MTL108
Common Discrete Random Variables

Rahul Singh

Discrete random variables are fundamental in probability theory, modeling countable outcomes such as the number of successes in trials or the number of events in a fixed interval.

Bernoulli Distribution

Motivation: The Bernoulli distribution arises naturally when modeling a single trial with two outcomes, such as a coin flip (heads or tails), a yes/no survey response, or a pass/fail test, where we are interested in whether a specific event (e.g., success) occurs.

Definition 1 (Bernoulli Distribution)**.** A RV X is said to have the Bernoulli distribution with parameter p if $\mathbb{P}(X = 1) = p$ and $\mathbb{P}(X = 0) = 1 - p$, where $0 < p < 1$. We write this as $X \sim \mathit{Bernoulli}(p)$; $\sim$ is read "is distributed as" or "follows". Alternatively, $X \sim \mathit{Bernoulli}(p)$ if PMF of X is given by

$$\mathbb{P}(X = k) = \begin{cases} p & \text{if } k = 1, \\ 1 - p & \text{if } k = 0, \\ 0 & \text{otherwise.} \end{cases}$$

Properties:

- Support: $\{0, 1\}$

- Mean: $\mathbb{E}[X] = 0 \cdot (1 - p) + 1 \cdot p = p$,

- Second Moment: $\mathbb{E}[X^2] = 0^2 \cdot (1 - p) + 1^2 \cdot p = p$,

- Variance: $\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = p - p^2 = p(1 - p)$,

- Standard Deviation: $\sigma = \sqrt{p(1 - p)}$.

Example 1. For $p = 0.3$, $\mathbb{E}[X] = 0.3$, $\text{Var}(X) = 0.3 \cdot 0.7 = 0.21$, $\sigma \approx 0.458$.

Remark 1 (Bernoulli trial)**.** An experiment that can result in either a "success" or a "failure" (but not both) is called a Bernoulli trial. A Bernoulli random variable can be thought of as the indicator of success in a Bernoulli trial: it equals 1 if success occurs and 0 if failure occurs in the trial.

Binomial Distribution

Motivation: The Binomial distribution is ideal for scenarios involving a fixed number of independent trials, such as counting the number of heads in multiple coin flips, the number of defective items in a batch of products, or the number of positive responses in a series of surveys, where each trial has the same success probability.

Remark 2 (Fact)**.** If $0 \le p \le 1$ and $n \in \mathbb{N}$ then binomial theorem gives

$$\binom{n}{k} p^k (1-p)^{n-k} = (p + (1-p))^n = 1^n = 1.$$

Definition 2 (Binomial Distribution)**.** Suppose that n independent Bernoulli trials are performed, each with the same success probability p. Let X be the number of successes. The distribution of X is called the Binomial distribution with parameters n and p. We write $X \sim Binomial(n, p)$ to mean that X has the Binomial distribution with parameters n and p, where n is a positive integer and $0 < p < 1$. Alternatively, $X \sim Binomial(n, p)$ if PMF of X is given by

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n.$$

Properties:

- Support: $\{0, 1, \ldots, n\}$

- Mean: $\mathbb{E}[X] = np$ (proved using $k\binom{n}{k} = n\binom{n-1}{k-1}$),

- Variance: $\mathrm{Var}(X) = np(1-p)$ (using $\mathbb{E}[X(X-1)] = n(n-1)p^2$),

- Standard Deviation: $\sigma = \sqrt{np(1-p)}$.

Theorem 1. *If* $X \sim \mathrm{Binomial}(n, p)$*, prove that* $\mathbb{E}(X) = np$ *and* $\mathrm{Var}(X) = np(1-p)$*.*

Proof. The PMF of a binomial random variable X is given by

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n,$$

where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ is the binomial coefficient.
Expectation $\mathbb{E}[X]$: The expectation of X is defined as

$$\mathbb{E}[X] = \sum_{k=0}^{n} k\, \mathbb{P}(X = k).$$

Substitute the PMF:

$$\mathbb{E}[X] = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k}.$$

For $k=0$, the term $k\binom{n}{k}=0$, so the sum can start from $k=1$:

$$\mathbb{E}[X]=\sum_{k=1}^{n} k\binom{n}{k} p^k(1-p)^{n-k}.$$

Note that $k\binom{n}{k}=n\binom{n-1}{k-1}$ for $k\geq 1$, because

$$k\binom{n}{k}=k\cdot\frac{n!}{k!(n-k)!}=n\cdot\frac{(n-1)!}{(k-1)!(n-k)!}=n\binom{n-1}{k-1}.$$

