Recall: a random variable X has a discrete distribution if the cdf F_X of X is a step function.

Note: Let X be a discrete variable. Then the values of X, denoted as $\mathscr{X}$, is a discrete set, a finite set or a countably infinite set.

For a examples, we will need to perform the following analysis:

1. Identify the pmf.
2. Calculate $E(X)$, $V(X)$, and $M_X(t)$.
3. Develop intuition for what the random variable does.

Bernoulli Distribution: We say that a random variable X has the Bernoulli distribution, denoted by $X \sim \text{Bernoulli}(p)$ with parameter p being the probability of success, provided that $\mathscr{X} = \{0,1\}$ and the pmf is defined by

$$p_X(x) = \begin{cases} 1-p & x = 0 \\ p & x = 1 \end{cases}.$$

Alternatively, p_X can also be defined by $p_X(x) = p^x \cdot (1-p)^{1-x}$, where $x \in \{0,1\}$ and $p \in (0,1)$.

Intuitively, the random variable X models a coin toss or any phenomena with a Yes/No answer where $P(\text{Yes}) = p = P(H)$.

Consider $X \sim \text{Bernoulli}(p)$.

1. To calculate $E(X)$, we write

$$E(X) = \sum_{x \in \mathscr{X}} x \cdot p_X(x) = 0 \cdot (1-p) + 1 \cdot p = p\,.$$

2. To calculate $V(X)$, we note that $V(X) = E(X^2) - (E(X))^2$. First we calculate

$$E(X^2) = \sum_{x \in \mathscr{X}} x^2 \cdot p_X(x) = 0 \cdot (1-p) + 1 \cdot p = p\,.$$

Now we see

$$V(X) = p - p^2 = p \cdot (1-p)\,.$$

3. To calculate $M_X(t)$, we write

$$M_X(t) = E(e^{tX}) = \sum_{x \in \mathscr{X}} e^{tx} \cdot p_X(x) = 1 \cdot (1-p) + e^t \cdot p = pe^t + (1-p)\,.$$

Here we observer that $M_X(t) = p \cdot e^t + (1-p)$, then we have

$$\frac{d}{dt}(M_X(t))|_{t=0} = pe^t|_{t=0} = p\,.$$

Thus we have checked

$$\frac{d}{dt}(M_X(t))|_{t=0} = E(X)\,, \qquad \frac{d^2}{dt^2}(M_X(t))|_{t=0} = p \cdot e^t|_{t=0} = p = E(X^2)\,.$$

Also, we observe that $E(X^2) \neq (E(X))^2$, and the k^{th} moments for $X \sim \text{Bernoulli}(p)$ satisfies

$$E(X^k) = \frac{d^k}{dt^k}(M_X(t))|_{t=0} = p\,.$$

Binomial Distribution: Suppose we are given n-indpependent Bernoulli(p) trials. Let X denote the number of successes among the n-trials, then X gives a Binomial distribution.

We say that X has the Binomial distribution with parameters n and p, where n denotes the number of trials and p denotes the probability of success of each trial, provided that $\mathscr{X} = \{0, 1, 2, 3, \cdots, n\}$ and the pmf of X is defined by

$$p_X(x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{1-x} .$$

with $x \in \mathscr{X}$ and $p \in (0, 1)$.

Intuitively, the binomial random variable counts the number of successes among n-independent Bernoulli(p) trials.

