Data Mining for Business Intelligence (IBM 312)

Sumit Kumar Yadav

Department of Management Studies

January 24, 2023

Random Sum of Random Numbers

- In a tea shop, the number of customers coming in a given day follows a Poisson distribution with parameter 500
- Each customer makes the purchase as per the following distribution - a.) No purchase with probability = 0.1, one cup of tea with probability = 0.8, two cups of tea with probability = 0.1
- Let X denote the number of tea cups sold in a day. What is E(X) and Var(X)

Conditional Expectation

$P(X=2 \mid Y=-1) = \frac{0.05}{0.2}$

Joint Probability		X		
		2	3	5
Y	-1	0.05	0.05	0.1
	0	0.2	0.1	0.05
	2	0.01	0.02	0.05
	5	0.07	0.25	0.05

$P(X=3 \mid Y=-1) = \frac{0.05}{0.2}$

$P(X=5 \mid Y=-1) = \frac{0.1}{0.2}$

$P(E(X \mid Y) = 3.75) = 0.2$

What is the value of E(X|Y) ?

Is it a function of Y ?

Is it a random variable?

What is E(X) ?

Is it same as $E_y(E_x(X|Y))$

$E(X \mid Y=-1) = \underline{3.75}$

$E(X \mid Y=0) =$ ____

$E(X \mid Y=2) =$ ____

$E(X \mid Y=5) =$ ____

Conditional Expectation

Joint Probability		X		
		2	3	5
Y	-1	0.05	0.05	0.1
	0	0.2	0.1	0.05
	2	0.01	0.02	0.05
	5	0.07	0.25	0.05

What is the value of E(X|Y) ?
Is it a function of Y ?
Is it a random variable?
What is E(X) ?
Is it same as $E_y(E_x(X|Y))$

Conditional Expectation

Joint Probability		X		
		2	3	5
Y	-1	0.05	0.05	0.1
	0	0.2	0.1	0.05
	2	0.01	0.02	0.05
	5	0.07	0.25	0.05

What is the value of E(X|Y) ?
Is it a function of Y ?
Is it a random variable?
What is E(X) ?
Is it same as $E_y(E_x(X|Y))$

Conditional Expectation

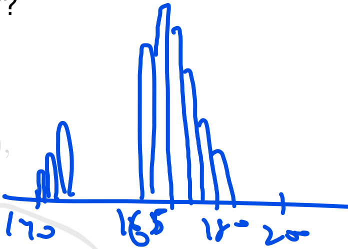
Joint Probability		X		
		2	3	5
Y	-1	0.05	0.05	0.1
	0	0.2	0.1	0.05
	2	0.01	0.02	0.05
	5	0.07	0.25	0.05

$\boxed{E_Y(E_X(X|Y)) = E(X)}$

What is the value of E(X|Y) ?
Is it a function of Y ?
Is it a random variable?
What is E(X) ?
Is it same as $E_y(E_x(X|Y))$

Continuous Random Variables

If a random variable X can take on any of a continuum of values, say, any value between 0 and 1, then we cannot define it by listing values x_i and giving the probability p_i that $X = x_i$; Why??



Two ways of defining -
the *cumulative distribution function*:

$$F(x) \equiv \mathrm{Prob}(X <= x),$$

or the *probability density function* (pdf):

$$\rho(x)\,dx \equiv \mathrm{Prob}(X \in [x, x + dx]) = F(x + dx) - F(x).$$

Letting $dx \to 0$, we find

$$\rho(x) = F'(x), \quad F(x) = \int_{-\infty}^{x} \rho(t)\,dt.$$

Continuous Random Variables

If a random variable X can take on any of a continuum of values, say, any value between 0 and 1, then we cannot define it by listing values x_i and giving the probability p_i that $X = x_i$; Why??
Two ways of defining -
the *cumulative distribution function*:

$$F(x) \equiv \text{Prob}(X <= x),$$

or the *probability density function* (pdf):

$$\rho(x)\, dx \equiv \text{Prob}(X \in [x, x + dx]) = F(x + dx) - F(x).$$

Letting $dx \to 0$, we find

$$\rho(x) = F'(x), \quad F(x) = \int_{-\infty}^{x} \rho(t)\, dt.$$

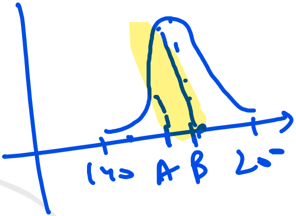
Continuous Random Variables

If a random variable X can take on any of a continuum of values, say, any value between 0 and 1, then we cannot define it by listing values x_i and giving the probability p_i that $X = x_i$; Why??

