1 Absolutely Continuous and Singularly Continuous Random Variables

As we observed earlier, there are continuous random variables that take every value in an interval. Such variables take uncountable values and we have developed the notion of $\mathbb{P}(A)$ for $A \in \mathcal{F}$ of the probability model $(\Omega, \mathcal{F}, \mathbb{P})$. Starting from this probability model, we can still write $F_X(a) = \mathbb{P}(X(\omega) \in (-\infty, a])$. So the cumulative distribution functions (cdfs) for the continuous random variables have the same properties as the cdfs of the discrete variables.

Continuity of random variable $X$ is equivalent to the continuity of $F_X$, where continuity of function $F_X$ is well-established. Absolute continuity of $X$ refers to the absolute continuity of the cdf $F_X$: That is, $\sum_{i=1}^{n} F_X(b_i) - F_X(a_i)$ can be bounded by $\epsilon$ for every given $\delta$ that bounds $\sum_{i=1}^{n} (b_i - a_i)$ for finitely many disjoint $n$ intervals $(a_i, b_i)$. Hence, an absolutely continuous function maps sufficiently tight intervals to arbitrarily tight intervals. In the special case, an absolutely continuous function maps intervals of length zero to intervals of length zero (p.413 Billingsley 1995). In other words, you cannot make something out of nothing by passing it through an absolutely continuous function.

In the definition of absolute continuity, for every given $\epsilon$ satisfying $\sum_{i=1}^{n} F_X(b_i) - F_X(a_i) < \epsilon$, a $\delta$ independent of the number $n$ of intervals is to be chosen to satisfy $\sum_{i=1}^{n} (b_i - a_i) < \delta$. When $\delta$ is chosen, the number of intervals is not known. The same $\delta$ should work regardless of the number of intervals. Moreover, only finite number of intervals are considered in the definition.

Example: Cantor function (also called Devil’s staircase) is defined along with the Cantor set. Cantor set is obtained by starting with $[0, 1]$ and removing the open interval in the middle $1/3$ of the remaining intervals; see Table 1. This set has uncountable number of points but its length (Lebesgue measure) is zero. Cantor function is a limit of series of functions $F_n$ defined over $[0, 1]$ where $F_n(0) = 0$ and $F_n(1) = 1$. Cantor function turns out to be continuous, increasing and constant outside Cantor set. Hence, Cantor function increases only inside Cantor set and increases from 0 to 1. That is, Cantor function maps Cantor set of length zero to the interval of $[0, 1]$ and it is not absolutely continuous. If you think of Cantor function as the cdf $F$ for random variable $X$, then $X$ becomes a continuous random variable but it is not absolutely continuous. For a different characterization of Cantor function, see Chalice (1991) which provides an operational description: Any increasing real-valued function on $[0, 1]$ that has a) $F(0) = 0$, b) $F(x/3) = F(x)/2$, c) $F(1-x) = 1 - F(x)$ is the Cantor function. ◦

<table>
<thead>
<tr>
<th>Iteration $n$</th>
<th>Available intervals</th>
<th>Length of available intervals</th>
<th>Length of removed intervals in this iteration</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$(0,1)$</td>
<td>1</td>
<td>$1/3$</td>
</tr>
<tr>
<td>1</td>
<td>$(0,1/3), (2/3,1)$</td>
<td>$2/3 = 1 - 1/3$</td>
<td>$(1/3)(2/3)$</td>
</tr>
<tr>
<td>2</td>
<td>$(0,1/9), (2/9,3/9), (6/9,7/9), (8/9,1)$</td>
<td>$(2/3)^2 = 2/3 - (1/3)(2/3)$</td>
<td>$(1/3)(2/3)^2$</td>
</tr>
<tr>
<td>3</td>
<td>$(0,1/27), (2/27,3/27), (6/27,7/27), (8/27,9/27)$</td>
<td>$(2/3)^3 = (2/3)^2 - (1/3)(2/3)^2$</td>
<td>$(1/3)(2/3)^3$</td>
</tr>
<tr>
<td>4</td>
<td>$(18/27,19/27), (20/27,21/27), (24/27,25/27), (26/27,1)$</td>
<td>$(2/3)^4 = (2/3)^3 - (1/3)(2/3)^3$</td>
<td>$(1/3)(2/3)^4$</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$n$</td>
<td>...</td>
<td>$(2/3)^n$</td>
<td>$(1/3)(2/3)^n$</td>
</tr>
</tbody>
</table>

Table 1: Size of Cantor’s set as $n \to \infty$ iterations is $\lim_{n \to \infty}(2/3)^n = 0$. 

Continuos Random Variables and Moment Generating Functions
OPRE 7310 Lecture Notes by Metin Çakanyıldırım
Compiled at 18:26 on Wednesday 27th September, 2017
As the last example shows, there are random variables that have continuous \(F_X\) at every point in their domain but fail the absolute continuity property for \(F_X\). These random variables are called singularly continuous. Their cdf’s being not absolutely continuous can map intervals of length zero to intervals of positive length. They really make something out of nothing. Singularly continuous random variables are pathological and require very special construction, so we ignore them throughout this course.

The discrete random variable \(X\) has the probability mass function \(P(X = a)\), which is mostly nonzero. For an absolutely continuous random variable, we must set \(P(X = a) = 0\). Otherwise, \(P(X = a) > 0\) implies that the set of Lebesgue measure \(A = [a, a]\) has \(P(A) > 0\), which contradicts the absolute continuity of \(X\) or its cdf \(F_X(a) = P(X \leq a)\). \(F_X\) is absolutely continuous if and only if there exists a nonnegative function \(f_X\) such that

\[
F_X(a) = \int_{-\infty}^{a} f_X(u) du.
\]

**Example:** Cantor function is differentiable almost everywhere in its domain \([0, 1]\) but it is not absolutely continuous. So differentiable almost everywhere is a weaker condition than absolute continuity. To obtain absolute continuity of \(F\) starting from differentiability almost everywhere of \(F\), we need two more conditions: i) the derivative \(f = F'\) is integrable and ii) \(F(b) - F(a) = \int_{a}^{b} f(u) du\). This is also known as the fundamental theorem of calculus for Lebesgue integral. Note that Cantor function does not satisfy ii). For more details on Cantor distribution, see §2.6 of Gut (2005). ○

In our discussion below, we focus on absolutely continuous random variables and presume that absolute continuity is equivalent to continuity. Hence, we drop “absolute” and write only continuous random variables. This convention implies that for every cdf \(F_X\) of a continuous random variable \(X\) there exists \(f_X\) satisfying \(F_X(a) = \int_{-\infty}^{a} f_X(u) du\) and \(dF(u) = f(u) du\). Since \(f_X\) exists, we call it the probability density function (pdf) of continuous random variable \(X\).

By assuming absolute continuity, we guarantee the existence of a density. However, we cannot guarantee the existence of a unique density as illustrated next.

**Example:** Consider the absolutely continuous cdf \(F(x) = \mathbb{I}_{0 \leq x \leq 1} + \mathbb{I}_{x > 1}\). Let \(f_0(x) = \mathbb{I}_{0 \leq x \leq 1}\) and let \(f_a(x) = f_0(x) + a \mathbb{I}_{x = a}\) for \(a \in [0, 1]\). Since an alteration at a single point or alterations at countably many points do not affect the integral, we have

\[
\int_{0}^{x} f_a(u) du = x = F(x) \quad \text{for } 0 \leq a, x \leq 1.
\]

Hence each function in the uncountable family \(\{f_a : a \in [0, 1]\}\) is a density for \(F(x) = \mathbb{I}_{0 \leq x \leq 1} + \mathbb{I}_{x > 1}\). ○

Although the last example provides uncountably many densities for single cdf, all densities differ from each other only at a single point. So all the densities are the same almost everywhere, except for countably many points. If you want a more interesting example with different and continuous densities, you must first note that two different continuous functions must differ over an interval \(I\); otherwise, one of them must be discontinuous. The question can be reframed as: Is it possible to find functions \(f_1, f_2\) and interval \(I\) such that \(f_1(u) \neq f_2(u)\) for \(u \in I\) and \(\int_{-\infty}^{x} f_1(u) du = \int_{-\infty}^{x} f_2(u) du\) for every \(x\)? Requiring \(\int_{-\infty}^{x} f_1(u) du = \int_{-\infty}^{x} f_2(u) du\) for every \(x\) is equivalent to requiring \(\int_{A} f_1(u) du = \int_{A} f_2(u) du\) for every set \(A\). The functions \(f_1, f_2\) that satisfy \(\int_{A} f_1(u) du = \int_{A} f_2(u) du\) for every set \(A\) must be the same almost everywhere, that is, they cannot differ over any interval. Hence, various densities associated with a cdf can differ only over a countable set and they cannot be all continuous.

Contrary to what is commonly thought, the derivation of probability models proceeds from \((\Omega, \mathcal{F}, P)\) to \(F(x) := P(X \leq x)\) without any reference to a density. A density is often an outcome of this process but not the input. The connection to a density, if it exists, is through integral. This puts integral at the center
of probability theory. For a practitioner though, the process can be reversed by assuming the existence of a
density and starting the formulation with a density. When the density exists, the probability manipulations
can be more intuitive.

