1 Moment Generating Functions

So far we have seen that the distribution of a random variable can always be encoded by the CDF. If we know that the random variable is discrete / continuous, then its distribution can alternatively be encoded by the PMF / PDF. We now show that it is also possible to encode the distribution through another object called the moment generating function provided that the moments of the random variable do not grow too fast.

Definition 1 (Moment Generating Function)**.** The *moment generating function* (MGF) of a random variable X is given by the function

$$M_X(t) = \mathbb{E}[e^{tX}],$$

provided the expression exists in a neighbourhood of zero, say for $t \in (-a, a)$.

Remark 1. The MGF might not always exist. For example, if the a moments of $\mathbb{E}[X^n]$ of X is infinite, then the MGF will be divergent near zero. In higher level courses, you will encounter characteristic functions or Fourier transforms which will always exist.

1.1 Properties

If X has finite support on $\{x_1, \dots, x_n\}$ with PMF p_X, then

$$M_X(t) = p_X(x_1)e^{tx_1} + \dots + p_X(x_n)e^{tx_n}$$

so it is easy to read off the PMF given the MGF. We see that even for more complicated distributions the MGF still encodes the distribution of the random variables.

Theorem 1 (*Uniqueness Theorem*)

If X and Y have MGFs $M_X(t)$ and $M_Y(t)$ defined in neighbourhoods of the origin, and $M_X(t) = M_Y(t)$ for all t where they are defined, then

$$X \overset{d}{=} Y.$$

The MGF is also computationally convenient since it can be used to recover the moments and compute the distribution of the sums of independent random variables easily.

1. **Moments:** The MGF encodes all the moments of X. Assuming that $M_X(t)$ is defined in a neighbourhood of $t = 0$,

   $$\frac{d^k}{dt^k} M_X(0) = \mathbb{E}[X^k] \text{ for all } k \geq 0.$$

   This follows from Taylor's theorem and linearity holds even with infinite sums provided that $M_X(t)$ is defined in a neighbourhood of $t = 0$

   $$M_X(t) = \mathbb{E}\left[\sum_{j=0}^{\infty} \frac{t^j X^j}{j!}\right] = \sum_{j=0}^{\infty} \frac{t^j\, \mathbb{E}[X^j]}{j!}.$$

   Formally taking the kth derivative at 0 gives us the formula (see Problem 1.9 for a careful proof).

2. **Independent Sums:** Suppose that X and Y are independent and have moment generating functions $M_X(t)$ and $M_Y(t)$ respectively. Then the moment generating function of $X + Y$ is

   $$M_{X+Y}(t) = \mathbb{E}\left(e^{t(X+Y)}\right) = \mathbb{E}\left(e^{tX}\right)\mathbb{E}\left(e^{tY}\right) = M_X(t)M_Y(t).$$

   This is a useful property because the distribution of the sum of random variables can now be easily computed (which required computing iterated sums).

1.2 Moment Generating Functions of Common Distributions

1. **Discrete Uniform:** If $X \sim \mathrm{DUnif}[a,b]$ (discrete uniform) then

$$M_X(t) = \frac{1}{b-a+1}\sum_{x=a}^{b} e^{tx} \text{ for } t \in \mathbb{R}$$

2. **Binomial:** If $X \sim \mathrm{Bin}(n,p)$ then

$$M_X(t) = (pe^t + (1-p))^n \text{ for } t \in \mathbb{R}$$

3. **Negative Binomial:** If $X \sim \mathrm{NegBin}(k,p)$ then

$$M_X(t) = \left(\frac{p}{1-(1-p)e^t}\right)^k \text{ for } t \leq -\ln(1-p)$$

4. **Poisson:** If $X \sim \mathrm{Poi}(\lambda)$ then

$$M_X(t) = e^{\lambda(e^t-1)} \text{ for } t \in \mathbb{R}$$

5. **Continuous Uniform:** If $X \sim \mathrm{Unif}[a,b]$ (continuous uniform) then

$$M_X(t) = \begin{cases} \frac{e^{bt}-e^{at}}{(b-a)t} & t \neq 0 \\ 1 & t = 0 \end{cases}$$

