Lecture 7:
Transform Methods
Integration

- Recall the definition of expectation:

\[ \mathbb{E}[X] = \sum_x xp_X(x), \quad X \text{ discrete with pmf } p_X \]

and

\[ \mathbb{E}[X] = \int x f_X(x) dx, \quad X \text{ continuous with pdf } f_X \]

- This can be compactly written as

\[ \mathbb{E}[X] = \int xdF_X(x) \]

where

\[ dF_X(x) = F_X(x) - \lim_{h \downarrow 0} F_X(x - h) \quad \text{or} \quad dF_X(x) = f_X(x) dx \]
This suggests a new notation that unifies the above two (and therefore applies to mixed-type RVs):

$$E[X] = \int x dF_X.$$ 

The notation extends to all cases we have seen so far. For example:

$$E[g(X)] = \int g(x) dF_X, \quad E[g(X)|Y] = \int g(x) dF_{X|Y}$$
Lebesgue-Stieltjes Integral

- \( F_X(x) \) yields a probability measure on the Borel \( \sigma \)-field \( \mathcal{B} \) on \( \mathbb{R} \) defined by

\[
\mu((a, b]) = F_X(b) - F_X(a), \quad \mu(B) = \int_B dF_X \quad \forall \ B \in \mathcal{B}
\]

This is particularly simple to understand if you consider that \( B \) is a (countable) union of intervals. As already noticed, \( (\mathbb{R}, \mathcal{B}, \mu) \) forms a probability space itself.

- For any \( \mathcal{B} \)-measurable function \( g : \mathbb{R} \rightarrow \mathbb{R} \), we write

\[
\mathbb{E}[g(X)] = \int g(x) dF_X
\]

- Let \( B \in \mathcal{B} \), and define \( g(x) = h(x)I_{\{x \in B\}} \), then

\[
\mathbb{E}[g(X)] = \int g(x) dF_X = \int_B h(x) dF_X
\]
Exchanging limits with expectation

- Consider a sequence of RVs \( \{X_n\} \), defined on a common probability space \((\Omega, \mathcal{F}, \mathbb{P})\), such that for all \( \omega \in \Omega \) we have \( \lim_{n \to \infty} X_n(\omega) = X(\omega) \), where \( X \) is a well-defined RV defined on the same probability space.

- We study the continuity of the operation \( \mathbb{E}[\cdot] \):
  a) Monotone convergence. If \( 0 \leq X_n \leq X_{n+1} \) for all \( n, \omega \), then \( \lim_{n \to \infty} \mathbb{E}[X_n] = \mathbb{E}[X] \).
  b) Dominated convergence. If \( |X_n| \leq Y \) for all \( n, \omega \) and \( \mathbb{E}[Y] < \infty \), then \( \lim_{n \to \infty} \mathbb{E}[X_n] = \mathbb{E}[X] \).
  c) Bounded convergence. If \( |X_n| \leq c \) for all \( n, \omega \), then \( \lim_{n \to \infty} \mathbb{E}[X_n] = \mathbb{E}[X] \).

- Since null events do not contribute to the expectation, the conditions of the above theorems can be relaxed from “for all \( \omega \)” to almost everywhere (also indicated as almost surely). This means that they must hold up to sets of probability measure zero.
• Example: sums of non-negative RVs. Consider \( \{Z_i\} \) such that \( Z_i \geq 0 \) a.s. (almost surely), and let \( X_n = \sum_{i=1}^{n} Z_i \). Clearly, \( \{X_n\} \) is a non-decreasing sequence (a.s.). Define \( X = \sum_{i=1}^{\infty} Z_i \), by the monotone convergence theorem we have

\[
\mathbb{E}[X] = \sum_{i=1}^{\infty} \mathbb{E}[Z_i]
\]
even though the sum may not converge.

• Example: suppose that \( \mathbb{E}[\sum_{i=1}^{n} |X_i|] < \infty \). Then, \( \mathbb{E}[\sum_{i=1}^{\infty} X_i] = \sum_{i=1}^{\infty} \mathbb{E}[X_i] \).

