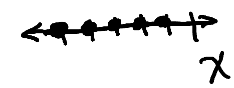
02/06/2023

Chapter 5 – Discrete Probability Distributions

Def: A random variable (rv) is a characteristic whose value may change from one subject to another in a population.

Def: A rv is discrete if it takes on a finite or countably infinite number of possible values. A rv is continuous if its possible values span an interval of real numbers.



Remark: Any rv, denoted by X, is a real-valued function having the sample space as its domain.

Def: The probability distribution of a discrete rv X can be represented by a probability function of X, denoted by $P(x)$.

$P(x)$ = probability the discrete rv takes on the value x.

Ex: Toss a fair coin 3 times

X = # of heads in 3 tosses of coin

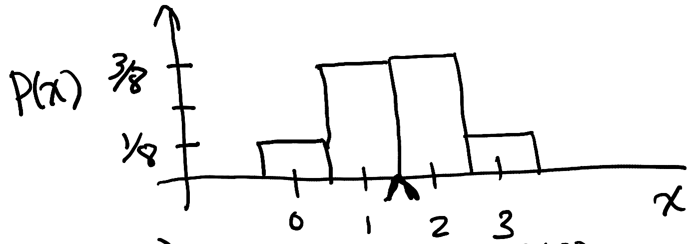
Sample space
{ HHH, HTH, HTT, TTT, THH, THT, HHT, TTH }

coin

HHT, TTH}

x	$P(x)$
0	$1/8$
1	$3/8$
2	$3/8$
3	$1/8$

Probability Histogram



$$P(\text{at least one head in 3 tosses}) = P(1) + P(2) + P(3) = \frac{7}{8} = P(X \geq 1)$$

$$P(X > 1) = \frac{4}{8} = 0.5$$

Properties of a Probability Distribution for a Discrete RV

1. $0 \leq P(x) \leq 1$

2. $\sum_{\text{all } x} P(x) = 1$

Ex: (continued)

X = # of heads in three tosses of fair coin

$$P(X \geq 1) = P(1) + P(2) + P(3) = \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = \frac{7}{8}$$

$$P(1) = P(X=1) = P(HTT) + P(HTH) + P(TTH) = \frac{3}{8}$$

$$P(X > 1) = P(X \geq 2) = P(2) + P(3) = \frac{4}{8} = \frac{1}{2} = 0.5$$

Mean, Variance, and Standard Deviation for a Discrete RV

Def: Let X be a discrete rv with probability $P(x)$.

The ... mean of X or expected value

Def: Let X be a discrete rv with probability $P(x)$.

Then the <u>mean of X</u>, or expected value of X is

summation occurs last

$$\text{Mean of } X = \mu = \sum_{\text{all } x} x \cdot P(x)$$

calculate first

Def: Let X be a discrete rv with probability $P(x)$.

The <u>variance of X</u> $= \sigma^2 = \sum_{\text{all } x} (x-\mu)^2 \cdot P(x)$

<u>standard deviation of X</u> is

$$\sigma = \sqrt{\sigma^2}$$

Ex: X = # of heads in 3 tosses of a fair coin.

Find μ, σ^2, and σ.

$$\begin{aligned}
\mu &= \sum_{x=0}^{3} x \cdot P(x) \\
&= 0 \cdot P(0) + 1 \cdot P(1) + 2 \cdot P(2) + 3 \cdot P(3) \\
&= 0 \cdot \tfrac{1}{8} + 1 \cdot \tfrac{3}{8} + 2 \cdot \tfrac{3}{8} + 3 \cdot \tfrac{1}{8} \\
&= \tfrac{12}{8} \\
&= 1.5
\end{aligned}$$

$$\begin{aligned}
\sigma^2 &= \sum_{x=0}^{3} (x-\mu)^2 \cdot P(x) \\
&= (0-1.5)^2 \tfrac{1}{8} + (1-1.5)^2 \tfrac{3}{8} + (2-1.5)^2 \tfrac{3}{8} + (3-1.5)^2 \tfrac{1}{8} \\
&= \tfrac{3}{4} \\
&= 0.75
\end{aligned}$$

and $\sigma = \sqrt{\sigma^2} = \sqrt{0.75} \approx 0.87$

Remarks:

1. μ = population mean of rv X
2. σ^2 = population variance of rv X

X = # heads in 4 tosses of fair coin

$$\mu = \sum_{x=0}^{4} x \cdot P(x) = 2$$

$$\sigma^2 = \sum_{x=0}^{4} (x-2)^2 \cdot P(x) = ?$$

$$\sigma = ?$$

The Binomial Distribution

02/13/2023

Def: A discrete variable X has a binomial distribution if X is the number of "successes" out of n items or trials of an experiment, where p = probability of success for item or trial. Assuming that p is the same for each independent trial, the probability function of X is

$P(x) = ?$

$$P(x) = \frac{n!}{x!\,(n-x)!} p^x (1-p)^{n-x}, \quad \text{for } x = 0, 1, 2, \ldots, n.$$

