

32

Special Discrete Dis

Bernoulli DistributionTwo events  $E$  and  $E'$   
often denoted as  $S$  &  $F$   
Success Failure $X \sim \text{BERNOULLI}(p)$ Let  $e$  be a success or a failure  $E'$ 

$$X(e) = \begin{cases} 1, & \text{if } e \in E \\ 0, & \text{if } e \in E' \end{cases}$$

Pdf

$$f(x) = p^x (1-p)^{1-x}, \quad x \in \{0, 1\}$$

 $p$  prob of successEX: Flip a coin  $\Rightarrow E = \{H\}$   $E' = \{T\}$ ,  $p = \frac{1}{2}$ Roll a die  $\Rightarrow E = \{2\}$   $E' = \{1, 3, 4, 5, 6\}$   
 $p = \frac{1}{6}$ 

$$E(X) = \sum_{x=0}^1 x f(x) = 0f(0) + 1f(1) = p$$

$$V(X) \Rightarrow \text{Need } E(X^2) \Rightarrow E(X^2) = 0^2 f(0) + 1^2 f(1) = p$$

$$V(X) = E(X^2) - \mu^2 = p - p^2 = p \underbrace{(1-p)}_q$$

$$\boxed{\text{Var}(X) = pq}$$

## Binomial Distribution

A sequence of independent Bernoulli Trials

The binomial dist is typically used for sampling with replacement

EX Multiple Choice Test

5 questions, each with a), b), c), d)

Find the prob of getting 3 out of 5 correct

$$P(RRRWW) = \frac{1}{4} \frac{1}{4} \frac{1}{4} \frac{3}{4} \frac{3}{4} = \left(\frac{1}{4}\right)^3 \left(\frac{3}{4}\right)^2$$

Is this the only way to get 3 right?

$$P(PWWRR) = \frac{1}{4} \frac{3}{4} \frac{3}{4} \frac{1}{4} \frac{1}{4} = \left(\frac{1}{4}\right)^3 \left(\frac{3}{4}\right)^2$$

Distinguishable Permutations of RRRWW

$$\frac{5!}{3!2!} = \binom{5}{3} = \frac{5!}{3!2!} = \frac{5 \cdot 4}{2} = 10$$

Let  $X = \#$  of questions correct

$$\begin{aligned} P[X=3] &= P(RRRWW) + P(RWWRR) \\ &\quad + \dots + P(\text{last one}) \\ &= \left(\frac{1}{4}\right)^3 \left(\frac{3}{4}\right)^2 + \left(\frac{1}{4}\right)^3 \left(\frac{3}{4}\right)^2 + \dots \\ &= 10 \left(\frac{1}{4}\right)^3 \left(\frac{3}{4}\right)^2 \end{aligned}$$

In general,  $p = \text{prob of success}$   
 $q = 1 - p = \text{prob of failure}$   
 $n$  independent Bernoulli, tr. etc.

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

The pdf of the binomial is

$$f(x) = b(x, n, p) = \binom{n}{x} p^x q^{n-x}, \quad x=0, 1, \dots, n$$

$$M_X(t) = (pe^t + q)^n$$

Note, the sum of the probabilities should be one.

$$\sum_{x=0}^n b(x, n, p) = \sum_{x=0}^n \binom{n}{x} p^x q^{n-x} = 1$$

CDF (for integer values)

$$B(x, n, p) = \sum_{k=0}^x b(k, n, p)$$

$x \geq 0$

## Hypergeometric Distribution

Suppose a population consists of a finite number of items  $N$ , and  $M$  are of Type 1, and

$N-M$  are of type 2

Suppose that  $n$  items are drawn without replacement. Denote

$X = \#$  of Type 1 items drawn

The pdf of  $X$  is

$$h(x, n, M, N) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}$$

, where  $\underbrace{\max(0, n-(N-M))}_c \leq x \leq \underbrace{\min(n, M)}_d$

The CDF is  $F(x) = \begin{cases} \sum_{i=c}^x h(i, n, M, N) & 0 < x < c \\ 1 & x \geq d \end{cases}$

$$X \sim \text{HYP}(n, M, N)$$

$$E(X) = n \frac{M}{N} \quad \text{Var}(X) = n \left( \frac{M}{N} \right) \left( 1 - \frac{M}{N} \right) \left( \frac{N-n}{N-1} \right)$$

Note the similarities with the binomial dist.

