STAT 230 SECTION 2 LECTURE 21

Today's Agenda

Last time:

- Illustration of sample mean in R
- Variance of a discrete random variable

Today (Lec 21, 06/20):

- More on variance
- Continuous random variables

Recap

- If X is a discrete rv with pf f, then

$$E(X) = \sum_{x \in X(S)} x\, f(x).$$

is the expected value of X and

$$Var(X) = E((X - E(X))^2) = E(X^2) - E(X)^2$$

is the variance of X.

Recap

- If X is a discrete rv with pf f, then

$$E(X) = \sum_{x \in X(S)} x\, f(x).$$

  is the expected value of X and

$$Var(X) = E((X - E(X))^2) = E(X^2) - E(X)^2$$

  is the variance of X.

- For constants $a, b \in \mathbb{R}$,

$$E(aX + b) = aE(X) + b$$

  and

$$Var(aX + b) = a^2 Var(X).$$

- The standard deviation is defined by

$$SD(X) = \sqrt{Var(X)}$$

Two important results

Theorem (Variance of linear combination)

For any random variable X and $a, b \in \mathbb{R}$,

$$Var(aX + b) = a^2 Var(X).$$

Two important results

Theorem (Variance of linear combination)

For any random variable X and $a, b \in \mathbb{R}$,

$$Var(aX + b) = a^2 Var(X).$$

Theorem

$Var(X) = 0$ *if and only if* $P(X = E(X)) = 1$.

Proof of Theorem 2

Some formulas

- If $X \sim Binomial(n, p)$, then

$$E(X) = np, \quad Var(X) = np(1-p).$$

- If $X \sim Poi(\lambda)$ for $\lambda > 0$ then

$$E(X) = Var(X) = \lambda.$$

- If $X \sim hyp(N, r, n)$, then

$$Var(X) = n\frac{r}{N}\left(1 - \frac{r}{N}\right)\left(\frac{N-n}{N-1}\right)$$

- If $X \sim NegBin(k, p)$, then

$$Var(X) = \frac{k(1-p)}{p^2}.$$

Example

Suppose that X_n is binomial with with parameters n and p_n so that $np_n \to \lambda$ as $n \to \infty$. If $Y \sim Poi(\lambda)$ show that

$$\lim_{n\to\infty} Var(X_n) = Var(Y).$$

Why do we care about the variance?

- The expectation and the variance give a simple summary of the distribution

Why do we care about the variance?

- The expectation and the variance give a simple summary of the distribution
- Other summaries include: (more later)
  - Skewness:

$$E\left(\frac{(X - E(X))}{\sqrt{Var(X)}}\right)^3$$

  - Kurtosis:

$$\frac{E(X - E(X))^4}{(E(X - E(X))^2)^2}$$

$$\vdots$$

Random variables whose mean does not exist

There exist distributions without expectation: Suppose X is a random variable with probability function

$$f_X(x) = \frac{6}{\pi^2} \frac{1}{x^2}, \quad x = 1, 2, \ldots$$

Then $E(X) = +\infty$ and $Var(X)$ is not defined.

Question

A person plays a game in which a fair coin is tossed until the first tail occurs. The person wins \$$2^x$ if x tosses are needed for $x = 1, 2, 3, 4, 5$, but loses \$256 if $x > 5$.

a) Determine the expected winnings.
b) Determine the variance of the winnings.

Chapter 8:
Continuous Random Variables

Recall...

- a random variable is a function $X : S \to \mathbb{R}$.
- If the range $X(S)$ is...
  - countable or finite, X is called a discrete random variable
  - an interval $(a, b) \subseteq \mathbb{R}$, X is called a continuous random variable

Cutting a stick of length 2

- Suppose you cut a stick of length 2 at random and denote by X the cutting point.
- Then X can take values in $(0, 2)$.
- What is the probability that the cut X is
  - ...between 0 and 1?
  - ...between 1 and 2?
  - ...between 0 and 1/2?
- The cdf $F(x) = P(X \leq x)$ is then
- And for any $x \in (0, 2)$,

$$P(X = x) =$$

So it seems that, for a continuous random variable, assigning probabilities to **intervals** is more “natural” than assigning probabilities to specific values.

So it seems that, for a continuous random variable, assigning probabilities to **intervals** is more “natural” than assigning probabilities to specific values.

We shall do exactly that by defining the cumulative distribution function for a continuous random variable.

