Prepared by: Dr. Haifa Abdul-Jawad & Dr. Zaid T. Al-Khaledi

Geometric distribution: A random variable X is said to have a geometric distribution, denoted by $X \sim Geometric(p)$, if its PMF has the following form:

$$f(x) = \begin{cases} pq^x & \text{for } x = 0, 1, 2, ... \\ 0 & \text{otherwise} \end{cases}$$

where $0 \leq p \leq 1$ is the probability of success, and $q = 1 - p$ is the probability of failure. This probability distribution models the behavior of a random variable X that represents the number of failures before the first success in a sequence of Bernoulli trials.

Sometimes the random variable X is regarded as the number of Bernoulli trials needed to get one success. In this case, the PMF of the geometric distribution has the following form:

$$f(x) = \begin{cases} pq^{x-1} & \text{for } x = 1, 2, 3, ...; \text{ and } 0 \leq p \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Mean, variance, MGF, and CDF: If X represents the number of failures before the first success, then by definition of the mean, variance, MGF, and CDF, we have

$$\begin{aligned}
E[X] &= \sum_{x=0}^{\infty} xpq^x \\
&= pq \sum_{x=1}^{\infty} xq^{x-1}, \text{ when } x = 0 \text{ the first term is zero} \\
&= pq \frac{d}{dq} \left(\sum_{x=0}^{\infty} q^x \right), \text{starts from 0 because } \frac{d}{dq}(1) = 0 \\
&= pq \frac{d}{dq} \left(\frac{1}{1-q} \right), \text{ by the geometric series} \\
&= \frac{pq}{(1-q)^2} \\
&= \frac{pq}{p^2} = \frac{q}{p}
\end{aligned}$$

to find $E[X^2]$, we obtain $E[X(X-1)]$ first. Then calculate $E[X^2] = E[X(X-1)] + E[X]$

$$\begin{aligned}
E[X(X-1)] &= \sum_{x=0}^{\infty} x(x-1)pq^x \\
&= pq^2 \sum_{x=2}^{\infty} x(x-1)q^{x-2}, \text{ when } x = 0 \text{ the first two terms are zeros} \\
&= pq^2 \frac{d^2}{dq^2} \left(\sum_{x=0}^{\infty} q^x \right), \text{starts from 0 because } \frac{d^2}{dq^2}(1+q) = 0 \\
&= pq^2 \frac{d^2}{dq^2} \left(\frac{1}{1-q} \right), \text{ by the geometric series} \\
&= 2\frac{pq^2}{(1-q)^3} \\
&= 2\frac{pq^2}{p^3} = 2\frac{q^2}{p^2}
\end{aligned}$$

$$
\begin{aligned}
E[X^2] &= E[X(X-1)] + E[X] \\
&= 2\frac{q^2}{p^2} + \frac{q}{p}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Var}(X) &= E[X^2] - E[X]^2 \\
&= 2\frac{q^2}{p^2} + \frac{q}{p} - \frac{q^2}{p^2} \\
&= \frac{q^2}{p^2} + \frac{q}{p} \\
&= \frac{q^2 + pq}{p^2} \\
&= \frac{q\left(q+p\right)}{p^2} = \frac{q}{p^2}
\end{aligned}
$$

$$
\begin{aligned}
M_X(t) &= \sum_{x=0}^{\infty} e^{tx} p q^x \\
&= p \sum_{x=0}^{\infty} \left(qe^t\right)^x = \frac{p}{1-qe^t}, \text{ by the geometric series}
\end{aligned}
$$

$$
\begin{aligned}
F(x) &= \sum_{k=0}^{x} p q^k \\
&= p \sum_{k=0}^{x} q^k \\
&= p \frac{1-q^{x+1}}{1-q} \\
&= 1 - q^{\lfloor x \rfloor + 1}
\end{aligned}
$$

$$F(x) = \begin{cases} 1 - q^{\lfloor x \rfloor + 1} & \text{for } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

where $\lfloor x \rfloor$ is the floor of x. For example if $x = 2.3$, then $\lfloor x \rfloor = 2$

