STAT 230 SECTION 2 LECTURE 34

Today's Agenda

Last time:

- Linear combinations

Today (Lec 34, 07/22):

- Linear combinations of independent normal random variables
- Indicator random variables

Definition

Suppose that $X_1, ..., X_n$ are jointly distributed RVs with joint probability function $f(x_1, ..., x_n)$. A **linear combination** of the RVs $X_1, ..., X_n$ is any random variable of the form

$$\sum_{i=1}^{n} a_i X_i$$

where $a_1, ..., a_n \in \mathbb{R}$. If $\mathbf{X} = (X_1, ..., X_n)^\top$, $\mathbf{a} = (a_1, ..., a_n)^\top$, then a linear combination is

$$\mathbf{X}^\top \mathbf{a}.$$

Example

Some "famous" linear combinations (ones you will see in STAT 231/241) are

a) The total

$$T = \sum_{i=1}^{n} X_i \quad a_i = 1, \;\; 1 \le i \le n$$

b) The sample mean

$$\bar{X} = \sum_{i=1}^{n} \frac{1}{n} X_i \quad a_i = \frac{1}{n}, \;\; 1 \le i \le n$$

Expected Value of a Linear Combination

Theorem

$$E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i)$$

This follows directly from the linearity of expected value.

Variance of a linear combination

The following result shows how the variance of a linear combination is "broken down" into pieces.

Theorem

Let X, Y be random variables, and $a, b \in \mathbb{R}$. Then,

$$Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2abCov(X, Y).$$

In general,

$$Var(\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i^2 Var(X_i) + 2 \sum_{1 \le i < j \le n} a_i a_j Cov(X_i, X_j)$$

An immediate, yet very useful, application of these results comes from a linear combination of normally-distributed random variables.

The following results about normal random variables are not only useful in STAT 230, but they will appear throughout your statistics education, and they are used in cutting-edge research very often.

Theorem ("A linear function of normal is normal")

Let $X \sim N(\mu, \sigma^2)$ *and* $Y = aX + b$*, where* $a, b \in \mathbb{R}$*. Then,*

$$Y \sim N(a\mu + b, a^2\sigma^2).$$

Theorem ("A linear combination of normal is normal")

Let $X_i \sim N(\mu_i, \sigma_i^2),\ i = 1, 2, \ldots, n$ **independently**, *and* $a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_n \in \mathbb{R}$. *Then,*

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

Theorem ("Sample mean of normal is normal")

Let $X_i \sim N(\mu, \sigma^2)$, $i = 1, 2, \ldots, n$ **independently**. *Then,*

$$\sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2),$$

and

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

It may not be obvious why we have $n\sigma^2$ and σ^2/n in the third theorem when $Var(aX) = a^2\sigma^2$. Let's see why.

Example

Suppose that $X_i \sim N(\mu, \sigma^2)$, and that $X_1, ..., X_n$ are independent. Compute $Var\left(\sum_{i=1}^{n} X_i\right)$ and $Var(\bar{X})$.

Example

Three cylindrical parts are joined end to end to make up a shaft in a machine: 2 type-A parts and 1 type-B part. The lengths of the parts vary a little, and have the following distributions:

$$A \sim N(6, 0.4), \quad B \sim N(35.2, 0.6).$$

The overall length of the assembled shaft must lie between 46.8 and 47.5 or else the shaft has to be scrapped. Assume the lengths of different parts are independent.

a) What percent of assembled shafts have to be scrapped?

b) What is the scrapping percentage if we reduce the variance of A and B by 50% each?

This kind of problem is an example of "process improvement" or "industrial engineering". The purpose of this field is to optimise the current process while minimising the cost of improvement.

Here's an example where process improvement techniques can be useful: `http://fortune.com/2018/07/16/recall-fda-valsartan-blood-pressure-medication/`

The statistics department offers a course on it: STAT 435 - Statistical Methods for Process Improvements.

Example

Suppose that the height of adult males in Canada is normally distributed with a mean of 70 inches and variance of 4^2 inches, and let $X_1, ..., X_{10}$ denote the heights of a random sample of adult males. Suppose $\bar{X}_{10}$ denotes the sample mean of these heights.

a) Compute the probability that X_3 exceeds 75.
b) Compute the probability that $\bar{X}_{10}$ exceeds 75

Definition

Let $A \subset S$ be an event. We say that $\mathbb{1}_A$ is the **indicator** random variable of the event A. $\mathbb{1}_A$ is defined by:

$$\mathbb{1}_A(\omega) = \begin{cases} 1 & \omega \in A, \\ 0 & \omega \in \bar{A} \end{cases}$$

Actually, it's another way of defining a Bernoulli random variable. Nothing new, really.

Indicator (Bernoulli) variable has a few interesting properties:

a) $\mathsf{E}[\mathbb{1}_A] = P(A)$
b) $Var(\mathbb{1}_A) = P(A)(1 - P(A))$
c) $\mathrm{cov}(\mathbb{1}_A, \mathbb{1}_B) = P(A \cap B) - P(A)P(B)$

Question

Let $p = p(A)$. At which value of p is $Var(\mathbb{1}_A)$ maximised?

A 0

B ϵ

C 0.1

D 0.5

E 1

Question: Why do we care whatsoever about indicator random variables??

Answer: To make many calculations (like computing the mean and variance) vastly easier, and to gain intuition about how random variables are constructed/behave.

Example

Suppose $X \sim Binomial(n, p)$. Show $E(X) = np$, and $Var(X) = np(1-p)$ using indicator random variables.

Indicator variables can be incredibly useful in solving some complex problems, and they play a crucial theoretical role in some parts of statistics.

- For problem-solving with indicators: STAT 333, STAT 433
- If you want to see indicators being theoretically important: STAT 332, STAT 454