Let  $p = \frac{M}{N}$ , then  $E(X) = n \frac{M}{N} = np$  and

$$V(X) = \underbrace{np(1-p)}_{\text{like the binomial}} \left( \frac{N-n}{N-1} \right)$$

Thm If  $X \sim \text{HYP}(n, M, N)$ , then  
 for each value of  $x$  between ctd  
 and as  $N \rightarrow \infty, M \rightarrow \infty, \frac{M}{N} \rightarrow p$ , a pos constant,

$$\lim_{\substack{N \rightarrow \infty \\ M \rightarrow \infty \\ \frac{M}{N} \rightarrow p}} \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} = \binom{n}{x} p^x (1-p)^{n-x}$$

$$\lim_{\substack{M \rightarrow \infty \\ N \rightarrow \infty}} h(x, n, M, N) = b(x, n, p)$$

proof extra credit

This provides an estimate to the hypergeometric when the number selected  $n$ , is small in relation to size of the population,  $N$ , and the number of Type 1 items,  $M$ .

In other words, sampling with replacement is equal to sampling without replacement when the population is infinite, but the proportion of type 1 items remains the same.

## Geometric Dist

Consider a sequence of Bernoulli trials with prob of success  $p$

$X = \#$  of trials required to get the first success

The pdf is

$$g(x, p) = p q^{x-1}, x=1, 2,$$

The dist is also known as the Pascal dist

The CDF is

$$G(x, p) = \sum_{l=1}^x p q^l = 1 - q^x$$

## Thm 3.2.2

No Memory Property.

If  $X \sim \text{GEO}(p)$ , then

$$P[X > j+k | X > j] = P[X > k]$$

proof: (In other words, it has no memory of where you start.)

$$P[X > j+k | X > j] = \frac{P[X > j+k \cap X > j]}{P[X > j]} = \frac{P[X > j+k]}{P[X > j]}$$
$$= \frac{1 - (1 - q^{j+k})}{1 - (1 - q^j)} = \frac{q^{j+k}}{q^j} = q^k$$

$$= 1 - (1 - q^k) = P[X > k]$$

$$E(X) = \frac{1}{p} \quad V(X) = \frac{q}{p^2} \quad M_X(t) = \frac{pe^t}{1 - qe^t}$$

## Negative Binomial (cont.)

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

$x$  is on the top and  $r \leq x < \infty$ .

$r + (x-r) = x$

$(x-1) - (r-1)$

Compare to Binomial:

$$P(X=x) = \binom{n}{x} p^x q^{n-x}$$

note the  $x$  is on the bottom and  $0 \leq x \leq n$ .

$x + n - x = n$

EX: Team A plays Team B in a seven game world series. The series is over when either team wins 4 games. Suppose  $P(\text{A wins}) = .6$ , and the games are independent. What is the probability the series lasts 6 games?

$p = .6, q = .4, x = 6, r = 4$   $x = \#$  of trials until the  $r^{\text{th}}$  success

$$P(\text{A wins in 6 games}) = \binom{5}{3} .6^4 .4^2 = .20736$$

$$p = .4, q = .6 \leftarrow P(\text{B wins in 6 games}) = \binom{5}{3} .4^4 .6^2 = .09216$$

$$P(\text{series lasts 6 games}) = .20736 + .09216 = .29952$$

To show the sum of the probabilities in the NB dist is 1, derive the Maclaurin series for

$$(1-q)^{-r} = \sum_{i=0}^{\infty} \binom{i+r-1}{r-1} q^i$$

Answer  $\Rightarrow$  derive it

Some Authors call  $Y = \#$  of failures until the  $r^{\text{th}}$  success the NB dist  $Y = X - r$  and

$$f_Y(y) = \binom{y+r-1}{r-1} p^r q^y, \quad y=0,1,2,\dots$$

The Mean, var, and mgf for NB are

$$E(X) = \frac{r}{p} \quad \text{Var}(X) = \frac{rq}{p^2} \quad M_X(t) = \left[ \frac{pe^{qt}}{1-qt} \right]^r$$