**Example:** Consider a continuous random variable $X$ whose domain is $[\underline{x}, \overline{x}]$ so that its density is positive
over $[\underline{x}, \overline{x}]$. For $\underline{x} \leq a < b \leq \overline{x}$, does strict inequality between $a$ and $b$ imply strict inequality between $F(a)$
and $F(b)$? We already know that the strict inequality implies $F(a) < F(b)$ for both discrete and continuous
variables. Now by using the density of the continuous random variable:

$$F(b) = \int_{\underline{x}}^{b} f(u)du = \int_{\underline{x}}^{a} f(u)du + \int_{a}^{b} f(u)du > \int_{\underline{x}}^{a} f(u)du = F(a).$$

With the existence of a density is assumed for a continuous random variable $X$, there is tendency to relate
the probability $P(X = x)$ with the density $f_X(x)$ for an arbitrary $x$; this tendency is perhaps carried over from
$P(Y = y) = p_Y(y)$ for the discrete random variable $Y$ and its pmf $p_Y$. With the existence of a density, $X$
is absolutely continuous and the probability must map each singleton $\{x\}$ to zero: $P(X = x) = 0$. On the other
hand, $f_X(x)$ is not necessarily unique and we do not know which one to relate to $P(X = x) = 0$. In summary,
we should not seek a relationship between the probability $P(X = x)$ and the density $f_X(x)$ for continuous
random variables.

### 2 Moments and Independence

The expectation of a continuous random variable is defined analogously

$$E(X) = \int_{-\infty}^{\infty} uf(u)du.$$  

The properties listed for discrete random variables continue to hold for continuous random variables:

- **i)** Linearity of expectation: For constant $c$, $E(cg(X)) = cE(g(X))$. For functions $g_1, g_2, \ldots, g_N$,
  $E(g_1(X) + g_2(X) + \cdots + g_N(X)) = E(g_1(X)) + E(g_2(X)) + \cdots + E(g_N(X))$.
- **ii)** For constant $c$, $E(c) = c$ and $V(c) = 0$.
- **iii)** Nonlinearity of variance: For constant $c$, $V(cg(X)) = c^2V(g(X))$.
- **iv)** $V(X) = E(X^2) - (E(X))^2$.

**Example:** For a continuous random variable $X$, let $f(u) = cu^2 + u$ for $0 \leq u \leq 1$; otherwise, $f(u) = 0$. What
value of $c$ makes $f$ a legitimate density so that it integrates to one? Find $E(X)$.

$$1 = F(1) = \int_{0}^{1} (cu^2 + u)du = \left(\frac{c}{3}\right)u^3 + \left(\frac{1}{2}\right)u^2\bigg|_{u=1}^{u=0} = c/3 + 1/2,$$

so $c = 3/2$. For the expected value

$$E(X) = \int_{0}^{1} u((3/2)u^2 + u)du = \left(\frac{3}{8}\right)u^4 + \left(\frac{1}{3}\right)u^3\bigg|_{u=1}^{u=0} = 3/8 + 1/3 = 17/24.$$  

The definition of independence extends to continuous random variables $X$ and $Y$. If we have

$$P(X \leq a, Y \leq b) = P(X \leq a)P(Y \leq b) \text{ for all } a, b \in \mathbb{R},$$

random variables $X$ and $Y$ are said to be independent.
Starting from $P(X \leq a, Y \leq b) = P(X \leq a)P(Y \leq b)$, we can write $P(X \leq a, Y \leq b)$ in terms of the pdfs of $X$ and $Y$.

\[ P(X \leq a, Y \leq b) = \int_{x=a}^{b} f_X(x)dx \int_{y=a}^{b} f_Y(y)dy = \int_{x=a}^{b} \int_{y=a}^{b} f_X(x)f_Y(y)dxdy. \]

This equality allows us to interpret $f_X(x)f_Y(y)$ as the pdf of $(X, Y)$. Thus, for independent continuous random variables $(X, Y)$ the density is the product of densities of $X$ and $Y$.

**Example:** For two independent random variables, we check the equality $E(X_1X_2) = E(X_1)E(X_2)$:

\[ E(X_1X_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1x_2)f_{X_1}(x_1)f_{X_2}(x_2)dx_1dx_2 = \int_{-\infty}^{\infty} x_1f_{X_1}(x_1)dx_1 \int_{-\infty}^{\infty} x_2f_{X_2}(x_2)dx_2 = E(X_1)E(X_2). \]

The first equality above uses the fact that the $f_X(x)f_Y(y)$ is the pdf of $(X, Y)$. ◦

For two independent continuous random variables, we can check the equality $V(X_1 + X_2) = V(X_1) + V(X_2)$. In summary, we have

v) For two independent random variables $X_1$ and $X_2$, $E(X_1X_2) = E(X_1)E(X_2)$.

vi) Additivity of variance for two independent random variables $X_1$ and $X_2$, $V(X_1 + X_2) = V(X_1) + V(X_2)$.

### 3 Common Continuous Random Variables

Some commonly used continuous random variables are introduced below.

#### 3.1 Uniform Random Variable

A uniform random variable $X$ takes values between $\underline{x}$ and $\overline{x}$, and the probability of $X$ being in any subinterval of $[\underline{x}, \overline{x}]$ of a given length $\delta$ is the same:

\[ P(a \leq X \leq a + \delta) = P(b \leq X \leq b + \delta) \quad \text{for} \quad \underline{x} \leq a, b \leq a + \delta, b + \delta \leq \overline{x}. \]

Uniform random variable over $[\underline{x}, \overline{x}]$ is denoted by $U(\underline{x}, \overline{x})$.

**Example:** What is the pdf of a uniform random variable over $[\underline{x}, \overline{x}]$? Since $P(a \leq X \leq a + \delta) = P(b \leq X \leq b + \delta)$, $f(u) = c$ must be constant over $\underline{x} \leq u \leq \overline{x}$. We need $1 = F(\overline{x})$:

\[ 1 = \int_{\underline{x}}^{\overline{x}} f(u)du = \int_{\underline{x}}^{\overline{x}} cdu = c(\overline{x} - \underline{x}). \]

Hence, $f(u) = 1/(\overline{x} - \underline{x})$ for $\underline{x} \leq u \leq \overline{x}$.

**Example:** Find $E(X)$. Since the probability distributed uniformly over $[\underline{x}, \overline{x}]$, the expected value is $(\underline{x} + \overline{x})/2$. You can confirm this by using integration. ◦

**Example:** For two independent uniform random variables $X_1, X_2 \sim U(0,1)$, find $P(X_1 \geq X_2)$.

\[ P(X_1 \geq X_2) = \int_{x_1 \geq x_2} (1)(1)dx_1dx_2 = \int_{x_2=0}^{1} \int_{x_1=0}^{x_2} dx_2dx_1 = \int_{x_2=0}^{1} x_1dx_1 = x_2^2/2|_{x_2=0}^{x_2=1} = 1/2. \]

**Example:** For two independent uniform random variables $X_1, X_2 \sim U(0,1)$, find the pdf of $\max\{X_1, X_2\}$.

$P(\max\{X_1, X_2\} \leq a) = P(X_1 \leq a, X_2 \leq a) = P(X_1 \leq a)P(X_2 \leq a) = a^2$ for $0 \leq a \leq 1$. Differentiating $a^2$, we obtain $f_{\max\{X_1, X_2\}} = 2a$ for $0 \leq a \leq 1$. ◦
3.2 Exponential Random Variable

An exponential random variable $X$ takes values between 0 and $\infty$, and the probability of $X$ being in any interval $[a, a + \delta]$ exponentially decreases as $a$ increases towards $b$:

$$P(a \leq X \leq a + \delta) = \exp(-\lambda(a - b))P(b \leq X \leq b + \delta) \quad \text{for } 0 \leq a \leq b \text{ and } \delta \geq 0,$$

where the decay parameter $\lambda$ is the parameter of the distribution. First letting $\delta \to \infty$ and then setting $a = 0$, we obtain the cdf

$$P(a \leq X) = \exp(-\lambda(a - b))P(b \leq X) \implies P(b \leq X) = \exp(-\lambda b) \implies F_X(b) = P(X \leq b) = 1 - \exp(-\lambda b).$$

$F_X$ is differentiable and its derivative yields the density

$$f_X(x) = \lambda \exp(-\lambda x) \quad \text{for } 0 \leq x.$$

We use $\text{Expo}(\lambda)$ to refer to exponential random variable.

**Example:** Find $E(X^k)$. For $k = 1$ and $k = 2$ we have

$$E(X) = \int_0^\infty x \lambda e^{-\lambda x} dx = -xe^{-\lambda x}|_0^\infty + \int_0^\infty e^{-\lambda x} dx = 0 - \frac{1}{\lambda} e^{-\lambda x}|_0^\infty = \frac{1}{\lambda}.$$

$$E(X^2) = \int_0^\infty x^2 \lambda e^{-\lambda x} dx = -x^2 e^{-\lambda x}|_0^\infty + \int_0^\infty 2xe^{-\lambda x} dx = 0 + \frac{2}{\lambda} \int_0^\infty x \lambda e^{-\lambda x} dx = \frac{2}{\lambda^2}.$$

Let us suppose that $E(X^k) = k! / \lambda^k$ as an induction hypothesis and find $E(X^{k+1}) = (k + 1)! / \lambda^{k+1}$:

$$E(X^{k+1}) = \int_0^\infty x^{k+1} \lambda e^{-\lambda x} dx = -x^{k+1} e^{-\lambda x}|_0^\infty + \int_0^\infty (k + 1)x^k \lambda e^{-\lambda x} dx = \frac{k + 1}{\lambda} \int_0^\infty x^k \lambda e^{-\lambda x} dx = \frac{k + 1}{\lambda} E(X^k)$$

$$= \frac{(k + 1)!}{\lambda^{k+1}}.$$

This confirms the induction hypothesis and we conclude $E(X^k) = k! / \lambda^k$. \textbf{⋄}

From the cdf, we can obtain the complementary cdf $\bar{F}(x) = \exp(-\lambda x)$. So we have $\bar{F}(a + b) / \bar{F}(a) = \bar{F}(b)$ which gives us the memoryless property:

$$P(X > a + b | X > a) = \frac{P(X > a + b)}{P(X > a)} = \frac{\bar{F}(a + b)}{\bar{F}(a)} = \bar{F}(b) = P(X > b).$$

Translated into lifetime of a machine, the memoryless property implies the following: If a machine is functioning at time $a$ (has not failed over $[0, a]$), the probability that it will be functioning at time $a + b$ (will not fail over $[0, a + b]$) is exactly the same as the probability that it functioned for the first $b$ time units after its installation. This implies that if the machine has not failed, it is as good as new. “As good as new” assumption is used in reliability theory along with other assumptions such as “new better than used”.

