GENERATING FUNCTIONS IN PROBABILITY

NEIL MAKUR

ABSTRACT. We discuss the moment generating function, which uniquely classifies a probability distribution.

1. Introduction

The goal of this paper is to determine what information we need to know about a probability distribution to completely determine it. We already have one tool to do so: expected value. Recall that the expected value of a random variable X is

$$\mathbb{E}(X) = \sum_{x \in \operatorname{im} X} x \cdot \mathbb{P}(X = x),$$

where we write $\operatorname{im} X$ for all possible values of X. Suppose that we have two random variables X and Y, with $\operatorname{im} X = \operatorname{im} Y = \{1, 2, 3, 4, 5, 6\}$, and $\mathbb{E}(X) = \mathbb{E}(Y) = 3.5$. Is this enough to completely classify the probability distribution? In other words, is $\mathbb{P}(X = a) = \mathbb{P}(Y = a)$ for all a? It turns out that the answer is no. For example, consider the following:

$$\mathbb{P}(X = d) = \frac{1}{6} \qquad d \in \{1, 2, 3, 4, 5, 6\}$$

$$\mathbb{P}(Y = d) = \begin{cases} \frac{1}{2} & d \in \{2, 5\} \\ 0 & d \in \{1, 3, 4, 6\} \end{cases}.$$

Then,

$$\mathbb{E}(X) = \sum_{x=1}^{6} \frac{1}{6} x = 3.5,$$

and

$$\mathbb{E}(Y) = \frac{1}{2} \cdot 2 + \frac{1}{2} \cdot 5 = 3.5.$$

Thus, despite having different probability distributions, X and Y have the same expected value.

2. Variance

2.1. **Introduction to Variance.** Now that we have seen the failure of expected value to completely classify a probability distribution, we define variance.

Definition 2.1. Let X be a random variable with expected value $\mathbb{E}(X) = \mu$. Then, the variance of X is

$$\operatorname{Var}(X) = \mathbb{E}((X - \mu)^2).$$

Definition 2.2. Let X be a random variable. Its standard deviation is $\operatorname{SD}(X) = \sqrt{\operatorname{Var}(X)}$.

Date: November 2020.

Remark 2.3. Both variance and standard deviation measure how much a random variable varies from its expected value on average. However, $\operatorname{Var}(X)$ measures the square of the difference. If we wanted a measure of the magnitude, it makes sense to define $\operatorname{SD}(X)$ as its square root. We will see also that $\operatorname{Var}(cX) = c^2 \operatorname{Var}(X)$, meaning that $\operatorname{SD}(cX) = c\operatorname{SD}(X)$, which makes intuitive sense. Despite this, we will not use $\operatorname{SD}(X)$ much in this paper, as there are larger upsides to using $\operatorname{Var}(X)$, such as the property that $\operatorname{Var}(X+Y) = \operatorname{Var}(X)+\operatorname{Var}(Y)$.

Example. Let us find the variance and standard deviation of the outcome of a standard die roll. If we let X represent the outcome, we know that $\mu = \mathbb{E}(X) = 3.5$. We can form the following table.

x	$\mathbb{P}(X = x)$	$(x-\mu)^2$
1	1/6	25/4
2	1/6	9/4
3	1/6	1/4
4	1/6	1/4
5	1/6	9/4
6	1/6	25/4

Therefore,

$$\operatorname{Var}(X) = \mathbb{E}((x-\mu)^2) = \frac{1}{6}\left(\frac{25}{4} + \frac{9}{4} + \frac{1}{4} + \frac{1}{4} + \frac{9}{4} + \frac{25}{4}\right) = \frac{35}{12},$$

and

$$\operatorname{SD}(X) = \sqrt{\operatorname{Var}(X)} = \sqrt{\frac{35}{12}} \approx 1.71.$$

2.2. **Properties of Variance.** We will prove some properties of variance that make its calculation easier.

