Mathematics For Computer Science Engineers

Linear Functions on Random Variables

Mamatha.H.R

Department of Computer Science and Engineering

Mathematics For Computer Science Engineers

Linear Functions on Random Variables

Mamatha H R

Department of Computer Science and Engineering

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
  - Add two or more random variables together.
    (Combine multiple random variables)

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
  - the variance and standard deviation remains unchanged

If X is a random variable and b is a constant, then

$$\mu_{X+b} = \mu_X + b$$

$$\sigma^2_{X+b} = \sigma^2_X$$

Multiplying by a Constant

- Often we need to multiply a random variable by a constant
- Multiplication by a constant affects the mean, variance, and standard deviation of a random variable:
  - Mean gets multiplied by the constant

If X is a random variable and a is a constant, then

$$\mu_{aX} = a\mu_X$$

  - Variance is multiplied by square of the constant
  - Standard Deviation gets multiplied by modulus value of the constant

If X is a random variable and a is a constant, then

$$\sigma^2_{aX} = a^2\sigma^2_X$$

$$\sigma_{aX} = |a|\sigma_X$$

Multiplying and Adding by a Constant

- Multiplication and Addition by a constant affects the mean, variance, and standard deviation of a random variable (Say constant ‘a’ is multiplied and ‘b’ is added):
  - Mean gets multiplied by the constant ‘a’ and the product value is increased by constant ‘b’
  - Variance is multiplied by square of the constant ‘a’
  - Standard Deviation gets multiplied by modulus value of the constant ‘a’

Mean: $\mu(aX+b)=a\mu_X+b \quad or \quad E(aX+b)=aE(X)+b$

Variance: $\sigma^2(aX+b)=a^2\sigma_X^2 \quad or \quad V(aX+b)=a^2V(X)$

Standard Deviation: $\sigma(aX+b)=|a|\sigma(X)$

Applying Transformations to Random Variable X

Tranformation	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 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

Example

- Marie has a part-time job walking dogs to earn money on weekends. The following probability distribution represents the probability of having a particular number of clients on any given day. If she earns $2.75 per client, how much could she expect to earn each day, on average, and what is the standard deviation of her expected earnings?

# clients	20	25	30	35	40
probability	.15	.35	.30	.15	.05

Example

- Start by finding the mean:
- $\boldsymbol{\mu}_X$=20×.15+25×.35+30×.3+35×.15+40×.05=28
- Use the mean to find the variance:
- $\sigma^2{}_x$=$(20-28)^2$×.15+$(25-28)^2$×.35+$(30-28)^2$×.30+$(35-28)^2$×.15+ $(40-28)^2$×.05
- =28.5
- Use the variance to find the standard deviation: $\sigma_x = \sqrt{28.5}$=5.3
- Now we can find her average income by multiplying the mean, 28 by Marie's rate, $2.75, to get her **average daily income of $77**
- Finally, we can multiply the calculated standard deviation, 5.3, by the rate, $2.75, to get the standard deviation of her income: 5.3×$2.75=$14.58
- What all this means is that Marie can expect to average $77 per day, on average, give or take about $14.58

Combining Random Variables

- There are many instances where it might require us to examine more than a single random variable at once
- We can form new distributions by combining random variables.

If $X_1, \ldots, X_n$ are random variables and $c_1, \ldots, c_n$ are constants, then the random variable

$$c_1X_1 + \cdots + c_nX_n$$

is called a **linear combination** of $X_1, \ldots, X_n$.

Combining Random Variables : Mean

- Consider the below scenario where X and Y are two random variables and let T be a combination of both the random variables X and Y then

	Mean
Adding: $T = X + Y$	$\mu_T = \mu_X + \mu_Y$
Subtracting: $D = X - Y$	$\mu_D = \mu_X - \mu_Y$

- More generally,

If $X_1, X_2, \ldots, X_n$ are random variables, then the mean of the sum $X_1 + X_2 + \cdots + X_n$ is given by

$$\mu_{X_1+X_2+\cdots+X_n} = \mu_{X_1} + \mu_{X_2} + \cdots + \mu_{X_n} \tag{2.47}$$

Combining Random Variables: Mean

If X and Y are random variables, and a and b are constants, then

$$\mu_{aX+bY} = \mu_{aX} + \mu_{bY} = a\mu_X + b\mu_Y \qquad (2.48)$$

More generally, if $X_1, X_2, \ldots, X_n$ are random variables and $c_1, c_2, \ldots, c_n$ are constants, then the mean of the linear combination $c_1X_1 + c_2X_2 + \cdots + c_nX_n$ is given by

$$\mu_{c_1X_1+c_2X_2+\cdots+c_nX_n} = c_1\mu_{X_1} + c_2\mu_{X_2} + \cdots + c_n\mu_{X_n} \qquad (2.49)$$

What are Independent random Variables?