Factorials

$0! = \quad = 1$

$1! = 1 = 1$

$2! = 2 \cdot 1 = 2$

$3! = 3 \cdot 2 \cdot 1 = 6$

$x! = x(x-1)(x-2) \cdots 1$

$$\frac{4!}{3!} = \frac{4 \cdot \cancel{3!}}{\cancel{3!}} = 4$$

$$4! = 24 = 4 \cdot 3 \cdot 2 \cdot 1 = 4 \cdot 3!$$

$$\frac{4!}{2!\,2!} = \frac{24}{4} = 6$$

"distributed as"

Notation: $X \sim \text{Binomial}(n, p)$

Ex: X = # of heads in 3 tosses of fair coin

"success"

$n = 3$

Ex: X = # of heads in 3 tosses of

↖ "success"

$n = 3$

$p = 0.5$

$P(TTT) = \frac{1}{8}$

$x=0, \quad P(0) = \frac{3!}{0!\,(3-0)!}(0.5)^0(1-0.5)^{3-0} = 1 \cdot 1 \cdot (0.5)^3$

$= 0.125$

$(0.5)^1(0.5)^2$

HTT
THT
TTH

$x=1, \quad P(1) = \frac{3!}{1!\,(3-1)!}(0.5)^1(1-0.5)^{3-1} = 3(0.5)^3$

$= 0.375$

$x=2, \quad P(2) = \frac{3!}{2!\,(3-2)!}(0.5)^2(1-0.5)^1 = 3(0.5)^3$

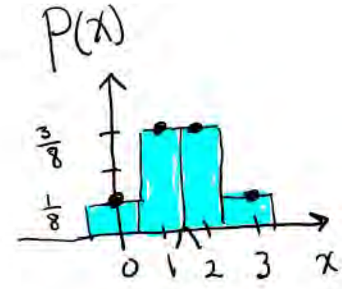
HHT
THH
HTH

$= 0.375$

$x=3, \quad P(3) = \frac{3!}{3!\,(3-3)!}(0.5)^3(1-0.5)^{3-3} = 1(0.5)^3$

$= 0.125$

NOTE: Probability histogram is symmetric in this case because $p = 0.5 = 1-p$.



Remark: x = # of "successes" out of n trials

$X \sim \text{Binomial}(n, p)$

n: # of trials

p: probability of "success" on a single trial

$\mu = np$ = population mean

$\sigma^2 = np(1-p) = npq$ = population variance

$\sigma = \sqrt{\sigma^2}$ = population standard deviation

$\sigma^2 = np(1-p) = npq$ = population variance

$\sigma = \sqrt{\sigma^2}$ = population standard deviation

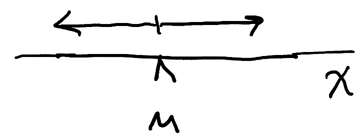
Ex: x = # of heads on 3 tosses of fair coin

Then $x \sim \text{Binomial}(3, 0.5)$

$\mu = np = 3(\tfrac{1}{2}) = 1.5$

$\sigma^2 = np(1-p) = 3(\tfrac{1}{2})(\tfrac{1}{2}) = 0.75$ so $\sigma = ?$

Remark: μ = population mean



$\mu - \sigma$ = one standard deviation to the left of μ

$\mu + \sigma$ = " " " " " right " "

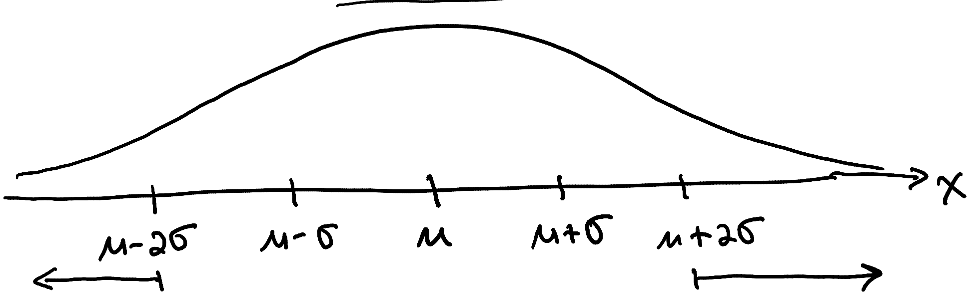
$\mu - 2\sigma$ = two standard deviations " " left of μ

$\mu + 2\sigma$ = " " " " " right of μ.

Identifying Significant Values

1. Range Rule of Thumb: Individual values for which $x \le \mu - 2\sigma$ or $x \ge \mu + 2\sigma$ are considered significantly low or high values of x, respectively.

Ex:



NOTE: For bell-shaped distributions,

$P(x \le \mu - 2\sigma) \approx 0.0228$ ✓

$P(x \ge \mu + 2\sigma) \approx 0.0228$ ✓

2. Using Probabilities

Consider calling x_{low} a <u>significantly low value</u> if $P(X \le x_{low}) \le 0.05$ and call x_{high} a <u>significantly high value</u> if $P(X \ge x_{high}) \le 0.05$.