Proposition 2.4. *If X is a random variable with $\mathbb{E}(X) = \mu$, then*

$$\operatorname{Var}(X) = \mathbb{E}(X^2) - \mu^2.$$

Proof. We have that

$$\begin{aligned}\operatorname{Var}(X) &= \mathbb{E}((x-\mu)^2) = \mathbb{E}(X^2 - 2X\mu + \mu^2) \\ &= \mathbb{E}(X^2) - \mathbb{E}(2\mu X) + \mathbb{E}(\mu^2) = \mathbb{E}(X^2) - 2\mu\mathbb{E}(X) + \mu^2 \\ &= \mathbb{E}(X^2) - \mu^2.\end{aligned}$$

■

Example. If X is the outcome of a die roll, then

$$\mathbb{E}(X^2) = \frac{1}{6}(1 + 4 + 9 + 16 + 25 + 36) = \frac{91}{6},$$

so

$$\operatorname{Var}(X) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}.$$

Proposition 2.5. *If X is a random variable and c is a constant, then*

$$\operatorname{Var}(cX) = c^2 \operatorname{Var}(X),$$

and

$$\operatorname{Var}(X + c) = \operatorname{Var}(X).$$

Proof. Let $\mu = \mathbb{E}(X)$. Then, $\mathbb{E}(cX) = c\mu$, so

$$\begin{aligned}\operatorname{Var}(cX) &= \mathbb{E}((cX - c\mu)^2) = \mathbb{E}(c^2(X-\mu)^2)\\ &= c^2\mathbb{E}((X-\mu)^2) = c^2 \operatorname{Var}(X).\end{aligned}$$

For the second part, we have $\mathbb{E}(X + c) = \mu + c$, so

$$\begin{aligned}\operatorname{Var}(X+c) &= \mathbb{E}(((X+c) - (\mu + c))^2) = \mathbb{E}((X-\mu)^2)\\ &= \operatorname{Var}(X).\end{aligned}$$

■

Proposition 2.6. *If X and Y are independent random variables, then*

$$\operatorname{Var}(X+Y) = \operatorname{Var}(X) + \operatorname{Var}(Y).$$

To prove this, we first need a lemma.

Lemma 2.7. *If X and Y are independent random variables, then*

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y).$$

Proof. We have that

$$\begin{aligned}\mathbb{E}(XY) &= \sum_{x\in \operatorname{im} X}\sum_{y \in \operatorname{im} Y} xy\mathbb{P}(X = x, Y = y)\\ &= \sum_{x\in \operatorname{im} X}\sum_{y \in \operatorname{im} Y} x\mathbb{P}(X = x) y\mathbb{P}(Y = y)\\ &= \left(\sum_{x\in \operatorname{im} X} x\mathbb{P}(X=x)\right)\left(\sum_{y \in \operatorname{im} Y} y\mathbb{P}(Y=y)\right)\\ &= \mathbb{E}(X)\mathbb{E}(Y).\end{aligned}$$

■

We can now prove Proposition 2.6.

Proof of Proposition 2.6. Note that

$$\begin{aligned}\operatorname{Var}(X+Y) &= \mathbb{E}((X+Y)^2) - (\mathbb{E}(X+Y))^2 = \mathbb{E}((X+Y)^2) - (\mathbb{E}(X) + \mathbb{E}(Y))^2\\ &= \mathbb{E}(X^2 + 2XY + Y^2) - \mathbb{E}(X)^2 - 2\mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(Y)^2\\ &= \mathbb{E}(X^2) + \mathbb{E}(2XY) + \mathbb{E}(Y^2) - \mathbb{E}(X)^2 - 2\mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(Y)^2\\ &= (\mathbb{E}(X^2) - \mathbb{E}(X)^2) + (\mathbb{E}(Y^2) - \mathbb{E}(Y)^2) + (2\mathbb{E}(XY) - 2\mathbb{E}(X)\mathbb{E}(Y))\\ &= \operatorname{Var}(X) + \operatorname{Var}(Y) + (2\mathbb{E}(X)\mathbb{E}(Y) - 2\mathbb{E}(X)\mathbb{E}(Y))\\ &= \operatorname{Var}(X) + \operatorname{Var}(Y).\end{aligned}$$

■

This result can easily be generalized to the sum of any amount of variables.