3.3 Pareto Random Variable

In the exponential random variable, the complementary cdf is $P(\text{Expo}(\lambda) > x) = \exp(-\lambda x)$ so the tail of the distribution decays exponentially. Can the decay be slower than exponential? We know $\exp(\lambda x) \geq x^a$ as $x \to \infty$ for any $a$, so $\exp(-\lambda x) \leq x^{-a}$. This inspires us to build a distribution based on the power of $x$ to have slower decay in the tail or a fatter (bigger) tail with respect to exponential distribution. We define
For \( \alpha \) equal to each other: \( 1/\lambda \) exponential random variable. To make the comparison sound, we consider either.

\[ P(X \geq a) \propto a^{-\alpha} \quad \text{for } a \geq x \text{ and } a \geq 0. \]

Since \( 1 = P(X \geq x) \), we have no choice but set

\[ P(X \geq a) = \left( \frac{x}{a} \right)^{\alpha} \quad \text{for } a \geq x. \]

This leads to \( F_X(x) = 1 - (x/x)^a \) and \( f_X(x) = \alpha x^a / x^{a+1} \). We can use \( \text{Par}(a, x) \) to denote a Pareto random variable.

**Example:** Find \( E(X) \). For \( \alpha > 1 \),

\[ E(X) = \int_{x}^{\infty} x \frac{ax^a}{x^{a+1}} \, dx = \alpha \left[ \frac{-\alpha}{\alpha - 1} \frac{x^a}{x^{a-1}} \right]_{x}^{\infty} = \frac{\alpha x}{\alpha - 1}. \]

For \( \alpha = 1 \),

\[ E(X) = \int_{x}^{\infty} x \frac{x}{x^2} \, dx = \int_{x}^{\infty} \frac{x}{x} \, dx = x \log x |_{x}^{\infty} \to \infty. \]

This also implies that \( E(X) \) diverges for any \( \alpha \leq 1 \). Pareto random variable presents an example where even the first moment does not exist. Note that when the first moment does not exist, the higher moments do not either. ◊

We can build some intuition about diverging moments of Pareto random variable by comparing it to exponential random variable. To make the comparison sound, we consider \( \alpha > 1 \) and set first moments equal to each other: \( 1/\lambda = \alpha x / (\alpha - 1) \). For further specification, suppose \( x = 1 \) so that \( \alpha = 1/(1 - \lambda) \) for \( \lambda < 1 \). In short, we compare \( \text{Expo}(\lambda) \) with \( \text{Par}(1/(1 - \lambda), 1) \). In our comparison, we want to find an interval where the tail of \( \text{Par}(1/(1 - \lambda), 1) \) is larger than the tail of \( \text{Expo}(\lambda) \):

\[ P(\text{Par}(1/(1 - \lambda), 1) \geq x) = x^{-1/(1-\lambda)} \geq e^{-\lambda x} = P(\text{Expo}(\lambda) \geq x). \]

For fixed \( \lambda \) and sufficiently large \( x \), \( \ln(x)/x \leq \lambda(1 - \lambda) \). This inequality implies \( (1/(1 - \lambda)) \ln(x) \leq \lambda x \) and in turn \( x^{-1/(1-\lambda)} \geq e^{-\lambda x} \). When \( \lambda = 0.2 \), the inequality \( \ln(x)/x \leq \lambda(1 - \lambda) \) holds for \( x \geq 19 \), e.g., \( \ln(19)/19 = 0.1549 \leq 0.16 = 0.2(1 - 0.2) \). When \( \lambda = 0.4 \), it holds for \( x \geq 10, \) e.g., \( \ln(10)/10 = 0.2302 \leq 0.24 = 0.4(1 - 0.4) \). For every \( \lambda \), we can find an interval of the form \([x, \infty)\) such that Pareto random variable places more probability into this interval than Exponential random variable. In summary, Pareto random variable has a heavier tail than Exponential random variable.

Distributions whose tails are heavier than Exponential are called heavy tail distributions. Pareto is an example of heavy tail distributions. Heavy tail distributions leave significant probability to their tails where the random variable takes large values. Multiplication of these large values by large tail probabilities can yield to infinite mean and variance. This may be disappointing but does not indicate any inconsistency within the probability theory. If anything, you should be careful when applying results requiring finite moments to arbitrary random variables.

A practitioner would wonder if heavy tail random variables show up in real life? The answer is definitely. For example, wealth distribution in a population tends to have a heavy tail: there are plenty of individuals with very large amount of wealth. The drop in the number of individuals possessing a particular level of wealth is decreasing in that wealth but not exponentially; it is decreasing according to a power law. Phone conversation duration tends to have heavy tails: there are plenty of phone conversations with a very long duration. Number of tweets sent by an individual can have heavy tails: there are plenty of individuals
Erlang random variable with parameters

the second floor had 4 routers and running the 5th now. Suppose that all these routers are identical with
immediately. We are told that the first floor had 7 failed routers and running the 8th now. On the other hand,
service must be provided without an interruption as much as practically possible, a failed router is replaced

Example

X

An Erlang random variable is obtained by summing up independent exponential random variables. For

3.4 Erlang and Gamma Random Variables

An Erlang random variable is obtained by summing up independent exponential random variables. For
example, let \( X_1 \sim \text{Expo}(\lambda) \), \( X_2 \sim \text{Expo}(\lambda) \) and \( X_1 \perp X_2 \), and define \( X := X_1 + X_2 \), then \( X \) has an Erlang
distribution with parameters \((2, \lambda)\).

Example: For \( X_1 \sim \text{Expo}(\lambda) \), \( X_2 \sim \text{Expo}(\lambda) \), \( X_1 \perp X_2 \) and \( X := X_1 + X_2 \), find the pdf of \( X \). We start with
the cdf of \( X_1 + X_2 \),

\[
P(X \leq x) = P(X_1 + X_2 \leq x) = \int_0^x P(X_2 \leq x-u)\lambda e^{-\lambda u} du = \int_0^x (1 - e^{-\lambda(x-u)})\lambda e^{-\lambda u} du
\]

\[
= \int_0^x \lambda e^{-\lambda u} du - \int_0^x \lambda e^{-\lambda x} du = -e^{-\lambda x} + 1 - \lambda x e^{-\lambda x}.
\]

Then

\[
f_X(x) = \frac{d}{dx} P(X \leq x) = \lambda e^{-\lambda x} - \lambda e^{-\lambda x} + \lambda^2 x e^{-\lambda x} = \lambda e^{-\lambda x}(\lambda x). \quad \diamond
\]

If we add up \( \alpha \) many independent Exponential random variables for integer \( \alpha \), the sum turns out to be an
Erlang random variable with parameters \((\alpha, \lambda)\). The pdf of Erlang can be extended to real valued \( \alpha \) by using
\( \Gamma \) function:

\[
f_X(x) = \frac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} \quad \text{for } x \geq 0 \text{ and } \alpha > 0.
\]

You can specialize this formula to \( \alpha = 2 \) to see that it is consistent with the result of the last exercise. We use
\textit{Gamma}(\alpha, \lambda) to denote a Gamma random variable.

Example: UTD’s SOM has routers at every floor of the building to provide wireless Internet service. Since the
service must be provided without an interruption as much as practically possible, a failed router is replaced
immediately. We are told that the first floor had 7 failed routers and running the 8th now. On the other hand,
the second floor had 4 routers and running the 5th now. Suppose that all these routers are identical with
\textit{Expo}(\lambda) lifetime with parameter \( \lambda \) measured in year\(^{-1}\). \( P(\text{Gamma}(8, \lambda) \geq 10) \) represents the unconditional probability that the routers on the first floor will last for 10 years. If we are told that the SOM building is in
operation for 8 years, then this probability can be revised as \( P(\text{Gamma}(8, \lambda) \geq 10 | \text{Gamma}(8, \lambda) \geq 8) \). Similar

...
probabilities for the second floor are \( P(\text{Gamma}(5, \lambda) \geq 10) \) and \( P(\text{Gamma}(5, \lambda) \geq 10 | \text{Gamma}(5, \lambda) \geq 8) \).