Substituting this in the above expression, we get

$$\begin{aligned}\mathbb{E}[X] &= \sum_{k=1}^{n} n\binom{n-1}{k-1} p^k(1-p)^{n-k}\\ &= np\sum_{k=1}^{n}\binom{n-1}{k-1}p^{k-1}(1-p)^{n-k}, \text{ factoring out } np.\end{aligned}$$

Let's change the index by denoting $j=k-1$, so when $k=1$, $j=0$, and when $k=n$, $j=n-1$. So,

$$\mathbb{E}[X]=np\sum_{j=0}^{n-1}\binom{n-1}{j}p^j(1-p)^{n-1-j}.$$

Now observe that,

$$\sum_{j=0}^{n-1}\binom{n-1}{j}p^j(1-p)^{n-1-j}=(p+(1-p))^{n-1}=1^{n-1}=1.$$

Thus, substituting in the above expression we get

$$\mathbb{E}[X]=np\cdot 1=np.$$

Variance Var(X) We know that $\mathrm{Var}(X)=\mathbb{E}[X^2]-(\mathbb{E}[X])^2$. We already have $\mathbb{E}[X]=np$, so we need $\mathbb{E}[X^2]$.

Note that $X^2=X(X-1)+X$, so using linearity of expectation

$$\mathbb{E}[X^2]=\mathbb{E}[X(X-1)]+\mathbb{E}[X].$$

Now we compute $\mathbb{E}[X(X-1)]$,

$$\mathbb{E}[X(X-1)]=\sum_{k=0}^{n}k(k-1)\mathbb{P}(X=k)=\sum_{k=2}^{n}k(k-1)\binom{n}{k}p^k(1-p)^{n-k},$$

since $k(k-1)=0$ for $k=0,1$.

Note that $k(k-1)\binom{n}{k}=n(n-1)\binom{n-2}{k-2}$ for $k\geq 2$, because

$$k(k-1)\binom{n}{k}=\frac{k(k-1)n!}{k!(n-k)!}=\frac{n!}{(k-2)!(n-k)!}\cdot\frac{k(k-1)}{k(k-1)}=n(n-1)\binom{n-2}{k-2}.$$

Thus,

$$\mathbb{E}[X(X-1)] = \sum_{k=2}^{n} n(n-1)\binom{n-2}{k-2} p^k (1-p)^{n-k}.$$

Factor out $n(n-1)p^2$, we have

$$\mathbb{E}[X(X-1)] = n(n-1)p^2 \sum_{k=2}^{n} \binom{n-2}{k-2} p^{k-2}(1-p)^{n-k}.$$

Let $j = k-2$, so when $k=2$, $j=0$, and when $k=n$, $j=n-2$, so

$$\begin{aligned}\mathbb{E}[X(X-1)] &= n(n-1)p^2 \sum_{j=0}^{n-2} \binom{n-2}{j} p^j (1-p)^{n-2-j} \\ &= n(n-1)p^2(p+(1-p))^{n-2} = n(n-1)p^2 \cdot 1 = n(n-1)p^2.\end{aligned}$$

Consequently,

$$\mathbb{E}[X^2] = \mathbb{E}[X(X-1)] + \mathbb{E}[X] = n(n-1)p^2 + np.$$

Therefore,

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = [n(n-1)p^2 + np] - (np)^2 \\ &= n(n-1)p^2 + np - n^2p^2 \\ &= np - n^2p^2 + n^2p^2 - np^2 = np - np^2 = np(1-p).\end{aligned}$$

∎

Example 2. For $n = 5$, $p = 0.4$, $\mathbb{E}[X] = 5 \cdot 0.4 = 2$, $\mathrm{Var}(X) = 5 \cdot 0.4 \cdot 0.6 = 1.2$, $\sigma \approx 1.095$.

Poisson Distribution

Motivation: The Poisson distribution is useful for modeling the number of rare events occurring in a fixed interval of time or space, such as the number of phone calls received at a call center in an hour, the number of typos on a page, or the number of accidents at an intersection, the number of emails received in an hour, assuming events occur at a constant average rate.