1. To calculate $E(X)$, first we define change of variable $y = x - 1$, and perform the calculation

$$\begin{aligned} E(X) = \sum_{x \in \mathscr{X}} x \cdot p_X(x) &= \sum_{x=0}^{n} x \cdot \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x} \\ &= \sum_{x=0}^{n} x \cdot \frac{n!}{(n-x)!\, x!} \cdot p^x \cdot (1-p)^{n-x} \\ &= \sum_{x=1}^{n} \frac{n!}{(n-x)!\, (x-1)!} \cdot p^x \cdot (1-p)^{n-x} \\ &= np \cdot \sum_{x=1}^{n} \frac{(n-1)!}{(n-x)!\, (x-1)} \cdot p^{x-1} \cdot (1-p)^{n-x} \\ &= np \cdot \sum_{y=0}^{n-1} \frac{(n-1)!}{(n-(y+1))!\, y!} p^y \cdot (1-p)^{n-(y+1} \\ &= np \cdot \sum_{y=0}^{n-1} \frac{(n-1)!}{((n-1)-y)!\, y!} p^y \cdot (1-p)^{(n-1)-y} \\ &= np \cdot \sum_{y=0}^{n-1} \binom{n-1}{y} \cdot p^y \cdot (1-p)^{(n-1)-y} \end{aligned}$$

Here the summands are pmf of Binom$(n-1, p)$, thus they add up to 1, and thus we conclude that $E(X) = np$. Now we have checked

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

thus we see that, with $a = p$ and $b = (1-p)$, we have the formula

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} \cdot a^k \cdot b^{n-k} .$$

2. To calculate $V(X)$, we first compute

$$\begin{aligned} E(X^2) = \sum_{x \in \mathscr{X}} x^2 \cdot p_X(x) &= \sum_{x \in \mathscr{X}} x^2 \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x} \\ &= \sum_{x \in \mathscr{X}} x^2 \cdot \frac{n!}{(n-x)!\, x!} \cdot p^x \cdot (1-p)^{n-x} \\ &= np \cdot \sum_{x=1}^{n} \frac{x \cdot (n-1)!}{(n-x!\, (x-1)!} \cdot p^{x-1} \cdot (1-p)^{n-x} \\ &= np \cdot \sum_{y=0}^{n-1} \frac{(y+1)\, (n-1)!}{((n-1)-y)!\, y!} \cdot p^y \cdot (1-p)^{(n-1)-y} \\ &= np \cdot \left(\sum_{y=0}^{n-1} y \cdot p_Y(y) + \sum_{y=1}^{n-1} p_Y(y) \right) \\ &= np \cdot (E(Y) + 1) \\ &= np \cdot (p \cdot (n-1) + 1) , \end{aligned}$$

where we have again used $y = x - 1$, and $Y = \text{Binom}(n-1, p)$. Thus we now have

$$V(X) = E(X^2) - (E(X))^2 = np(np - p + 1) - (np)^2 = np - np^2 \,.$$

We conclude that we have

$$V(X) = np(1-p) \,.$$

3. Lastly, we shall compute $M_X(t)$.

$$\begin{aligned}
M_X(t) = \sum_{x=0}^{n} e^{tx} \cdot p_X(x) &= \sum_{x=0}^{n} e^{tx} \cdot \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x} \\
&= \sum_{x=0}^{n} \binom{n}{x} \cdot e^{tx} \cdot p^x \cdot (1-p)^{n-x} \\
&= \sum_{x=0}^{n} \binom{n}{x} \cdot (pe^t)^x \cdot (1-p)^{n-x} \\
&= (pe^t + (1-p))^n \,,
\end{aligned}$$

where we have used the binomial formula in the last equality. Now we can check

$$\begin{aligned}
\frac{d}{dt}(M_X(t))|_{t=0} = \frac{d}{dt}(pe^t + (1-p))^n|_{t=0} &= \left(n \cdot (pe^t + (1-p))^{n-1} \cdot pe^t\right)\Big|_{t=0} \\
&= n \cdot (p + (1-p))^{n-1} \cdot p = np = E(X) \,.
\end{aligned}$$

Hypergeometric Distribution: Consider a bag of N balls, M of which are blue, and the other $N-M$ balls are not blue. We would like to choose n balls from this bag, order does not matter. There are two ways of choosing the balls:

1. Choosing with replacement. In this case $X \sim$Binom(n,p) with $p = M/N$ is the random variable for number of blue balls among the n chosen balls.

2. Choosing without replacement. In this case, we shall use the hypergoemetric distribution to describe the number of blue balls among the n chosen balls.