**Example:** For integer \( \alpha \), find \( E(\text{Gamma}(\alpha, \lambda)) \) and \( V(\text{Gamma}(\alpha, \lambda)) \). We can start with the first two moments of Exponential random variable: \( E(\text{Expo}(\lambda)) = 1/\lambda \) and \( E(\text{Expo}(\lambda)^2) = 2/\lambda^2 \), which imply \( V(\text{Expo}(\lambda)) = 1/\lambda^2 \). Since \( \alpha \) is integer, we can think of Gamma as sum of \( \alpha \) independent exponential random variables so that we can apply the formulas for the expected value and variance of sum of independent random variables. \( E(\text{Gamma}(\alpha, \lambda)) = \alpha/\lambda \) and \( V(\text{Gamma}(\alpha, \lambda)) = \alpha/\lambda^2 \).

Lastly, \( \text{Gamma}(\nu/2, 1/2) \) gets a special name and it is called a \( \chi^2 \) random variable with \( \nu \) degrees of freedom, whose density is

\[
f_{\chi^2}(x) = \frac{x^{\nu/2-1}e^{-x/2}}{2^{\nu/2}\Gamma(\nu/2)} \text{ for } x \geq 0.
\]

In \( \text{Gamma}(\alpha, \lambda) \), \( \alpha \) is called the shape parameter while \( 1/\lambda \) is called the scale parameter, whereas \( \lambda \) can be called the rate parameter. Gamma density has many shapes that provide modelling flexibility; see Figure 1.

![Figure 1: Densities of Gamma(1,1) in black, Gamma(2,1) in red, Gamma(0.5, 1) in green and Gamma(2,2) in blue produced by R commands x <- seq(0, 8, by=.001); y <- dgamma(x, shape=1, scale=1); plot(x,y, type="l", ylab="Density", col=1); lines(x, dgamma(x, 2, 1), col=2); lines(x, dgamma(x, 0.5, 1), col=3); lines(x, dgamma(x, 2, 0.5), col=4).](image)

### 3.5 Beta Random Variable

To describe the Beta random variable, let us first define the Beta function \( B(m, k) \)

\[
B(m, k) = \int_0^1 u^{m-1}(1-u)^{k-1}du = \int_0^1 u^{k-1}(1-u)^{m-1}du = B(k, m).
\]

As we show in the Appendix, \( B(k, m) = \Gamma(k)\Gamma(m)/\Gamma(k+m) \).

Having defined the Beta function, let us consider an experiment which helps us to relate Beta random variable to the other random variables we have already studied. We consider Bernoulli trials where the
success probability is constant but unknown. That is, let $P$ be a random variable denoting the success probability and suppose that the success probabilities in each of the Bernoulli trials is $P$. After $n$ trials, suppose that we observe $m$ successes. Given these $m$ successes we can update the distribution of $P$, we are seeking $P(P = p | Bin(n, P) = m)$:

$$P(P = p | Bin(n, P) = m) = \frac{P(P = p, Bin(n, P) = m)}{P(Bin(n, P) = m)} = \frac{P(P = p, Bin(n, P) = m)}{\sum_{0 < u < 1} P(P = u, Bin(n, P) = m)}.$$  

We approximate $P(P = p, Bin(n, P) = m)$ by $f_P(p)P(Bin(n, P) = m)$ which leads to $\sum_n P(P = u, Bin(n, P) = m) \approx \int_0^1 f_P(u)P(Bin(n, u) = m)du$ where $f_P$ is the pdf of the success probability $P$. These lead to

$$f_{P|Bin(n,P)=m}(p) = \frac{f_P(p)P(Bin(n, P) = m)}{\int_0^1 f_P(u)P(Bin(n, u) = m)du} = \frac{f_P(p)\sum_{m} p^m (1-p)^{n-m}}{\int_0^1 f_P(u)\sum_{m} u^m (1-u)^{n-m}du}.$$  

To simplify further, suppose that $f_P$ is the uniform density, then

$$f_{P|Bin(n,P)=m}(p) = \frac{p^m (1-p)^{n-m}}{\int_0^1 u^m (1-u)^{n-m}du} = \frac{p^m (1-p)^{n-m}}{B(m+1, n - m + 1)}$$  

The density on the right-hand side is pdf of a new random variable, which is the success probability conditioned on $m$ successes in $n$ trials when the success probability before the trials is thought to be uniformly distributed. This random variable is called a Beta random variable.

In general, a Beta random variable $X$ can be denoted as $Beta(\alpha, \lambda)$ and its density is

$$f_X(x) = \frac{x^{\alpha-1}(1-x)^{\lambda-1}}{B(\alpha, \lambda)}$$ for $\alpha, \lambda > 0$ and $0 \leq x \leq 1$.

Beta random variable can be used to represent percentages and proportions.

**Example**: Find $E(X)$.

$$E(X) = \int_0^1 x \frac{x^{\alpha-1}(1-x)^{\lambda-1}}{B(\alpha, \lambda)} dx = \frac{1}{B(\alpha, \lambda)} \int_0^1 x^{\alpha}(1-x)^{\lambda-1} dx = \frac{B(\alpha + 1, \lambda)}{B(\alpha, \lambda)} = \frac{\Gamma(\alpha + 1)\Gamma(\lambda)}{\Gamma(\alpha + \lambda + 1)}$$

$$= \frac{\alpha}{\alpha + \lambda}. \diamond$$

**Example**: Monthly maintenance cost for a house, when measured in thousands, has a pdf $f_X(x) = 3(1-x)^2$. How much money should be budgeted each month for maintenance so that the actual cost exceeds the budget only with 5% chance. The cost has $Beta(1,3)$ random variable and we want to find budget $B$ such that $P(Beta(1,3) \geq B) = 0.05$ or $P(Beta(1,3) \leq B) = 0.95$.

$$0.95 = \int_0^B 3(1-x)^2 dx = -(1-x)^3|_0^B = 1 - (1-B)^3 \implies B = 1 - 0.05^{1/3} = 0.63$$

or $630 needs to be reserved. $ \diamond$

### 3.6 Normal Random Variable

Normal random variable plays a central role in probability and statistics. Its density has light tails so many observations are expected to be distributed around its mean. The density is a symmetric and has a range over the entire real line, so the normal random variable can take negative values. The density for normally distributed $X$ is

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$ for $\sigma > 0$ and $\mu, x \in \mathbb{R}$. 
The cumulative distribution function at an arbitrary \( x \) cannot be found analytically so we refer to published tables or software for probability computation. Nevertheless, we can analytically integrate over the entire range. We use \( \text{Normal}(\mu, \sigma^2) \) to denote a normal random variable with mean \( \mu \) and standard deviation \( \sigma \).

**Example:** Establish that the normal density integrates to 1 over \( \Re \). We first establish by using \( z = (x - \mu) / \sigma \) that

\[
\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi \sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \, dz
\]

Equivalent to checking that the last integral is 1 is to check that the multiplication of integrals below is 1:

\[
\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z_1^2/2} \, dz_1 \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z_2^2/2} \, dz_2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{(z_1^2 + z_2^2)}{2}} \, dz_1 \, dz_2.
\]

Now consider the transformation \( r^2 = z_1^2 + z_2^2 \) and \( \theta = \arccos(z_1/r) \). Hence, \( z_1 = r \cos \theta \) and \( z_2 = r \sin \theta \). The Jacobian for this transformation is

\[
J(r, \theta) = \begin{bmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{bmatrix},
\]

whose determinant is just \( r \). Thus,

\[
\frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{(z_1^2 + z_2^2)}{2}} \, dz_1 \, dz_2 = \frac{1}{2\pi} \int_{0}^{2\pi} \int_{0}^{\infty} r e^{-r^2/2} \, dr \, d\theta = \frac{1}{2\pi} \int_{0}^{2\pi} d\theta \int_{0}^{\infty} \frac{re^{-r^2/2}}{2\pi} \, dr = 1.
\]

**Example:** Find \( E(X) \). Using \( z = (x - \mu) / \sigma \)

\[
E(X) = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi \sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx = \left[ \int_{-\infty}^{\infty} (\mu + z\sigma) \frac{1}{\sqrt{2\pi \sigma^2}} e^{-\frac{z^2}{2\sigma^2}} \, dz \right] \mu + \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \, dz + \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{0} ze^{-z^2/2} \, dz = \mu.
\]

The last integral is zero because \( ue^{-u^2/2} = (-u)e^{(-u)^2/2} \) and so \( \int_{-\infty}^{\infty} ze^{-z^2/2} \, dz = \int_{\infty}^{0} (-u)e^{-(u)^2/2} \, (-du) = \int_{0}^{\infty} ue^{-u^2/2} \, du = -\int_{\infty}^{0} ue^{-u^2/2} \, du \).

**Example:** What is the distribution of \( aX + b \) if \( X \sim \text{Normal}(\mu, \sigma^2) \). We can consider \( P(aX + b \leq u) = P(X \leq (u - b) / a) \) and differentiate this with respect to \( u \):

\[
\frac{d}{du} P(X \leq (u - b) / a) = \frac{d}{du} \int_{-\infty}^{(u-b)/a} \frac{1}{\sqrt{2\pi \sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx = \frac{1}{a} \frac{1}{\sqrt{2\pi \sigma^2}} e^{-\frac{((u-b)/a-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi a\sigma^2}} e^{-\frac{(u-b)^2}{2a^2\sigma^2}}.
\]

\( aX + b \) has \( \text{Normal}(\mu a, b, a^2 \sigma^2) \) distribution.

From the last example, we can see that any random variable can be standardized by deducting its mean and dividing by its standard deviation, the standard normal random variable is \( \text{Normal}(0, 1) \).