Proof: Let \( Y = \sum_{i=1}^{\infty} |X_i| \), and notice that, for any \( n \) and \( \omega \in \Omega \), \( \sum_{i=1}^{n} X_i \leq Y \) and that, by assumption, \( \mathbb{E}[Y] < \infty \). Hence, by dominated convergence:

\[
\mathbb{E} \left[ \sum_{i=1}^{\infty} X_i \right] = \mathbb{E} \left[ \lim_{n \to \infty} \sum_{i=1}^{n} X_i \right] = \lim_{n \to \infty} \mathbb{E} \left[ \sum_{i=1}^{n} X_i \right] = \lim_{n \to \infty} \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{\infty} \mathbb{E}[X_i]
\]
Liminf and limsup

• For a set $A \subset \mathbb{R}$, $a = \inf \{ A \}$ is the largest value such that $a \leq x$, for all $x \in A$, and $b = \sup \{ A \}$ is the smallest value such that $b \geq x$ for all $x \in A$.

• Let $\{ A_n \}$ denote a sequence of sets (not necessarily nested). We define

$$\liminf_{n \to \infty} A_n = \lim_{n \to \infty} \left\{ \bigcap_{m \geq n} A_m \right\} = \bigcup_{n=1}^{\infty} \bigcap_{m \geq n} A_m$$

and

$$\limsup_{n \to \infty} A_n = \lim_{n \to \infty} \left\{ \bigcup_{m \geq n} A_m \right\} = \bigcap_{n=1}^{\infty} \bigcup_{m \geq n} A_m$$

• Notice that the new sequence $B_n = \bigcap_{m \geq n} A_m$ is nested, in fact, $B_n \subseteq B_{n+1} \subseteq B_{n+2} \cdots$. Also, $B_n = \bigcup_{m \geq n} A_m$ is nested, in fact, $B_n \supseteq B_{n+1} \supseteq B_{n+2} \cdots$. Hence, the limits are well-defined.
• For a sequence of numbers \( \{a_n\} \), we define

\[
\liminf_{n \to \infty} a_n = \lim_{n \to \infty} \inf \{a_m : m \geq n\}
\]

and

\[
\limsup_{n \to \infty} a_n = \lim_{n \to \infty} \sup \{a_m : m \geq n\}
\]

• \( \{a_n\} \) has a limit if and only if

\[
\liminf_{n \to \infty} a_n = \limsup_{n \to \infty} a_n = \lim_{n \to \infty} a_n
\]
\textbf{Fatou’s Lemma:} Let \( \{X_n\} \) be a sequence of RVs such that \( X_n \geq Y \) a.s., where \( Y \) is a RV with well-defined expectation (i.e., such that \( \mathbb{E}[|Y|] < \infty \)), then

\[
\mathbb{E} \left( \liminf_{n \to \infty} X_n \right) \leq \liminf_{n \to \infty} \mathbb{E}[X_n]
\]

In particular, if \( X_n \geq 0 \) a.s., the inequality can be applied with \( Y = 0 \).
Moment generating function

**Definition 30.** The moment generating function of a RV $X$ is defined as

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

$M_X(t)$ is a function $\mathbb{R} \to \mathbb{R}_+$. $\diamond$

- MGFs are related to the Laplace transform of the pdf (for continuous RVs), in fact,

$$\mathbb{E} [e^{tX}] = \int e^{tx} f_X(x) dx$$
• If $M_X(t) < \infty$ for $t$ in an open interval containing the origin, then

1. $\mathbb{E}[X^k] = \frac{d^k}{dt^k}M_X(t)\bigg|_{t=0} = M_X^{(k)}(0)$;
2. The function $M_X$ can be expanded in Taylor series as

$$M_X(t) = \sum_{k=0}^{\infty} \frac{\mathbb{E}[X^k]}{k!} t^k$$

within its convergence domain.
3. Convolution – product property: if $X$ and $Y$ are independent, then

$$M_{X+Y}(t) = M_X(t)M_Y(t)$$
Definition 31. The characteristic function of a RV $X$ is defined as

$$\phi_X(u) = \mathbb{E}[e^{juX}] = \int e^{jux} dF_X$$

where $j = \sqrt{-1}$.