Definition (CDF of continuous random variable)

The cumulative distribution function of a continuous random variable X, $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$, is a function such that

Definition (CDF of continuous random variable)

The cumulative distribution function of a continuous random variable X, $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$, is a function such that

1. $F(x)$ is defined for all $x \in \mathbb{R}$,

Definition (CDF of continuous random variable)

The cumulative distribution function of a continuous random variable X, $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$, is a function such that

1. $F(x)$ is defined for all $x \in \mathbb{R}$,
2. $F(x)$ is a non-decreasing function of x for all $x \in \mathbb{R}$,

Definition (CDF of continuous random variable)

The cumulative distribution function of a continuous random variable X, $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$, is a function such that

1. $F(x)$ is defined for all $x \in \mathbb{R}$,
2. $F(x)$ is a non-decreasing function of x for all $x \in \mathbb{R}$,
3. $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$,

Definition (CDF of continuous random variable)

The cumulative distribution function of a continuous random variable X, $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$, is a function such that

1. $F(x)$ is defined for all $x \in \mathbb{R}$,
2. $F(x)$ is a non-decreasing function of x for all $x \in \mathbb{R}$,
3. $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$,
4. $P(a < X \leq b) = F(b) - F(a)$,

Note: $P(X = x) = 0$ for all $x \in \mathbb{R}$!

Strict inequalities don't matter...

If X is a continuous random variable, then

$$P(a < X \leq b) = F(b) - F(a)$$

$$P(a \leq X \leq b) = P(a < X \leq b) + P(X = a) = F(b) - F(a) + 0$$

$$P(a < X < b) = P(a < X \leq b) - P(X = b) = F(b) - F(a) - 0$$

$$P(a \leq X < b) = P(a < X \leq b) + P(X = a) - P(X = b) = F(b) - F(a)$$

so if X is continuous, all these probabilities coincide!

If X was discrete, these 4 probabilities could all be different.

So we now have a tool for computing probabilities of a continuous random variable. However, the cdf doesn't show directly which values of X are more likely or less likely.

So we now have a tool for computing probabilities of a continuous random variable. However, the cdf doesn't show directly which values of X are more likely or less likely.

To find that out, let's see how the CDF changes as the value of x changes over a small interval $[x, x + \Delta x]$.

We have

$$P(x \le X \le x + \Delta x) = F(x + \Delta x) - F(x).$$

So we now have a tool for computing probabilities of a continuous random variable. However, the cdf doesn't show directly which values of X are more likely or less likely.

To find that out, let's see how the CDF changes as the value of x changes over a small interval $[x, x + \Delta x]$.

We have

$$P(x \leq X \leq x + \Delta x) = F(x + \Delta x) - F(x).$$

To study how CDF changes, we want to look at its rate of change as Δx becomes really small. That is,

So we now have a tool for computing probabilities of a continuous random variable. However, the cdf doesn't show directly which values of X are more likely or less likely.

To find that out, let's see how the CDF changes as the value of x changes over a small interval $[x, x + \Delta x]$.

We have

$$P(x \leq X \leq x + \Delta x) = F(x + \Delta x) - F(x).$$

To study how CDF changes, we want to look at its rate of change as Δx becomes really small. That is,

$$\lim_{\Delta x \to 0} \frac{P(x \leq X \leq x + \Delta x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x}.$$

So we now have a tool for computing probabilities of a continuous random variable. However, the cdf doesn't show directly which values of X are more likely or less likely.

To find that out, let's see how the CDF changes as the value of x changes over a small interval $[x, x+\Delta x]$.

We have

$$P(x \leq X \leq x + \Delta x) = F(x + \Delta x) - F(x).$$

To study how CDF changes, we want to look at its rate of change as Δx becomes really small. That is,

$$\lim_{\Delta x \to 0} \frac{P(x \leq X \leq x + \Delta x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x}.$$

Looks familiar? That's the derivative!

Definition (Probability density function)

The probability density function (pdf) of a continuous random variable X is the derivative of the cdf

$$f(x) = \frac{d}{dx}F(x),$$

where $f(x)$ is continuous.

Definition (Probability density function)

The probability density function (pdf) of a continuous random variable X is the derivative of the cdf

$$f(x) = \frac{d}{dx} F(x),$$

where $f(x)$ is continuous.

Definition (Support of pdf)

The support of a pdf $f(x)$ is defined as

$$supp(f) = \{x \in \mathbb{R} : f(x) \neq 0\}.$$

Integrals of a pdf f over a domain D may be computed on $D \cap supp(f)$.

Properties of the pdf

Properties of the pdf

a)

$$f(x) \geq 0 \quad \forall x \in \mathbb{R}$$

b)

$$\int_{-\infty}^{\infty} f(x) dx = 1$$

c)

$$P(a \leq X \leq b) = \int_{a}^{b} f(x) dx \quad \text{for } a < b \in \mathbb{R}$$

Example

Suppose that X is a continuous random variable with probability density function

$$f(x) = \begin{cases} cx(1-x) & \text{if } 0 \leq x \leq 1, \\ 0 & \text{otherwise} \end{cases}$$

Example

Suppose that X is a continuous random variable with probability density function

$$f(x) = \begin{cases} cx(1-x) & \text{if } 0 \leq x \leq 1, \\ 0 & \text{otherwise} \end{cases}$$

a) Compute c so that this is a valid pdf
b) Graph $f(x)$
c) Compute $P(X \geq 1/2)$
d) Compute $P(1/4 \leq X \leq 3/4)$
e) Compute $P(X = 1/2)$