**Example:** We know \( E(\text{Normal}(\mu, \sigma^2)) = \mu \) obtained from integrating over \( (-\infty, \infty) \), find the partial integration over \( (-\infty, a) \).

\[
E(\text{Normal}(\mu, \sigma^2) \mathbb{1}_{\text{Normal}(\mu, \sigma^2) \leq a}) = \int_{-\infty}^{a} x \frac{1}{\sqrt{2\pi \sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx = \int_{-\infty}^{(a-\mu)/\sigma} (\mu + z\sigma) \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \, dz
\]

\[
= \mu F_{\text{Normal}(0,1)} \left( \frac{a-\mu}{\sigma} \right) + \sigma \int_{\infty}^{(a-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} e^{-u^2} \, du
\]

\[
= \mu F_{\text{Normal}(0,1)} \left( \frac{a-\mu}{\sigma} \right) - \sigma f_{\text{Normal}(0,1)} \left( \frac{a-\mu}{\sigma} \right)
\]
Above we use $u = z^2/2$ substitution. This formula can be used as a starting point to compute expected values of some other functions involving $X$ and $a$. ♦

### 3.7 Lognormal Random Variable

$X$ is a lognormally distributed random variable if $\ln X$ is $\text{Normal}(\mu, \sigma^2)$. That is, $X$ can be expressed as

$$X = e^Y \text{ for } Y \sim \text{Normal}(\mu, \sigma^2).$$

So the range for $X$ is $[0, \infty)$. Having a nonnegative range can make the lognormal distribution more useful than the normal distribution to model quantities such as income, city sizes, marriage age and number of letters in spoken words, which are all nonnegative.

**Example:** Find the pdf of $X$.

$$P(X \leq x) = P(Y \leq \ln x) = \int_{-\infty}^{\ln x} \frac{1}{\sqrt{2\pi \sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy.$$  

Taking the derivative with respect to $x$

$$f_X(x) = \frac{d}{dx}P(X \leq x) = \frac{1}{x} \frac{1}{\sqrt{2\pi \sigma^2}} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}.$$  

**Example:** Find $E(X^k)$

$$E(X^k) = E((e^Y)^k) = E(e^{kY}) = \int_{-\infty}^{\infty} e^{ky} \frac{1}{\sqrt{2\pi \sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy = e^{k\mu} \int_{-\infty}^{\infty} e^{k\mu u} \frac{1}{\sqrt{2\pi \sigma^2}} e^{-\frac{u^2}{2\sigma^2}} du = e^{k\mu} e^{k^2\sigma^2/2}.$$  

### 3.8 Mixtures of Random Variables

Given two random variable $X$ and $Y$, their mixture $Z$ can be obtained as

$$Z = \left\{ \begin{array}{ll} X & \text{with probability } p \\ Y & \text{with probability } 1 - p \end{array} \right\}.$$  

Then we have

$$F_Z(z) = P(Z \leq z) = pP(X \leq z) + (1 - p)P(Y \leq z) = pF_X(z) + (1 - p)F_Y(z).$$  

We can easily obtain the pdf of $Z$ if both $X$ and $Y$ are continuous random variables.

When one of the mixing distributions is discrete and another one is continuous, we do not have a pdf.

**Example:** Consider the mixture of a mass at $[Z = 0]$ and an exponential distribution

$$F_Z(z) = \left\{ \begin{array}{ll} q & \text{if } z = 0 \\ \frac{q}{q + (1 - q)(1 - \exp(-\lambda z))} & \text{if } z > 0 \end{array} \right\}.$$  

This cdf is not continuous at 0, so $Z$ is not a continuous random variable. $Z$ takes values in $\mathbb{R}$, its range is uncountable, so $Z$ is not a discrete random variable either. ♦
Random variables derived from others through max or min operators can have masses at the end of their ranges. Product returns due to unmet expectations to a retailer can be treated as negative demand, then the demand \( D \) takes values in \( \mathbb{R} \). If you want only the new customer demand for products, you should consider \( \max\{0, D\} \) which has a mass at 0. When a person buys a product, you know that his willingness to pay \( W \) is more than the price \( p \). Willingness-to-pay of customers who already bought the product is \( \max\{p, W\} \). When you have a capacity of \( c \) and accept only demand that is less than the capacity, the accepted demand becomes \( \min\{c, D\} \). The last two derived random variables have masses at \( p \) and \( c \).

4 Moment Generating Functions

A variable \( X \) can be summarized by its mean \( \mathbb{E}(X) \), second moment \( \mathbb{E}(X^2) \) and variance \( \text{Var}(X) \). But many random variables have the same mean and variance (so the same second moment). To distinguish random variables we can use their moment generating functions: \( m_X(t) := \mathbb{E}(\exp(tX)) \). For discrete and continuous nonnegative random variables, the moment generating functions are obtained in analogous ways:

\[
m_X(t) = \sum_{x=\infty} \exp(tx)p_X(x) \quad \text{and} \quad m_X(t) = \int_{-\infty}^{\infty} \exp(tx)f_X(x)dx.
\]

Although the moment generating function is defined for every \( t \), we need it to be finite in a neighborhood of \( t = 0 \). This technical condition holds for most of the random variables.

By using Taylor expansion of \( \exp(t) \) at \( t = 0 \),

\[
\exp(tx) = 1 + tx + \frac{(tx)^2}{2!} + \frac{(tx)^3}{3!} + \cdots + \frac{(tx)^i}{i!} + \cdots
\]

Inserting this into \( m_X(t) \) for a discrete random variable,

\[
m_X(t) = \sum_{x=-\infty} \left( 1 + tx + \frac{(tx)^2}{2!} + \cdots + \frac{(tx)^i}{i!} + \cdots \right) p_X(x)
\]

\[
= 1 + t \mathbb{E}(X) + \frac{t^2}{2!} \mathbb{E}(X^2) + \cdots + \frac{t^i}{i!} \mathbb{E}(X^i) + \cdots
\]

Form the last equality, we obtain \( m_X(t = 0) = 1 \). To pick up only the \( \mathbb{E}(X) \) term on the right-hand side, we consider the derivative of \( m_X(t) \) with respect to \( t \). Provided that derivative can be pushed into the summation in the last equation above, we can obtain

\[
\frac{d^k}{dt^k} m_X(t) = \sum_{i=0}^{\infty} \mathbb{E}(X^i) \frac{d^k}{dt^k} \frac{t^i}{i!} = \mathbb{E}(X^k) + \sum_{i=k+1}^{\infty} \mathbb{E}(X^i) \frac{(i-1) \cdots (i-k+1) t^{i-k}}{(i-k)!} = \mathbb{E}(X^k) + \sum_{i=k+1}^{\infty} \mathbb{E}(X^i) \frac{t^{i-k}}{(i-k)!}.
\]

For example,

\[
\frac{d}{dt} m_X(t) = \mathbb{E}(X^1) + \sum_{i=2}^{\infty} \mathbb{E}(X^i) \frac{t^{i-1}}{(i-1)!} \quad \text{and} \quad \frac{d^2}{dt^2} m_X(t) = \mathbb{E}(X^2) + \sum_{i=3}^{\infty} \mathbb{E}(X^i) \frac{t^{i-2}}{(i-2)!}.
\]

Then

\[
\mathbb{E}(X^k) = \left. \frac{d^k}{dt^k} m_X(t) \right|_{t=0}.
\]
The last equation makes it clear why \( m_X(t) \) is called a moment generating function.

If two moment generating functions are equal, what can we say about the corresponding random variables? Moment generating functions are related to power series in the appendix to motivate that \( m_X(t) = m_Y(t) \) implies \( X \sim Y \). So moment generating functions uniquely identify random variables.

**Example:** If \( X \) and \( Y \) are two independent random variables with \( m_X \) and \( m_Y \), what is the moment generating function \( m_{X+Y} \) of the sum of these variables? We proceed directly as

\[
m_{X+Y}(t) = E(\exp(t(X+Y))) = E(\exp(tX))E(\exp(tY)) = m_X(t)m_Y(t),
\]

where independence of \( \exp(tX) \) and \( \exp(tY) \) is used in the middle equality. ◦

**Example:** Find the moment generating function for the Bernoulli random variable. Use this function to obtain the moment generating function of the Binomial random variable. For a Bernoulli random variable \( X_i \) with success probability \( p \),

\[
m_{X_i}(t) = E(\exp(tX_i)) = \exp(t0)(1-p) + \exp(t1)p = (1-p) + p \exp(t).
\]

Binomial random variable \( B(n,p) \) is a sum of \( n \) Bernoulli random variables so

\[
m_{\sum_{i=1}^{n} X_i}(t) = \prod_{i=1}^{n} m_{X_i}(t) = ((1-p) + p \exp(t))^n. \quad ◦
\]

**Example:** For an exponential random variable \( X \) find the moment generating function \( m_X(t) \) and the moments \( E(X^k) \).

\[
m_X(t) = E(e^{tX}) = \int_{0}^{\infty} e^{tx} \lambda e^{-\lambda x} dx = \lambda \int_{0}^{\infty} e^{-(\lambda - t)x} dx = \lambda \left(-\frac{1}{\lambda - t} e^{-(\lambda - t)x}\right) \bigg|_{0}^{\infty} = \lambda \left(-\frac{1}{\lambda - t}\right) = \lambda/(\lambda - t).
\]