Remark 3 (Fact)**.** The sum $\sum_{j=0}^{\infty} \lambda^j/j!$ is the Taylor series expansion of e^{λ}, i.e.,

$$e^{\lambda} = \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \quad \Rightarrow \quad \sum_{j=0}^{\infty} \frac{e^{-\lambda}\lambda^j}{j!} = 1.$$

Definition 3 (Poisson Distribution)**.** A RV X is said to follow Poisson distribution with parameter $\lambda > 0$, denoted by $X \sim Poisson(\lambda)$, if the PMF of X is given by

$$\mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \ldots.$$

Properties:

- Support: $\{0, 1, 2, \dots\}$
- Mean: $\mathbb{E}[X] = \lambda$,
- Variance: $\text{Var}(X) = \lambda$,
- Standard Deviation: $\sigma = \sqrt{\lambda}$.

Theorem 2. *If $X \sim Poisson(\lambda)$ with $\lambda > 0$, then*

$$\mathbb{E}[X] = \lambda, \quad Var(X) = \lambda.$$

Proof. The PMF of a Poisson random variable is

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \dots.$$

Mean $\mathbb{E}[X]$

The mean is defined as

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} kP(X = k) = \sum_{k=0}^{\infty} k \cdot \frac{\lambda^k e^{-\lambda}}{k!}.$$

For $k = 0$, the term $k \cdot \frac{\lambda^k e^{-\lambda}}{k!} = 0$, so the sum starts from $k = 1$, that is,

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} k \cdot \frac{\lambda^k e^{-\lambda}}{k!}.$$

Since $k! = k \cdot (k-1)!$ for $k \geq 1$, we can write

$$k \cdot \frac{\lambda^k}{k!} = \frac{\lambda^k}{(k-1)!},$$

so

$$\mathbb{E}[X] = e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!}.$$

Change the index by letting $j = k - 1$, so when $k = 1$, $j = 0$, and factoring out λ, we have

$$\sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} = \sum_{j=0}^{\infty} \frac{\lambda^{j+1}}{j!} = \lambda \sum_{j=0}^{\infty} \frac{\lambda^j}{j!}$$

Recall $e^{\lambda} = \sum_{j=0}^{\infty} \lambda^j / j!$, therefore

$$\mathbb{E}[X] = e^{-\lambda} \cdot \lambda e^{\lambda} = \lambda.$$

Variance Var(X)

The variance is given by $\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$. We already have $\mathbb{E}[X] = \lambda$, so we need to compute $\mathbb{E}[X^2]$.

We have $k^2 = k(k-1) + k$, so using linearity of expectation

$$\mathbb{E}[X^2] = \mathbb{E}[X(X-1) + X] = \mathbb{E}[X(X-1)] + \mathbb{E}[X].$$

We know $\mathbb{E}[X] = \lambda$, so we need to copute $\mathbb{E}[X(X-1)]$,

$$\mathbb{E}[X(X-1)] = \sum_{k=0}^{\infty} k(k-1)P(X=k) = \sum_{k=2}^{\infty} k(k-1) \cdot \frac{\lambda^k e^{-\lambda}}{k!},$$

since $k(k-1) = 0$ for $k = 0, 1$.

Note that $k(k-1) \cdot \frac{\lambda^k}{k!} = \frac{\lambda^k}{(k-2)!}$ for $k \geq 2$, so

$$k(k-1) \cdot \frac{\lambda^k}{k!} = \frac{k(k-1)\lambda^k}{k \cdot (k-1) \cdot (k-2)!} = \frac{\lambda^k}{(k-2)!},$$

consequently

$$\mathbb{E}[X(X-1)] = e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^k}{(k-2)!}.$$

Let $j = k-2$, so when $k = 2$, $j = 0$, and

$$\sum_{k=2}^{\infty} \frac{\lambda^k}{(k-2)!} = \sum_{j=0}^{\infty} \frac{\lambda^{j+2}}{j!} = \lambda^2 \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = \lambda^2 e^{\lambda},$$

thus

$$\mathbb{E}[X(X-1)] = e^{-\lambda} \cdot \lambda^2 e^{\lambda} = \lambda^2.$$

Now,

$$\mathbb{E}[X^2] = \mathbb{E}[X(X-1)] + \mathbb{E}[X] = \lambda^2 + \lambda.$$

The variance is

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = (\lambda^2 + \lambda) - \lambda^2 = \lambda.$$

□

Example 3. For $\lambda = 3$, $\mathbb{E}[X] = 3$, $\mathrm{Var}(X) = 3$, $\sigma \approx 1.732$, $M_X(t) = e^{3(e^t - 1)}$.

