Chapter 3 Some Special Distributions

Section 3.1 The Binomial, Trinomial, and Multinomial Distributions

Binomial: If n is a positive integer, recall that, $(a+b)^n = \sum_{x=0}^{n} \binom{n}{x} b^x a^{n-x}$.

Now, consider the function $f(x) = \binom{n}{x} p^x (1-p)^{n-x}$; $x = 0, 1, 2, \ldots, n$; zero elsewhere

where n is a positive integer and $0 < p < 1$.

Question: Is $f(x)$ a probability density function (pdf) ?

1. $f(x) \geq 0$
2. $\sum_{x=0}^{n} f(x) = 1$. Since $\sum_{x=0}^{n} f(x) = \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = (p + (1-P))^n = (1)^n = 1$.

Yes, $f(x)$ is a pdf and **X** is a binomial random variable. Notation: $X \sim b(n,p)$ or $X \sim B(n,p)$ where n and p are the parameters of the distribution.

Derivation of the Binomial Distribution

Conditions for a binomial random experiment; i.e .

1. The experiment is repeated n independent, identical trials.
2. There are two outcomes possible on each trial -- success or failure.
3. The probability of success, p , remains the same from trial to trial.

Let Y be the number of successes in n trials. Find the pdf of Y i.e. $P(Y = y) = f(y)$.

If the probability of success(S) is p then the probability of failure(F) is **1-*P***.

We know that the number of successes is **y** and the number of failures is ***n*-y;** i.e.

$\overbrace{SSSSS\ldots S}^{y} \; \overbrace{FFFFF\ldots F}^{n-y} = S^y F^{n-y}$; i.e. $\overbrace{ppppp\ldots p}^{y} \; \overbrace{(1-p)(1-p)(1-p)(1-p)\ldots(1-p)}^{n-y} = p^y (1-p)^{n-y}$.

This is the probability of one of the combinations to get **y** successes and **n-y** failures.

The number of different ways to obtain **y** successes in n trials is given by $\binom{n}{y}$

Hence, the pdf of **y** is: $f(y) = \binom{n}{y} p^y (1-p)^{n-y}$; $y = 0, 1, 2, \ldots, n$. Notation: $Y \sim b(n,p)$

The mgf for the binomial random variable is:

$$M(t)=\sum_{x=0}^{n} e^{tx} f(x)=\sum_{x=0}^{n} e^{tx}\binom{n}{x} p^x (1-p)^{n-x}=\sum_{x=0}^{n}\binom{n}{x}\left(pe^t\right)^x (1-p)^{n-x}=\left((1-p)+pe^t\right)^n.$$

$$M(t)=\left((1-p)+pe^t\right)^n$$

$$\mu=E(X)=M'(t)=n\left((1-p)+pe^t\right)^{n-1} pe^t\big|_{t=0}=n(1-p+p)^{n-1}p=np.\quad \mu=E(x)=np$$

$$E\left(X^2\right)=M''(t)=n(n-1)\left((1-p)+pe^t\right)^{n-2}(pe^t)^2+\ n\left((1-p)+pe^t\right)^{n-1} pe^t\big|_{t=0}$$
$$=n(n-1)p^2+np\ =\ (np)^2-np^2+np.$$

$$\sigma^2=E(X^2)-\left(E(X)\right)^2=(np)^2-np^2+np-(np)^2=np-np^2=np(1-p).\quad \sigma^2=np(1-p)$$

Example 1 : If $X \sim b\left(4,\frac{1}{9}\right)$, what is the mgf? $M(t)=\left((1-\frac{1}{9})+\frac{1}{9}e^t\right)^4=\left(\frac{8}{9}+\frac{1}{9}e^t\right)^4$.

Example 2 : Given the mgf $M(t)=\left(\frac{1}{6}+\frac{5}{6}e^t\right)^{10}$, what is the pdf of **X** ? $X \sim b\left(10,\frac{5}{6}\right)$.

The Negative Binomial Distribution:

Binomial random Variable: X is the number of successes in **n** trials?

Now, consider a sequence of trials. Let the random variable **Y** denote the total number of failures before the **r^th^** success; that is, **Y + r** is equal to the total number of trials necessary to produce exactly **r** successes. Note: **r** is a fixed positive integer.

Let $Y=0, 1, 2, 3, \ldots$. What is the pdf of **Y** ?

Derivation of the Negative Binomial Distribution

Consider the case of obtaining **r - 1** successes in the first **Y + (r – 1)** trials and the **r^th** success on the **Y + r** trial.