Then

\[
E(X^k) = \frac{d^k}{dt^k} m_X(t) \bigg|_{t=0} = \lambda \frac{d^k}{dt^k} (\lambda - t)^{-1} \bigg|_{t=0} = \lambda(k!) (\lambda - 1)^{-k} \bigg|_{t=0} = \frac{k!}{\lambda^k}. \quad ◦
\]

**Example:** Find the moment generating function for \( Gamma(n, \lambda) \) for integer \( n \geq 1 \).

\[
m_{Gamma(n, \lambda)}(t) = \prod_{i=1}^{n} m_{Exp(\lambda)}(t) = \left(\frac{\lambda}{\lambda - t}\right)^n. \quad ◦
\]

**Example:** Find the moment generating function for \( X = Normal(\mu, \sigma^2) \) and use this moment generating function to obtain the distribution of two independent Normal random variables \( X_1 + X_2 \), where \( X_1 = Normal(\mu_1, \sigma_1^2) \) and \( X_2 = Normal(\mu_2, \sigma_2^2) \). We have

\[
m_X(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-0.5(x-\mu)^2/\sigma^2} dx = e^{t\mu} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{tz} e^{-0.5z^2} dz = e^{t\mu + 0.5\sigma^2t^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-0.5(z-\sigma t)^2} dz = 1
\]

where the second equality follows from \( x = \mu + \sigma z \) and the third equality follows from \(-0.5z^2 + \sigma tz = -0.5(z-\sigma t)^2 + 0.5\sigma^2t^2 \).

\[
m_{X_1+X_2}(t) = m_{X_1}(t)m_{X_2}(t) = e^{t\mu_1 + 0.5\sigma_1^2t^2} e^{t\mu_2 + 0.5\sigma_2^2t^2} = e^{(\mu_1+\mu_2)t + 0.5(\sigma_1^2 + \sigma_2^2)t^2},
\]
so \( X_1 + X_2 \sim Normal(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2) \).

**Example:** For a random variable \( X \), the moment generating function is argued to be
\[
m_X(t) = 1 + \sum_{i=1}^{\infty} \frac{t^i}{(i-1)!}.
\]
Find \( E(X), E(X^2), E(X^k) \) for \( k \geq 3 \) and determine if \( m_X(t) \) can be a valid moment generating function.
\[
E(X) = \frac{d}{dt} m_X(t) \bigg|_{t=0} = 1 + \frac{2t}{1!} + \frac{3t^2}{2!} + \frac{4t^3}{3!} + \ldots \bigg|_{t=0} = 1.
\]
\[
E(X^2) = \frac{d^2}{dt^2} m_X(t) \bigg|_{t=0} = 2 + \frac{3(2)t}{2!} + \frac{4(3)t^2}{3!} + \ldots \bigg|_{t=0} = 2.
\]
\[
E(X^3) = \frac{d^3}{dt^3} m_X(t) \bigg|_{t=0} = 3 + \frac{4(3)(2)t}{3!} + \ldots \bigg|_{t=0} = 3.
\]
\[
E(X^4) = \frac{d^4}{dt^4} m_X(t) \bigg|_{t=0} = 4 + \ldots \bigg|_{t=0} = 4.
\]
\[
E(X^k) = \frac{d^k}{dt^k} m_X(t) \bigg|_{t=0} = k! \frac{k!}{(k-1)!} + \frac{k\ldots(k-2)t}{k!} + \frac{3t^2}{2} + \ldots \bigg|_{t=0} = k.
\]
Let \( Y = X^3 \) a new random variable. We have \( V(Y) = E(Y^2) - E(Y)^2 = E(X^6) - E(X^3)^2 = d^6m_X(t)/dt^6 - (d^3m_X(t)/dt^3)^2 \bigg|_{t=0} = 6 - 3^2 = -3. \) This points to a contradiction and \( m_X(t) \) cannot be moment generating function for any random variable. A moment generating function needs to yield a “positive” value when it is differentiated, differenced and evaluated in certain ways.

As the last example shows, an arbitrary function cannot be a moment generating function. The last example hints at some sort of “positivity” property for a moment generating function. This can be a starting point to answer the question: Is there a set of conditions to check on \( m_X(t) \) to decide if it corresponds to random variable \( X \) or its pmf \( p_X \) / pdf \( f_X \)? This question has a complicated answer that goes well beyond the scope of this note. It is answered fully for a similar transformation (characteristic function rather than moment generating function). A characteristic function must be positive definite; see appendix for details.

### 5 Application: From Lognormal Distribution to Geometric Brownian Motion

If \( S_t \) denotes the price of a security (a stock, a bond, or their combinations) at time \( t \geq 0 \) for given \( S_0 = s_0 \) known now, we can consider
\[
\frac{S_t}{S_0} = \frac{S_t}{S_{t-1}} \cdot \frac{S_{t-1}}{S_{t-2}} \cdot \ldots \cdot \frac{S_2}{S_1} \cdot \frac{S_1}{S_0}.
\]
Suppose that each ratio \( S_t/S_{t-1} \) is independent of the history \( S_{t-1}, S_{t-2}, \ldots, S_1, S_0 \) and is given by a Lognormal distribution with parameters \( \mu \) and \( \sigma \).

Since \( S_{i+1}/S_i \) and \( S_i/S_{i-1} \) are independent and lognormally distributed with parameters \( \mu \) and \( \sigma^2 \), we have \( \ln S_{i+1} - \ln S_i \sim Normal(\mu, \sigma^2) \) and \( \ln S_i - \ln S_{i-1} \sim Normal(\mu, \sigma^2) \). Their sum is \( \ln S_{i+1} - \ln S_i = \ln S_{i+1} - \ln S_i + \ln S_i - \ln S_{i-1} \sim Normal(2\mu, 2\sigma^2) \). Hence, \( S_{i+1}/S_{i-1} \) has lognormal distribution with parameters \( 2\mu \) and \( 2\sigma^2 \). This line of logic yields that \( S_t/S_0 \) has lognormal distribution with parameters \( t\mu \) and \( t\sigma^2 \).
Another way to represent security price at time $t$ is to write

$$S_t = s_0 X_t,$$

where $X_t$ is a lognormal random variable with parameters $t\mu$ and $t\sigma^2$. Then we can immediately obtain the mean and variance of the security price

$$
E(S_t) = s_0 E(X_t) = s_0 e^{t\mu + t\sigma^2 / 2},
$$

$$
V(S_t) = s_0^2 V(X_t) = s_0^2 (e^{2t\mu + 2t\sigma^2} - e^{2t\mu + t\sigma^2}) = s_0^2 e^{2t\mu + t\sigma^2} (e^{t\sigma^2} - 1)
$$

$S_t$ as constructed above is a Geometric Brownian Motion process. Note we have not constructed the process at noninteger times, so strictly speaking we have constructed a discrete-time process that coincides with the Geometric Brownian Motion at integer times.
6 Exercises

1. Find $E(Beta(a, b)^2)$.

2. Let us consider the following transform $n_X(t) = \ln E(\exp(tX))$. Clearly, $n_X(0) = 0$. Determine what moments $(d/dx)n_X(x)|_{x=0}$ and $(d^2/dx^2)n_X(x)|_{x=0}$ correspond to. That is, express these in terms of $E(X^k)$ for integer $k$.

3. For the transform $n_X(t) = \log E(\exp(tX))$, what is $n_{X_i}(t) = \log E(\exp(tX_i))$ for Bernoulli distributed $X_i$? Can you use $n_{X_i}(t)$ to easily compute the $n_X(t)$ for the Binomial random variable $X = \sum_{i=1}^{n} X_i$, how?

4. For two independent random variables $X$ and $Y$, let $X \sim U(0,1)$ and $Y \sim U(0,1)$ and find the pdf of $X + Y$.

5. Let $Y \sim Expo(\lambda)$ and set $X := \lfloor Y \rfloor + 1$. $\lfloor y \rfloor$ is the floor function for a possibly fractional number $y$, it returns the largest integer smaller than or equal to $y$. What is the distribution of $X$; is it one of the common distributions?

6. Which random variable has the moment generating function $m(t) = \exp((\mu + at))(1 - \exp(-2at))/(2at)$ for parameters $\mu, a$?

7. When $X$ is demand and $a$ is inventory, $E(1_{X\geq a}(X - a))$ is the expected inventory shortage. For $X = Normal(\mu, \sigma^2)$ and $Z = Normal(0,1)$, establish that

   Inventory Shortage: $E(1_{X\geq a}(X - a)) = \sigma f_Z\left(\frac{a-\mu}{\sigma}\right) - (a - \mu) \left(1 - F_Z\left(\frac{a-\mu}{\sigma}\right)\right)$.

8. When $X$ is demand and $a$ is inventory, $E(1_{X\leq a}(a - X))$ is the expected leftover inventory. For $X = Normal(\mu, \sigma^2)$ and $Z = Normal(0,1)$, establish that

   Leftover Inventory: $E(1_{X\leq a}(a - X)) = (a - \mu)F_Z\left(\frac{a-\mu}{\sigma}\right) + \sigma f_Z\left(\frac{a-\mu}{\sigma}\right)$.

9. When $X$ is demand and $a$ is inventory, express $E(1_{X\leq a}(a - X)) - E(1_{X\geq a}(X - a))$ in English. For $X = Normal(\mu, \sigma^2)$, find this quantity in terms of $a, \mu, \sigma, \text{cdf and pdf of } Z = Normal(0,1)$.