Geometric Distribution

Motivation: The Geometric distribution is appropriate for modeling the number of trials needed to achieve the first success in a sequence of independent trials, such as the number of times you need to roll a die to get a six or the number of sales calls before the first sale, assuming a constant success probability.

Remark 4. The geometric series sum is

$$\sum_{k=0}^{\infty} q^k = \frac{1}{1-q}, \quad |q| < 1.$$

Definition 4 (Geometric Distribution)**.** A RV X is said to follow geometric distribution with parameter $p > 0$, denoted by $X \sim \textit{Geometric}(p)$, if the PMF of X is given by

$$\mathbb{P}(X = k) = (1-p)^{k-1}p, \quad k = 1, 2, \ldots.$$

Properties:

- Support: $\{1, 2, \ldots\}$

- Mean: $\mathbb{E}[X] = \frac{1}{p}$ (using $\sum k(1-p)^{k-1}p = \frac{1}{p^2}$),

- Variance: $\text{Var}(X) = \frac{1-p}{p^2}$,

- Standard Deviation: $\sigma = \frac{\sqrt{1-p}}{p}$.

Example 4. For $p = 0.2$, $\mathbb{E}[X] = 5$, $\text{Var}(X) = \frac{0.8}{0.04} = 20$, $\sigma \approx 4.472$.

Theorem 3. *If $X \sim Geometric(p)$ with success probability p (where $0 < p \leq 1$), and X is the number of trials until the first success (support starting at 1), then*

$$\mathbb{E}[X] = \frac{1}{p}, \quad Var(X) = \frac{1-p}{p^2}.$$

Proof. The PMF is

$$\mathbb{P}(X = k) = (1-p)^{k-1}p, \quad k = 1, 2, 3, \ldots.$$

Mean $\mathbb{E}[X]$

The expected value is

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} k(1-p)^{k-1}p = p\sum_{k=1}^{\infty} k(1-p)^{k-1}.$$

Denote $q = 1-p$, then the sum $\sum_{k=1}^{\infty} kq^{k-1}$ is the derivative of the geometric series. Recall that the geometric series sum is

$$\sum_{k=0}^{\infty} q^k = \frac{1}{1-q}, \quad |q| < 1.$$

Differentiate both sides with respect to q, we have

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

Substitute $q = 1 - p$, we get

$$\sum_{k=1}^{\infty} k(1-p)^{k-1} = \frac{1}{p^2},$$

therefore

$$\mathbb{E}[X] = p \cdot \frac{1}{p^2} = \frac{1}{p}.$$

Variance Var(X)

The variance is $\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$. We have $\mathbb{E}[X] = 1/p$, so we need $\mathbb{E}[X^2]$.

First, compute $\mathbb{E}[X^2]$,

$$\mathbb{E}[X^2] = \sum_{k=1}^{\infty} k^2(1-p)^{k-1}p = p\sum_{k=1}^{\infty} k^2 q^{k-1}, \text{ where } q = 1-p.$$

The sum $\sum_{k=1}^{\infty} k^2 q^{k-1}$ can be found by differentiating the geometric series twice. To derive it, start from $\sum_{k=1}^{\infty} kq^{k-1} = 1/(1-q)^2$, differentiate again, we have

$$\sum_{k=2}^{\infty} k(k-1)q^{k-2} = \frac{2}{(1-q)^3} \quad \Rightarrow \quad \sum k(k-1)q^{k-1} = \frac{2q}{(1-q)^3}.$$

Next,

$$\begin{aligned}
\sum k^2 q^{k-1} &= \sum [k(k-1)+k]q^{k-1} = \sum k(k-1)q^{k-1} + \sum kq^{k-1} = \frac{2q}{(1-q)^3} + \frac{1}{(1-q)^2} \\
&= \frac{2q+1-q}{(1-q)^3} = \frac{q+1}{(1-q)^3} \\
\Rightarrow \sum_{k=1}^{\infty} k^2 q^{k-1} &= \sum k^2 q^{k-1} = \frac{1+q}{(1-q)^3}.
\end{aligned}$$

Substitute $q = 1 - p$,

$$\sum_{k=1}^{\infty} k^2(1-p)^{k-1} = \frac{1+(1-p)}{p^3} = \frac{2-p}{p^3},$$

so

$$\mathbb{E}[X^2] = p \cdot \frac{2-p}{p^3} = \frac{2-p}{p^2}.$$

Now,

$$\text{Var}(X) = \frac{2-p}{p^2} - \left(\frac{1}{p}\right)^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{2-p-1}{p^2} = \frac{1-p}{p^2}.$$

□

Negative Binomial Distribution

Motivation: The Negative Binomial distribution extends the Geometric case to count the number of failures before the r-th success, useful in scenarios like the number of defective items produced before the third non-defective item or the number of attempts before passing a test r times.

