Derivations for Common Discrete Distributions

Xiankang Wang

November 7, 2025

Contents

1 Introduction and Setup

This document provides a detailed, step-by-step derivation of the Expectation, Variance, and Moment Generating Function (MGF) for several common discrete distributions. For each distribution, we will derive these properties using the MGF and also directly from the Probability Mass Function (PMF).

Let $(Z_t)_{t\geq 1}$ be a sequence of independent and identically distributed (i.i.d.) Bernoulli random variables where:

$$\mathbb{P}(Z_t = 1) = p \in (0, 1), \qquad \mathbb{P}(Z_t = 0) = q := 1 - p.$$

We refer to $Z_t = 1$ as a "success" and $Z_t = 0$ as a "failure".

2 The Bernoulli Distribution

The Bernoulli distribution is the fundamental building block for many other discrete distributions. It models a single trial with two possible outcomes.

Distribution 1 (Bernoulli)**.** *Let X be a Bernoulli random variable.*

- ***Support:*** $k \in \{0, 1\}$
- ***Probability Mass Function (PMF):***

$$\mathbb{P}(X = k) = p^k (1 - p)^{1-k}$$

Where $\mathbb{P}(X = 1) = p$ (success) and $\mathbb{P}(X = 0) = 1 - p = q$ (failure).

2.1 Derivations from MGF

2.1.1 Moment Generating Function (MGF)

The MGF is defined as $M_X(t) = \mathbb{E}[e^{tX}]$.

$$\begin{aligned} M_X(t) &= \sum_{k=0}^{1} e^{tk} \mathbb{P}(X = k) \\ &= e^{t(0)} \mathbb{P}(X = 0) + e^{t(1)} \mathbb{P}(X = 1) \\ &= 1 \cdot (1 - p) + e^t \cdot p \end{aligned}$$

$$\boxed{M_X(t) = q + pe^t}$$

2.1.2 Expectation

The expectation is the first derivative of the MGF evaluated at $t = 0$.

$$\begin{aligned} M'_X(t) &= \frac{d}{dt}(q + pe^t) = pe^t \\ \mathbb{E}[X] &= M'_X(0) = pe^0 = p \end{aligned}$$

$$\boxed{\mathbb{E}[X] = p}$$

2.1.3 Variance

We find the second moment $\mathbb{E}[X^2] = M''_X(0)$ first.

$$M''_X(t) = \frac{d}{dt}(pe^t) = pe^t$$
$$\mathbb{E}[X^2] = M''_X(0) = pe^0 = p$$

Now, we use the formula $\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.

$$\text{Var}(X) = p - p^2 = p(1-p)$$

$\text{Var}(X) = pq$

2.2 Derivations from PMF

2.2.1 Expectation

By definition, $\mathbb{E}[X] = \sum k \cdot \mathbb{P}(X = k)$.

$$\begin{aligned} \mathbb{E}[X] &= (0) \cdot \mathbb{P}(X = 0) + (1) \cdot \mathbb{P}(X = 1) \\ &= 0 \cdot q + 1 \cdot p = p \end{aligned}$$

2.2.2 Variance

First, we calculate the second moment, $\mathbb{E}[X^2] = \sum k^2 \cdot \mathbb{P}(X = k)$.

$$\begin{aligned} \mathbb{E}[X^2] &= (0^2) \cdot \mathbb{P}(X = 0) + (1^2) \cdot \mathbb{P}(X = 1) \\ &= 0 \cdot q + 1 \cdot p = p \end{aligned}$$

The variance is then:

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

3 The Binomial Distribution

The Binomial distribution models the number of successes in a fixed number of independent Bernoulli trials. It can be viewed as the sum of n i.i.d. Bernoulli random variables.