6. **Exponential:** If $X \sim \mathrm{Exp}(\theta)$ (waiting time parametrization) then

$$M_X(t) = \frac{1}{1-\theta t} \text{ for } t < \frac{1}{\theta}$$

7. **Normally Distributed:** If $X \sim \mathrm{N}(\mu,\sigma^2)$, then

$$M_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}} \text{ for } t \in \mathbb{R}$$

1.3 Example Problems

Problem 1.1. Suppose that X has MGF

$$M_X(t) = 0.3 + 0.2e^t + 0.5e^{2t}$$

What is the PMF of X?

Solution 1.1. We have

$$M_X(t) = 0.3 + 0.2e^t + 0.5e^{2t} = 0.3e^{0\cdot t} + 0.2e^{1\cdot t} + 0.5e^{2\cdot t}.$$

Then reverse engineering the PMF from the MGF implies that

$$p_X(0) = \mathbb{P}(X=0) = 0.3, \qquad p_X(1) = \mathbb{P}(X=1) = 0.2 \qquad p_X(2) = \mathbb{P}(X=2) = 0.5.$$

It is clear that at least for finitely supported PMF, the MGF is simply another way of writing encoding the distribution.

Problem 1.2. Use the MGFs to show that if $X \sim \mathrm{Poi}(\lambda)$, then $\mathbb{E}[X] = \lambda$ and $\mathrm{Var}(X) = \lambda$

Solution 1.2. The MGF is
$$M_X(t) = e^{\lambda(e^t-1)}.$$
Taking derivatives, we have
$$M_X'(t) = e^{\lambda(e^t-1)}\lambda e^t \Rightarrow \mathbb{E}(X) = M_X'(0) = \lambda.$$
To compute the variance, we take the second derivative
$$M_X''(t) = e^{\lambda(e^t-1)}\lambda e^t + e^{\lambda(e^t-1)}(\lambda e^t)^2 \Rightarrow \mathbb{E}(X^2) = M_X''(0) = \lambda^2 + \lambda$$
and conclude
$$\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Problem 1.3. Use MGFs to show that if $X \sim \mathrm{Poi}(\lambda)$ and $Y \sim \mathrm{Poi}(\mu)$ are independent, then $X+Y \sim \mathrm{Poi}(\lambda+\mu)$.

Solution 1.3. We know that $M_X(t) = e^{\lambda(e^t-1)}$ and $M_Y(t) = e^{\mu(e^t-1)}$. Since X and Y are independent, the MGF of $X+Y$ is
$$M_{X+Y}(t) = M_X(t)M_Y(t) = e^{\lambda(e^t-1}e^{\mu(e^t-1)} = e^{(\lambda+\mu)(e^t-1)}$$
which we recognize as the MGF of a $\mathrm{Poi}(\lambda+\mu)$ random variables. By uniqueness of the MGF, we conclude $X+Y \sim \mathrm{Poi}(\lambda+\mu)$.

Problem 1.4. Use MGFs to show that if $X \sim \mathrm{N}(\mu_1, \sigma_1^2)$ and $Y \sim \mathrm{N}(\mu_2, \sigma_2^2)$ are independent, then $X+Y \sim \mathrm{N}(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$.

Solution 1.4. We know that $M_X(t) = e^{\mu_1 t + \frac{\sigma_1^2 t^2}{2}}$ and $M_Y(t) = e^{\mu_2 t + \frac{\sigma_2^2 t^2}{2}}$. Since X and Y are independent, the MGF of $X+Y$ is
$$M_{X+Y}(t) = M_X(t)M_Y(t) = e^{\mu_1 t + \frac{\sigma_1^2 t^2}{2}} e^{\mu_2 t + \frac{\sigma_2^2 t^2}{2}} = e^{(\mu_1+\mu_2)t + \frac{(\sigma_1^2\sigma_2^2)t^2}{2}}$$
which we recognize as the MGF of a $\mathrm{N}(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$ random variables. By uniqueness of the MGF, we conclude $X+Y \sim \mathrm{N}(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$.