- Ch.Fs are related to the Fourier transform of the pdf (for continuous RVs), in fact,

$$\mathbb{E}[e^{juX}] = \int e^{jux} f_X(x) dx$$
**Theorem 15.** The characteristic function $\phi_X$ satisfies:

1. $\phi_X(0) = 1, |\phi_X(u)| \leq 1$ for all $u \in \mathbb{R}$.

2. $\phi_X$ is uniformly continuous on $\mathbb{R}$.

3. $\phi_X$ is non-negative definite, which is to say that

$$\sum_{i,k} \phi_X(u_i - u_k) z_i z_k^* \geq 0$$

for all reals $u_1, u_2, \ldots, u_n$ and complex coefficients $z_1, z_2, \ldots, z_n$. 
Theorem 16. a) If the $k$-th derivative of $\phi_X$ at $u = 0$ exists, then

$$
\begin{cases}
\mathbb{E}[|X|^k] < \infty, & k \text{ even} \\
\mathbb{E}[|X|^{k-1}] < \infty, & k \text{ odd}
\end{cases}
$$

b) If $\mathbb{E}[|X|^k] < \infty$, then

$$
\phi_X(u) = \sum_{i=0}^{k} \frac{\mathbb{E}[X^i]}{i!} (ju)^i + o(u^k)
$$

and also $\mathbb{E}[X^k] = j^{-k} \frac{d^k}{du^k} \phi_X \bigg|_{u=0}$.

Proof. It follows from Taylor’s series theorem for a function of complex variable.
Theorem 17. Let $M_X(t)$ and $\phi_X(u)$ denote the MGF and the Ch.F of $X$. For any $a > 0$ the following statements are equivalent:

1. $|M(t)| < \infty$ for all $|t| < a$.

2. $\phi_X(s)$, with $s \in \mathbb{C}$, is analytic on the strip $|\text{Im}\{s\}| < a$.

3. The moments $m_k = \mathbb{E}[X^k]$ exist for all $k = 1, 2, \ldots$ and satisfy
   \[
   \limsup_{k \to \infty} \left\{ \frac{|m_k|}{k!} \right\}^{1/k} \leq a^{-1}.
   \]

If the above conditions hold, then $M_X(t)$ may be extended analytically to the strip $|\text{Im}\{s\}| < a$ and $M_X(jt) = \phi_X(t)$. 
More properties of the Ch.F

**Theorem 18.** If $X$ and $Y$ are independent, then

$$
\phi_{X+Y}(u) = \phi_X(u)\phi_Y(u)
$$

**Theorem 19.** For $a, b \in \mathbb{R}$ and $Y = aX + b$, then

$$
\phi_Y(u) = e^{ibu}\phi_X(au)
$$
Ch.F of random vectors

**Definition 32.** Let $X$ be an $n$-dimensional random vector with joint cdf $F_X(x)$. Then, its joint characteristic function is defined as

$$
\phi_X(u) = \mathbb{E} \left[ e^{j u^T X} \right] = \int_{\mathbb{R}^n} e^{j \sum_{i=1}^{n} u_i x_i} dF_X
$$

- In particular, for a pair of jointly distributed RVs $X$ and $Y$, we have

$$
\phi_{X,Y}(u, v) = \mathbb{E} \left[ e^{j(uX+vY)} \right]
$$

- If $X$ and $Y$ are independent, then

$$
\phi_{X,Y}(u, v) = \phi_X(u) \phi_Y(v)
$$
Inversion formula

- From Fourier transform theory: if $X$ is continuous with density $f_X$, then

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi_X(u)e^{-jux} \, du$$

for all $x \in \mathbb{R}$ where $f_X$ is differentiable.

- If the above integral is not absolutely convergent, we interpret it as the Cauchy principal value integral.

- Sufficient (but not necessary) condition such that $\phi_X$ is the Ch.F of a continuous RV is that

$$\int_{-\infty}^{\infty} |\phi_X(u)| \, du < \infty$$
The general case is more complicated and it is stated as follows:

**Inversion theorem.** Let $X$ have cdf $F_X$ and Ch.F. $\phi_X$. Define $\overline{F}_X$ as

$$
\overline{F}_X(x) = \frac{1}{2} \left[ F_X(x) + \lim_{h \downarrow 0} F_X(x - h) \right]
$$

Then,

$$
\overline{F}_X(b) - \overline{F}_X(a) = \lim_{T \to \infty} \int_{-T}^{T} \frac{e^{-jau} - e^{-jbu}}{2\pi ju} \phi_X(u) \, du
$$

**Notice:** analogous inversion formulas hold of the multidimensional (random vector) case, from the theory of multidimensional Fourier transform.