Definition 5 (Negative Binomial Distribution)**.** A RV X is said to follow Negative Binomial distribution with parameters $r \in \mathbb{N}$ and $p > 0$, denoted by $X \sim NegBin(r, p)$, if the PMF of X is given by

$$\mathbb{P}(X = k) = \binom{k+r-1}{k} p^r (1-p)^k, \quad k = 0, 1, 2, \ldots.$$

Properties:

- Support: $\{0, 1, 2, \ldots\}$

- Mean: $\mathbb{E}[X] = \frac{r(1-p)}{p}$,

- Variance: $\mathrm{Var}(X) = \frac{r(1-p)}{p^2}$,

- Standard Deviation: $\sigma = \frac{\sqrt{r(1-p)}}{p}$.

Example 5. For $r = 3$, $p = 0.4$, $\mathbb{E}[X] = \frac{3 \cdot 0.6}{0.4} = 4.5$, $\mathrm{Var}(X) = \frac{3 \cdot 0.6}{0.16} = 11.25$, $\sigma \approx 3.354$.

Hypergeometric Distribution

Motivation: The Hypergeometric distribution is relevant for sampling without replacement, such as determining the number of defective items in a sample from a finite batch, the number of aces drawn from a deck of cards, or the number of voters favoring a candidate in a small poll, where the population size and success count are fixed.

Definition 6 (Hypergeometric Distribution)**.** A RV X is said to follow hypergeometric distribution with parameters $M \in \mathbb{N}$, $N \in \mathbb{N}$ and $n \in \mathbb{N}$ denoted by $X \sim HG(M, N, n)$, if the PMF of X is given by

$$\mathbb{P}(X = k) = \frac{\binom{M}{k}\binom{N}{n-k}}{\binom{M+N}{n}}, \quad k = \max(0, n-N), \ldots, \min(n, M).$$

- Moments:

- Mean: $\mathbb{E}[X] = n \cdot \frac{M}{M+N}$,

- Variance: $\mathrm{Var}(X) = n \cdot \frac{M}{M+N} \cdot \frac{N}{M+N} \cdot \frac{M+N-n}{M+N-1}$,

- Standard Deviation: $\sigma = \sqrt{n \cdot \frac{M}{M+N} \cdot \frac{N}{M+N} \cdot \frac{M+N-n}{M+N-1}}$.

Example 6. For $N = 20$, $K = 7$, $n = 5$, $\mathbb{E}[X] = 5 \cdot \frac{7}{20} = 1.75$, $\mathrm{Var}(X) = 5 \cdot \frac{7}{20} \cdot \frac{13}{20} \cdot \frac{15}{19} \approx 0.898$, $\sigma \approx 0.948$.

Uniform Discrete Distribution

The uniform discrete distribution models outcomes that are equally likely, such as rolling a fair die.

Definition 7 (Uniform Discrete Distribution)**.** A random variable X is said to follow discrete uniform distribution over $\Omega = \{x_1, x_2, \dots, x_m\}$, if its probability mass function (PMF) is

$$\mathbb{P}(X = x_i) = \begin{cases} \dfrac{1}{m}, & \text{if} \quad i = 1, 2, \dots, m \\ 0, & \text{otherwise.} \end{cases}$$

Theorem 4 (Mean of Uniform Discrete Random Variable)**.** *The expected value (mean) of X is*

$$\mathbb{E}[X] = \frac{1}{m} \sum_{i=1}^{m} x_i.$$

Proof. By definition, $\mathbb{E}[X] = \sum_{i=1}^{m} x_i \mathbb{P}(X = x_i) = \sum_{i=1}^{m} x_i \cdot \frac{1}{m} = \frac{1}{m} \sum_{i=1}^{m} x_i$. □