Distribution 2 (Binomial)**.** *Let $X \sim Binomial(n,p)$.*

- ***Support:*** $k \in \{0, 1, 2, \ldots, n\}$
- ***PMF:*** $\mathbb{P}(X = k) = \binom{n}{k} p^k q^{n-k}$
- ***Sum Representation:*** *Let $Z_1, \ldots, Z_n$ be i.i.d. Bernoulli(p) random variables. Then $X = \sum_{i=1}^{n} Z_i$.*

3.1 Derivations from MGF

3.1.1 Moment Generating Function (MGF)

Using the sum representation is most elegant.

$$\begin{aligned} M_X(t) &= M_{\sum Z_i}(t) = \mathbb{E}[e^{t(Z_1 + \cdots + Z_n)}] \\ &= \prod_{i=1}^{n} \mathbb{E}[e^{tZ_i}] \quad \text{(due to independence)} \\ &= (M_Z(t))^n = (q + pe^t)^n \end{aligned}$$

$$M_X(t) = (q + pe^t)^n$$

3.1.2 Expectation

$$\begin{aligned} M_X'(t) &= \frac{d}{dt}(q + pe^t)^n = n(q + pe^t)^{n-1} \cdot (pe^t) \\ \mathbb{E}[X] &= M_X'(0) = n(q + p)^{n-1} \cdot p = n(1)^{n-1} p \end{aligned}$$

$$\mathbb{E}[X] = np$$

3.1.3 Variance

$$\begin{aligned} M_X''(t) &= \frac{d}{dt}\left(n(pe^t)(q + pe^t)^{n-1}\right) \\ &= n(pe^t)(q + pe^t)^{n-1} + n(pe^t)(n-1)(q + pe^t)^{n-2}(pe^t) \quad \text{(Product Rule)} \\ \mathbb{E}[X^2] &= M_X''(0) = np(q + p)^{n-1} + n(n-1)p^2(q + p)^{n-2} \\ &= np(1)^{n-1} + n(n-1)p^2(1)^{n-2} = np + n(n-1)p^2 \end{aligned}$$

Then, the variance is:

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

$$\text{Var}(X) = npq$$

3.2 Derivations from PMF

3.2.1 Expectation

We use the identity $k\binom{n}{k} = n\binom{n-1}{k-1}$.

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{k=0}^{n} k\binom{n}{k}p^k q^{n-k} = \sum_{k=1}^{n} k\binom{n}{k}p^k q^{n-k} \\
&= \sum_{k=1}^{n} n\binom{n-1}{k-1}p^k q^{n-k} \\
&= np\sum_{k=1}^{n} \binom{n-1}{k-1}p^{k-1} q^{n-k} \quad \text{(Let } j = k-1\text{)} \\
&= np\sum_{j=0}^{n-1} \binom{n-1}{j}p^{j} q^{(n-1)-j} \\
&= np(p+q)^{n-1} \quad \text{(Binomial Theorem)} \\
&= np(1)^{n-1} = np
\end{aligned}$$

3.2.2 Variance

We first calculate the factorial moment $\mathbb{E}[X(X-1)]$.

$$\begin{aligned}
\mathbb{E}[X(X-1)] &= \sum_{k=0}^{n} k(k-1)\binom{n}{k}p^k q^{n-k} = \sum_{k=2}^{n} k(k-1)\binom{n}{k}p^k q^{n-k} \\
&= \sum_{k=2}^{n} k(k-1)\frac{n!}{k!(n-k)!}p^k q^{n-k} = \sum_{k=2}^{n} \frac{n!}{(k-2)!(n-k)!}p^k q^{n-k} \\
&= n(n-1)\sum_{k=2}^{n} \frac{(n-2)!}{(k-2)!(n-k)!}p^k q^{n-k} \\
&= n(n-1)p^2\sum_{k=2}^{n} \binom{n-2}{k-2}p^{k-2} q^{n-k} \quad \text{(Let } j = k-2\text{)} \\
&= n(n-1)p^2\sum_{j=0}^{n-2} \binom{n-2}{j}p^{j} q^{(n-2)-j} = n(n-1)p^2(p+q)^{n-2} = n(n-1)p^2
\end{aligned}$$

The second moment is $\mathbb{E}[X^2] = \mathbb{E}[X(X-1)] + \mathbb{E}[X] = n(n-1)p^2 + np$. The variance is:

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

4 The Geometric Distribution

The Geometric distribution is the fundamental building block for the Negative Binomial distribution. It models the waiting time for the *first* success. It is the only discrete distribution that possesses the memoryless property.

4.1 Definition 1: Number of Trials

Let X be the random variable representing the **total number of trials required to obtain the first success**.