10. For $X = Normal(\mu, \sigma^2)$ find a formula to compute $E(\min\{X, a\})$ in terms of only $a, \mu, \sigma, \text{cdf and pdf of } Z = Normal(0,1)$.

11. We set October as month 0, and use $S_1$ and $S_2$ to denote stock prices for months of November and December. We assume that $S_2/S_1$ and $S_1/S_0$ have independent Lognormal distributions. In other words, $\log(S_2/S_1)$ and $\log(S_1/S_0)$ both have normal distributions with mean 0.09 and standard deviation 0.06. Find the probability that the stock price increases either in October or November. Find the probability that the stock price increases at the end of two months with respect to the initial price in October.

12. A bond’s price is normalized to $S_0 = 1$ last year. The price changes every quarter by a multiple $M_t$ so the price $S_t$ at the end of quarter $t$ is given by $S_t = M_t S_{t-1}$ for $S_0 = s_0$. The logarithm of the multiple $M_t$ has normal distribution whose variance remains constant at $\sigma^2$ over the quarters but its mean increases proportionally $E(\ln M_t) = \mu t$. What is the distribution of $S_4$ at the end of the year? Suppose that $\mu = 0.03$ and $\sigma = 0.1$, what is the probability that bond’s price at least doubles in a year?
13. A nonnegative random variable \( X \) has finite \( k \)th moment if \( E(X^k) < \infty \). For a given positive integer \( k \), we can compute the \( k \)th moment with an alternative formula by using the tail probability as

\[
E(X^k) = m \int_0^\infty x^n P(X > x)dx
\]

for appropriate numbers \( m, n \leq k \). Find out what \( m, n \) are in terms of \( k \). Specialize your formula to \( k = 1 \) and \( k = 2 \) and provide these special versions.

14. A lake hosts \( N = 3 \) frogs, which start taking sunbathes together everyday. Frogs are very particular about the dryness of their skin so they jump into water after a while and finish their sunbathing. Each frog takes a sunbathe for \( \text{Exp}(\lambda) \) amount of time independent of other frogs. Let \( X_n \) be the duration of the time at least \( n \) frogs are sunbathing together. Find the cdfs of \( X_1 \) and \( X_3 \). Does \( X_1 \) or \( X_3 \) have one of the common distributions?

15. \(^1\) Let \( X_1, X_2, \ldots, X_n \) be random variables and their sum be \( S_n = X_1 + X_2 + \cdots + X_n \). Furthermore, suppose that \( S_n \) is a Gamma random variable with parameters \((n, \lambda)\).
   a) Is it always true that each \( X_i \) is an exponential random variable with parameter \( \lambda \)? If yes, prove this statement. Otherwise, provide a counterexample.
   
   b) Suppose that \( X_i \)'s are identically distributed. Is it always true that each \( X_i \) is an exponential random variable with parameter \( \lambda \)? If yes, prove this statement. Otherwise, provide a counterexample.
   
   c) Suppose that \( X_i \)'s are independent and identically distributed. Is it always true that each \( X_i \) is an exponential random variable with parameter \( \lambda \)? If yes, prove this statement. Otherwise, provide a counterexample.
   
   d) Suppose that \( X_i \)'s are independent. Is it always true that each \( X_i \) is an exponential random variable with parameter \( \lambda \)? If yes, prove this statement. Otherwise, provide a counterexample.

16. Suppose that \( X_i \) is the wealth of individuals in country \( i \), whose population is \( n_i \), for \( i = 1, 2 \). Let \( X_1, X_2 \) be iid Pareto random variables such that \( P(X_i \geq a) = 1/a \) for \( a \geq 1 \). \( P(X_i \geq a) \) can also be interpreted as the percentage of the population with wealth \( a \) or more. We make some groups of individuals by picking \( n_1 \) individual from country 1 and \( n_2 \) individuals country 2 such that \( n_1/n_2 = n_1/n_2 \): The representation of each country in these groups is proportional to its sizes. Let \( X \) be the average wealth of individuals in any one of these groups: \( X = (\tilde{n}_1 X_1 + \tilde{n}_2 X_2) / (\tilde{n}_1 + \tilde{n}_2) \). Groups have identically distributed average wealths and we are interested in the distribution of this wealth.
   a) The tail probability of \( X \) can be expressed as

\[
P(X \geq a) = \frac{1}{a} + \left( \frac{a}{\rho} \right)^{-1} \left( \frac{a}{1-\rho} \right)^{-1} \left\{ \ln \left( \frac{a}{\rho} - 1 \right) + \ln \left( \frac{a}{1-\rho} - 1 \right) \right\}
\]

for an appropriate parameter \( \rho \). Derive this tail probability expression and find what exactly \( \rho \) is. Hint: For \( a - mx \geq 0 \), \( \int (1/(a - mx))(1/x^2)dx = (m/a^2)\left( \ln \left( x/(a - mx) \right) \right) - 1/(ax) \).
   
   b) Evaluate the tail probability formula obtained above at extreme values of range of \( X \) to see if you get 0 and 1.

\(^1\)Inspired by a question from Mehdi Hosseinalabadi, Probability Class 2015.
17. Suppose that $X_i$ is the wealth of individuals in country $i$, whose population is $n_i$, for $i = 1, 2$. Let $X_1, X_2$ be iid Pareto random variables such that $P(X_i \geq a) = 1/a$ for $a \geq 1$. $P(X_i \geq a)$ can also be interpreted as the percentage of the population with wealth $a$ or more. We make some groups of individuals by picking $\tilde{n}_1$ individual from country 1 and $\tilde{n}_2$ individuals country 2 such that $\tilde{n}_1/\tilde{n}_2 = n_1/n_2$: The representation of each country in these groups is proportional to its sizes. Let $X$ be the average wealth of individuals in any one of these groups: $X = (\tilde{n}_1X_1 + \tilde{n}_2X_2) / (\tilde{n}_1 + \tilde{n}_2)$. The tail probability of $X$ can be expressed as

$$
P(X \geq a) = \frac{1}{a} + \left(\frac{a}{\rho}\right)^{-1} \left(\frac{a}{1-\rho}\right)^{-1} \left\{ \ln \left(\frac{a}{\rho} - 1\right) + \ln \left(\frac{a}{1-\rho} - 1\right) \right\}.
$$

a) Evaluate the tail probability as one of the countries becomes very large, i.e., $n_i \to \infty$.

b) Evaluate the tail probability when the countries have the same size.

c) When countries have the same size, what is the largest difference between $P(X \geq a)$ and $P(X_1 \geq a)$, i.e., what is the maximum value of $|P(X \geq a) - P(X_1 \geq a)|$?

18. Suppose that $X_i$ is the wealth of individuals in country $i$, whose population is $n_i$, for $i = 1, 2$. Let $X_1$ be Pareto random variable such that $P(X_i \geq a) = 1/a$ for $a \geq 1$. People in country 2 are twice as rich as the people in country 1 and their wealth also has a Pareto distribution.

a) Find the tail probability of $X_2$, i.e., $P(X_2 \geq a)$.

b) Suppose $X_1$ and $X_2$ are independent. If we randomly pick an individual from either country 1 or 2, what is the probability that he has wealth $a$ or more. Does the wealth of a randomly chosen individual have a Pareto distribution? Does the tail of this wealth distribution resemble the tail of a Pareto distribution for some range of values? Identify that range, if there is one.
7 Appendix: Change of Variables in Double Integrals and Beta Function

Suppose that we want to write \( \int_S f(x,y)dx\,dy \) as \( \int_T F(u,v)dudv \). This is changing variables from \((x,y)\) to \((u,v)\). The initial region \( S \) is in the \( xy \) plane while its transform \( T \) is in \( uv \) plane. We assume that the transformation between \( S \) and \( T \) is one-to-one. Then there will be functions \( U, V : S \to T \), explicitly written as

\[
u = U(x,y) \text{ and } v = V(x,y) \quad \text{for} \quad (x,y) \in S, \quad (u,v) \in T,
\]
as well as functions \( X, Y : T \to S \)

\[
x = X(u,v) \text{ and } y = Y(u,v) \quad \text{for} \quad (x,y) \in S, \quad (u,v) \in T.
\]

Note that \( X \) and \( Y \) are functions of \((u,v)\), so we can think of partial derivatives of the form \( \partial X / \partial u, \partial X / \partial v, \partial Y / \partial u, \partial Y / \partial v \). If the partials do not exist at certain \((x,y)\) and differ in some subsets of \( S \), you can consider the transformation over each subset separately from the other subsets. With the partials, Jacobian \( J \) matrix can be made up

\[
J(u,v) = \begin{vmatrix} \frac{\partial X}{\partial u} & \frac{\partial X}{\partial v} \\ \frac{\partial Y}{\partial u} & \frac{\partial Y}{\partial v} \end{vmatrix}.
\]

What we need is the determinant \( det(J(u,v)) \) of the Jacobian. Note that there is no precedence between \( X \) and \( Y \), so the Jacobian for the desired transformation might be constructed as

\[
J'(u,v) = \begin{vmatrix} \frac{\partial Y}{\partial u} & \frac{\partial Y}{\partial v} \\ \frac{\partial X}{\partial u} & \frac{\partial X}{\partial v} \end{vmatrix}.
\]

We have \( det(J(u,v)) = -det(J'(u,v)) \), so which determinant should we use? The answer is that shuffling rows or columns of the Jacobian does not matter because what we actually use is the absolute value of the determinant of the Jacobian. Then we have the change-of-variable formula

\[
\int_S f(x,y)dx\,dy = \int_T f(X(u,v), Y(u,v)) |J(u,v)| dudv.
\]

