
THANK GOD IT'S FRIDAY?
FALSE. THANK POPE GREGORY VII FOR THE GREGORIAN CALENDAR DECREE
makeameme.org

Today's Agenda

Last time:

- Expected value of Binomial and Poisson

Today (Lec 20, 06/17):

- Illustration of sample mean in R
- Variance

Note that there will be a R tutorial on Monday!

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.

- If $g : \mathbb{R} \to \mathbb{R}$ is a function, then

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

"Law of unconscious statistician"

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

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

"Linearity of the expectation"

- Some people use round brackets() and some people use [], and they both mean the same, so $E(X) = E[X]$.

With the expectation, we now know about the "weighted average" value of the distribution.

On that note, think about this:

- X is a r.v. representing the outcome of one fair 6-sided die roll
- Y is a r.v. representing the number of phone calls over 1 minute at Lenovo call centre, with the rate of 3.5 calls per minute

Are they similar in any ways? How are they different?

X and Y have the same expectation:

$$\mu_X = \mathrm{E}(X) = 3.5 = \mathrm{E}(Y) = \mu_Y.$$

but the extent of deviation from the expectations are different:

- For X, 1 or 6 are just as likely to appear as 2 or 5.
- For Y, values closer to 3.5 are more likely to appear, but you could possibly get an insanely high number too. (ex: order backlog from Black Friday)

This is illustrated in the R script `lec20_rscript.R` on Learn.

As seen in the example, expectation alone may not be enough often times; we may want to study how much the random variable tends to deviate from its mean. Let $\mathsf{E}(X) = \mu$.

a) Deviation

$$\mathsf{E}((X - \mu)) = \mathsf{E}(X) - \mu = 0$$

b) Absolute deviation

$$\mathsf{E}\left(|X - \mu|\right)$$

c) Squared deviation

$$\mathsf{E}\left((X - \mu)^2\right)$$

The squared deviation turns out to be a particularly useful measure of variability.

Definition

The **variance** of a random variable X is denoted $Var(X)$, and is defined by

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

- A simple calculation gives the "short cut formula" to compute $Var(X)$:

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

- Another formula useful when the probability function involves $x!$ is

$$Var(X) = E(X(X-1)) + E(X) - (E(X))^2$$

- Since the random variable $(X - E(X))^2 \geq 0$, we have that $Var(X) \geq 0$.

Question

Let X denote the outcome of a fair six sided die roll. Compute $Var(X)$.

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* $P(X = E(X)) = 1$.

Question

Suppose that X has variance $Var(X) = 2$. Compute the variance of Y, where $Y = -2X + 3$.

A $Var(Y) = 2$

B $Var(Y) = -8$

C $Var(Y) = 8$

D $Var(Y) = 5$

E $Var(Y) = 6$

Standard deviation

Note that $Var(X)$ is in the squared unit (e.g., X in *meters* $\Rightarrow Var(X)$ is in *meters*^{2}. To recover the original unit, we take the square root of variance.

Definition

The **standard deviation** of a random variable X is denoted $SD(X)$, and defined by

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

Theorem (Variance of Binomial)

Suppose that $X \sim Binomial(n, p)$, then

$$Var(X) = np(1-p).$$

Proof.

Use the formula $Var(X) = E(X(X-1)) + E(X) - (E(X))^2$ and proceed similarly as in the derivation of $E(X)$, i.e., tweek the sum to apply the binomial formula. $\square$

Example

Suppose a fair coin is flipped 10,000 times, and let X denote the number of heads observed. Calculate the mean, variance, and standard deviation of X.

Theorem (Variance of Poisson)

Suppose that $X \sim Poi(\lambda)$, then

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

Proof.

Use the formula $Var(X) = E(X(X-1)) + E(X) - (E(X))^2$ and proceed similarly as in the derivation of $E(X)$, i.e., tweek the sum to apply the exponential series. $\square$

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

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

Variance of Hypergeometric and Negative Binomial

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 $Z \sim NB(k, p)$, then

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

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.