3. Moments

3.1. **Failure of Expected Value and Variance.** We now have two tools to classify a probability distribution; expected value and variance. However, it turns out that these are not enough to classify one completely. For example, consider the following distributions.

t	1	2	3	4	5	6
$\mathbb{P}(X = t)$	0	1/4	1/2	0	0	1/4
$\mathbb{P}(Y = t)$	1/4	0	0	1/2	1/4	0

Then, $\mathbb{E}(X) = \mathbb{E}(Y) = 3.5$, and $\mathrm{Var}(X) = \mathrm{Var}(Y) = 2.25$. Thus, given $\mathbb{E}(X)$ and $\mathrm{Var}(X)$, we cannot completely classify X's probability distribution.

3.2. **Moments.** This failure of expected value and variance may inspire us to define moments.

Definition 3.1. Let X be a random variable. The k^{th} moment of X is

$$\mu_k = \mathbb{E}(X^k) = \sum_{x \in \operatorname{im} X} x^k \mathbb{P}(X = x),$$

provided that the sum converges.

It turns out that expected value and variance are encoded in moments.

Proposition 3.2. *Let X be a random variable. Then,*

$$\mathbb{E}(X) = \mu_1,$$

and

$$\mathrm{Var}(X) = \mu_2 - \mu_1^2.$$

Proof. We have $\mu_1 = \mathbb{E}(X^1) = \mathbb{E}(X)$, and $\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \mu_2 - \mu_1^2$. ∎

4. Moment Generating Functions

4.1. **Introduction.** Now that we have discussed moments, we encode them into a single object.

Definition 4.1. Let X be a random variable. We define the moment generating function for X as

$$g_X(t) = \mathbb{E}(e^{tX}) = \mathbb{E}\left(\sum_{k=0}^{\infty} \frac{X^k t^k}{k!}\right) = \sum_{k=0}^{\infty} \frac{\mu_k t^k}{k!}.$$

The reason that we say we have encoded the moments is that we can in fact derive them from the function, since

$$\left.\frac{d^n}{dt^n} g_X(t)\right|_{t=0} = \left.\frac{d^n}{dt^n} \sum_{k=0}^{\infty} \frac{\mu_k t^k}{k!}\right|_{t=0} = \left.\sum_{k=n}^{\infty} \frac{k!}{(n-k)!} \cdot \frac{\mu_k t^{k-n}}{k!}\right|_{t=0} = \mu_n.$$

4.2. **Examples.**

Example. Suppose that $\operatorname{im} X = \{1, 2, \ldots, n\}$, and that X is distributed according to the uniform distribution (i.e. $\mathbb{P}(X = j) = n^{-1}$ for all $j \in \operatorname{im} X$). Let us compute the moment generating function of X. We have

$$\begin{aligned} g_X(t) = \mathbb{E}(e^{tX}) &= \sum_{j=1}^{n} \frac{1}{n} e^{jt} \\ &= \frac{1}{n} \sum_{i=1}^{n} e^{jt} \\ &= \frac{1}{n} \cdot \frac{e^{(n+1)t} - e^t}{e^t - 1}. \end{aligned}$$

Using the second line, we can see that

$$\mu_1 = g'(0) = \frac{1}{n} \sum_{j=1}^{n} j = \frac{n+1}{2},$$

that

$$\mu_2 = g''(0) = \frac{1}{n} \sum_{j=1}^{n} j^2 = \frac{(n+1)(2n+1)}{6},$$

and that in general,

$$\mu_k = g^{(k)}(0) = \frac{1}{n} \sum_{j=1}^{n} j^k.$$

In particular, $\mathbb{E}(X) = \mu_1 = (n+1)/2$, and $\operatorname{Var}(X) = \mu_2 - \mu_1^2 = (n^2 - 1)/12$.

Example. Suppose that $\operatorname{im} X = \{1, 2, \ldots\}$, $0 < p < 1$, $q = 1 - p$, and that

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

In this case, X represents the first heads in an unfair coin (where heads has probability p). Then,

$$g_X(t) = \sum_{j=1}^{\infty} e^{tj} q^{j-1} p = pe^t \sum_{j=0}^{\infty} e^{tj} q^j = \frac{pe^t}{1 - qe^t}.$$

We also have that

$$\begin{aligned} \mu_1 = g_X'(0) &= \left.\frac{pe^t}{(1 - qe^t)^2}\right|_{t=0} = \frac{1}{p} \\ \mu_2 = g_X''(0) &= \left.\frac{pe^t + pqe^{2t}}{(1 - qe^t)^3}\right|_{t=0} = \frac{1+q}{p^2}. \end{aligned}$$

Thus, $\mathbb{E}(X) = \mu_1 = 1/p$, and $\operatorname{Var}(X) = \mu_2 - \mu_1^2 = q/p^2$.