Suppose X denotes the number of blue balls among the n chosen balls, choosing without replacement. We want to have $p_X(x) = P(X = x)$. We observe that $\mathscr{X} = \{0, 1, 2, 3, \cdots, \min(n, M)\}$. Notice that

$$|\{X = x\}| = |\{\text{exactly } x \text{ blue balls among } n \text{ spots}\}| = \binom{M}{x} \cdot \binom{N-M}{n-x}$$

Thus it is easy to see

$$p_X(x) = \frac{\binom{M}{x} \cdot \binom{N-M}{n-x}}{\binom{N}{n}}. \tag{*}$$

We say that $X \sim$Hyper(N, M, n), where N is the population size, M is the number of successes in the population, and n is the sample size, provided that X has the pmf defined by (*).

In the following we will discuss examples of discrete random variables with infinite sample space. First we focus on the examples of "waiting" distribution, waiting for something to happen.

Geometric distribution: Consider the experiment where we keep performing a Bernoulli(p) trial until a success shows. That is, waiting for a success. We say X has the geometric distribution, with parameter p being the probability of success, provided that $\mathscr{X} = \{1, 2, 3, \cdot, \}$ and the pmf of X is given by

$$p_X(x) = p \cdot (1-p)^{x-1}$$

for $x \in \{0, 1, 2, \cdots\}$.

For instance, we consider the experiment of tossing a coin with $P(H) = p$, until the first head H shows up. Then the sample space is $S = \{H, TH, TTH, TTTH, \cdots\}$. Let X denote the number of trials to get the first head.

$$p_X(x) = P(\{X = x\}) = P(\{TTT \cdots TH\}) = (1-p)^{x-1} \cdot p\,,$$

as we have got the tail $x-1$ times (each with probability $1-p$) and the head 1 time (with probability p). One can check that we have

$$\sum_{x \in \mathscr{X}} p_X(x) = \sum_{x=1}^{\infty} p \cdot (1-p)^{x-1} = p \cdot \sum_{x=1}^{\infty} (1-p)^{x-1} = p \cdot \sum_{y=0}^{\infty} (1-p)^y = p \cdot \frac{1}{1-(1-p)} = p \cdot \frac{1}{p} = 1\,,$$

where we have denoted $y = x - 1$. To calculate $E(X)$, we first write

$$E(X) = \sum_{x=1}^{\infty} x \cdot p \cdot (1-p)^{x-1} = p \cdot \sum_{x=1}^{\infty} x \cdot (1-p)^{x-1}\,.$$

Recall that $(1-q)^{-1} = \sum_{n=0}^{\infty} q^n$ for all $|q| < 1$, differentiating both sides we get

$$\frac{d}{dq}\left(\frac{1}{1-q}\right) = \frac{1}{(1-q)^2} = \sum_{n=1}^{\infty} nq^{(n-1)}\,.$$

Thus we see

$$E(X) = p \cdot \frac{1}{(1-(1-p))^2} = \frac{p}{p^2} = \frac{1}{p}\,,$$

which gives the expected number of trials until the first head shows up, and is inversely related to $P(H)$.

To calculate $V(X)$, we notice that

$$\begin{aligned}
E(X^2) = \sum_{x=1}^{\infty} x^2 \cdot p \cdot (1-p)^{x-1} &= p \cdot \sum_{x=1}^{\infty} x \cdot (x-1+1) \cdot (1-p)^{x-1} \\
&= p \cdot \sum_{x=1}^{\infty} (x(x-1) + x) \cdot q^{x-1} \\
&= p \cdot \left(q \sum_{x=2}^{\infty} x(x-1) \cdot q^{x-2} + \sum_{x=1}^{\infty} x \cdot q^{x-1} \right)\,,
\end{aligned}$$

where we have denoted $q = 1 - p$. Now we define

$$f(q) = \frac{1}{1-q} = \sum_{n=0}^{\infty} q^n\,,$$

then

$$f'(q) = \frac{1}{(1-q)^2} = \sum_{n=1}^{\infty} nq^{n-1}\,, \qquad f''(q) = \frac{2}{(1-q)^3} = \sum_{n=2}^{\infty} n \cdot (n-1) \cdot q^{n-2}\,,$$

then combining we see that

$$E(X^2) = p \cdot \left(\frac{2q}{(1-q)^3} + \frac{1}{(1-q)^2} \right) = p \cdot \left(\frac{2(1-p)}{p^3} + \frac{1}{p^2} \right) = p \cdot \left(\frac{2(1-p)}{p^3} + \frac{p}{p^3} \right) = \frac{2-p}{p^2}\,.$$

