \infty \\ 0 & \text{otherwise} \end{cases}.$

What is $F_{X|Y}(x|y)$?

Use this to find $P(X > 1|Y = y)$.

The End