Functions of Random Variables

Keegan Korthauer
Department of Statistics
UW Madison

LINEAR FUNCTIONS OF RANDOM VARIABLES

Mean

Variance

Addition of a Constant

- The addition of a constant to a random variable shifts its mean by the value of that constant and the variance/standard deviation remain unchanged

- Let X be a random variable and let *b* be a constant
  - Then **E(X + *b*) = E(X) + *b*** or $\mu_{X+b} = \mu_X + b$
  - And **Var(X + *b*) = Var(X)** or $\sigma^2_{X+b} = \sigma^2_X$
  - And **sd(X + *b*) = sd(X)** or $\sigma_{X+b} = \sigma_X$

Multiplying by a Constant

- The multiplication of a random variable by a constant also multiplies its mean by the value of that constant and multiplies the variance by the square of that constant

- Let X be a random variable and let *a* be a constant
  - Then **$E(aX) = aE(X)$** or **$\mu_{aX} = a\mu_X$**
  - And **$Var(aX) = a^2Var(X)$** or **$\sigma^2_{aX} = a^2\sigma^2_X$**
  - And **$sd(aX) = |a|sd(X)$** or **$\sigma_{aX} = |a|\sigma_X$**

In Summary

Let X be a random variable and let *a* and *b* be constants

- Then $\mathbf{E(}\boldsymbol{a}\mathbf{X + }\boldsymbol{b}\mathbf{) = }\boldsymbol{a}\mathbf{E(X) + }\boldsymbol{b}$ or $\mu_{aX+b} = a\mu_X + b$
- And $\mathbf{Var(}\boldsymbol{a}\mathbf{X + }\boldsymbol{b}\mathbf{) = }\boldsymbol{a}^{\mathbf{2}}\mathbf{Var(X)}$ or $\sigma^2_{aX+b} = a^2\sigma^2_X$
- And $\mathbf{sd(}\boldsymbol{a}\mathbf{X + }\boldsymbol{b}\mathbf{) = |}\boldsymbol{a}\mathbf{|sd(X)}$ or $\sigma_{aX+b} = |a|\sigma_X$

Example

- Let X be a RV which represents the height of corn stalks growing in a field. Assume they currently measure 38 inches tall on average, with a standard deviation of 4 inches.
- If they all grow exactly 3 more inches, what will be the mean and standard deviation of their height in **centimeters**?

Currently:

μ_X = 38 inches and σ_X = 4 inches

After growing three more inches:

μ_{X+3} = 38+3 = 41 inches and σ_{X+3} = 4 inches (unchanged)

In centimeters (using 2.54 cm per inch):

$\mu_{2.54(X+3)}$ = 2.54(38+3) = 104.14 centimeters and

$\sigma_{2.54(X+3)}$ = |2.54|*4 = 10.16 centimeters

Linear Combinations of RVs

Let $X_1,\ldots, X_n$ be a random variables and $c_1,\ldots, c_n$ be constants. Then the random variable

$$c_1X_1 + c_2X_2 + \ldots + c_nX_n$$

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

Means of Linear Combinations of RVs

Let $X_1, \ldots, X_n$ be a random variables and $c_1, \ldots, c_n$ be constants. Then the mean of the linear combination $c_1X_1 + c_2X_2 + \ldots + c_nX_n$ is

$$E(c_1X_1 + \ldots + c_nX_n) = c_1E(X_1) + \ldots + c_n E(X_n)$$

Example – Adding two RVs

- Let X_1 represent the time it takes (in minutes) to walk from my house to the bus stop. Assume $E(X_1)=3$, $Var(X_1)=1$.
- Let X_2 represent the time it takes the bus to travel between the bus stop and campus. Assume $E(X_2)=8$, $Var(X_2)=4$.
- We are interested in the total time it takes to get from home to campus
- Let $Y = X_1 + X_2$ represent to total commute time
- Then the average commute time is:

  $E(Y) = E(X_1) + E(X_2) = 3 + 8 = 11$ min

- What about the variance of the commute time?

Independence of RVs

If $X_1,\ldots, X_n$ are **independent** random variables and $S_1,\ldots, S_n$ are sets of numbers, then

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

$$P(X_1 \in S_1)P(X_2 \in S_2) \ldots P(X_n \in S_n)$$

Variance of Independent Linear Combinations of RVs

If $X_1, \ldots, X_n$ are **independent** random variables then the variance of the sum $X_1 + X_2 + \ldots + X_n$ is

$$\mathrm{Var}(X_1 + \ldots + X_n) = \mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_n)$$