Thus we conclude

$$V(X) = E(X^2) - (E(X))^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{(1-p)}{p^2}\,.$$

Lastly, we would like to calculate $M_X(t)$,

$$M_X(t) = \sum_{x\in\mathscr{X}} e^{tx}\cdot p\cdot(1-p)^{x-1} = \sum_{x=1}^{\infty} p\cdot e^{e(x-1+1)}\cdot(1-p)^{x-1} = pe^t\sum_{x=1}^{\infty}\left(e^t\cdot(1-p)\right)^{x-1} = pe^t\cdot\sum_{y=0}^{\infty}\left(e^t\cdot(1-p)\right)^y ,$$

where $y = x-1$, the last summing term gives a geometric series with common ratio $r = e^t\cdot(1-p)$, which converges to $1/(1-r)$ whenever $|r|<1$. Equivalently, we requires $e^t(1-p)<1$, or $e^t<(1-p)^{-1}$, or $t<\ln((1-p)^{-1})$. Thus we write

$$M_X(t) = \frac{pe^t}{1-(1-p)e^t} \qquad \text{for } t<\ln\left(\frac{1}{1-p}\right).$$

Now suppose X ~Geom(p), and consider $Y = X-1$. Then X denotes the number of trials until the first success, and Y denotes the number of failures until the first success. For instance $X(FFFFFS) = 6$, and $Y(FFFFFS) = 5$, where F denotes fails and S denotes success. Here we see that

$$E(Y) = E(X-1) = E(X)-1 = \frac{1}{p}-1 = \frac{1-p}{p}.$$

We also have

$$V(Y) = V(X-1) = V(X) = \frac{1-p}{p^2}.$$

Negative Binomial Distribution: Here we keep performing the Bernoulli(p) trials until r success shows up. Let X denote the number failures until the r-th successes. The sample space in this case is $\{SSS\cdots S, FSSS\cdots S, SFSS\cdots S,\cdots\}$. To calculate the pmf of X, we first notice that $\mathscr{X} = \{0,1,2,3,\cdots\}$. $p_X(x) = P(\{X=x\}) = P(\text{Exactly } x \text{ failures until the } r\text{-th success})$. Notice that

$$\{X=x\} = \{x+r \text{ spots;Exactly } r \text{ of which are sucesses and } x \text{ of which are failures.}\}.$$

Thus we see

$$|\{X=x\}| = \binom{x+r-1}{r-1} = \binom{x+r-1}{x}.$$

Also, for $\omega\in\{X=x\}$, we have $P(\omega) = p^r\cdot(1-p)^x$, so every outcome in $\{X=x\}$ has the same probability. That is,

$$P(X=x) = |\{X=x\}|\cdot p^r\cdot(1-p)^x ,$$

from which we conclude that

$$p_X(x) = \binom{x+r-1}{x}\cdot p^r\cdot(1-p)^x. \tag{**}$$

We say that X has the negative binomial distribution with parameters r and p, where r denotes the number of successes that we are waiting for, and p denotes the probability of each success, provided that $\mathscr{X} = \{0,1,2,3,\cdots\}$ and the pmf of X is defined by (**). In this setting, we can calculate

$$E(X) = \frac{r\cdot(1-p)}{p}, \qquad V(X) = \frac{r(1-p)}{p^2}, \qquad M_X(t) = \left(\frac{p}{1-(1-p)e^t}\right)^r \text{ for } t<\ln\left(\frac{1}{1-p}\right).$$

Poisson Distribution: We say that a random variable X has the Poisson distribution with parameter $\lambda>0$, λ represents the rate, provided that $\mathscr{X} = \{0,1,2,3,\cdots\}$ and the pmf of X is defined by

$$p_X(x) = e^{-\lambda}\cdot\frac{\lambda^x}{x!},$$

with $x\in\{0,1,2,\cdots\}$ and $\lambda\in(0,\infty)$.