Theorem 5 (Variance of Uniform Discrete Random Variable)**.** *The variance of X is*

$$Var(X) = \frac{1}{m} \sum_{i=1}^{m} x_i^2 - \left(\frac{1}{m} \sum_{i=1}^{m} x_i \right)^2.$$

Proof. The variance is $\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$. First, compute the second moment:

$$\mathbb{E}[X^2] = \sum_{i=1}^{m} x_i^2 \mathbb{P}(X = x_i) = \sum_{i=1}^{m} x_i^2 \cdot \frac{1}{m} = \frac{1}{m} \sum_{i=1}^{m} x_i^2.$$

Thus,

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \big(\mathbb{E}[X]\big)^2 = \frac{1}{m} \sum_{i=1}^{m} x_i^2 - \left(\frac{1}{m} \sum_{i=1}^{m} x_i \right)^2.$$

□

Theorem 6. *If X is a uniform discrete random variable over $\Omega = \{1, 2, \dots, m\}$, then*

1. Mean: $\mathbb{E}[X] = (1 + m)/2$,
2. Variance: $\mathrm{Var}(X) = (m^2 - 1)/12$.

Proof. We have X is uniform discrete RV over $\Omega = \{1, 2, \dots, m\}$, so

$$\mathbb{E}[X] = \frac{1}{m} \sum_{i=1}^{m} i = \frac{1}{m} \cdot \frac{m(m+1)}{2} = \frac{m+1}{2},$$

and

$$\mathbb{E}[X^2] = \frac{1}{m} \sum_{i=1}^{m} i^2 = \frac{1}{m} \cdot \frac{m(m+1)(2m+1)}{6} = \frac{(m+1)(2m+1)}{6}.$$

Thus,

$$\begin{aligned}\mathrm{Var}(X) &= \frac{(m+1)(2m+1)}{6} - \frac{(m+1)^2}{4} = \frac{2(m+1)(2m+1) - 3(m+1)^2}{12} \\ &= \frac{(m+1)[2(2m+1) - 3(m+1)]}{12} = \frac{(m+1)[4m+2-3m-3]}{12} \\ &= \frac{(m+1)(m-1)}{12} = \frac{m^2-1}{12}.\end{aligned}$$

□

Summary

Distribution	Notation	PMF $P(X=k)$	Mean	Variance	MGF $^*M_X(t)$
Bernoulli	$X \sim \mathrm{Bernoulli}(p)$	$\begin{cases} p, & k=1 \\ 1-p, & k=0 \end{cases}$	p	$p(1-p)$	$(1-p)+pe^t$
Binomial	$X \sim \mathrm{Binomial}(n,p)$	$\binom{n}{k}p^k(1-p)^{n-k},\ k=0,\dots,n$	np	$np(1-p)$	$\big((1-p)+pe^t\big)^n$
Poisson	$X \sim \mathrm{Poisson}(\lambda)$	$\frac{e^{-\lambda}\lambda^k}{k!},\ k=0,1,\dots$	λ	λ	$\exp\big(\lambda(e^t-1)\big)$
Geometric	$X \sim \mathrm{Geometric}(p)$	$(1-p)^{k-1}p,\ k=1,2,\dots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{pe^t}{1-(1-p)e^t},\ t<-\ln(1-p)$
Negative Binomial	$X \sim \mathrm{NegBin}(r,p)$	$\binom{k+r-1}{k}p^r(1-p)^k,\quad k=0,1,2,\dots$	$\frac{r(1-p)}{p}$	$\frac{r(1-p)}{p^2}$	$\left(\frac{p}{1-(1-p)e^t}\right)^r,\ t<-\ln(1-p).$

* MGF will be discussed later.

Remark 5. Geometric distribution is also defined with **PMF:**

$$f(x) = \begin{cases} p(1-p)^x, & \text{if } x \in \{0,1,2,\dots\} \\ 0, & \text{otherwise} \end{cases}$$

Here parameter p is the same, i.e., probability of success on each trial ($0 < p < 1$), and **Expectation:** $\mathbb{E}[X] = \frac{1-p}{p}$, **Variance:** $Var(X) = \frac{1-p}{p^2}$ for $t < -\ln(1-p)$. Similarly, negative binomial distribution can be defined with a different PMF.

References

[1] Blitzstein, J. K., & Hwang, J. (2019). *Introduction to probability.* Chapman and Hall/CRC.

Disclaimer