Sec 3.1 - Binomial and Related Distributions

1. Hypergeometric Distribution.

→ use when sampling done without replacement
→ used a lot in acceptance sampling

ex) Suppose there are 100 fuses in a lot.
20 are defective.
Suppose 5 are chosen at random.
If all are working, then lot is accepted.

We are interested in the probability that the lot is rejected.

sol) X = # defective fuses in sample.

$$D_X = \{0, 1, 2, 3, 4, 5\}$$

$$\mathbb{P}[X = x] = \frac{\underbrace{\binom{20}{x}}_{\text{success}} \underbrace{\binom{80}{5-x}}_{\text{failures}}}{\binom{100}{5}}, \quad x = 0, 1, 2, 3, 4, 5$$

$$\mathbb{P}[X = 2] = \frac{\binom{20}{2}\binom{80}{3}}{\binom{100}{5}}$$

def) Hypergeometric Dist.

X = # of defectives in a sample of size n

N = total items
D = # total defectives.

The pmf of X is

$$\mathbb{P}[X=x] = p_X(x)$$

$$= \begin{cases} \dfrac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}}, & x=0,1,\dots,n \\ 0, & \text{o.w.} \end{cases}$$

We say

$$X \sim \text{HyperGeom}\underbrace{(N, D, n)}_{\text{parameter values.}}$$

↑ follows a

$$\mathbb{E}[X] = n \cdot \frac{D}{N} \qquad Var[X] = n \cdot \frac{D}{N} \cdot \frac{N-D}{N} \cdot \frac{N-n}{N-1}$$

2 sample with replacement

Bernoulli Experiment

- random experiment
- outcomes classified as a success or a failure.
- prob. of success = p
- prob. of failure = 1 - p

Let X be the result of a Bernoulli experiment such that

$X(\text{success}) = 1$

$X(\text{failure}) = 0$

The pmf of X is

$$p_X(x) = \begin{cases} p^x (1-p)^{1-x}, & x = 0,1 \\ 0, & \text{o.w.} \end{cases}$$

only for 1 trial/experiment

We say $X \sim Ber(p)$

$E[X] = p = \sum_{x=0}^{1} x \cdot p^x (1-p)^{1-x}$

$Var[X] = p(1-p)$

3 Binomial Distribution

Take n Bernoulli trials.

Let X_i = Bernoulli RV associated with the i^{th} trial.

Interested in the total # of successes, not the order of occurrence.

→ consider each trial independent.
→ prob of success $= p$ is constant.

Define X = # of observed successes in n Bernoulli trials.

$$p_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & x = 0, \dots, n \\ 0, & \text{o.w.} \end{cases}$$

$X \sim Bin(n,p)$ $\qquad$ $X \sim b(n,p)$

$\mathbb{E}[X] = np$ $\qquad$ $Var[X] = np(1-p)$

$$M_X(t) = \left[(1-p) + pe^t\right]^n$$

Thm) Let $X_1, X_2, \ldots, X_m$ be indep. RVs.

Take $X_i \sim Bin(n_i, p)$ for $i = 1, 2, \ldots, m$

Let $Y = \sum_{i=1}^{m} X_i$

Then $Y \sim Bin\left(\sum_{i=1}^{m} n_i, p\right)$

Proof)

$$M_{X_i}(t) = (1-p+pe^t)^{n_i}$$

$$M_Y(t) = M_{X_1+\cdots+X_m}(t)$$

$$\overset{indep}{=} M_{X_1}(t) \cdots M_{X_m}(t)$$

$$= (1-p+pe^t)^{n_1} \cdot \ldots \cdot (1-p+pe^t)^{n_m}$$

$$= (1-p+pe^t)^{n_1+\cdots+n_m}$$

4 Multinomial Dist.

→ extension of Binomial.
→ instead of success/failure, have k outcomes.

↙ mutually exclusive and exhaustive.

Setup:

- Random experiment repeated n independent times
- On each repetition, result is one of

$$\underbrace{C_1, C_2, \ldots, C_k}_{\text{mutually exclusive and exhaustive list of outcomes.}}$$

- Let p_i = prob. that an outcome is an element of C_i

↙ these stay constant throughout the repetitions.

- Define X_i = # outcomes that are elements of C_i where $i = 1, \ldots, k-1$

- Let $x_1, x_2, \ldots, x_{k-1}$ be non-neg integers so that

$$x_1 + x_2 + \cdots + x_{k-1} \leq n$$

The pmf of $X_1, \ldots, X_{k-1}$ is

$$\frac{n!}{x_1! \cdot \ldots \cdot x_{k-1}! \, x_k!} p_1^{x_1} \cdot \ldots \cdot p_{k-1}^{x_{k-1}} \cdot p_k^{x_k}$$

$$x_k = n - (x_1 + \cdots + x_{k-1})$$

MGF: $M(t_1, \ldots, t_{k-1}) = \left(p_1 e^{t_1} + \cdots + p_{k-1} e^{t_{k-1}} + p_k\right)^n$

(5) Trinomial Dist (special case where $k=3$)

Let $X_1 = X$
$X_2 = Y$
$X_3 = n - X - Y$

Joint pmf of X and Y.

$$p(x,y) = \frac{n!}{x!\,y!\,(n-x-y)!} p_1^x p_2^y p_3^{n-x-y}$$

where x, y are nonneg integers.

$$x + y \le n$$

$$p_1 + p_2 + p_3 = 1$$

MGF. $M(t_1, t_2) = \left(p_1 e^{t_1} + p_2 e^{t_2} + p_3\right)^n$

MGF of Marginal Dist.

For X : $M(t_1, 0)$

$$= \left(p_1 e^{t_1} + p_2 e^{0} + p_3\right)^n$$

$$= \left(p_1 e^{t_1} + \underbrace{p_2 + p_3}_{1-p_1}\right)^n$$

$$X \sim Bin(n, p_1)$$

For Y : $M(0, t_2) = \left(p_2 e^{t_2} + 1 - p_2\right)^n$

$Y \sim Bin(n, p_2)$

X and Y are Dependent.

since
$M(t_1, t_2) \neq M(t_1) M(t_2)$