If X represents the number of Bernoulli trials needed to get one success, then by definition of the mean, variance, MGF, and CDF, we have

$$\begin{aligned}
E[X] &= \sum_{x=1}^{\infty} xpq^{x-1} \\
&= p\sum_{x=1}^{\infty} xq^{x-1} \\
&= p\frac{d}{dq}\left(\sum_{x=0}^{\infty} q^x\right) \text{, starts from 0 because } \frac{d}{dq}(1) = 0 \\
&= p\frac{d}{dq}\left(\frac{1}{1-q}\right), \text{ by the geometric series} \\
&= \frac{p}{(1-q)^2} \\
&= \frac{p}{p^2} = \frac{1}{p}
\end{aligned}$$

to find $E[X^2]$, we obtain $E[X(X-1)]$ first. Then calculate $E[X^2] =$

$E[X(X-1)] + E[X]$

$$\begin{aligned}
E[X(X-1)] &= \sum_{x=1}^{\infty} x(x-1)pq^{x-1} \\
&= pq\sum_{x=2}^{\infty} x(x-1)q^{x-2}, \text{ when } x=0 \text{ the first term is zero} \\
&= pq\frac{d^2}{dq^2}\left(\sum_{x=0}^{\infty} q^x\right), \text{starts from 0 because } \frac{d^2}{dq^2}(1+q)=0 \\
&= pq\frac{d^2}{dq^2}\left(\frac{1}{1-q}\right), \text{ by the geometric series} \\
&= 2\frac{pq}{(1-q)^3} \\
&= 2\frac{pq}{p^3} = 2\frac{q}{p^2}
\end{aligned}$$

$$\begin{aligned}
E[X^2] &= E[X(X-1)] + E[X] \\
&= 2\frac{q}{p^2} + \frac{1}{p}
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}(X) &= E[X^2] - E[X]^2 \\
&= 2\frac{q}{p^2} + \frac{1}{p} - \frac{1}{p^2} \\
&= \frac{2q+p-1}{p^2} \\
&= \frac{q+q+p-1}{p^2} = \frac{q}{p^2}, \text{ because } p+q=1
\end{aligned}$$

$$
\begin{aligned}
M_X(t) &= \sum_{x=1}^{\infty} e^{tx} pq^{x-1} \\
&= pe^t \sum_{x=1}^{\infty} \left(qe^t\right)^{x-1} \\
&= pe^t \sum_{k=0}^{\infty} \left(qe^t\right)^{k}, \text{ for } k = x-1 \\
&= \frac{pe^t}{1-qe^t}, \text{ by the geometric series}
\end{aligned}
$$

$$
\begin{aligned}
F(x) &= \sum_{k=1}^{x} pq^{k-1} \\
&= p \sum_{k=1}^{x} q^{k-1}
\end{aligned}
$$

let $m = k - 1$, then m takes the values $0, 1, 2, ..., x - 1$. Hence

$$
\begin{aligned}
F(x) &= \sum_{m=0}^{m} q^{m} \\
&= p\left(\frac{1-q^x}{1-q}\right) \\
&= 1 - q^{\lfloor x \rfloor}
\end{aligned}
$$

$$
F(x) = \begin{cases} 1 - q^{\lfloor x \rfloor} & \text{for } x \geq 1 \\ \\ 0 & \text{otherwise} \end{cases}
$$

Example 2.1: Let X represents the number of Bernoulli trials needed to get one success with probability of success $p = 0.6$.

1- Write down the PMF of X.

2- Obtain the mean, the variance, the MGF, and the CDF of X.

3- Calculate $P(X = 0)$, $P(X \geq 3)$, and $P(X \leq 3.4)$.

4- Let $Y = X - 1$. Find the mean and variance of Y, then find $P(Y = 0)$.

