[3.2] Special Discrete Distributions

Bernoulli Distribution

Two events E or E'

often denoted as S & F (Success, Failure)

$X \sim \text{BERNOULLI}(p)$

Let e be a success (E) or a failure (E')

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

Pdf

$$f(x) = p^x(1-p)^{1-x}; x = 0, 1$$

p is the probability of success

Ex: Flip a coin $\Rightarrow$ $E = \{H\}$ $E' = \{T\}$, $p = \frac{1}{2}$

Roll a die $\Rightarrow$ $E = \{2\}$ $E' = \{1,3,4,5,6\}$ $p = \frac{1}{6}$

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

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

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

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

$X = \#$ of successes in n trials

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?

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

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

$$P[X=3] = P[RRRWW] + P[RWWRR] + \cdots + P[\text{last one}]$$

$$= \left(\frac{1}{4}\right)^3\left(\frac{3}{4}\right)^2 + \left(\frac{1}{4}\right)^3\left(\frac{3}{4}\right)^2 + \cdots$$

$$= 10\left(\frac{1}{4}\right)^3\left(\frac{3}{4}\right)^2$$

In general, p = prob. of success

q = 1-p = prob. of failure

n independent Bernoulli trials

$X \sim BIN(n, p)$

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

The pdf of the binomial is

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

Note, the sum of the probabilities should be one.

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

CDF

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

where $\lfloor x \rfloor$ is the greatest integer less than or equal to x

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

$$h(x, n, M, N) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, \text{ where } \underbrace{\max(0, n-(M-N))}_{c} \leqslant x \leqslant \underbrace{\min(n, M)}_{d}$$

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

$X \sim HYP(n, M, N)$

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

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 $X \sim HYP(n, M, N)$, then for each value of x between c&d and as $N \to \infty$, $M \to \infty$, $\frac{M}{N} \to p$, a pos constant,

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

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

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 Distribution

Consider a sequence of Bernoulli trials with prob. of success p on each trial

X = # of trials required to get the first success

The pdf is

$$g(x, p) = pq^{x-1}, \; x = 1, 2, \ldots$$

The dis is also known as the Pascal dist

The CDF is

$$G(x, p) = \sum_{k=1}^{\lfloor x \rfloor} pq^{x-1} = 1 - q^{\lfloor x \rfloor}$$

thm 3.2.2

No Memory Property.

If $X \sim GEO(p)$, then

$$P[X > j+k \mid X > j] = P[X > k]$$

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

proof:

$$P[X > j+k \mid X > j] = \frac{P[X > j+k \cap X > j]}{P[X > j]} = \frac{P[X > j+k]}{P[X > j]}$$

$$= \frac{1-(1-q^{j+k})}{1-(1-q^{j})} = \frac{q^{j+k}}{q^{j}} = q^{k}$$

$$= 1-(1-q^{k}) = P(X > k)$$

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

Negative Binomial

$X = \#$ of trials until the r^{th} success

$\boxed{r+(x-r)=x}$

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

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

x is on the top and $r \le x \le \infty$.

Compare to Binomial:

$x+n-x=n.$

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

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

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

X = # of trials until the r^{th} success

$p=.6$, $q=.4$, $x=6$, $r=4$

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

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

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

Note that this is not a negative binomial distribution. However, counting the number of games until the 4th success for Team A or Team B is. Combining these together creates a distribution from 4 to 7 games. Try adding up the probabilities! You will find it sums to 1!

The Mean, Variance, and mgf for NB are

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

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

$$(1-q)^{-r} \overset{\text{Answer}}{=} \underbrace{\sum_{i=0}^{\infty} \binom{i+r-1}{r-1} q^i}_{\text{Answer} \Rightarrow \text{derive it}}$$

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

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

Binomial relationship to NB

The NB dist. is sometimes referred to as <u>inverse binomial sampling</u>

Suppose $X \sim NB(r, p)$ $\quad W \sim BIN(n, p)$

It follows that

$$P[X \le n] = P[W \ge r]$$

$W \ge r \Rightarrow$ 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 $\quad F(x; r, p) = 1 - B(r-1; x, p) = B(x-r; x, q)$

Poisson Dist

$X \sim POI(\mu)$

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

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

The CDF is

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

The mean and variance

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

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

Thm: Relationship to Binomial:

If $X \sim BIN(n,p)$, then for each $x = 0, 1, 2, \cdots$ and as $n \to \infty$, $p \to 0$, $np \to \mu$

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

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 $n \geq 100$, $p \leq .01$ and when x is close to np

This distribution shows up in a lot of places. In particular, when we have a Homogeneous Poisson Process being follows, the dist is Poisson

Poisson Process

Thm 3 2 4

Homogeneous Poisson Process

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

Consider the following properties

1) $X(0) = 0$ (starts with no arrivals)

2) $P[X(t+h) - X(t) = n | X(s) = m] = P[X(t+h) - X(t) = n]$ for all $0 \leqslant s \leqslant t$ and $0 < h$

3) $P[X(t+\Delta t) - X(t) = 1] = \lambda \Delta t + o(\Delta t)$

4) $P[X(t+\Delta t) - X(t) \geqslant 2] = o(\Delta t)$

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

Thus, $X(t) \sim POI(\lambda t)$,

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

λ = rate of occurrence or the intensity

Because λ is constant, this is why it's called homogeneous

Discrete Uniform Dist

$X \sim DU(N)$

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

Ex: Rolling a die

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

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

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

Think of this one as N sided die. For example, if N=6, then we have f(x)=1/6. The mean is

E(X)=(6+1)/2 = 3.5 and

V(X) = (6^2-1)/12 = 35/12