There are $\binom{y+r-1}{r-1}$ different ways of obtaining the **r - 1** successes out of **Y + (r – 1)** trials.

The probability in each way is $p^{r-1}(1-p)^{y+(r-1)-(r-1)} = p^{r-1}(1-p)^{y}$, that is, $\binom{y+r-1}{r-1} p^{r-1}(1-p)^{y}$.

Now, the probability of success on the **Y + r** trial is p .

That is $f(y) = \binom{y+r-1}{r-1} p^{r-1}(1-p)^{y}\, p = \binom{y+r-1}{r-1} p^{r}(1-p)^{y}$. Hence,

$f(y) = \binom{y+r-1}{r-1} p^{r}(1-p)^{y}$; $y = 0, 1, 2, 3, \ldots$. Notation: $Y \sim Negb(Y+r, p)$

Find the mgf for the negative binomial random variable.

$$M(t) = \sum_{y=0}^{n} e^{ty} f(x) = \sum_{y=0}^{n} e^{ty} \binom{y+r-1}{r-1} p^{r}(1-p)^{y} = \sum_{y=0}^{n} \binom{y+r-1}{r-1} p^{r}\left((1-p)e^{t}\right)^{y}$$

$$= \frac{p^{r}}{\left(1-(1-p)e^{t}\right)^{r}} \sum_{y=0}^{n} \binom{y+r-1}{r-1} \left(1-(1-p)e^{t}\right)^{r} \left((1-p)e^{t}\right)^{y} = \frac{p^{r}}{\left(1-(1-p)e^{t}\right)^{r}} = p^{r}\left(1-(1-p)e^{t}\right)^{-r};$$

$$where\ 1-(1-p)e^{t} > 0 \;\Rightarrow t < -\ln(1-p)$$

$$M(t) = p^{r}\left(1-(1-p)e^{t}\right)^{-r}; \quad for \quad t < -\ln(1-p)$$

Also note that the negative binomial can be presented in another way. Currently, the random variable **Y** is the number of failures necessary to produce **r** Successes. Now let the random variable **X** be the total number of trials necessary to produce **r** Successes. Note: **X = Y + r .** The probability distribution of **X** is

$f(x) = \binom{x-1}{r-1} p^{r}(1-p)^{x-r}$; $x = r, r+1, r+2, r+3, \ldots$; zero elsewhere .

Question: What is the mean and variance of **X** ? Since X is a linear combination of Y, we don't have to do much work. $E(X) = E(Y+r) = E(Y) + r = \frac{r(1-p)}{p} + r = \frac{r-rp+rp}{p} = \frac{r}{p}$. The variance is $Var(X) = Var(Y+r) = Var(Y) + 0 = \frac{r(1-p)}{p^2}$.

The Trinomial Distribution:

Trinomial : If **n** is a positive integer and **a_1, a_2, a_3** are fixed constants, we have

$(a_1 + a_2 + a_3)^n = \sum_{x=0}^{n}\sum_{y=0}^{n-x} \frac{n!}{x!y!(n-x-y)!} a_1^x a_2^y a_3^{n-x-y}$. Let $a_1 = p_1$, $a_2 = p_2$, $a_3 = p_3$.

If $p_1 + p_2 + p_3 = 1$ then $\sum_{x=0}^{n}\sum_{y=0}^{n-x} \frac{n!}{x!y!(n-x-y)!} p_1^x p_2^y p_3^{n-x-y} = 1$ and $f(x,y) = \frac{n!}{x!y!(n-x-y)!} p_1^x p_2^y p_3^{n-x-y}$; $x + y \le n$, is a pdf.

1. $f(x,y) \ge 0$
2. $\sum_{x=0}^{n}\sum_{y=0}^{n-x} f(x,y) = (p_1 + p_2 + p_3)^n = (1)^n = 1$.

Example: Roll a die 5 times. Let the random variable X be the number of terminations in the set$\{x : x = 1, 2\}$ and let the random variable Y be the number of terminations in the set $\{y : y = 3,4,5\}$.

a. What is $f(x,y)$, b. Compute $P(x = 2, y = 1)$.

a. $f(x,y) = \left(\frac{n!}{x!y!(n-x-y)!}\right) p_1^x p_2^y p_3^{n-x-y}$; $x + y \le n$

$$f(x,y) = \left(\frac{5!}{x!y!(5-x-y)!}\right)\left(\tfrac{2}{6}\right)^x \left(\tfrac{3}{6}\right)^y \left(\tfrac{1}{6}\right)^{5-x-y} ;\ x + y \le 5$$

b. $P(x = 2, y = 1) = f(x = 2, y = 1) = \left(\frac{5!}{2!1!(5-2-1)!}\right)\left(\tfrac{2}{6}\right)^2 \left(\tfrac{3}{6}\right)^1 \left(\tfrac{1}{6}\right)^{5-2-1} = 0.0463$.