Example: Use change of variable rule above to establish the connection between Beta function and Gamma function:

\[
B(k,m) = \int_0^1 u^{k-1}(1-u)^{m-1}du = \frac{\Gamma(k)\Gamma(m)}{\Gamma(k+m)}.
\]

We start with \( \Gamma(k)\Gamma(m) \) and proceed as follows

\[
\Gamma(k)\Gamma(m) = \int_0^\infty e^{-x}x^{k-1}dx \int_0^\infty e^{-y}y^{k-1}dy = \int_0^\infty \int_0^\infty e^{-(x+y)}x^{k-1}y^{k-1}dx\,dy.
\]

Let \( u = x/(x+y) \) and \( v = x+y \) so that \( x = X(u,v) = uv \) and \( Y(u,v) = v(1-u) \). Then the Jacobian matrix is

\[
J(u,v) = \begin{vmatrix} v & u \\ -v & 1-u \end{vmatrix},
\]

whose determinant is \( v(1-u) + vu = v \). Hence,

\[
\Gamma(k)\Gamma(m) = \int_0^1 \int_0^1 e^{-v(1-u)}(1-u)^{m-1}u^{k-1}dudv = \int_0^\infty e^{-v}v^{m+k-1}dv \int_0^1 u^{k-1}(1-u)^{m-1}du,
\]

which can be written as \( \Gamma(k)\Gamma(m) = \Gamma(k+m) \int_0^1 u^{k-1}(1-u)^{m-1}du \) to establish the desired equality. \( \diamond \)
8 Appendix: Induction over Natural Numbers

In these notes, we take natural numbers as \( \mathcal{N} = \{0, 1, 2, \ldots \} \) by choosing to include 0 in \( \mathcal{N} \). However, \{0, 1, 2, \ldots \} is not a workable definition for natural numbers and needs formalization. This formalization is not trivial and achieved in late 19th century by Peano. It requires 5 axioms to define the set of natural numbers.

\( P1: \) 0 \( \in \mathcal{N} \).

\( P2: \) There is a successor function \( \delta : \mathcal{N} \to \mathcal{N} \). For each \( x \in \mathcal{N} \), there is a successor \( \delta(x) \) of \( x \) and \( \delta(x) \in \mathcal{N} \).

\( P3: \) For every \( x \in \mathcal{N} \), we have \( \delta(x) \neq 0 \).

\( P4: \) If \( \delta(x) = \delta(y) \) for \( x, y \in \mathcal{N} \), then \( x = y \).

\( P5: \) If \( S \subseteq \mathcal{N} \) is such that i) 0 \( \in S \) and ii) \( x \in \mathcal{N} \cap S \implies \delta(x) \in S \), then \( S = \mathcal{N} \).

By \( P1 \) and \( P3 \), 0 can be thought as the initial element of natural numbers in the sense that 0 is not a successor of any other natural numbers. \( P2 \) introduces the successor function and \( P4 \) requires that if the successors of two numbers are equal, then these numbers are equal. \( P5.i \) is the base step for initializing \( S \) and \( P5.ii \) is the induction step for enlargening \( S \).

These axioms are adapted from p.23 of Gunderson (2011) – “Handbook of Mathematical Induction”. In the next few pages of his book, Gunderson uses axioms \( P1 – P5 \) to prove some properties of natural numbers such as:

- A natural number cannot be a successor of its own: If \( x \in \mathcal{N} \), then \( \delta(x) \neq x \).
- Each natural number except 0 has a unique predecessor: If \( x \in \mathcal{N} \setminus \{0\} \), then there exists \( y \) such that \( x = \delta(y) \).
- There exists a unique function \( f : \mathcal{N} \times \mathcal{N} \to \mathcal{N} \) such that \( f(x, 1) = \delta(x) \) and \( f(x, \delta(y)) = \delta(f(x, y)) \) for \( x, y \in \mathcal{N} \).

This function \( f \) turns out to be summation: \( f(x, y) = x + y \). Hence, \( \delta(x) = f(x, 1) = x + 1 \) is the successor of \( x \). Hence, natural numbers start with 0, includes its successor 1\( = 0+1 \), includes 2\( = 1+1 \), and so on.

We can now present the idea of induction as a theorem and prove it.

**Induction Theorem:** If \( S(n) \) is a statement involving \( n \) and if i) \( S(0) \) holds and ii) For every \( k \in \mathcal{N} \), \( S(k) \implies S(k+1) \), then \( S(n) \) holds for every \( n \in \mathcal{N} \).

One cannot help but realize the similarity between the theorem statement and axiom \( P5 \). So the proof is based on \( P5 \). Let \( A = \{ n \in \mathcal{N} : S(n) \text{ is true} \} \). Then by condition i) in the theorem statement, 0 \( \in A \). By ii) in the theorem statement and \( S(k = 0) \) true, we obtain that \( S(k + 1) = 1 \) is true. Then 1 \( \in A \). In general, \( k \in A \) implies \( k + 1 \in A \). Hence \( A = \mathcal{N} \) by \( P5 \). That is, \( S(n) \) holds for every \( n \in \mathcal{N} \).

Does the induction extend to infinity? To answer this, we ask a more fundamental question: Is infinity a natural number? We need to be more specific about what infinity “\( \infty \)” is. Infinity is a symbol that denotes something intuitively larger than all the numbers and can be operated under summation: \( \infty + a = \infty \) for \( a \in \mathcal{N} \). In particular, \( \infty + 1 = \infty \). We can use this last equality to check whether \( \infty \in \mathcal{N} \).

Actually, \( \infty \notin \mathcal{N} \). To use proof by contradiction suppose that \( \infty \in \mathcal{N} \). Then it has a successor \( \delta(\infty) \).

Infinity cannot have itself as its successor, so \( \infty \neq \delta(\infty) = \infty + 1 \). Contradicting to this is the property of infinity \( \infty = \infty + 1 \) presented above. This contradiction establishes \( \infty \notin \mathcal{N} \). In sum, the induction theorem presented above does not apply to infinity \( \infty \).

There can be other versions of the induction theorem (by initializing with \( S(m) \) for \( m > 0 \) or by assuming \( S(m) \) for all \( m \leq k \) in the \( k \)th induction step). I have even found a description of induction over real numbers by considering intervals and proceeding from a tighter interval towards a broader one (see The Instructor’s Guide to Real Induction by P.L.Clark at [http://arxiv.org/abs/1208.0973](http://arxiv.org/abs/1208.0973)) but I am not in a position to validate/illustrate this approach.
9 Appendix: From Power Series to Transforms and Back

Moment generating function can be thought as a power series

\[ m(t) = \sum_{n=0}^{\infty} p(n) \exp(nt) = \sum_{n=0}^{\infty} a_n x^n =: f(x), \]

where the series coefficient \( a_n = p(n) \) of the pmf of the random variable and \( x = \exp(t) \).

For a power series that has a positive radius of convergence (convergence for some interval of \( R \)), it is known that the power series is infinitely differentiable around \( x = 0 \) and

\[ a_n = \frac{f^{(n)}(0)}{n!}. \]

If another series \( g(x) = \sum_n b_n x^n \) has non-zero radius of convergence and \( f(x) = g(x) \) for an interval around 0, we obtain

\[ a_n = \frac{f^{(n)}(0)}{n!} = \frac{f^{(n)}(0)}{n!} = b_n. \]

That is, all of the coefficients of the power series must be equal for the series to be equal.

This result establishes a one-to-one correspondence between power series and their coefficients. It also establishes a one-to-one correspondence between moment generating function of a random variable and its pmf but proving that a given moment generating function can belong to a single random variable is involved. One of the transforms is the moment generating function. Another one is generating function (sometimes pmf but proving that a given moment generating function can belong to a single random variable is involved.

That is puzzlingly similar to the transformation formula. Every characteristic function may not correspond to a probability density. The following theorem has necessary and sufficient conditions and settles this issue.

**Example:** Find the characteristic function for an exponential random variable with rate 1.

\[ \phi_X(t) = \int_{-\infty}^{\infty} \exp(itx) f_X(x) dx = \int_{0}^{\infty} \exp(itx) \exp(-x) dx = \int_{0}^{\infty} \exp(-(1-it)x) dx \]

\[ = -\frac{1}{1-it} \exp(-(1-it)x) \bigg|_0^\infty = \frac{1}{1-it}. \]

A characteristic function can be inverted to obtain the corresponding density. The inversion formula

\[ f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-itx) \phi_X(t) dt \]

is puzzlingly similar to the transformation formula. Every characteristic function may not correspond to a probability density. The following theorem has necessary and sufficient conditions and settles this issue.

**Theorem 1. Bochner’s Theorem** A function \( \phi(t) \) is the characteristic function of a random variable if and only if it has \( \phi(0) = 1 \), it is continuous and positive definite.
A function $\varphi : \mathbb{R} \to \mathbb{R}$ is positive definite if for all real numbers $x_1, x_2, \ldots, x_n$ and all complex numbers $z_1, z_2, \ldots, z_n$, it satisfies

$$\sum_{j=1}^{n} \sum_{k=1}^{n} \varphi(x_j - x_k)z_jz_k^* \geq 0,$$

where $z_k^*$ is the conjugate of $z_k$. For more details, see chapter 7 of Borovkov (2013).

References


