Discrete Probability Distribution

Recap on Basic Probability

Example 1:

- Take the example of 5 coin tosses. What's the probability that you flip exactly 3 heads in 5 coin tosses?

Example 1 Solution:

- One way to get exactly 3 heads: HHHTT. What's the probability of this exact arrangement?

  $P(heads).P(heads).P(heads).P(tails).P(tails) = P(heads)^3 P(tails)^2 = 0.5^3 0.5^2$
  $= 0.03125$

- Another way to get exactly 3 heads: THHHT. Probability of this exact outcome :

  $P(tails).P(heads).P(heads).P(heads).P(tails) = P(heads)^3 P(tails)^2 = 0.5^3 0.5^2$
  $= 0.03125$

Example 1 Solution:

In fact, $(0.5)^3.(0.5)^2$ is the probability of each unique outcome that has exactly 3 heads and 2 tails.

So, the overall probability of 3 heads and 2 tails is:

$(0.5)^3 \; x \; (0.5)^2 + (0.5)^3 \; x \; (0.5)^2 + (0.5)^3 \; x \; (0.5)^2 + \ldots..$ for as many unique arrangements as there are—but how many are there??

Example 1 Solution:

$\binom{5}{3}$ ways to arrange 3 heads in 5 trials

$_5C_3 = 5!/3!2! = 10$

Outcome	Probability
THHHT	$(1/2)^3\ x\ (1/2)^2$
HHHTT	$(1/2)^3\ x\ (1/2)^2$
TTHHH	$(1/2)^3\ x\ (1/2)^2$
HTTHH	$(1/2)^3\ x\ (1/2)^2$
HHTTH	$(1/2)^3\ x\ (1/2)^2$
THTHH	$(1/2)^3\ x\ (1/2)^2$
HTHTH	$(1/2)^3\ x\ (1/2)^2$
HHTHT	$(1/2)^3\ x\ (1/2)^2$
THHTH	$(1/2)^3\ x\ (1/2)^2$
HTHHT	$(1/2)^3\ x\ (1/2)^2$

10 arrangements $x\ (1/2)^3\ x\ (1/2)^2$

The probability of each unique outcome (note: they are all equal)

Example 1 Solution:

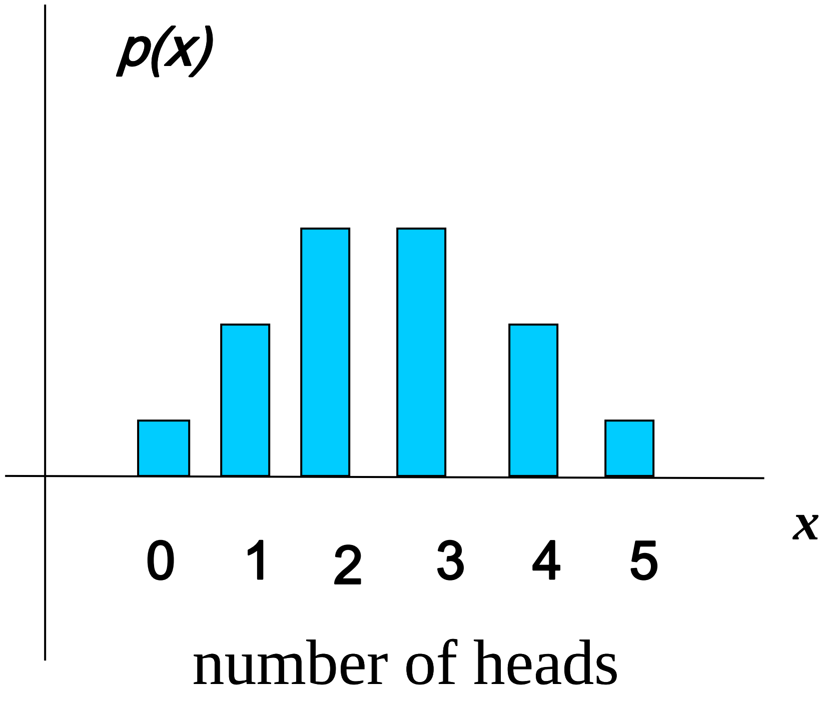
Therefore the probability that you flip exactly 3 heads in 5 coin tosses is:

$$P(3\ heads\ and\ 2\ tails) = {}_5C_3\, P(heads)^3\, P(tails)^2$$

$$= 10\ x\ (0.5)^5 = 31.25\%$$

Example 1 Binomial distribution function:

X= the number of heads tossed in 5 coin tosses



Example 2:

As voters exit the polls, you ask a representative random sample of 6 voters if they voted for proposition 100. If the true percentage of voters who vote for the proposition is 55.1%, what is the probability that, in your sample, exactly 2 voted for the proposition and 4 did not?