4.3. **Ordinary Generating Functions.** We now turn to the case where $\operatorname{im} X = \{0, 1, 2, \ldots, n\}$. In this case, we may define an ordinary generating function.

Definition 4.2. Suppose that $\operatorname{im} X = \{0, 1, 2, \ldots, n\}$. Then the ordinary generating function of X is

$$h_X(z) = \sum_{j=0}^{n} z^j \mathbb{P}(X = j).$$

The function $h_X(z)$ is very closely related to $g_X(t)$, and they contain the same information. More precisely:

$$h_X(z) = g_X(\log z)$$
$$g_X(t) = h_X(e^t).$$

Thus, if we know $g_X(t)$, we can determine $h_X(z)$, and vice versa. It also turns out that $h_X(z)$, and thus $g_X(t)$, can answer our question in this special case.

Theorem 4.3. *Suppose that X and Y are random variables with $\operatorname{im} X = \operatorname{im} Y = \{0, 1, 2, \ldots, n\}$. Then, $h_X(z) = h_Y(z)$ if and only if $\mathbb{P}(X = j) = \mathbb{P}(Y = j)$ for all $0 \leq j \leq n$.*

Proof. Suppose that $h_X(z) = h_Y(z)$ for all z. Then,

$$\sum_{j=0}^{n} z^j \mathbb{P}(X = j) = h_X(z) = h_Y(z) = \sum_{j=0}^{n} z^j \mathbb{P}(Y = j)$$

for all j. Thus,

$$\sum_{j=0}^{n} z^j (\mathbb{P}(X = j) - \mathbb{P}(Y = j)) = 0.$$

This is a polynomial in z of degree n, so it can have at most n zeros, unless it is the 0 polynomial. It clearly has more than n roots, so $\mathbb{P}(X = j) - \mathbb{P}(Y = j)$ must be 0 for all j. Therefore, $\mathbb{P}(X = j) = \mathbb{P}(Y = j)$. Now, suppose that $\mathbb{P}(X = j) = \mathbb{P}(Y = j)$ for all j. Then,

$$h_X(z) = \sum_{j=0}^{n} z^n \mathbb{P}(X = j) = \sum_{j=0}^{n} z^n \mathbb{P}(Y = j) = h_Y(z).$$

■

This proof does not give us a nice way to extract the probability distribution from $h_X(z)$. However, it is possible to do this as well.

Proposition 4.4. *Let X be a random variable with $\operatorname{im} X = \{0, 1, 2, \ldots, n\}$. Then,*

$$\mathbb{P}(X = j) = \frac{1}{j!} \frac{d^j}{dz^j} h_X(z) \bigg|_{t=0}.$$

Proof. We have that

$$h_X^{(j)}(z) = \sum_{i=j}^{n} \frac{i!}{(i-j)!} z^{i-j} \mathbb{P}(X = j) = j! \mathbb{P}(X = j) + \sum_{i=j+1}^{n} \frac{i!}{(i-j)!} z^{i-j} \mathbb{P}(X = j).$$

Thus,

$$h_X^{(j)}(0) = j! \mathbb{P}(X = j),$$

so

$$\frac{1}{j!}h_X^{(j)}(0) = \mathbb{P}(X = j).$$

■

4.4. **Properties of Moment Generating Functions.** Similar to how we have formulas for $\mathbb{E}(X + a)$ and $\operatorname{Var}(X + a)$, there is also a formula for $g_{X+a}(t)$.

Proposition 4.5. *Let X be a random variable. If a and b are constant, then,*

$$g_{X+a}(t) = e^{ta}g_X(t) \qquad \text{and} \qquad g_{bX}(t) = g_X(bt).$$

Proof. We have that

$$g_{X+a}(t) = \mathbb{E}(e^{t(X+a)}) = \mathbb{E}(e^{tX}e^{ta}) = e^{ta}\mathbb{E}(e^{tX}) = e^{ta}g_X(t),$$

and that

$$g_{bX} = \mathbb{E}(e^{tbX}) = \mathbb{E}(e^{(tb)X}) = g_X(bt).$$

■

Corollary 4.6. *Let X be a random variable, and let a and b be constants. Then,*

$$g_{aX+b}(t) = e^{bt/a}g_X(at).$$

Proof. Notice that

$$g_{aX+b}(t) = g_{a(X+b/a)}(t) = g_{X+b/a}(at) = e^{bt/a}g_X(at).$$

■

We can also generalize to the case of the sum of random variables.