Problem 1.5. Use MGFs to show that for large n and small p such that $\mu = np$, we can approximate the $\mathrm{Bin}(n,p)$ distribution with a $\mathrm{Poi}(\mu)$ distribution.

Solution 1.5. By the uniqueness theorem, we can prove this by showing that the MGF of $X \sim \mathrm{Bin}(n,p)$ converges to the MGF of $Y \sim \mathrm{Poi}(\mu)$ as $n \to \infty$ where $\mu = np \Leftrightarrow p = \frac{\mu}{n}$. Indeed,
$$M_X(t) = (pe^t + 1 - p)^n = (1 + p(e^t-1))^n = \Big(1 + \frac{\mu}{n}(e^t-1)\Big)^n.$$
Since for every x
$$\lim_{n\to\infty}\Big(1+\frac{x}{n}\Big)^n = e^x$$
we have
$$\lim_{n\to\infty} M_X(t) = \lim_{n\to\infty}\Big(1+\frac{\mu}{n}(e^t-1)\Big)^n = e^{\mu(e^t-1)} = M_Y(t),$$

Problem 1.6. Let $Z \sim \mathrm{N}(0,1)$. What is the distribution of $X = \sigma Z + \mu$?

Solution 1.6. This result is an alternative proof of the de-standardization argument. If $Z \sim \mathrm{N}(0,1)$, then

$$M_Z(t) = e^{\frac{t^2}{2}}.$$

If we define $X = \sigma Z + \mu$, then the MGF of X satisfies

$$M_X(t) = M_{\sigma Z+\mu}(t) = \mathbb{E}[e^{t\sigma Z + t\mu}] = e^{\mu t} M_Z(\sigma t) = e^{\mu t} e^{\frac{\sigma^2}{2}t^2}$$

which we recognize as the MGF of $X \sim \mathrm{N}(0,1)$.

Problem 1.7. Let $X \sim \mathrm{Exp}(1)$. What is the distribution of $Y = \theta X$?

Solution 1.7. If $X \sim \mathrm{Exp}(1)$, then

$$M_X(t) = \frac{1}{1-t}.$$

If we define $Y = \theta X$, then the MGF of Y satisfies

$$M_Y(t) = M_{\theta X}(t) = \mathbb{E}[e^{t\theta X}] = M_X(\theta t) = \frac{1}{1-\theta t}$$

which we recognize as the MGF of $Y \sim \mathrm{Exp}(\theta)$.

1.4 Proofs of Key Results

Problem 1.8. Prove the following formulas for the MGFs of common distributions

1. **Poisson:** If $X \sim \mathrm{Poi}(\lambda)$ then
$$M_X(t) = e^{\lambda(e^t-1)} \text{ for } t \in \mathbb{R}$$
2. **Normal Distribution:** If $X \sim \mathrm{N}(\mu, \sigma^2)$, then
$$M_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}} \text{ for } t \in \mathbb{R}$$
3. **Negative Binomial:** If $X \sim \mathrm{NegBin}(k,p)$ then
$$M_X(t) = \left(\frac{p}{1-(1-p)e^t}\right)^k \text{ for } t \le -\ln(1-p)$$

Solution 1.8. The computation of all the MGFs use similar tricks we have seen before (summation formulas, reducing to sums of PMF/integral PDF, etc)

Poisson: The MGF is computed using the exponential sum

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

Normal Distribution: The MGF is computed by completing the square. We have

$$M_X(t) = \mathbb{E}(e^{tX}) = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} e^{tx}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} e^{-\frac{x^2-2(\mu+t\sigma^2)x+\mu^2}{2\sigma^2}}\,dx.$$