- Two random variables are independent if **knowledge concerning one of them does not affect the probabilities of the other**.
- The notion of independence for random variables is very much like the notion of independence for events.
- 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)

Independent Random Variables

If X and Y are **independent** random variables, and S and T are sets of numbers, then

$$P(X \in S \text{ and } Y \in T) = P(X \in S)P(Y \in T) \tag{2.50}$$

More generally, if $X_1, \ldots, X_n$ are independent random variables, and $S_1, \ldots, S_n$ are sets, then

$$P(X_1 \in S_1 \text{ and } X_2 \in S_2 \text{ and} \cdots \text{ and } X_n \in S_n) = \\ P(X_1 \in S_1)P(X_2 \in S_2) \cdots P(X_n \in S_n) \tag{2.51}$$

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

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

Combining Random Variables: Variance

If $X_1, X_2, \ldots, X_n$ **are independent, then the variance of** $Y = c_1X_1 + c_2X_2 + \cdots + c_nX_n$ **is**

$$V(Y) = c_1^2\sigma_1^2 + c_2^2\sigma_2^2 + \cdots + c_n^2\sigma_n^2,$$

When all Constants c1,c2,…,cn = 1 then

$$\sigma^2_{X_1+X_2+\cdots+X_n} = \sigma^2_{X_1} + \sigma^2_{X_2} + \cdots + \sigma^2_{X_n}$$

Combining Random Variables: Variance

If X and Y are *independent* random variables with variances σ_X^2 and σ_Y^2, then the variance of the sum $X + Y$ is

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 \tag{2.54}$$

The variance of the difference $X - Y$ is

$$\sigma_{X-Y}^2 = \sigma_X^2 + \sigma_Y^2 \tag{2.55}$$

The fact that the variance of the difference is the *sum* of the variances may seem counterintuitive. However, it follows from Equation (2.53) by setting $c_1 = 1$ and $c_2 = -1$.

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:

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

Combining Random Variables : Important Points

- Here's a few important points about combining random variables:
  - Make sure that the **random variables are independent** or that it's reasonable to assume independence, before combining variances.
  - **Even when we subtract random variables, we still add their variances**; subtracting two variables increases the overall variability in the outcomes.
  - We can find the standard deviation of the combined distributions by taking the square root of the combined variances.

Independence and Simple Random Samples

- When a simple random sample of numerical values is drawn from a population, each item in the sample can be considered as random variables.
- Also the items in a simple random sample may be treated as **independent** when the sample is a **small proportion** (<= 5%) of a finite population.

$If\ X_1, X_2, \ldots \ldots X_n$ is a simple random sample then $X_1, X_2, \ldots \ldots X_n$ can be treated as independent random variables all with the same distribution

Independence and Simple Random Samples

- When $X_1, X_2, \ldots \ldots X_n$ *are independent random variables, all with the same distribution, it is* sometimes said that $X_1, X_2, \ldots \ldots X_n$ *are* ***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.**

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,

- We can’t add standard deviations. square root of variance is the standard deviation.
- Add variances even if subtracting the random variable

Problems for Practice

Example 1 (Discrete Random Variable)

Find $E(Y)$ If $Y = X^2 - 2X + 6$ given that X is a random variable with the following probability distribution:

x	0	1	2	3
$f(x)$	$\frac{1}{6}$	$\frac{1}{2}$	0	$\frac{1}{3}$

Example 1: Solution

$$E(Y) = E(X^2 - 2X + 6) \to E(Y) = E(X^2 - 2X + 6) \to E(Y) = E(X^2) - 2E(X) + 6$$

$$E(X^2) = 0^2\left(\frac{1}{6}\right) + 1^2\left(\frac{1}{2}\right) + 2^2(0) + 3^2\left(\frac{1}{3}\right) = \frac{7}{2}$$

$$E(X) = 0\left(\frac{1}{6}\right) + 1\left(\frac{1}{2}\right) + 2(0) + 3\left(\frac{1}{3}\right) = \frac{3}{2}$$

$$E(Y) = E(X^2 - 2X + 6) \to E(Y) = \frac{7}{2} - 2\left(\frac{3}{2}\right) + 6 = \frac{13}{2}$$

$$E(Y) = \frac{13}{2}$$

Example 2 (Continuous Random Variable)

If $f(x)=\frac{1}{3}x^2, -1\le x\le 2$ and $g(X)=4X+5$, find $E(g(X))$:

Solution:

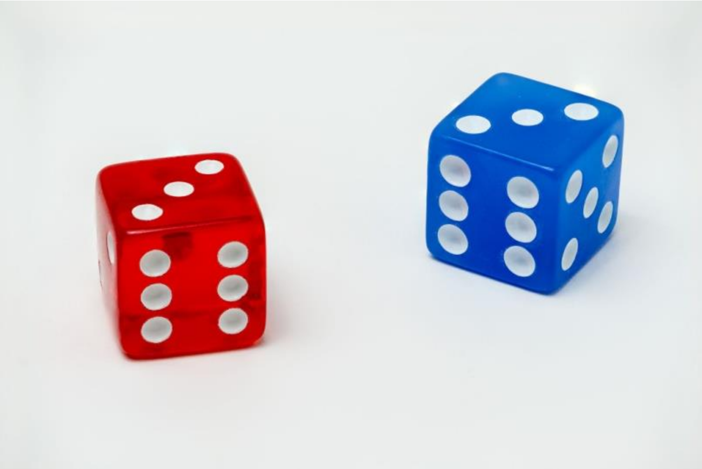
$$E(g(X))=E(4X+5)=4E(X)+5$$ where $E(X)=\int_a^b xf(x)dx$

$$E(g(X))=4\int_{-1}^{2}\left(\frac{1}{3}x^3\right)+5=\quad 4\left(\frac{1}{12}x^4\right]_{-1}^{2}+5=\quad 4\left(\frac{5}{4}\right)+5$$

$$E(g(X))=E(4X+5)=10$$

Example 3

let X represent the number that occurs when a blue die is tossed and Y, the number that happens when an orange die is tossed. Calculate the expected value of the linear combination 3X – Y of the random variables X and Y



Example 3: Solution

$X = Blue = \{1,2,3,4,5,6\}$

X	1	2	3	4	5	6
$f(x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$Y = Orange = \{1,2,3,4,5,6\}$

Y	1	2	3	4	5	6
$f(y)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$$E(X) = \sum xf(x) = 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right) = \frac{7}{2}$$

$$E(Y) = \sum yf(y) = 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right) = \frac{7}{2}$$

$$E(3X - Y) = 3E(X) - E(Y)$$

$$E(3X - Y) = 3\left(\frac{7}{2}\right) - \left(\frac{7}{2}\right) = 7$$

Example 4

A certain bowling league has weekly matches where two teams compete against each other. Each team has four players whose individual scores are combined to form their team's overall score that week. Based on data from previous matches, the table below shows summary statistics for the weekly scores of two teams.

Team	Mean	Standard deviation
Bowling Buds	$\mu_B = 610$	$\sigma_B = 60$
Alleycats	$\mu_A = 630$	$\sigma_A = 80$

Let A represent the Alleycats' score on a randomly chosen week, B represent the Bowling Buds' score on a randomly chosen week, and D represent the difference between their scores $(D = A - B)$. Assume that the team's scores are independent from each other.

Find the standard deviation of D.

Example 4: Solution

The variance of the difference of two **independent** random variables X and Y is equal to the sum of their variances.

That is, if X and Y are **independent** random variables, $\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y$.

Since we can assume that their scores are independent, the variance of the difference between their scores is

$$\sigma^2_D = \sigma^2_A + \sigma^2_B$$

$$= 80^2 + 60^2$$

$$= 10{,}000$$

and the standard deviation is

$$\sigma_D = \sqrt{\sigma^2_A + \sigma^2_B} = \sqrt{10{,}000} = 100$$

Example 5

A factory process requires 3 steps to finish a product. The steps each have a mean completion time of $\mu = 10$ seconds and a standard deviation of $\sigma = 5$ seconds. The time it takes to complete each step is independent from the other steps. Let T be the total completion time for the 3 steps.

Find the standard deviation of T.

Example 5: Solution

The variance of the sum of **independent** random variables X, Y, and Z is equal to the sum of their variances.

That is, if X, Y, and Z are **independent** random variables, $\sigma^2_{X+Y+Z} = \sigma^2_X + \sigma^2_Y + \sigma^2_Z$.

Since we were told that the times are independent, the variance of the total time is

$$\sigma_T^2 = \sigma_1^2 + \sigma_2^2 + \sigma_3^2$$

$$= 5^2 + 5^2 + 5^2$$

$$= 75$$

and the standard deviation is

$$\sigma_T = \sqrt{\sigma_1^2 + \sigma_2^2 + \sigma_3^2} = \sqrt{75}$$

Example 6

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 6: Solution

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 6: Solution

	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	

The expectation of a linear function E[aX+b] is equal to

A) aE(X)+b

B) E(a)E(X)+b

C) E(X+a+b)

D) aE(X+b)a

If X is a random variable and Y=3X+5 then Var(Y)=

A) 3·Var(X)

B) 9·Var(X)

C) Var(X)+5

D) 3+Var(X)+5

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

Dr.Mamatha H R

Department of Computer Science and Engineering