Proposition 4.7. *Suppose that X and Y are independent random variables. Then,*

$$g_{X+Y}(t) = g_X(t)g_Y(t).$$

Proof. Note that $\mathbb{P}(e^{tX} = e^{tx}) = \mathbb{P}(X = x)$. Thus, if X and Y are independent, so are e^{tX} and e^{tY}. Therefore,

$$g_{X+Y}(t) = \mathbb{E}(e^{t(X+Y)}) = \mathbb{E}(e^{tX}e^{tY}) = \mathbb{E}(e^{tX})\mathbb{E}(e^{tY}) = g_X(t)g_Y(t).$$

■

In the case that $\operatorname{im} X = \{0, 1, 2, \ldots, n\}$ and $\operatorname{im} Y = \{0, 1, 2, \ldots, m\}$, then $\operatorname{im} X + Y = \{0, 1, 2, \ldots, n + m\}$. We also have that

$$h_{X+Y}(z) = h_X(z)h_Y(z).$$

Thus,

$$\frac{d^j}{dz^j}h_{X+Y}(z) = \sum_{i=0}^{j}\binom{j}{i}h_X^{(i)}(z)h_Y^{(j-i)}(z),$$

so

$$\frac{h_{X+Y}^{(j)}(z)}{j!} = j!\sum_{i=0}^{j}\binom{j}{i}\frac{h_X^{(i)}(z)}{j!}\frac{h_Y^{(j-i)}(z)}{j!} = j!\sum_{i=0}^{j}\binom{j}{i}\frac{h_X^{(i)}(z)}{i!}\frac{h_Y^{(j-i)}(z)}{(j-i)!}\cdot\frac{i!(j-i)!}{j!^2}.$$

Evaluating at 0,

$$\mathbb{P}(X + Y = j) = \sum_{i=0}^{j}\mathbb{P}(X = i)\mathbb{P}(Y = j - i),$$

an expected formula.

4.5. **Random Walks on $\mathbb{Z}$.** Let us consider a random walk on $\mathbb{Z}$. We determine how we move by flipping a coin that has a probability p of heads occurring, and a probability $q = 1 - p$ of tails occurring. We define X_k, which is how we move, as

$$X_k = \begin{cases} +1 & k^{\text{th}} \text{ toss is heads} \\ -1 & k^{\text{th}} \text{ toss is tails} \end{cases}.$$

We define

$$S_n = X_1 + X_2 + \cdots + X_n$$

to be our position at time n. We investigate when we first go to $\mathbb{Z}_{>0}$. Let

$$r_n = \begin{cases} 0 & n \equiv 0 \pmod 2 \\ p & n = 1 \\ q(r_1 r_{n-2} + r_3 r_{n-4} + \cdots + r_{n-2} r_1) & 1 \neq n \equiv 1 \pmod 2 \end{cases}.$$

Further, let T be a random variable representing the first time we are positive. We claim that $\mathbb{P}(T = j) = r_j$. This is clear in the first two cases, so we only check the last case. Our method of proof will be induction, with base case $n = 1$, which is trivial. Now, we must hit 0 at least once after our initial starting position. Our first move must also be in the negative direction. Suppose that we first hit 0 at time $2k$. The probability of this happening, along with first becoming positive at time n, is

$$q\mathbb{P}(T = 2k - 1)\mathbb{P}(T = n - 2k).$$

This is equal to $qr_{2k-1}r_{n-2k}$, by the induction hypothesis. Summing over all possible values of $2k$ gives the desired result. Thus, the ordinary generating function of T is

$$h_T(z) = \sum_{n=0}^{\infty} r_n z^n.$$

Now,

$$h_T^2(z) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} r_n r_m z^{n+m}.$$