We want to rewrite the integral in terms of the integral of a PDF, so we complete the square

$$\begin{aligned} x^2 - 2(\mu + t\sigma^2)x + \mu^2 &= x^2 - 2(\mu + t\sigma^2)x + (\mu + t\sigma^2)^2 - (\mu + t\sigma^2)^2 + \mu^2 \\ &= (x - (\mu + t\sigma^2))^2 - 2t\sigma^2\mu - t^2\sigma^4 \end{aligned}$$

so

$$\begin{aligned} \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{x^2 - 2(\mu+t\sigma^2)x + \mu^2}{2\sigma^2}}\, dx &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{(x-(\mu+t\sigma^2))^2 - 2t\sigma^2\mu - t^2\sigma^4}{2\sigma^2}}\, dx \\ &= e^{\mu t + \frac{\sigma^2 t^2}{2}} \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{(x-(\mu+t\sigma^2))^2}{2\sigma^2}}\, dx}_{=1 \text{ sum of PDF of } \mathrm{N}(\mu + t\sigma^2, \sigma^2)}. \end{aligned}$$

Negative Binomial: We first consider the case when $Y \sim \mathrm{NegBin}(1, p) = \mathrm{Geo}(p)$. In this case, using the geometric series (which exists for $|(1-p)e^t| < 1 \implies t < -\ln(1-p)$)

$$M_Y(t) = \sum_{x=0}^{\infty} e^{tx} p(1-p)^x = \sum_{x=0}^{\infty} p((1-p)e^t)^x = \frac{p}{1-(1-p)e^t}.$$

We recall that $X \sim \mathrm{NegBin}(k, p)$ is the sum of k independent $\mathrm{Geo}(p)$ random variables. Therefore, $X = Y_1 + \cdots + Y_k$ and the sum of independent random variables formula for MGFs give

$$M_X(t) = M_{Y_1+\cdots+Y_k}(t) = M_{Y_1}(t)\cdots M_{Y_k}(t) = M_Y^k(t) = \left(\frac{p}{1-(1-p)e^t}\right)^k,$$

which exists under the same condition as $M_Y(t)$.

Problem 1.9. Suppose that $M_X(t)$ is defined for all $t \in [-a, a]$ for some $a > 0$. Prove that

$$\mathbb{E}[X^k] = \frac{d^k}{dt^k} M_X(0) = \frac{d^k}{dt^k} M_X(t)\Big|_{t=0} \qquad \forall k \geq 0.$$

Solution 1.9. Assuming that we can interchange the differentiation and integration, we have

$$\frac{d^k}{dt^k} M_X(t) = \frac{d^k}{dt^k} \mathbb{E}[e^{tX}] = \mathbb{E}\left[\frac{d^k}{dt^k} e^{tX}\right] = \mathbb{E}[X^k e^{tX}]$$

Therefore,

$$\frac{d^k}{dt^k} M_X(0) = \mathbb{E}[X^k e^{0\cdot X}] = \mathbb{E}[X^k].$$

The interchange between the differentiation and expectation is justified by a result called the dominated convergence theorem. The technical assumption that $M_X(t)$ is defined for all $t \in [-a, a]$ for some $a > 0$ is required to find a dominating function. In the case when X has finite support, then the interchange of differentiation and expectation simply follows from the linearity of the differentiation operator.

2 Probability Generating Functions

When we are working with random variables taking non-negative integer values, then there is yet another generating function that encodes the PMF. This provides us a powerful tool to compute the PMF of much more complicated discrete random variables representing counts.

Definition 2 (Probability Generating Function)**.** The *probability generating function* (PGF) of a non-negative random variable X taking integer values is given by the function

$$G_X(s) = \mathbb{E}[s^X].$$

Unlike the MGF, the PGF always converges for all $s \in [-1, 1]$ to a value in $[-1, 1]$ by the dominated convergence theorem, so it always exists.

Remark 2. The probability generating function is actually a special case of the MGF. It is defined by $G_X(s) = \mathbb{E}[s^X]$ and it is just a change of variables of the MGF since

$$G_X(s) = \mathbb{E}[s^X] = \mathbb{E}[e^{\ln(s^X)}] = \mathbb{E}[e^{\ln(s)X}] = M_X(\ln(s)).$$

2.1 Properties

Since the PGF is essentially a MGF under a change of variables, it obeys very similar properties.