It is easy to check that $p_X(x)>0$. Furthermore,

$$\sum_{x=0}^{\infty} p_X(x) = \sum_{x=0}^{\infty} e^{-\lambda}\cdot\frac{\lambda^x}{x!} = e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^x}{x!} = e^{-\lambda}e^{\lambda} = 1.$$

To calculate $E(X)$, we see that

$$E(X) = \sum_{x=0}^{\infty} x\cdot e^{-\lambda}\cdot\frac{\lambda^x}{x!} = \sum_{x=0}^{\infty}\frac{x}{x}\cdot e^{-\lambda}\cdot\frac{\lambda^x}{(x-1)!} = \lambda\cdot\sum_{x=1}^{\infty} e^{-\lambda}\cdot\frac{\lambda^{x-1}}{(x-1)!} = \lambda\cdot\sum_{y=0}^{\infty} e^{-\lambda}\cdot\frac{\lambda^y}{y!} = \lambda,$$

where we have set $y = x-1$. Similarly, one would find that $V(X) = \lambda$. Now we calculate

$$M_X(t) = \sum_{x=0}^{\infty} e^{tx}\cdot e^{-\lambda}\cdot\frac{\lambda^x}{x!} = \sum_{x=0}^{\infty} e^{-\lambda}\cdot\frac{(e^t\lambda)^x}{x!} = e^{-\lambda}e^{\lambda e^t}\sum_{x=0}^{\infty} e^{-\lambda e^t}\cdot\frac{(\lambda e^t)^x}{x!} = e^{\lambda(e^t-1)}.$$

Properties of Moment Generating Functions

Definition 0.1
Let X_n be a sequence of random variables, we say X_n converges in distribution to X provided that $F_{X_n}(u)$ converges pointwise to $F_X(u)$ for all u.

1. Let X be a random variable, we consider $Y = aX + b$, then $M_Y(t) = e^{bt} \cdot M_X(at)$.
2. Let X_n be a sequence of random variables, $X_n \to X$ in distribution iff $M_{X_n}(t) \to M_X(t)$ pointwise.
3. For random variables X and Y, X is identically distributed to Y iff $F_X(u) = F_Y(u)$ for all u, iff $M_X(t) = M_Y(t)$ for all $t \in (-\delta, \delta)$.

Theorem 0.2 (Poisson Approximation to the Binomial)
Let $X_n \sim$Binom(n, p) and $Y \sim$Pois(λ), then Y can be approximated by X_n if $np \to \lambda$.

Proof. The proof of this theorem is argued by $M_{X_n}(t) \to M_Y(t)$. □

Theorem 0.3
Consider a convergent sequence $a_n \to a$. We have

$$\lim_{n\to\infty} \left(1 + \frac{a_n}{n}\right)^n = e^a \,.$$

Here we denote $b_n = (1 + a_n/n)^n$.

Proof. Here we notice

$$\ln\left(1 + \frac{a_n}{n}\right)^n = n \cdot \ln\left(1 + \frac{a_n}{n}\right) = \frac{\ln(1 + a_n/n)}{1/n} = \frac{(1 + a_n/n)^{-1} \cdot a_n \cdot \frac{d}{dn}\left(\frac{1}{n}\right)}{\frac{d}{dn}\left(\frac{1}{n}\right)} = \frac{a_n}{1 + a_n/n} \,,$$

then we see that

$$\lim_{n\to\infty} \ln(b_n) = \lim_{n\to\infty} \frac{a_n}{1 + a_n/n} = a \,,$$

and as ln is a continuous function, we can write

$$\ln\left(\lim_{n\to\infty} b_n\right) = \lim_{n\to\infty} \ln(b_n) = a \,,$$

from which we conclude

$$\lim_{n\to\infty} b_n = e^b \,.$$

□