The coefficient of z^a is

$$\sum_{n=0}^{a} r_n r_{a-n}.$$

Thus, the coefficient of z^a in $qzh_T^2(z)$ is

$$\sum_{n=0}^{a} q r_n r_{a-1-n}.$$

Note that we can disregard all values of n that are even. Note also that if a is even (meaning $a - 1$ is odd), we can also disregard all odd values of n. Thus, the coefficient of z^a is r_a. However, there is no z term in this. We may add pz to obtain

$$h_T(z) = pz + qzh_T^2(z).$$

Solving, we get that

$$h_T(z) = \frac{1 \pm \sqrt{1-4pqz^2}}{2qz} = \frac{2pz}{1 \mp \sqrt{1-4pqz^2}}.$$

We know that $h_T(0)$ exists, so we choose the one with a limit as z approaches 0. Then,

$$h_T(z) = \frac{1 - \sqrt{1-4pqz^2}}{2qz} = \frac{2pz}{1 + \sqrt{1-4pqz^2}}.$$

Now, the probability that we are ever positive is

$$\sum_{n=0}^{\infty} r_n = h_T(1) = \frac{1-\sqrt{1-4pq}}{2q}.$$

We have that

$$\begin{aligned} \frac{1-\sqrt{1-4pq}}{2q} &= \frac{1-\sqrt{1-4p+4p^2}}{2q} \\ &= \frac{1-|p-q|}{2q} \\ &= \begin{cases} p/q & p < q \\ 1 & p \geq q \end{cases}. \end{aligned}$$

Now, let us find the expected value of T. This is

$$\begin{aligned} \mathbb{E}(T) = h_T'(1) &= \frac{d}{dz}\left(\frac{1}{2qz} - \sqrt{\frac{1}{4q^2z^2} - \frac{p}{q}}\right)\Bigg|_{z=1} \\ &= \frac{-1}{2qz^2} - \frac{(-2)/(4q^2z^3)}{2\sqrt{1/(4q^2z^2) - p/q}}\Bigg|_{z=1} = \frac{-1}{2q} + \frac{1}{2q\sqrt{1-4pq}} \\ &= \frac{1-|p-q|}{2q|p-q|} \\ &= \begin{cases} 1/(p-q) & p > q \\ \infty & p = q \\ p/(q \cdot (q-p)) & p < q \end{cases}. \end{aligned}$$

4.6. **Complete Classification.** We now return to our problem of classifying a probability distribution with the following theorem.

Theorem 4.8. *Let X be a random variable with* $\operatorname{im} X = \{x_1, x_2, \ldots, x_n\}$ *with probability distribution p. Then g_X is uniquely determined by p, and vice versa.*

Proof. It is clear that p uniquely determines g_X, as

$$g_X(z) = \sum_{j=1}^{n} e^{tx_j} \mathbb{P}(X = x_j).$$

Now, suppose that $\operatorname{im} X = \operatorname{im} Y$, and $g_X = g_Y$. Then,

$$\begin{aligned}\sum_{j=1}^{n} e^{tx_j}\mathbb{P}(X = x_j) &= \sum_{j=1}^{n} e^{ty_j}\mathbb{P}(Y = y_j) \\ &= \sum_{j=1}^{n} e^{tx_j}\mathbb{P}(Y = x_j).\end{aligned}$$

Thus,

$$\sum_{j=1}^{n} e^{tx_j}(\mathbb{P}(X = x_j) - \mathbb{P}(Y = x_j)) = 0.$$

Since $e^{tx_1}, e^{tx_2}, \ldots, e^{tx_n}$ are linearly independent, $\mathbb{P}(X = x_j) - \mathbb{P}(Y = x_j) = 0$ for all j, so X and Y have the same probability distribution. ■

Once again, this proof is not illuminating as to how to extract the probability distribution, so we show how to do this as well. Reordering if necessary, we may assume that $x_1 < x_2 < \cdots < x_n$. Note that $g_X(t)$ is differentiable. We have that

$$\frac{g_X'(t)}{g_X(t)} = \frac{x_1\mathbb{P}(X = x_1)e^{tx_1} + \cdots + x_n\mathbb{P}(X = x_n)e^{tx_n}}{\mathbb{P}(X = x_1)e^{tx_1} + \cdots + \mathbb{P}(X = x_n)e^{tx_n}}.$$

