Discrete Distributions: Expected Value, Variance, Standard Deviation

BIOS 6611

CU Anschutz

Week 2

Random Variables

Random Variable

Random variables and probability distributions are the theoretical or mathematical representations of data values and frequency distributions.

- **Random Variable**: variable whose values depend on a random phenomenon. Possible values represent possible outcomes of a yet-to-be-performed experiment.
- We call these possible values the **sample space**
- Notation:
  - X, Y, Z denotes a random variable
  - x, y, z denotes a value of the random variable

Probability Distribution

The **probability distribution** describes the probabilities for each outcome for

- a discrete random variable (probability mass function, *pmf*)
- the probabilities of values in a range for a continuous random variable (probability density function, *pdf*)

The pmf/pdf is a useful tool for describing the frequency distributions of random variables for the entire population of interest and for providing probability statements about events involving random variables.

Discrete Random Variable

- **Discrete random variable**: sample space is a discrete list of values
  - Ex: Let X denote random experiment of flipping a coin.The possible outcomes are $\{Heads, Tails\}$.
- Characterized by a **probability mass function (PMF)**: function that gives probability that X is equal to values in sample space.
  - Values must be non-negative and sum to 1
  - Notation:

$$\begin{aligned} P(X = k) &= \text{function that depends on k,} \\ k &= \{\text{sample space}\} \end{aligned}$$

PMFs and CDFs

PMF

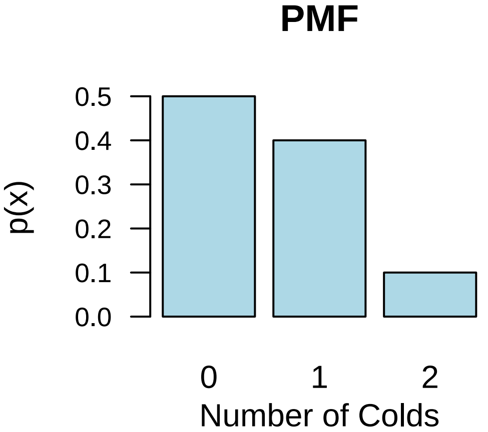
Let X be a random variable and let x represent the values that X can take on. The probability distribution of X is $p(x) = P(X = x)$.

For example, if $X =$ the number of colds in a year caught by a healthy adult:

$$\begin{aligned} P(X=0) &= 0.5 \\ P(X=1) &= 0.4 \\ P(X=2) &= 0.1 \end{aligned}$$

$$x = \{0, 1, 2\}$$

Note: All probabilities are nonnegative and sum to 1.



Cumulative Distribution Function (CDF)

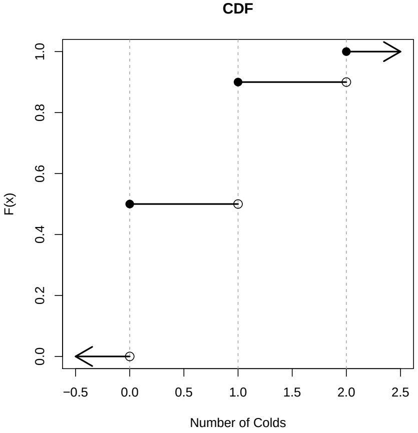
The **cumulative probability distribution** of X:

$F_X(x) = P(X \leq x) = \sum_{x=0}^{k} P(X = x)$ (accumulate probabilities from lowest to highest)

The CDF is monotone increasing, $F(-\infty) = 0,\ F(\infty) = 1.$

$$\begin{aligned} P(X = 0) &= 0.5 \\ P(X = 1) &= 0.4 \\ P(X = 2) &= 0.1 \end{aligned}$$

F(0)$=$ 0.5 ($\#$ of colds ≤ 0)
F(1) $= 0.5 + 0.4$ ($\#$ of colds ≤ 1)
F(2) $= 0.5 + 0.4 + 0.1 = 1.0$

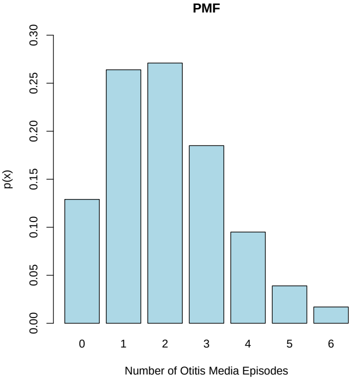


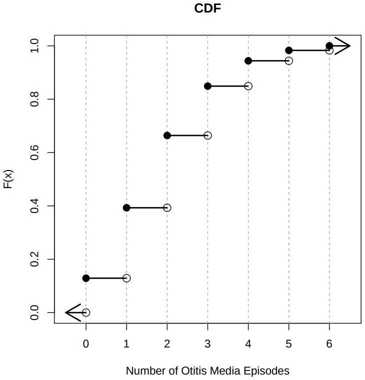
Another PMF/CDF Example

X = number of episodes of ear infections in first 2 years of life:

x	0	1	2	3	4	5	6
$P(X = x)$	0.129	0.264	0.271	0.185	0.05	0.039	0.017

PMF: $P(X = x)$ CDF: $P(X \leq x)$





Expected Value (Mean), Variance, Standard Deviation (SD)

Expected Value

There are different summary values for random variables. The **expected value** (i.e., mean or average) is one of the most common.

For a discrete r.v. X:

$$
\begin{aligned}
E(X) &= \sum_x xP(X = x) = \mu \\
&= \sum_x xp(x) \qquad\quad = \mu
\end{aligned}
$$

Expected Value Example

For our number of colds example:

x	0	1	2
$P(X=x)$	0.5	0.4	0.1

$$\mu = E(X) = \sum_x xP(X=x) =$$

Variance

In addition to summarizing the center (i.e., mean) of the distribution, the variability is also helpful to quantify. The **variance** can be described as the expected value of our r.v. X minus the true mean, μ, squared:

$$Var(X) = V(X) = \sigma^2 = E\left[(X - E(X))^2\right] = E\left[(X - \mu)^2\right]$$

Fortunately, like $E(X)$, we have specific formulas we can use for a discrete r.v. X:

$$\begin{aligned} Var(X) &= \sum_x (x-\mu)^2 P(X = x) = \sigma^2 \\ &= \sum_x (x-\mu)^2 p(x) \qquad = \sigma^2 \end{aligned}$$

The units for the variance are the square of the units of the variable itself (e.g., ft^2 or oz^2).

Standard Deviation

Another common way to summarize the variability is with the **standard deviation**, which is simply the square root of the variance:

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

One advantage of the variance is has the same units as the variable itself instead of being in squared units.

Variance and Standard Deviation Example

For our number of colds example:

x	0	1	2
$P(X = x)$	0.5	0.4	0.1

$$\sigma^2 = V(X) = \sum_x (x - \mu)^2 p(x) =$$

Computational Formula for Variance

In addition to the definition of the variance previously, we can also calculate it by knowing the expected value of our random variable X and X^2:

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

Formula Review for Discrete Random Variables

$$E(X) = \mu = \sum_x xP(X = x) \\ = \sum_x xp(x)$$

$$Var(X) = \sigma^2 = \sum_x (x - \mu)^2 P(X = x) \\ = \sum_x (x - \mu)^2 p(x) \\ = E[X^2] - (E[X])^2$$

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