**Corollary 5.** $X$ and $Y$ have the same Ch.Fs if and only if they have the same cdfs.
Example

Let $X_i$ be i.i.d. RVs $\sim \mathcal{N}(0, 1)$ and $Y = \sum_{i=1}^{n} X_i^2$. The RV $Y$ is said to be “central chi-squared” with $n$ degrees of freedom (we say $Y \sim \chi^2(n)$). For $n = 2m$ (even integer) find the pdf and the cdf of $Y$.

Solution:

First, we find the Ch.F of $Z = X^2$. We have

$$
\Phi_Z(u) = \mathbb{E}[e^{juX^2}] = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{1}{2} - ju\right)x^2} dx
$$

Notice that this integral is absolutely convergent for all $u \in \mathbb{R}$, and in particular, replacing $\frac{1}{2} - ju$ with a complex variable $s$, we notice that

$$
g_Z(s) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-sx^2} dx
$$

converges for all $\text{Re}\{s\} > 0$ (right half-plane).
It follows that we can use analytic continuation, compute $g_Z(s)$ for real $s$ in an interval $[a, b]$ containing the point $1/2$, and then extend the result to the vertical strip $a < \Re\{s\} < b$ and in particular to the line $s = \frac{1}{2} - js$.

We have immediately that, for real positive $s$,

$$g_Z(s) = \frac{1}{\sqrt{2s}}$$

Hence, by the analytic continuation argument above, we have that

$$\Phi_Z(u) = g_Z(1/2 - ju) = \frac{1}{\sqrt{1 - j2u}}$$

Now, consider $Y = \sum_{i=1}^{2m} X_i^2$. By the Sum-of-RVs theorem we have that

$$\Phi_Y(u) = \Phi_Z(u)^{2m} = \frac{1}{(1 - j2u)^m}$$
The pdf of $Y$ is found by using the inversion formula (similar to the inverse Fourier transform) applied to $\Phi_Y(u)$. To this purpose, we notice that the Laplace transform of the pdf of $Y$ is obtained as follows:

$$\mathcal{L}_Y(s) = \mathbb{E}[e^{-sY}] = \Phi_Y(u)|_{u=js} = \frac{1}{(1 + 2s)^m}$$

This is analytic in the right half-plane $\text{Re}\{s\} > -1/2$, and it has a pole of multiplicity $m$ at $s_0 = -1/2$. Since the ROC of $\mathcal{L}_Y(s)$ contains the imaginary axis, it follows that the inverse transform of $\Phi_Y(u)$ and the inverse Laplace transform of $\mathcal{L}_Y(s)$ coincide. The latter can be easily calculated using contour integration and the Cauchy residue theorem.
In particular, for a suitable integration path $c + ju$ with $c > -1/2$ and $u \in (-\infty, \infty)$, we have:

$$f_Y(y) = \mathcal{L}^{-1} \left\{ \frac{1}{(1 + 2s)^m} \right\}$$

$$= \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \frac{e^{sy}}{(1 + 2s)^m} ds$$

$$= \sum \text{Res} \left( \frac{2^{-m}e^{sy}}{(1/2 + s)^m}, \text{poles on the left half of the complex plane} \right)$$

$$= \text{Res} \left( \frac{2^{-m}e^{sy}}{(1/2 + s)^m}, \frac{-1}{2} \right)$$

The residue at the pole with multiplicity $m$ is computed as follows:
Let \( f(s) = \frac{g(s)}{(s-s_0)^m} \) be a function with a pole at \( s_0 \) with multiplicity \( m \) (assume \( g(s) \) is analytic at \( s_0 \)), then

\[
\text{Res} \left\{ f(s), s_0 \right\} = \frac{1}{(m-1)!} \left. \frac{d^{m-1} g(s)}{ds^{m-1}} \right|_{s=s_0}
\]