If $X_1, \ldots, X_n$ are **independent** random variables and $c_1, \ldots, c_n$ are constants, then the variance of the linear combination $c_1X_1 + c_2X_2 + \ldots + c_nX_n$ is

$$\mathrm{Var}(c_1X_1 + \ldots + c_nX_n) = c_1^2\,\mathrm{Var}(X_1) + \ldots + c_n^2\,\mathrm{Var}(X_n)$$

Example – Adding two RVs Continued

If the two legs of the trip are independent, we can get the variance of the commute time:

$$Var(Y) = Var(X_1) + Var(X_2) = 1 + 4 = 5 \text{ minutes}$$

and sd(Y) = sqrt(5) = 2.24 minutes

Note that we can **NOT** find the standard deviation by the summing the standard deviations of the individual legs:

$$sd(X_1) + sd(X_2) = 1 + 2 = 3 \neq sd(Y) \,!$$

Example – Light bulb lifetimes

Let X be the RV that represents the lifetime of a light bulb, where the probability distribution function is:

$$f(x) = \begin{cases} \dfrac{1}{500} & 250 < x < 750 \\ 0 & \text{otherwise} \end{cases}$$

Find the mean and standard deviation of the total lifetime of two (independent) light bulbs from this distribution.

Example – Light bulb lifetimes

Let X be the RV that represents the lifetime of a light bulb, where the probability distribution function is:

$$f(x) = \begin{cases} \dfrac{1}{500} & 250 < x < 750 \\ 0 & \text{otherwise} \end{cases}$$

Find the mean and standard deviation of the total lifetime of two (independent) light bulbs from this distribution.

$E(X_1 + X_2) = 1000$

$sd(X_1 + X_2) = 204.12$

Identically Distributed

- A set of random variables $X_1,\ldots,X_n$ that have the same probability distribution are called **identically distributed**
- If they are also independent, then they are called **independent and identically distributed (i.i.d.)**
- Previous example: X_1 and X_2 are i.i.d. because they both have the same PDF f(x) and they are independent of one another

Expectation of the Sample Mean

- What if we want to compute the **mean of the sample mean** (for a simple random sample from a population with mean μ)
- The sample mean is a linear combination:

$$E(\bar{X}) = \mu_{\bar{X}} = E\left[\frac{1}{n}\left(X_1 + X_2 + \ldots + X_n\right)\right]$$

$$= \frac{1}{n}\left(E(X_1) + E(X_2) + \ldots + E(X_n)\right)$$

$$= \frac{1}{n}(\mu + \mu + \ldots + \mu) = \mu$$

Variance of the Sample Mean

- If we want to compute the **variance of the sample mean** (for a simple random sample from a population with variance σ^2)
- The sample mean is a linear combination:

$$Var(\bar{X}) = \sigma_{\bar{X}}^2 = Var\left[\frac{1}{n}\left(X_1 + X_2 + \ldots + X_n\right)\right]$$

$$= \frac{1}{n^2}\left(Var(X_1) + Var(X_2) + \ldots + Var(X_n)\right)$$

$$= \frac{1}{n^2}(\sigma^2 + \sigma^2 + \ldots + \sigma^2) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Example

- Suppose the mean mass of shipping containers arriving at a warehouse is 47.2 kg and the standard deviation is 8.3 kg.
- For a simple random sample of 10 containers, find the mean and variance of the average mass.

Let M_1, M_2,…, M_{10} be the masses of the random sample of 10 containers.

The average is E(M) = $\overline{M}$ = (M_1 + M_2 +… + M_{10})/n

Then E($\overline{M}$) = E(M) = 47.2 kg and

sd($\overline{M}$) = sqrt(Var(M)/n) = sd(M)/sqrt(n) = 8.3/sqrt(10) = 2.625

Functions of Random Variables

Let X be a random variable and let h(X) be an arbitrary function of X (can be nonlinear)

- If X is discrete with PMF p(x), the mean of h(X) is

$$E[h(X)] = \sum_{x} h(x)p(x)$$

- If X is continuous with PDF f(x), the mean of h(X) is

$$E[h(X)] = \int_{-\infty}^{\infty} h(x)f(x)dx$$

Example 2.58

Let X represent the bore diameter of a cylinder in mm. Assume the PDF of X is

$$f(x) = \begin{cases} 10 & 80.5 < x < 80.6 \\ 0 & \text{otherwise} \end{cases}$$

Let A = $\pi X^2/4$ represent the area of the bore. Find the mean of A.

Example 2.58

Let X represent the bore diameter of a cylinder in mm. Assume the PDF of X is

$$f(x) = \begin{cases} 10 & 80.5 < x < 80.6 \\ 0 & \text{otherwise} \end{cases}$$

Let A = $\pi X^2/4$ represent the area of the bore. Find the mean of A.

$E[A]=E[\pi X^2/4] = 5096\text{mm}^2$

Next

- Joint distributions of two random variables (2.6)