STATISTICS FOR DATA SCIENCE

Linear Functions of Random Variables

Prof. Uma D

Department of Computer Science and Engineering

STATISTICS FOR DATA SCIENCE

Linear Functions of Random Variables

Prof. Uma D

Topics to be covered...

- Different Transformations
- Means of Linear Combinations of Random Variables
- Variance of Linear Combinations of Random Variables
- Independent Random Variables

Linear Functions of Random variables

- We often **construct new random variables** by performing arithmetic operations on other random variables.

- For example, we might add a constant to a random variable, multiply a random variable by a constant, or add two or more random variables together.

Linear Functions of Random variables – Different Transformations

Addition – Adding a constant to each value of X.

Subtraction – Subtracting a constant from each value of X.

Multiplication – Multiplying each value of X by a constant.

Division – Dividing each value of X by a constant.

where, X represents a Random Variable.

Adding a Constant

- When a **constant is added** to a random variable, the **mean** is **increased** by the **value of the constant**, but the **variance** and **standard deviation** are **unchanged.**

Multiplying by a Constant

- Often we need to multiply a random variable by a constant.
- For example, to convert to a more convenient set of units.
- Multiplication by a constant affects the mean, variance, and standard deviation of a random variable.

Multiplying by a Constant

- In general, when a random variable is multiplied by a constant, its mean is multiplied by the same constant.

- In general, when a random variable is multiplied by a constant, its variance is multiplied by the square of the constant.

Multiplying by a Constant

- If a random variable is multiplied by a constant and then added to another constant, the effect on the mean and variance.

Applying Transformations to a random variable X

Tranformatio n	Effect on mean	Effect on Variance	Effect on shape of probability histogram
+ or - a Constant	√ Changes	X Doesn't change	X Doesn't change
* or / by a Constant	√ Changes	√ Changes	X Doesn't change

Applying Transformations to a random variable X

Tranformation	Effect on mean	Effect on Variance	Effect on SD	Effect on shape of probability histogram
Add Ex : $Y = X + 2$	$E(Y) = E(X) + 2$	$Var(Y) = Var(X)$	$SD(Y) = SD(X)$	Doesn't change
Subtract Ex : $Y = X - 2$	$E(Y) = E(X) - 2$	$Var(Y) = Var(X)$	$SD(Y) = SD(X)$	Doesn't change
Multiplying by a constant Ex: $Y = X * 2$	$E(Y) = E(X) * 2$	$Var(Y) = Var(X) * 2^2$	$SD(Y) = SD(X) * 2$	Doesn't change
Dividing by a constant Ex: $Y = X/2$	$E(Y) = E(X) / 2$	$Var(Y) = Var(X) /2^2$	$SD(Y) = SD(X) / 2$	Doesn't change

Combining Variables

Many interesting statistics problems require us to examine two or more random variables.

Ex.: **Casino games**

- Roulette games
- Deal of shuffled cards
- Roll of Dice

Means of Linear Combinations of Random Variables

Consider the case of adding two random variables.

Means of Linear Combinations of Random Variables

To find the mean of a linear combination of random variables.

Variance for Linear Combinations of IRV

Example

SRS travels offers a half-day trip in a tourist area. There must be at least 2 passengers for the visit to run. The vehicle provided by SRS travels can hold up to 6 passengers.

SRS travels charges Rs. 150 per passenger. The amount spent on petrol and permit by SRS travels per trip is Rs. 100. Number of passengers that turn up on a randomly selected day(X) and the corresponding probabilities are given below.

X	2	3	4	5	6
p(x)	0.15	0.25	0.35	0.20	0.05

Example

X	2	3	4	5	6
p(x)	0.15	0.25	0.35	0.20	0.05

Define new random variables for the following:

1. The amount SRS travels collects on a randomly selected day and write the probability distribution function **(Let the r.v. be Y).**

2. Profit made by SRS travels on a randomly selected day. **(Let the r.v. be Z).**

Example

	X	p(x)	Y (= 150 * X)	p(y)	Z (= Y – 100)	p(z)
Probability Distribution	2	0.15	300	0.15	200	0.15
	3	0.25	450	0.25	350	0.25
	4	0.35	600	0.35	500	0.35
	5	0.20	750	0.20	650	0.20
	6	0.05	900	0.05	800	0.05
Mean	E(X) = 3.75		E(Y) = 150 * E(X) = 562.5		E(Z) = E(Y) – 100 = 462.5	
SD	SD(X) = 1.09		SD(Y) = 150 * SD(X) = 163.5		SD(Z) = SD(Y) = 163.5	

Independent Random Variables

- The notion of independence for random variables is very much like the notion of independence for events.
- Two random variables are independent if knowledge concerning one of them does not affect the probabilities of the other.
- When two events are **independent**, the probability that both occur is found by multiplying the probabilities for each event.

  i.e. **$P(X=a$** and **$Y=b) = P(X=a) * P(Y=b)$**

Note: If they are **not independent** then a **covariance term** has to be introduced.