1. **Uniqueness Theorem:** If X and Y have PGFs $G_X(s)$ and $G_Y(s)$ and there exists a δ such that $G_X(s) = G_Y(s)$ for all $s \in (-\delta, \delta)$, then
$$X \overset{d}{=} Y.$$
2. **Probabilities:** The MGF encodes the PMF of X. We have that
$$\frac{d^k}{ds^k} G_X(0) = k!\, \mathbb{P}(X = k) = k! p_X(k) \text{ for all } k \geq 0.$$
This follows from the definition of the expected value
$$G_X(s) = \mathbb{E}[s^X] = \sum_{j=0}^{\infty} s^j p_X(j) = \sum_{j=0}^{\infty} s^j\, \mathbb{P}(X = j).$$
Taking the kth derivative at 0 gives us the formula.
3. **Independent Sums:** Suppose that X and Y are independent and with PGFs $G_X(s)$ and $G_Y(s)$ respectively. Then the moment generating function of $X + Y$ is
$$G_{X+Y}(s) = \mathbb{E}\left(s^{X+Y}\right) = \mathbb{E}\left(s^X\right) \mathbb{E}\left(s^Y\right) = G_X(s) G_Y(s).$$
This also allows us to compute the distribution of sums of independent random variables in a simple way just like the MGFs.

2.2 Example Problems

Problem 2.1. n passengers board a plane with n seats, where $n \geq 1$. Despite every passenger having an assigned seat, when they board the plane they sit in one of the remaining available seats at random. Let N denote the number of passengers that are sitting in the assigned seats. What is the distribution of N?

Solution 2.1. We computed the expected value and variance of the matching using linearity of expectation without having to compute the PMF of N. In this problem, we will find the PGF and use it to compute the PMF.

Let N_k denote the number of passengers that are sitting in the assigned seats when there are k passengers in total. We will show that the PGF satisfies the following recursion: for all $k \geq 1$,

$$G'_{N_{k+1}}(s) = G_{N_k}(s) \qquad \text{and} \qquad G_{N_k}(1) = 1 \tag{1}$$

The fact that $G_k(1) = \mathbb{E}[1^{N_k}] = 1$ always holds by definition, because probabilities sum to 1. To see the second identity, we have

$$G'_{N_{k+1}}(s) = \frac{d}{ds}\sum_{j=0}^{k+1} s^j \,\mathbb{P}(N_{k+1} = j) = \sum_{j=1}^{k+1} j s^{j-1}\,\mathbb{P}(N_{k+1} = j) = \sum_{i=0}^{k}(i+1)s^i\,\mathbb{P}(N_{k+1} = i+1)$$

We now claim that

$$(i+1)\,\mathbb{P}(N_{k+1} = i+1) = \mathbb{P}(N_k = i) \iff \mathbb{P}(N_{k+1} = i+1) = \frac{\mathbb{P}(N_k = i)}{(i+1)}. \tag{2}$$

This is natural, because if there are $i+1$ correctly seated passengers, then there are $i+1$ ways to fix a particular corrected seated passenger. Having fixed one correctly seated passenger, the remaining k passengers are seated uniformly according to seating arrangements on the k remaining sets and i of these must be seated correctly. Substituting (2) immediately implies (1).

We can now compute the generating functions recursively form (1).

1. If there is 1 passenger, then it is clear $p_{N_1}(1) = 1$, so $G_{N_1}(s) = s$.
2. Since $G'_{N_2}(s) = G_{N_1}(s)$, we can integrate the above to conclude that

$$G_{N_2}(s) = \int G_{N_1}(s)\,ds = \frac{s^2}{2} + C_2.$$

   Since $G_{N_2}(1) = 1$, we must have that $C_2 = \frac{1}{2}$. This implies that

$$p_{N_2}(0) = \frac{1}{2} \quad p_{N_2}(2) = \frac{1}{2}.$$

3. Integrating this again, we have

$$G_{N_3}(s) = \frac{s}{2} + \frac{s^3}{6} + C_3 \overset{G_{N_3}(1)=1}{\Longrightarrow} G_{N_3}(s) = \frac{1}{3} + \frac{s}{2} + \frac{s^3}{6}$$

   so

$$p_{N_3}(0) = \frac{1}{3} \quad p_{N_3}(1) = \frac{1}{2} \quad p_{N_3}(2) = \frac{1}{6}.$$