Applying the above formula to our case, we obtain

\[
\text{Res} \left( \frac{2^{-m} e^{sy}}{(1/2 + s)^m}, \frac{-1}{2} \right) = \frac{1}{(m-1)!} \left. \frac{d^{m-1}}{ds^{m-1}} 2^{-m} e^{sy} \right|_{s=-1/2}
\]

\[
= \frac{2^{-m}}{(m-1)!} \left. y^{m-1} e^{sy} \right|_{s=-1/2}
\]

\[
= \frac{2^{-m}}{(m-1)!} y^{m-1} e^{-y/2}
\]

This holds for \( y \geq 0 \). Clearly, for \( y < 0 \) the pdf of \( Y \) is zero since \( Y \geq 0 \) with probability 1 (it is a sum of squares).
**Continuity**

**Definition 33.** We say that the sequence of cdfs $F_1, F_2, \ldots$ converges to the cdf $F$ (we write $F_n \to F$) if

$$\lim_{n \to \infty} F_n(x) = F(x)$$

for all $x \in \mathbb{R}$ where $F$ is continuous.

**Theorem 20.** Let $F_1, F_2, \ldots$ be a sequence of cdfs with Ch.Fs $\phi_1, \phi_2, \ldots$, respectively.

a) If $F_n \to F$, where $F$ is a cdf with Ch.F $\phi$, then

$$\lim_{n \to \infty} \phi_n(u) = \phi(u), \quad \forall \ u \in \mathbb{R}$$

b) Conversely, if $\phi(u) = \lim_{n \to \infty} \phi_n(u)$ exists and is continuous at $u = 0$, then $\phi$ is a valid characteristic function of some cdf $F$ such that $F_n \to F$. 
The Law of Large Numbers (LLN)

Definition 34. (Convergence in distribution). A sequence of RVs $X_1, X_2, \ldots$ with cdfs $F_1, F_2, \ldots$ is said to converge in distribution to a RV $X$ with cdf $F$, if $F_n \to F$ as $n \to \infty$.

In this case, we write $X_n \xrightarrow{D} X$.

Theorem 21. Law of Large Numbers. Let $X_1, X_2, \ldots$ be a sequence of i.i.d. RVs with finite mean $\mu = \mathbb{E}[X_1]$ and define their partial sum $S_n = \sum_{i=1}^{n} X_i$. Then,

$$\frac{1}{n} S_n \xrightarrow{D} \mu$$
The Central Limit Theorem (CLT)

**Theorem 22. Central Limit Theorem.** Let $X_1, X_2, \ldots$ be a sequence of i.i.d. RVs with finite mean $\mu = \mathbb{E}[X_1]$ and variance $\sigma^2 = \text{var}(X_1)$, and define their partial sum $S_n = \sum_{i=1}^{n} X_i$. Then,

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{D} \mathcal{N}(0, 1)$$

Notice the abuse of notation ...here $\mathcal{N}(0, 1)$ indicates a Gaussian(0,1) variable.
Convergence of densities

**Theorem 23.** *Local CLT.* Let $X_1, X_2, \ldots$ be a sequence of i.i.d. RVs with finite mean $\mathbb{E}[X_1] = 0$ and variance $\text{var}(X_1) = 1$. Suppose that their common Ch.F. satisfies

$$\int |\Phi(u)|^r du < \infty$$

for some integer $r \geq 1$. The pdf $f_n$ of $Y_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$ exists for all $n \geq r$ and furthermore

$$f_{Y_n}(y) \to \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$$

uniformly for all $y \in \mathbb{R}$. 
An application of the MGF: Large Deviations

- The LLN states that \( S_n = \sum_{i=1}^{n} X_i \), for large \( n \), is typically close to \( n\mu \).

- The CLT states that with high probability we may find \( S_n \) between the values \( n\mu - \beta \sqrt{n\sigma^2} \) and \( n\mu + \beta \sqrt{n\sigma^2} \) (e.g., the probability is larger than 0.98 if \( \beta \geq 3 \)).

- Notice that the size of this high probability interval is \( O(\sqrt{n}) \), which is “much less than \( n \”).

- A “large deviation” from this typical behavior is when \( S_n \) is found far apart from its expected value \( n\mu \).