Example 2: Solution

Outcome	Probability
YYNNNN	$= (.551)^2 \, x \, (.449)^4$
NYYNNN $(.449)^1 \, x \, (.551)^2 \, x \, (.449)^3$	$= (.551)^2 \, x \, (.449)^4$
NNYYNN $(.449)^2 \, x \, (.551)^2 \, x \, (.449)^2$	$= (.551)^2 \, x \, (.449)^4$
NNNYYN $(.449)^3 \, x \, (.551)^2 \, x \, (.449)^1$	$= (.551)^2 \, x \, (.449)^4$
NNNNYY $(.449)^4 \, x \, (.551)^2$	$= (.551)^2 \, x \, (.449)^4$
.	
.	

$\binom{6}{2}$ ways to arrange 2 propotition votes among 6 voters

$$15\, arrangements\, .\, (0.551)^2 .\, (0.449)^4$$

$$\therefore \text{P(2 yes votes exactly)} = \binom{6}{2} (0.551)^2 \; (0.449)^4 = 18.5\%$$

Binomial Distribution

Binomial Probability Distribution

- A fixed number of observations (trials), n
  - e.g., 15 tosses of a coin; 20 patients; 1000 people surveyed
- A binary random variable
  - e.g., head or tail in each toss of a coin; defective or not defective light bulb
  - Generally called “success” and “failure”
  - Probability of success is p, probability of failure is $1 - p$
- Constant probability for each observation
  - e.g., Probability of getting a tail is the same each time we toss the coin

Binomial Distribution

- Note the general pattern emerging □ if you have only two possible outcomes (call them 1/0 or yes/no or success/failure) in n independent trials, then the probability of exactly X "successes"=

n = number of trials

$$\boxed{\binom{n}{X} p^{X}(1-p)^{n-X}}$$

X = # successes out of *n* trials

p = probability of success

1-p = probability of failure

Definitions: Binomial Experiment

- **Binomial:** Suppose that *n* independent experiments, or trials, are performed, where *n* is a fixed number, and that each experiment results in a "success" with probability *p* and a "failure" with probability *1-p*. The total number of successes, *X*, is a binomial random variable with parameters *n* and *p*.
- We write: ***X ~ Bin (n, p)*** *{reads: "X is distributed binomially with parameters n and p}*
- And the probability that *X=r* (i.e., that there are exactly *r* successes) is:

$$\boxed{P(X=r)=\binom{n}{r}p^r(1-p)^{n-r}}$$

Definitions: Bernoulli Experiment

Bernouilli trial: If there is only 1 trial with probability of success *p* and probability of failure *1-p*, this is called a Bernouilli distribution. (special case of the binomial with *n*=1)

Probability of success:

$$P(X=1)=\binom{1}{1}p^1(1-p)^{1-1}=p$$

Probability of failure:

$$P(X=0)=\binom{1}{0}p^0(1-p)^{1-0}=1-p$$

Example 3: Binomial Distribution

- If I toss a coin 20 times, what's the probability of getting exactly 10 heads?

Solution:

$$P(X = r) = \binom{n}{r} p^r (1-p)^{n-r}$$

$$N = 20, r = 10$$

$$\binom{20}{10} 0.5^{10} (1 - 0.5)^{20-10} = 0.176$$

Example 4: Binomial DIstribution

- If I toss a coin 20 times, what's the probability of getting of getting 2 or fewer heads?

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$$

$$P(X \leq 2) = \binom{20}{0} 0.5^0 (1 - 0.5)^{20-0} + \binom{20}{1} 0.5^1 (1 - 0.5)^{20-1} + \binom{20}{2} 0.5^2 (1 - 0.5)^{20-2}$$

$$P(X \leq 2) \approx 1.8 \times 10^{-4}$$

Characteristic of Binomial Distribution

- If X follows a binomial distribution with parameters n and p: $X \sim Bin\ (n, p)$
- Then:

$$\mu_x = E(X) = np$$

$$\sigma_x^2 = Var\ (X) = np(1-p)$$

$$\sigma_x = SD\ (X) = \sqrt{np(1-p)}$$

Note: the variance will always lie between

0*N-.25 *N

p(1-p) reaches maximum at p=.5

P(1-p)=.25

Characteristic of Bernoulli Distribution

- If X follows a bernoulli distribution with parameters $n = 1$ and p: $X \sim Bernoulli\ (p)$
- Then:

$$\mu_x = E(X) = p$$

$$\sigma_x{}^2 = Var\ (X) = p(1-p)$$

$$\sigma_x = SD\ (X) = \sqrt{p(1-p)}$$

Example 5: Problem

1. You are performing a cohort study. If the probability of developing disease in the exposed group is .05 for the study duration, then if you sample (randomly) 500 exposed people, how many do you expect to develop the disease? Give a margin of error (+/- 1 standard deviation) for your estimate.
2. What's the probability that **at most** 10 exposed people develop the disease?