Now, we have that

$$\begin{aligned}\lim_{t\to\infty} \frac{g_X'(t)}{g_X(t)} &= \lim_{t\to\infty} \frac{x_1\mathbb{P}(X = x_1)e^{tx_1} + \cdots + x_n\mathbb{P}(X = x_n)e^{tx_n}}{\mathbb{P}(X = x_1)e^{tx_1} + \cdots + \mathbb{P}(X = x_n)e^{tx_n}} \\ &= \lim_{t\to\infty} \frac{x_1\mathbb{P}(X = x_1)e^{t(x_1-x_n)} + \cdots + x_n\mathbb{P}(X = x_n)}{\mathbb{P}(X = x_1)e^{t(x_1-x_n)} + \cdots + \mathbb{P}(X = x_n)} \\ &= x_n,\end{aligned}$$

where the last equality holds because $x_n > x_j$ for all $j < n$. Thus, we have extracted x_n from $g_X(t)$. It is not hard to see that $\lim_{t\to\infty} e^{-tx_n}g_X(t) = \mathbb{P}(X = x_n)$, so we can extract this as well. Now, we may subtract $\mathbb{P}(X = x_n)e^{tx_n}$ from $g_X(t)$ and repeat to obtain $x_{n-1}, x_{n-2}, \ldots, x_1$ and $\mathbb{P}(X = x_{n-1}), \mathbb{P}(X = x_{n-2}), \ldots, \mathbb{P}(X = x_1)$. Note that we have shown that we actually do not need to know $\operatorname{im} X$ beforehand, as it can be obtained from g_X.

5. Applications

5.1. **Markov's Inequality.** We turn now to applications of moment generating functions. We first state and prove Markov's Inequality.

Proposition 5.1 (Markov's Inequality)**.** *Let X be a non-negative random variable. Then, for every constant $a > 0$,*

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}.$$

Proof. Notice that

$$\mathbb{E}(X) = \sum_{x \in \operatorname{im} X} x\mathbb{P}(X = x) = \sum_{\substack{x \in \operatorname{im} X \\ x < a}} x\mathbb{P}(X = x) + \sum_{\substack{x \in \operatorname{im} X \\ x \geq a}} x\mathbb{P}(X = x)$$

$$\geq \sum_{\substack{x \in \operatorname{im} X \\ x \geq a}} x\mathbb{P}(X = x) \geq \sum_{\substack{x \in \operatorname{im} X \\ x \geq a}} a\mathbb{P}(X = x) = a\mathbb{P}(X \geq a).$$

Dividing by a gives the desired result. ■

5.2. **Chernoff Bounds.** We can now use moment generating functions to obtain bounds on $\mathbb{P}(X \geq a)$ and $\mathbb{P}(X \leq a)$.

Proposition 5.2. *Let X be a random variable. Then,*

$$\mathbb{P}(X \geq a) \leq \inf_{t>0} g_X(t)e^{-ta},$$

and

$$\mathbb{P}(X \leq a) \leq \inf_{t<0} g_X(t)e^{-ta}.$$

Proof. Note that

$$\mathbb{P}(X \geq a) = \mathbb{P}(e^{tX} \geq e^{ta}) \leq \frac{\mathbb{E}(e^{tX})}{e^{ta}} = g_X(t)e^{-ta},$$

in the case where $t > 0$. We also have that

$$\mathbb{P}(X \leq a) = \mathbb{P}(e^{tX} \geq e^{ta}) \leq \frac{\mathbb{E}(e^{tX})}{e^{ta}} = g_X(t)e^{-ta},$$

where $t < 0$. We may take infimums in both cases. ■

It is worth noting that this bound is always at least as good as the simple bound of 1, as $g_X(t)e^{-ta}|_{t=0} = 1$, and moment generating functions are continuous. This is not true of Markov's Inequality. For example, let X represent the outcome of a die roll. Then,

$$\mathbb{P}(X \geq 1) \leq \frac{\mathbb{E}(X)}{1} = 3.5,$$

which is useless

Moment Generating Functions can also be used to prove the Weak Law of Large Numbers and the Central Limit Theorem, see [GS12, §10.3] for the latter.

References

[GS12] Charles Miller Grinstead and James Laurie Snell. *Introduction to probability.* American Mathematical Soc., 2012.