- We are interested in studying the probability

\[
P(S_n > na), \quad \text{for } a > \mu
\]
MGF and log-MGF

- The log-moment generating function is defined as

\[ m_X(t) = \log M_X(t) = \log \mathbb{E}[e^{tX}] \]

- Properties:

  1. \( m_X(0) = \log M_X(0) = 0. \)
  2. \( m'_X(0) = \frac{M'_X(0)}{M_X(0)} = \mu. \)
  3. \( m_X(t) \) is convex in its domain (where it is finite), in fact,

\[
m''_X(t) = \frac{M''_X(t)M_X(t) - (M'_X(t))^2}{M^2_X(t)} = \frac{\mathbb{E}[X^2e^{tX}]\mathbb{E}[e^{tX}] - \left(\mathbb{E}[Xe^{tX}]\right)^2}{M^2_X(t)} \geq 0
\]

where the latter follows from Cauchy-Schwarz, applied to the RVs \( Y = Xe^{tX/2} \) and \( Z = e^{tX/2} \).

Copyright G. Caire (Sample Lectures) 209
Definition 35. **Fenchel-Legendre Transform.** The F-L transform of $m_X(t)$ is the function

$$m^*_X(a) = \sup_{t \in \mathbb{R}} \{at - m_X(t)\}$$

\[\diamondsuit\]

Theorem 24. **Large deviations.** Let $X_1, X_2, \ldots$ be a sequence of i.i.d. RVs with mean $\mu$ and suppose that their common MGF, $M(t)$, exists and it is finite in a neighborhood of the origin ($t = 0$). Consider $a > \mu$ such that $\mathbb{P}(X_1 > a) > 0$ and let $m(t) = \log M(t)$. Then, $m^*(a) > 0$ and

$$-\frac{1}{n} \log \mathbb{P} \left( \sum_{i=1}^{n} X_i > na \right) \to m^*(a), \quad \text{as } n \to \infty$$
Notice: in practice, this means that the probability \( P \left( \sum_{i=1}^{n} X_i > na \right) \) vanishes exponentially as

\[
P \left( \sum_{i=1}^{n} X_i > na \right) \asymp e^{-nm^*(a)}
\]

The exponential rate (or simply “exponent”) at which the large deviation probability vanishes is given by the F-L transform \( m^*(a) \) of the log-MGF \( m(t) \).
Generating functions

• Let \( \{a_i : i = 0, 1, 2, \ldots \} \) denote a sequence of real numbers. Its associated generating function is defined as

\[
G_a(s) = \sum_{i=0}^{\infty} a_i s^i, \quad s \in \mathbb{C}
\]

• \( G_a(s) \) is meaningful inside its region of convergence (ROC), defined as the subset of \( \mathbb{C} \) for which the sum is absolutely convergent.

• The intersection of the ROC with the real line \( \mathbb{R} \) is called interval of convergence.

• The elements of the sequence \( a \) can be recovered from \( G_a(s) \) by setting

\[
a_i = G_a^{(i)}(0)/i!.
\]
Generating functions are often easier to deal with than directly handling the associated sequences.

**Convolution property:** Given \( a = \{a_i : i \geq 0\} \) and \( b = \{b_i : i \geq 0\} \), the convolution sequence \( c = a \ast b \) is defined as

\[
c_i = \sum_{j=0}^{i} a_j b_{i-j} = a_0 b_i + a_1 b_{i-1} + a_2 b_{i-2} + \cdots + a_i b_0
\]

Then, the associated generating functions satisfy

\[
G_c(s) = G_a(s)G_b(s)
\]
• Example: show the combinatorial identity \[ \sum_{i=0}^{n} \binom{n}{i}^2 = \binom{2n}{n}. \]

• Example: Let \( X \) and \( Y \) be independent Poisson RVs with parameters \( \lambda \) and \( \mu \), respectively. Find the distribution of \( Z = X + Y \).

• Convergence: power series of the type \( G_a(s) = \sum_{i=0}^{\infty} a_i s^i \) (i.e., involving only non-negative powers of \( s \)) converge in a circle of given radius \( R \). That is: the sum is absolutely convergent for all \( |s| < R \), and diverges for \( |s| > R \). Furthermore, for all \( R' < R \) the sum is uniformly convergent on sets of the form \( \{ s \in \mathbb{C} : |s| \leq R' \} \).
• **Differentiation:** \( G_a(s) \) can be differentiated (any number of times) or integrated term by term for all \(|s| < R\).