Solution:

1- The PMF of X:

$$f(x) = \begin{cases} 0.6 \times 0.4^{x-1} & \text{for } x = 1, 2, 3, ... \\ 0 & \text{otherwise} \end{cases}$$

2- The mean, the variance, the MGF, the CDF of X

$$\begin{aligned} E[X] &= \frac{1}{p} = \frac{1}{0.6} = 1.667 \\ \text{Var}(X) &= \frac{q}{p^2} = \frac{0.4}{(0.6)^2} = 1.111 \\ M_X(t) &= \frac{pe^t}{1 - qe^t} = \frac{0.6e^t}{1 - 0.4e^t} \end{aligned}$$

$$F(x) = \begin{cases} 1 - 0.4^{\lfloor x \rfloor} & \text{for } x \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

3- Calculate $P(X = 0)$, $P(X \geq 3)$, and $P(X \leq 3.4)$

$P(X = 0) = 0$, because the PMF is defined as zero when $x = 0$

$$
\begin{aligned}
P(X \geq 3) &= 1 - P(X \leq 2) \\
&= 1 - F(2) \\
&= 1 - \left(1 - 0.4^{\lfloor 2 \rfloor}\right) \\
&= 0.4^2 = 0.16
\end{aligned}
$$

$$
\begin{aligned}
P(X \leq 3.4) &= F(3.4) \\
&= 1 - 0.4^{\lfloor 3.4 \rfloor} \\
&= 1 - 0.4^3 = 0.936
\end{aligned}
$$

4- Let $Y = X - 1$. Find the mean and variance of Y, then find $P(Y = 0)$.

$$E[Y] = E[X - 1] = E[X] - 1 = 1.667 - 1 = 0.667$$

$$\text{Var}(Y) = \text{Var}(X - 1) = \text{Var}(X) = 1.111$$

$$P(Y = 0) = P(X - 1 = 0) = P(X = 1) = 0.6(0.4)^{1-1} = 0.6$$

Negative binomial distribution: A random variable X is said to have a negative binomial distribution, denoted by $X \sim NB(r, p)$, if its PMF has the following form:

$$f(x) = \begin{cases} \binom{x+r-1}{r-1} p^r q^x & \text{for } x = 0, 1, 2, ... \\ 0 & \text{otherwise} \end{cases}$$

where $0 \leq p \leq 1$ is the probability of success, $q = 1 - p$ is the probability of failure, and $r > 0$ is the targeted number of successes. This probability distribution models the behavior of a random variable X that represents the number of failures before the rth success in a sequence of Bernoulli trials. Notice that the geometric distribution is identical to the negative binomial distribution with $r = 1$.

$$E[X] = \frac{rq}{p}$$

$$\text{Var}(X) = \frac{rq}{p^2}$$

$$M_X(t) = \left(\frac{p}{1 - qe^t}\right)^r$$

Example 2.2: Let $X \sim NB(4, 0.3)$. Find the following:

1- Write down the PMF of X.

3- Calculate $P(X \geq 1)$.

2- Find the mean, the variance, and the MGF of $Y = 4 + 5X$.

Solution:

$$f(x) = \begin{cases} \binom{x+3}{3} 0.3^4 0.7^x & \text{for } x = 0, 1, 2, ... \\ \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} P(X \geq 1) &= 1 - P(X = 0) \\ &= 1 - \binom{0+3}{3} 0.3^4 0.7^0 \\ &= 1 - 0.0081 = 0.9919 \end{aligned}$$

$$\begin{aligned} E[X] &= \frac{rq}{p} = \frac{4(0.7)}{0.3} = 9.333 \\ \text{Var}(X) &= \frac{rq}{p^2} = \frac{4(0.7)}{(0.3)^2} = 31.111 \\ M_X(t) &= \left(\frac{p}{1 - qe^t}\right)^r = \left(\frac{0.3}{1 - 0.7e^t}\right)^4 \end{aligned}$$

$$\begin{aligned} E[Y] &= E[4 + 5X] \\ &= 4 + 5E[X] \\ &= 4 + 5(9.333) = 50.665 \end{aligned}$$

$$\begin{aligned} \text{Var}(Y) &= \text{Var}(4 + 5X) \\ &= 25\text{Var}(X) \\ &= 25(31.111) = 777.775 \end{aligned}$$

$$
\begin{aligned}
M_Y(t) &= E[e^{tY}] \\
&= E\left[e^{t(4+5X)}\right] \\
&= e^{4t}E\left[e^{5tX)}\right] \\
&= e^{4t}M_X(5t) \\
&= \left(\frac{0.3e^t}{1-0.7e^{5t}}\right)^4
\end{aligned}
$$