Two ways of defining -

the *cumulative distribution function*:

$$F(x) \equiv \text{Prob}(X <= x),$$



or the *probability density function* (pdf):

$$\rho(x)\,dx \equiv \text{Prob}(X \in [x, x + dx]) = F(x + dx) - F(x).$$

Letting $dx \to 0$, we find

$$\rho(x) = F'(x), \quad F(x) = \int_{-\infty}^{x} \rho(t)\,dt.$$

Expected Value

The expected value of a continuous random variable X is then defined by

$$E(X) = \int_{-\infty}^{\infty} x\rho(x)\,dx.$$

Note that by definition, $\int_{-\infty}^{\infty} \rho(x)\,dx = 1$. The expected value of X^2 is

$$E(X^2) = \int_{-\infty}^{\infty} x^2\rho(x)\,dx,$$

and the variance is again defined as $E(X^2) - (E(X))^2$.

Expected Value

The expected value of a continuous random variable X is then defined by

$$E(X) = \int_{-\infty}^{\infty} x\rho(x)\,dx.$$

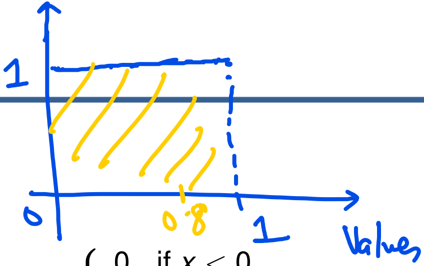
Note that by definition, $\int_{-\infty}^{\infty} \rho(x)\,dx = 1$. The expected value of X^2 is

$$E(X^2) = \int_{-\infty}^{\infty} x^2\rho(x)\,dx,$$

and the variance is again defined as $E(X^2) - (E(X))^2$.

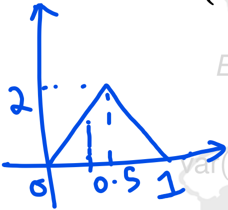
Uniform Distribution

- ~~Prob~~ Density



Example: Uniform Distribution in [0, 1].

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \le x \le 1 \\ 1 & \text{if } x > 1 \end{cases}, \quad \rho(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } 0 \le x \le 1 \\ 0 & \text{if } x > 1 \end{cases}$$



$$E(X) = \int_{-\infty}^{\infty} x\rho(x)\,dx = \int_0^1 x\,dx = \frac{1}{2},$$

$$\int_0^{0.3} 4x\,dx = 2x^2\Big|_0^{0.3} = 0.18$$

$$\mathrm{Var}(X) = \int_0^1 x^2\,dx - \left(\frac{1}{2}\right)^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}.$$

$P(X > 0.8) = 0.2$

$P(x < 0.3)$

Uniform Distribution

Example: Uniform Distribution in $[0, 1]$.

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \le x \le 1 \\ 1 & \text{if } x > 1 \end{cases}, \quad \rho(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } 0 \le x \le 1 \\ 0 & \text{if } x > 1 \end{cases}$$

$$E(X) = \int_{-\infty}^{\infty} x\rho(x)\, dx = \int_0^1 x\, dx = \frac{1}{2},$$

$$\mathrm{var}(X) = \int_0^1 x^2\, dx - \left(\frac{1}{2}\right)^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}.$$

Normal Distribution

Example: Normal (Gaussian) Distribution, Mean μ, Variance σ^2.

$$\rho(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),$$

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) dt$$

Simulation of Random numbers in Python

Practice Problem

Random Variable X follows uniform distribution from [0,1], Random Variable Y follows same distribution and is independent of X. What is the distribution of X+Y?