Moment generating function for the trinomial distribution.

$$M(t_1,t_2) = E(e^{xt_1+yt_2}) = \sum_{x=0}^{n}\sum_{y=0}^{n-x} e^{xt_1+yt_2} \frac{n!}{x!y!(n-x-y)!} p_1^x p_2^y p_3^{n-x-y}$$

$$= \sum_{x=0}^{n}\sum_{y=0}^{n-x} \frac{n!}{x!y!(n-x-y)!} \left(p_1e^{t_1}\right)^x \left(p_2e^{t_2}\right)^y p_3^{n-x-y} = \left(p_1e^{t_1} + p_2e^{t_2} + p_3\right)^n .$$ $M(t_1,t_2) = \left(p_1e^{t_1} + p_2e^{t_2} + p_3\right)^n$

$M(t_1,0) = \left(p_1e^{t_1} + p_2e^{0} + p_3\right)^n = \left(p_2 + p_3 + p_1e^{t_1}\right)^n$ Since

$p_1 + p_2 + p_3 = 1 \Rightarrow p_2 + p_3 = 1 - p_1$ then $\left(1 - p_1 + p_1e^{t_1}\right)^n$. Note $X \sim b(n, p_1)$.

Similarly, $M(0,t_2) = \left(1 - p_2 + p_2e^{t_2}\right)^n$. Note $Y \sim b(n, p_2)$.

Further more, **X** and **Y** are not independent since $M(t_1,t_2) \neq M(t_1,0)M(0,t_2)$.
Since **X** and **Y** are not independent, we would like to know the conditional pdfs of $f(y|x)$ *and* $f(x|y)$. Mathematics show that

$$f(y|x)=\binom{n-x}{y}\left(\frac{p_2}{1-p_1}\right)^y\left(\frac{p_3}{1-p_1}\right)^{n-x-y} \quad ; y=0,1,2,\ldots,n-x.$$ Note: $Y|X \sim b\left(n-x, \frac{p_2}{1-P_1}\right)$.

$$f(x|y)=\binom{n-y}{x}\left(\frac{p_1}{1-p_2}\right)^x\left(\frac{p_3}{1-p_2}\right)^{n-y-x} \quad ; x=0,1,2,\ldots,n-y.$$ Note: $X|Y \sim b\left(n-y, \frac{p_1}{1-P_2}\right)$.

Since **X** and **Y** are dependent random variables compute the correlation, $\rho_{1,2}$.

We know the conditional distributions are binomial random variables.

$E(Y|X)=(n-x)\left(\frac{p_2}{1-P_1}\right)$ and $E(X|Y)=(n-y)\left(\frac{p_1}{1-P_2}\right)$

$$\rho_{1,2}^2=\left(-\frac{p_2}{1-p_1}\right)\left(-\frac{p_1}{1-p_2}\right)=\frac{p_1p_2}{(1-p_1)(1-p_2)} \quad \text{and} \quad \rho_{1,2}=-\sqrt{\frac{p_1p_2}{(1-p_1)(1-p_2)}}$$

Homework 3.1

1. For each distribution provide the pdf, mean, variance, and mgf. Make sure you include the limits for each pdf. a. Binomial b. Negative Binomial

2. For the Trinomial distribution provide the $f(x,y)$, $M(t_1,t_2)$, $f(x)$, $f(y)$, $f(x|y)$, $f(y|x)$, $E(Y|X)$, $E(X|Y)$, and $\rho_{1,2}$.

3. Derive the Binomial Distribution.

4. Given X is a binomial random variable find $M(t)$ and then compute $E(X)$ *and* σ^2.
5. Given Y is a Negative binomial random variable find $M(t)$ and then compute $E(Y)$ *and* σ^2.
6. Given the mgf of the trinomial distribution, $M(t_1,t_2)=\left(p_1e^{t_1}+p_2e^{t_2}+p_3\right)^n$, compute the covariance of X and Y.