- **Support**: $k \in \{1, 2, 3, \dots\}$
- **Probability Mass Function (PMF)**: For the k-th trial to be the first success, we must have $k-1$ failures followed by one success.

$$\mathbb{P}(X = k) = q^{k-1} p$$

4.1.1 The Memoryless Property

The memoryless property states that the probability of waiting for additional time is independent of the time already spent waiting. Formally, for any positive integers m and n:

$$\mathbb{P}(X > m + n \mid X > m) = \mathbb{P}(X > n)$$

Intuitively, this means that given we have already observed m failures, the probability of observing at least n more failures is the same as if we were starting from the beginning. The process "forgets" its history.

Proof: To prove this, we first need a formula for the survival function, $\mathbb{P}(X > k)$. The event $\{X > k\}$ means that the first k trials were all failures.

$$\mathbb{P}(X > k) = \overbrace{q \cdot q \cdots q}^{k \text{ times}} = q^k$$

Now we use the definition of conditional probability:

$$\begin{aligned}
\mathbb{P}(X > m + n \mid X > m) &= \frac{\mathbb{P}(\{X > m + n\} \cap \{X > m\})}{\mathbb{P}(X > m)} \\
&= \frac{\mathbb{P}(X > m + n)}{\mathbb{P}(X > m)} \\
&= \frac{q^{m+n}}{q^m} \\
&= q^n \\
&= \mathbb{P}(X > n)
\end{aligned}$$

This completes the proof of the memoryless property.

4.1.2 Moment Generating Function (MGF)

The MGF is defined as $M_X(t) = \mathbb{E}[e^{tX}]$.

$$\begin{aligned}
M_X(t) &= \sum_{k=1}^{\infty} e^{tk} \mathbb{P}(X = k) = \sum_{k=1}^{\infty} e^{tk} q^{k-1} p \\
&= pe^t \sum_{k=1}^{\infty} e^{t(k-1)} q^{k-1} = pe^t \sum_{j=0}^{\infty} (qe^t)^j \quad (\text{let } j = k - 1)
\end{aligned}$$

This is a geometric series which converges for $|qe^t| < 1$.

$$\boxed{M_X(t) = \frac{pe^t}{1 - qe^t}}$$

4.1.3 Expectation

The expectation is the first derivative of the MGF evaluated at $t = 0$.

$$\begin{aligned}
M'_X(t) &= \frac{d}{dt}\left(\frac{pe^t}{1 - qe^t}\right) = \frac{(pe^t)(1 - qe^t) - (pe^t)(-qe^t)}{(1 - qe^t)^2} = \frac{pe^t}{(1 - qe^t)^2} \\
\mathbb{E}[X] = M'_X(0) &= \frac{pe^0}{(1 - qe^0)^2} = \frac{p}{(1 - q)^2} = \frac{p}{p^2}
\end{aligned}$$

$$\boxed{\mathbb{E}[X] = \frac{1}{p}}$$

4.1.4 Variance

We find the second moment $\mathbb{E}[X^2] = M''_X(0)$ first.

$$\begin{aligned}
M''_X(t) &= \frac{d}{dt}\left(\frac{pe^t}{(1 - qe^t)^2}\right) = \frac{pe^t(1 - qe^t)^2 - pe^t \cdot 2(1 - qe^t)(-qe^t)}{(1 - qe^t)^4} \\
&= \frac{pe^t(1 - qe^t) + 2pq(e^t)^2}{(1 - qe^t)^3} \\
\mathbb{E}[X^2] = M''_X(0) &= \frac{p(1 - q) + 2pq}{(1 - q)^3} = \frac{p^2 + 2pq}{p^3} = \frac{p + 2q}{p^2}
\end{aligned}$$

Now, we use the formula $\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.

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

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

4.2 Definition 2: Number of Failures

Let Y be the random variable representing the **number of failures before the first success**.

- **Support**: $k \in \{0, 1, 2, \dots\}$
- **Relationship**: $Y = X - 1$
- **Probability Mass Function (PMF)**: $\mathbb{P}(Y = k) = \mathbb{P}(X = k + 1) = q^{(k+1)-1}p = q^k p$

We can derive its properties using the relationship $Y = X - 1$.