Example 5: Solution

1.

$$X \sim binomial\ (500, 0.05)$$
$$E(X)\ =\ 500\ (0.05)\ =\ 25$$
$$Var(X)\ =\ 500\ (0.05)\ (0.95)\ =\ 23.75$$
$$StdDev(X)\ =\ \sqrt{23.75} =\ 4.87$$

Therefore the expectation of people to develop the disesase

$$E(X) \pm 1\ .\ StdDev(X)$$
$$25 \pm 4.87\ people$$

Example 5: Solution

2. What's the probability that at most 10 exposed subjects develop the disease?

This is asking for a CUMULATIVE PROBABILITY: the probability of 0 getting the disease or 1 or 2 or 3 or 4 or up to 10.

$P(X\leq 10) = P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4)+\ldots.+ P(X=10)=$

$$\binom{500}{0}(.05)^0(.95)^{500}+\binom{500}{1}(.05)^1(.95)^{499}+\binom{500}{2}(.05)^2(.95)^{498}+\ldots+\binom{500}{10}(.05)^{10}(.95)^{490}<.01$$

Example 6:

- If Stanford tickets in the medical center ‘A’ lot approximately twice a week (2/5 weekdays), if you want to park in the ‘A’ lot twice a week for the year, are you financially better off buying a parking sticker (which costs $726 for the year) or parking illegally (tickets are $35 each)?

Example 6: Solution

- Use Binomial→ Let X be a random variable that is the number of tickets you receive in a year.
- Assuming 2 weeks vacation, there are 50x2 days (twice a week for 50 weeks) you'll be parking illegally. p=.40 is the chance of receiving a ticket on a given day:

X~binomial (100, .40)

E(X) = 100x.40 = 40 tickets expected (with std dev of about 5)

40 x $35 = $1400 in tickets (+/- $200);

better to buy the sticker!

Poisson Distribution

Poisson Distribution

- The poisson distribution is a discrete probability distribution that applies to occourence of some event over a specified interval. The random variable x is the number of occurrences of the event in an interval. The interval can be time, distance, area, volume, or some similar unit.

$$\boxed{p(X=k)=\frac{\lambda^k e^{-\lambda}}{k!}}$$

- Where μ = mean number of occurrences of the event over the interval

Requirement of the Poisson distribibition

- The random variable x is the number of occurrences of an event over some interval
- The occurrences must be random
- The occurrences must be independent of each other
- The occurrences must be uniformly distributed over the interval being used

Differences from a Binomial Distribution

- The poisson distribution differs from the binomial distribution in these fundamental ways:
- The binomial distribution is affected by the sample size n and the probability p, whereas the Poisson distribution is affected only by mean μ
- In a binomial distribution the possible values of the random variable x are $0, 1,2, \ldots, n$, but a Poisson distribution has x values of $0,1,2, \ldots .$, with no upper limit.

Example 7:

- For a recent period of 100 years, there were 530 Atlantic hurricanes. Assume the Poisson distribution is a suitable model.

a) Find μ, the mean number of hurricanes per year

b) If $P(X = k)$ is the probability of k hurricanes in a randomly selected year, find $P(X = 2)$

Example 7:Solution

a) Find μ, the mean number of hurricanes per year

$$\mu = \frac{number\ of\ hurricanes}{number\ of\ years} = \frac{530}{100} = 5.3$$

b) If $P(X = k)$ is the probability of k hurricanes in a randomly selected year, find $P(X = 2)$

$$P(X = 2) = \frac{\mu^k e^{-k}}{k!} = \frac{5.3^2(2.71828)^{-5.3}}{2!} = 0.0701$$

Example 8:

- Suppose that a rare disease has an incidence of 1 in 1000 person-years. Assuming that members of the population are affected independently, find the probability of k cases in a population of 10,000 (followed over 1 year) for k=0,1,2.
- The expected value (mean) =λ = 0.001*10,000 = 10
- 10 new cases expected in this population per year→

$$P(X=0)=\frac{(10)^0 e^{-(10)}}{0!}=.0000454$$

$$P(X=1)=\frac{(10)^1 e^{-(10)}}{1!}=.000454$$

$$P(X=2)=\frac{(10)^2 e^{-(10)}}{2!}=.00227$$

More on Poisson

"Poisson Process" (rates)

Note that the Poisson parameter λ can be given as the mean number of events that occur in a defined time period OR, equivalently, λ can be given as a rate, such as λ=2/month (2 events per 1 month) that must be multiplied by *t*=time (called a "Poisson Process") →

$$X \sim Poisson\ (\lambda)\ \text{->}$$

$$P(X=k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$$

$$E(X) = \lambda t$$

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

Poisson as an Approximation to the Binomial Distribution

- The Poisson Distribution is sometimes used to approximate the binomial distribution when $\boldsymbol{n}$ is large and $\boldsymbol{p}$ is small

- Rule of Thumb to use the Poisson to approximate the Binomial, where:

$$\boldsymbol{n \geq 100}$$
$$\boldsymbol{np \leq 10}$$

End of This Lecture