Ex. : Flips of a coin

Rolls of a dice

Independent and Identically Distributed Random Variables

IID Variables

If $X_1, X_2, X_3, \ldots, X_n$ are **independent random variables** all with same **Distribution** then, they are called independent and identically distributed (**i. i. d.**)

Independent : outcome of one observation does not affect the outcome of other observation.

Identically Distributed: They have **same mean** and **variance**.

Ex.: Casino games

- Roulette games
- Deal of shuffled cards
- Roll of Dice

Example

If X and ,Y are independent random variables such that $E(X) = 9.5$ and $E(Y) = 6.8$, $SD(X) = 0.4$ and $SD(Y) = 0.1$

Find Means and SD of the following:

1) $3X$
2) $Y - X$
3) $X + 4Y$

Example

Solution:

	Mean	SD
X	9.5	0.4
Y	6.8	0.1
3X	3 * 9.5 = 28.5	3 * 0.4 = 1.2
Y – X	6.8 – 9.5 = -2.7	Sqrt($0.1^2 + 0.4^2$) = 0.4123
X + 4Y	9.5 + 4 * 6.8 = 36.7	Sqrt(0.4^2 + ($4^2 * 0.1^2$) = 0.5656

Combinations of Two Random Variables

- Means combine very easily with addition or subtraction.
- We can’t add standard deviations.
- Variances can be added,
- Add variances even if subtracting the random variable.

Note: Attention!!!

X + X not same as 2X

Example

Rectangular plastic covers for a compact disc (CD) tray have specifications regarding length and width. Let X be the length and Y be the width, each measured to the nearest millimeter, of a randomly sampled cover. The probability mass function of X is given by $P(X = 129) = 0.2$, $P(X = 130) = 0.7$, and $P(X = 131) = 0.1$. The probability mass function of Y is given by $P(Y = 15) = 0.6$ and $P(Y = 16) = 0.4$. The area of a cover is given by $A = XY$. Assume X and Y are independent. Find the probability that the area is 1935 mm^2.

Example

Solution

The area will be equal to 1935 if $X = 129$ and $Y = 15$. Therefore

$$\begin{aligned} P(A = 1935) &= P(X = 129 \text{ and } Y = 15) \\ &= P(X = 129)P(Y = 15) \qquad \text{since } X \text{ and } Y \text{ are independent} \\ &= (0.2)(0.6) \\ &= 0.12 \end{aligned}$$

Example

A piston is placed inside a cylinder. The clearance is the distance between the edge of the piston and the wall of the cylinder and is equal to one-half the difference between the cylinder diameter and the piston diameter. Assume the piston diameter has a mean of 80.85 cm with a standard deviation of 0.02 cm. Assume the cylinder diameter has a mean of 80.95 cm with a standard deviation of 0.03 cm. Find the mean clearance. Assuming that the piston and cylinder are chosen independently, find the standard deviation of the clearance.

Example

Solution

Let X_1 represent the diameter of the cylinder and let X_2 the diameter of the piston. The clearance is given by $C = 0.5X_1 - 0.5X_2$. Using Equation (2.49), the mean clearance is

$$\begin{aligned}\mu_C &= \mu_{0.5X_1-0.5X_2}\\ &= 0.5\mu_{X_1} - 0.5\mu_{X_2}\\ &= 0.5(80.95) - 0.5(80.85)\\ &= 0.050\end{aligned}$$

Since X_1 and X_2 are independent, we can use Equation (2.53) to find the standard deviation σ_C:

$$\begin{aligned}\sigma_C &= \sqrt{\sigma^2_{0.5X_1-0.5X_2}}\\ &= \sqrt{(0.5)^2\sigma^2_{X_1} + (-0.5)^2\sigma^2_{X_2}}\\ &= \sqrt{0.25(0.02)^2 + 0.25(0.03)^2}\\ &= 0.018\end{aligned}$$

Example

Do It Yourself !!!

X – represents the number of passengers on a randomly selected trip with SRS travels.

Y – represents the number of passengers on a randomly selected trip with VRL Logistics.

The probability distributions of X and Y are given below:

X	2	3	4	5	6
p(x)	0.15	0.25	0.35	0.20	0.05

Y	2	3	4	5
p(y)	0.3	0.4	0.2	0.1

Define a new random variable T that represents number of passengers that SRS and VRL can expect on a randomly selected day.

Example

Do It Yourself !!!

1. Define a new random variable T that represents number of passengers that SRS and VRL can expect on a randomly selected day.

2. Find mean and variance of T.

THANK YOU

Prof. Uma D

Department of Computer Science and Engineering