4.2.1 Expectation, Variance, and MGF via Relationship

- **Expectation**: $\mathbb{E}[Y] = \mathbb{E}[X-1] = \mathbb{E}[X] - 1 = \frac{1}{p} - 1 = \frac{1-p}{p}$

$$\boxed{\mathbb{E}[Y] = \frac{q}{p}}$$

- **Variance**: $\mathrm{Var}(Y) = \mathrm{Var}(X-1) = \mathrm{Var}(X)$

$$\boxed{\mathrm{Var}(Y) = \frac{q}{p^2}}$$

- **MGF**: $M_Y(t) = \mathbb{E}[e^{tY}] = \mathbb{E}[e^{t(X-1)}] = e^{-t}\mathbb{E}[e^{tX}] = e^{-t}M_X(t)$

$$M_Y(t) = e^{-t}\frac{pe^t}{1-qe^t} = \frac{p}{1-qe^t}$$

$$\boxed{M_Y(t) = \frac{p}{1-qe^t}}$$

5 The Negative Binomial Distribution

This distribution generalizes the Geometric, modeling the waiting time for the r-*th* success. It is best understood as the sum of r independent Geometric random variables.

5.1 Definition 1: Number of Failures before r Successes

Let X_r be the **number of failures before the r-th success**.

Distribution 3 (Negative Binomial, Failures).

- ***Support:*** $k \in \{0, 1, 2, \ldots\}$
- ***PMF:*** $\mathbb{P}(X_r = k) = \binom{k+r-1}{k} p^r q^k$
- ***Sum Representation:*** *Let Y_i be i.i.d. Geometric (failures) random variables. Then $X_r = \sum_{i=1}^{r} Y_i$.*

Using the sum representation and properties of $Y \sim \text{Geometric}(p)$ (failures):

- **Expectation**: $\mathbb{E}[X_r] = \sum_{i=1}^{r} \mathbb{E}[Y_i] = r \cdot \mathbb{E}[Y] = r\frac{q}{p}$

$$\mathbb{E}[X_r] = \frac{rq}{p}$$

- **Variance**: $\text{Var}(X_r) = \sum_{i=1}^{r} \text{Var}(Y_i)$ (by independence)

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

- **MGF**: $M_{X_r}(t) = \prod_{i=1}^{r} M_{Y_i}(t) = (M_Y(t))^r$

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

5.2 Definition 2: Number of Trials for r Successes

Let Y_r be the **total number of trials to obtain r successes**.

Distribution 4 (Negative Binomial, Trials).

- ***Support:*** $k \in \{r, r+1, \ldots\}$
- ***PMF:*** $\mathbb{P}(Y_r = k) = \binom{k-1}{r-1} p^r q^{k-r}$
- ***Sum Representation:*** *Let X_i be i.i.d. Geometric (trials) random variables. Then $Y_r = \sum_{i=1}^{r} X_i$.*

Using the sum representation and properties of $X \sim \text{Geometric}(p)$ (trials):

- **Expectation**: $\mathbb{E}[Y_r] = \sum_{i=1}^{r} \mathbb{E}[X_i] = r \cdot \mathbb{E}[X] = r\frac{1}{p}$

$$\mathbb{E}[Y_r] = \frac{r}{p}$$

- **Variance**: $\text{Var}(Y_r) = \sum_{i=1}^{r} \text{Var}(X_i)$

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

- **MGF**: $M_{Y_r}(t) = (M_X(t))^r$

$$M_{Y_r}(t) = \left(\frac{pe^t}{1-qe^t}\right)^r$$

6 The Poisson Distribution

The Poisson distribution models the number of events in a fixed interval, given a constant average rate.

Distribution 5 (Poisson)**.** *Let* $X \sim Poisson(\lambda)$*.*

- ***Support:*** $k \in \{0, 1, 2, \dots\}$
- ***Parameter:*** $\lambda > 0$ *is the average rate.*
- ***PMF:*** $\mathbb{P}(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$

6.1 Derivations from MGF

6.1.1 Moment Generating Function (MGF)

Derived using the Taylor series for $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$.

