Math 3070, Applied Statistics

Section 1

October 18, 2019

Lecture Outline, 10/18

Section 5.5

- Linear Combinations

Linear Combination, Definition

Consider $X_1, X_2, \ldots, X_n$ random variables and $a_1, a_2, \ldots, a_n$.

$$Y = a_1X_1 + a_2X_2 + \ldots + a_nX_n$$

is a **linear combination** of the X_i's.

For example, setting $a_i = 1/n$ yields the mean.

Linear Combination, Properties

Random variables $X_1, X_2, \ldots, X_n$ have means $\mu_1, \mu_2, \ldots, \mu_n$ and variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$, respectively.

- It is always true that

$$E(Y) = a_1\mu_1 + \ldots + a_n\mu_n$$

- Assuming the X_i's are independent

$$Var(Y) = a_1^2\sigma_1^2 + \ldots + a_n^2\sigma_n^2$$

- It is always true that

$$Var(Y) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j Cov(X_i, X_j)$$

Linear Combination, Properties

$$
\begin{aligned}
E(Y) &= E(a_1X_1 + a_2X_2 + \ldots + a_nX_n) \\
&= a_1E[X_1] + a_2E[X_2] + \ldots + a_nE[X_n] \\
&= a_1\mu_1 + a_2\mu_2 + \ldots + a_n\mu_n \\
Var(Y) &= E\left[\left(\sum_{i=1}^{n} a_iX_i - E\left[\sum_{j=1}^{n} a_jX_j\right]\right)^2\right] \\
&= E\left[\left(\sum_{i=1}^{n} a_iX_i - \sum_{j=1}^{n} a_jE[X_j]\right)^2\right] \\
&= E\left[\left(\sum_{i=1}^{n} (a_iX_i - a_iE[X_i])\right)^2\right] \\
&= E\left[\left(\sum_{i=1}^{n} (a_iX_i - a_iE[X_i])\right)\left(\sum_{j=1}^{n} (a_jX_j - a_jE[X_j])\right)\right]
\end{aligned}
$$

Linear Combination, Properties

$$= E\left[\sum_{i=1}^{n}\sum_{j=1}^{n}\left(a_i X_i - a_i E[X_i]\right)\left(a_j X_j - a_j E[X_j]\right)\right]$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} E\left[\left(a_i X_i - a_i E[X_i]\right)\left(a_j X_j - a_j E[X_j]\right)\right]$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} Cov(a_i X_i, a_j X_j)$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j Cov(X_i, X_j)$$

independence $\rightarrow$ $Cov(X_i, X_j) = 0$ if $i \neq j$

$$Var(Y) = \sum_{i=1}^{n} a_i^2 Cov(X_i, X_i) = \sum_{i=1}^{n} a_i^2 Var(X_i)$$, assuming independence

Linear Combination, Example

The number of bus arrivals at any stop are independent follow a Poisson distribution and average 5 hours an hour. You need to take 3 buses to reach Lagoon. Each bus ride takes 10 minutes. You are currently waiting for the first bus. Compute the mean and standard deviation of the total time of the trip.

$$X_i = \# \text{ of buses in an hour at stop } i, X_i \sim Poisson(5)$$

$$T_i = \# \text{ waiting time at bus stop } i,\ T_i \sim \exp(5)$$

Due to the memoryless property, it does not matter when you arrive at a bus stop. The next wait time has the same distribution. Not true in reality.

$$\begin{aligned} T &= \text{Total trip time} \\ &= T_1 + 1/6 + T_2 + 1/6 + T_3 + 1/6 \\ &= T_1 + T_2 + T_3 + 1/2 \end{aligned}$$

Linear Combination, Example

$$
\begin{aligned}
E[T] &= E[T_1 + T_2 + T_3 + 1/2] \\
&= E[T_1] + E[T_2] + E[T_3] + \frac{1}{2} \\
&= \frac{1}{\lambda} + \frac{1}{\lambda} + \frac{1}{\lambda} + \frac{1}{2} \\
&= \frac{1}{5} + \frac{1}{5} + \frac{1}{5} + \frac{1}{2} = \frac{11}{10} \\
&= \text{an hour and ten minutes} \\
Var(T) &= V(T_1 + T_2 + T_3 + \frac{1}{2}) \\
&= V(T_1) + V(T_2) + V(T_3) \\
&= \frac{1}{\lambda^2} + \frac{1}{\lambda^2} + \frac{1}{\lambda^2} = \frac{1}{25} + \frac{1}{25} + \frac{1}{25} \\
&\approx 7.2\text{minutes}
\end{aligned}
$$

Linear Cominations of Normal Random Variables

If $X_i \sim N(\mu_i, \sigma_i)$, then

$$Y = \sum_{i=1}^{n} a_i X_i \sim N\left(\sum_{i=1}^{n} a_i \mu_i, \sqrt{\sum_{i=1}^{n} a_i^2 \sigma_i^2} \right)$$

Using variance notation instead of standard deviation notation looks more tidy.

$$X_i \sim N(\mu_i, \sigma_i^2) \Rightarrow Y = \sum_{i=1}^{n} a_i X_i \sim N\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2 \right)$$

I plan to use variance notation in the notes from here and onward.

Example, Difference of Random Variables

A machine produces X grams of chemical X. Another produces Y grams of another chemical Y. Every two grams of chemical Y reacts one gram of chemical X when mixed. Assume that X is normally distributed with a mean of 150 grams and a standard deviation of 2 grams, Y is normally distributed with a mean of 300 grams and variance of 1 gram, and that X and Y are independent. What is the probability that more than one gram of chemical Y remains after mixing?

$$X \sim N(150, 4), Y \sim N(300, 1)$$

$L = \frac{1}{2}Y - X$ normally distributed. Just need the mean and standard deviation. There is leftover chemical Y when $L > 0$.

$$E[L] = \frac{1}{2}E[Y] - E[X] = \frac{1}{2}300 - 150 = 0$$

$$Var(L) = \left(\frac{1}{2}\right)^2 Var(Y) + (-1)^2 Var(X) = 1 + 1 = 2$$

Example, Difference of Random Variables

$$L \sim N(0, 2)$$

Want $P(L > 1)$ which has an event contained inside the event that there is leftover chemical Y, $\{L > 0\}$.

$$\begin{aligned} P(L > 1) &= P\left(\frac{L}{\sqrt{2}} > \frac{1}{\sqrt{2}}\right) \\ &= P\left(Z > \frac{1}{\sqrt{2}}\right) \\ &= \Phi(1/\sqrt{2}) \approx 0.2398 \end{aligned}$$

Summary

- $Y = a_1 X_1 + a_2 X_2 + \ldots + a_n X_n$
- $E[Y] = \sum_{i=1}^{n} a_i E[X_i]$
- Assuming independence, $Var(Y) = \sum_{i=1}^{n} a_i^2 Var(X_i)$
- Generally, $Var(Y) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j Cov(X_i, X_j)$
- $X_i \sim N(\mu_i, \sigma_i^2) \Rightarrow$
  $Y = \sum_{i=1}^{n} a_i X_i \sim N\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2 \right)$