4. And similarly,

$$G_{N_4}(s) = \frac{3}{8} + \frac{1}{3}s + \frac{1}{4}s^2 + \frac{1}{24}s^4$$

   so

$$p_{N_4}(0) = \frac{3}{8} \quad p_{N_4}(1) = \frac{1}{3} \quad p_{N_4}(2) = \frac{1}{4} \quad p_{N_4}(4) = \frac{1}{24}.$$

Remark 3. By an induction argument, we can show that

$$\mathbb{P}(N_n = k) = \frac{1}{k!}\,\mathbb{P}(N_{n-k} = 0) = \frac{1}{k!}\sum_{j=0}^{n-k}\frac{(-1)^j}{j!}.$$

Remark 4. We can carefully show (2). Let A denote the event that the $k+1$st passenger is in the right seat. If we know that the $k+1$ passenger is in the right seat, then

$$\mathbb{P}(N_{k+1} = i+1 \mid A) = \mathbb{P}(N_k = i)$$

since there are k passengers remaining and we want to compute the probability there are i correctly seated out of the remaining passengers. By Bayes' theorem, we have

$$\mathbb{P}(N_{k+1} = i+1 \mid A) = \frac{\mathbb{P}(A \mid N_{k+1} = i+1)\,\mathbb{P}(N_{k+1} = i+1)}{\mathbb{P}(A)} = (i+1)\,\mathbb{P}(N_{k+1} = i+1)$$

since $\mathbb{P}(A) = \frac{1}{k+1}$ by symmetry (see Week 5 Problem 1.1) and $\mathbb{P}(A \mid N_{k+1} = i+1) = \frac{i+1}{k+1}$ since on the event $\{N_{k+1} = i+1\}$ we know that there are $i+1$ correctly seated passengers out of $k+1$ so the probability that any particular passenger is seated correctly is $\frac{i+1}{k+1}$.

2.3 Proofs of Key Results

Problem 2.2. Show that the PGF always exist.

Solution 2.2. Let X have PMF p_X. Since X is supported on non-negative integers

$$G_X(s) = \sum_{j=0}^{\infty} s^j p_X(j)$$

For any $s \in [-1,1]$ we have that $|s^j p_X(j)| \le p_X(j)$ so

$$\sum_{j=0}^{\infty} |s^j p_X(j)| \le \sum_{j=0}^{\infty} p_X(j) = 1.$$

The infinite series is absolutely convergent, so its infinite sum exists.

Problem 2.3. Show that

$$\frac{d^k}{ds^k} G_X(0) = k!\,\mathbb{P}(X = k) = k! p_X(k) \text{ for all } k \ge 0.$$

Solution 2.3. We have that

$$G_X(0) = \sum_{j=0}^{\infty} s^j p_X(j)\Big|_{s=0} = p_X(0).$$

Likewise, since we can differentiate the series term by term (since 0 is in the radius of convergence of the power series) so

$$G_X'(0) = \sum_{j=1}^{\infty} j s^{j-1} p_X(j)\Big|_{s=0} = p_X(1)$$

Continuing inductively gives us the formula

$$\frac{d^k}{ds^k} G_X(0) = k!\,\mathbb{P}(X = k) = k! p_X(k) \text{ for all } k \ge 0.$$

Problem 2.4. Prove the uniqueness theorem for the MGFs.

Solution 2.4. This is a direct consequence from the derivative characterization. If we know the PGF $G_X(s)$, then it uniquely determines the PMF, i.e.

$$p_X(k) = \mathbb{P}(X = k) = \frac{1}{k!} \frac{d^k}{ds^k} G_X(0).$$

Therefore, if X and Y have the same PGF, then for all $k \geq 0$,

$$p_Y(k) = \mathbb{P}(Y = k) = \frac{1}{k!} \frac{d^k}{ds^k} G_Y(0) = \frac{1}{k!} \frac{d^k}{ds^k} G_X(0) = p_X(k)$$

so X and Y have the PMF and hence the same distribution.