$$\begin{aligned} M_X(t) = \mathbb{E}[e^{tX}] &= \sum_{k=0}^{\infty} e^{tk} \frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^t)^k}{k!} \\ &= e^{-\lambda} e^{\lambda e^t} \quad \text{(using the Taylor series for } e^{\lambda e^t}\text{)} \end{aligned}$$

$$M_X(t) = e^{\lambda(e^t - 1)}$$

6.1.2 Expectation

$$\begin{aligned} M'_X(t) &= \frac{d}{dt}\left(e^{\lambda(e^t-1)}\right) = e^{\lambda(e^t-1)} \cdot (\lambda e^t) \\ \mathbb{E}[X] &= M'_X(0) = e^{\lambda(e^0-1)} \cdot (\lambda e^0) = e^0 \cdot \lambda \end{aligned}$$

$$\mathbb{E}[X] = \lambda$$

6.1.3 Variance

$$\begin{aligned} M''_X(t) &= \frac{d}{dt}\left((\lambda e^t) e^{\lambda(e^t-1)}\right) \\ &= (\lambda e^t) e^{\lambda(e^t-1)} + (\lambda e^t) \cdot e^{\lambda(e^t-1)} \cdot (\lambda e^t) \quad \text{(Product Rule)} \\ \mathbb{E}[X^2] &= M''_X(0) = (\lambda)e^0 + (\lambda)^2 e^0 = \lambda + \lambda^2 \end{aligned}$$

Then, the variance is:

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

$$\text{Var}(X) = \lambda$$

6.2 Derivations from PMF

6.2.1 Expectation

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{k=0}^{\infty} k \frac{e^{-\lambda}\lambda^k}{k!} = \sum_{k=1}^{\infty} k \frac{e^{-\lambda}\lambda^k}{k!} \\
&= \sum_{k=1}^{\infty} \frac{e^{-\lambda}\lambda^k}{(k-1)!} \quad \text{(Let } j = k-1\text{)} \\
&= \lambda \sum_{j=0}^{\infty} \frac{e^{-\lambda}\lambda^j}{j!} \\
&= \lambda \cdot 1 \quad \text{(Sum of Poisson PMF is 1)} \\
&= \lambda
\end{aligned}$$

6.2.2 Variance

We first calculate the factorial moment $\mathbb{E}[X(X-1)]$.

$$\begin{aligned}
\mathbb{E}[X(X-1)] &= \sum_{k=0}^{\infty} k(k-1) \frac{e^{-\lambda}\lambda^k}{k!} = \sum_{k=2}^{\infty} k(k-1) \frac{e^{-\lambda}\lambda^k}{k!} \\
&= \sum_{k=2}^{\infty} \frac{e^{-\lambda}\lambda^k}{(k-2)!} \quad \text{(Let } j = k-2\text{)} \\
&= \lambda^2 \sum_{j=0}^{\infty} \frac{e^{-\lambda}\lambda^j}{j!} = \lambda^2 \cdot 1 = \lambda^2
\end{aligned}$$

The second moment is $\mathbb{E}[X^2] = \mathbb{E}[X(X-1)] + \mathbb{E}[X] = \lambda^2 + \lambda$. The variance is:

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

7 Relationship: Binomial to Poisson via MGFs

The Poisson distribution can be derived as a limiting case of the Binomial distribution when the number of trials n is very large and the probability of success p is very small. A more formal and direct way to show this is by taking the limit of the Moment Generating Function (MGF).

We consider a Binomial(n, p) distribution and let $n \to \infty$ and $p \to 0$ in such a way that the mean, $\lambda = np$, remains constant. This implies we should set $p = \lambda/n$. Let $X_n \sim$ Binomial$(n, p = \lambda/n)$.

Step 1: State the Binomial and Poisson MGFs. From previous sections, we know the MGFs for the Binomial and Poisson distributions:

- **Binomial MGF**: $M_{X_n}(t) = (q + pe^t)^n = ((1-p) + pe^t)^n$
- **Poisson MGF**: $M_Y(t) = e^{\lambda(e^t-1)}$ for $Y \sim$ Poisson(λ)

Step 2: Substitute $p = \lambda/n$ into the Binomial MGF. Replacing p with λ/n in the Binomial MGF gives:

$$\begin{aligned} M_{X_n}(t) &= \left(\left(1 - \frac{\lambda}{n}\right) + \frac{\lambda}{n}e^t\right)^n \\ &= \left(1 + \frac{\lambda}{n}(e^t - 1)\right)^n \end{aligned}$$

Step 3: Take the limit as $n \to \infty$. Now, we take the limit of this expression as n approaches infinity. We use the well-known limit identity from calculus, which states that for any real number x:

$$\lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n = e^x$$

In our case, the term x is $\lambda(e^t - 1)$. Applying this identity:

$$\lim_{n\to\infty} M_{X_n}(t) = \lim_{n\to\infty} \left(1 + \frac{\lambda(e^t - 1)}{n}\right)^n = e^{\lambda(e^t-1)}$$

Conclusion The resulting expression, $e^{\lambda(e^t-1)}$, is exactly the MGF of a Poisson distribution with parameter λ. Since the MGF uniquely determines the distribution, this shows that the Binomial distribution converges to the Poisson distribution under these limiting conditions.