Binomial relationship to NB

The NB dist. is sometimes referred to as inverse binomial sampling

Suppose  $X \sim \text{NB}(r, p)$      $W \sim \text{BIN}(n, p)$   
 It follows that

$$P[X \leq n] = P[W \geq r]$$

$W \geq r =$  having  $r$  or more successes in  $n$  trials, and that means that  $n$  or fewer trials will be needed to obtain the first  $r$  successes.

In terms of CDF's  $F(X, r, p) = 1 - B(r-1, X, p) = B(X-r+1, X, p)$



## Poisson Dist

$$X \sim \text{POI}(\mu)$$

A discrete R.V.  $X$  is said to have the Poisson dist with parameter  $\mu > 0$  if it has the pdf

$$f(x; \mu) = \frac{e^{-\mu} \mu^x}{x!} \quad x=0, 1, 2, \dots$$

The CDF is

$$F(x, \mu) = \sum_{k=0}^x f(k; \mu)$$

The mean and variance

$$E(X) = \mu \quad V(X) = \mu$$

$$M_X(t) = e^{\mu(e^t - 1)}, \quad -\infty < t < \infty$$

Thm: Relationship to Binomial:

If  $X \sim \text{BIN}(n, p)$ , then for each  $x=0, 1, 2, \dots$

and as  $n \rightarrow \infty, p \rightarrow 0, np \rightarrow \mu$

$$\lim_{\substack{n \rightarrow \infty \\ p \rightarrow 0 \\ np \rightarrow \mu}} \binom{n}{x} p^x q^{n-x} = \frac{e^{-\mu} \mu^x}{x!}$$

Note that if  $n$  is large and  $p$  is small, then the Poisson pdf provides an approx to  $b(x, n, p)$

(As a rule of thumb, ~~this~~ gives reasonable results if  $n \geq 100, p \leq 0.1$  and when  $x$  is close to  $np$ )

# Poisson Process

Thm 3.2.4

## Homogeneous Poisson Process

Let  $X(t)$  denote the number of occurrences in the interval  $[0, t]$ , and  $P_n(t) = P\{n \text{ occurrences in an interval } [0, t]\}$

Consider the following properties

- 1)  $X(0) = 0$  ( $t=0$  with no arrivals)
- 2)  $P\{X(t+h) - X(t) = n \mid X(s) = m\} = P\{X(t+h) - X(t) = n\}$

Arrivals at disjoint time periods are independent and only depend on length for all  $0 < s < t$ , and  $h > 0$

$$P\{X(t+\Delta t) - X(t) = 1\} = \lambda \Delta t + o(\Delta t)$$

$$P\{X(t+\Delta t) - X(t) = 0\} = 1 - \lambda \Delta t + o(\Delta t)$$

If prop 1-4 holds, then for all  $t > 0$ ,

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

(Note  $\lim_{\Delta t \rightarrow 0} \frac{o(\Delta t)}{\Delta t} = 0$ )

Thus,  $X(t) \sim \text{POI}(\lambda t)$ ,

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

$\lambda$  = rate of occurrence or the intensity

Because  $\lambda$  is constant, this is why it's called homogeneous

## Discrete Uniform Dist

$$X \sim \text{DU}(N)$$

The pdf is  $f(x) = \frac{1}{N}$ ,  $x=1, 2, \dots, N$

Ex: Rolling a die

$$f(x) = \frac{1}{6}$$

$$f(1) = \frac{1}{6} = f(2) = f(3) = f(4) = f(5) = f(6)$$

$$E(X) = \frac{N+1}{2} \quad V(X) = \frac{N^2-1}{12}$$