This result justifies using the Poisson distribution as a practical approximation for the Binomial distribution when n is large and p is small. A common rule of thumb is that the approximation is good if $n \geq 100$ and $np \leq 10$.

8 Summary Table

Table 1: Comparison of Properties for Common Discrete Distributions

Distribution	**PMF** $P(X=k)$	**Support** $\mathcal{X}$	**Expectation** $\mathbb{E}[X]$	**Variance** $\mathrm{Var}(X)$	**MGF** $M_X(t)$
Bernoulli	$p^k q^{1-k}$	$k \in \{0,1\}$	p	pq	$q + pe^t$
Binomial	$\binom{n}{k} p^k q^{n-k}$	$k \in \{0,\dots,n\}$	np	npq	$(q+pe^t)^n$
Poisson	$\frac{e^{-\lambda}\lambda^k}{k!}$	$k \in \{0,1,\dots\}$	λ	λ	$e^{\lambda(e^t-1)}$
Geometric					
Trials for 1st success	$q^{k-1}p$	$k \in \{1,2,\dots\}$	$\frac{1}{p}$	$\frac{q}{p^2}$	$\frac{pe^t}{1-qe^t}$
Failures before 1st success	$q^k p$	$k \in \{0,1,\dots\}$	$\frac{q}{p}$	$\frac{q}{p^2}$	$\frac{p}{1-qe^t}$
Negative Binomial					
Trials for r-th success	$\binom{k-1}{r-1} p^r q^{k-r}$	$k \in \{r, r+1, \dots\}$	$\frac{r}{p}$	$\frac{rq}{p^2}$	$\left(\frac{pe^t}{1-qe^t}\right)^r$
Failures before r-th success	$\binom{k+r-1}{k} p^r q^k$	$k \in \{0,1,\dots\}$	$\frac{rq}{p}$	$\frac{rq}{p^2}$	$\left(\frac{p}{1-qe^t}\right)^r$

A Derivation of the Negative Binomial Series

The key to deriving the MGF for the negative binomial distribution directly from its PMF is the following series identity, often called the negative binomial series. For any real number $r > 0$ and $|x| < 1$:

$$\sum_{k=0}^{\infty} \binom{k+r-1}{k} x^k = (1-x)^{-r}$$

We prove this using a Maclaurin series expansion for the function $f(x) = (1-x)^{-r}$. The Maclaurin series is given by $f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} x^k$.

Step 1: Compute derivatives of $f(x)$.

$$\begin{aligned} f(x) &= (1-x)^{-r} \\ f'(x) &= (-r)(1-x)^{-r-1}(-1) = r(1-x)^{-r-1} \\ f''(x) &= r(-r-1)(1-x)^{-r-2}(-1) = r(r+1)(1-x)^{-r-2} \\ f^{(k)}(x) &= r(r+1)(r+2)\cdots(r+k-1)(1-x)^{-r-k} \end{aligned}$$

Step 2: Evaluate derivatives at $x = 0$.

$$f^{(k)}(0) = r(r+1)(r+2)\cdots(r+k-1)$$

Step 3: Determine the Maclaurin series coefficients. The coefficients are $\frac{f^{(k)}(0)}{k!} = \frac{r(r+1)\cdots(r+k-1)}{k!}$. This expression is exactly the definition of the binomial coefficient $\binom{r+k-1}{k}$.

Step 4: Assemble the Maclaurin series. Substituting the coefficients back into the series formula, we get:

$$f(x) = (1-x)^{-r} = \sum_{k=0}^{\infty} \binom{k+r-1}{k} x^k$$

This completes the proof. This identity is a special case of the generalized binomial theorem.