• **Uniqueness:** if \( G_a(s) = G_b(s) \) for all \( s \) such that \(|s| < R' \) with \( R' \leq R \), then \( a_i = b_i \) for all \( i = 0, 1, 2, \ldots \) and \( a_i = \frac{1}{i!} G_a^{(i)}(0) \).

• **Abel’s theorem:** if \( a_i \geq 0 \) for all \( i = 0, 1, 2, \ldots \) and \( |G_a(s)| < \infty \) for all \(|s| < 1\), then

\[
\lim_{s \uparrow 1} G_a(s) = \sum_{i=0}^{\infty} a_i
\]

where the sum in the left-hand side may be finite or infinite.
Definition 36. Let $X$ denote a discrete RV taking on values in $\{0, 1, 2, \ldots \}$, with pmf $p_X[0], p_X[1], p_X[2], \ldots$. The probability generating function of $X$ is given by

$$G_X(s) = \sum_{i=0}^{\infty} p_X[i] s^i = \mathbb{E}[s^X]$$

Notice: by letting $s = e^t$, we have that $G_X(e^t) = M_X(t)$. Also, by letting $s = e^{ju}$ we have that $G_X(e^{ju}) = \Phi_X(u)$.

- Directly from the definition we have that $G_X(0) = p_X[0]$, and that $G_X(1) = 1$.
- Directly from the convolution property we have that if $X$ and $Y$ are independent, then

$$G_{X+Y}(s) = G_X(s)G_Y(s)$$
Examples

- **Constant RV.** \( G_X(s) = \mathbb{E}[s^X] = s^c. \)

- **Bernoulli RV.** \( G_X(s) = (1 - p)s^0 + ps = 1 - p + ps. \)

- **Geometric RV.** \( G_X(s) = \sum_{i=1}^{\infty} p(1 - p)^{i-1}s^i = \frac{ps}{1-s(1-p)}. \)

- **Poisson RV.** \( G_X(s) = \sum_{i=0}^{\infty} \frac{\lambda^i s^i}{i!} e^{-\lambda} = e^{(s-1)\lambda}. \)

- **Binomial RV.** Recalling that \( X \) can be written as \( \sum_{i=0}^{n} X_i \) where the \( X_i \) are Bernoulli i.i.d., we obtain

\[
G_X(s) = \prod_{i=1}^{n} G_{X_i}(s) = G_{X_1}(s)^n = (1 - p + ps)^n
\]
**Theorem 25.** Let $X$ have probability generating function $G_X(s)$. Then,

a) $E[X] = G'_X(1)$.

b) $E[X(X - 1)(X - 2)\cdots(X - k + 1)] = G^{(k)}_X(1)$.

**Notice:** The quantity $E[X(X - 1)(X - 2)\cdots(X - k + 1)]$ is known as the “$k$-th factorial moment” of $X$. Also, $G^{(k)}_X(1)$ should be intended as the limit for $s \uparrow 1$ when the ROC of $G_X$ is $|s| < 1$.

**Example:** in order to calculate the variance of $X$, we write

$$\text{var}(X) = E[X^2] - E[X]^2 = E[X(X - 1) + X] - E[X]^2 = G''_X(1) + G'_X(1) - G'(1)^2$$
Compounding

**Theorem 26.** Let $X_1, X_2, \ldots$ denote a sequence of i.i.d. discrete RVs with common generating function $G_X(s)$, and let $N$ denote a RV taking on values in $0, 1, 2, \ldots$, independent of the $X_i$’s, with generating function $G_N(s)$. Then, the new random variable $S = \sum_{i=1}^{N} X_i$ has generating function

\[
G_S(s) = G_N(G_X(s))
\]

Notice: if $N = n$ is a constant, then $G_N(s) = s^n$ and we get back the well-known convolution theorem, where $G_N(G_X(s)) = G_X(s)^n$. 
Example: eggs and chicks

A hen lays $N$ eggs, where $N$ is Poisson with parameter $\lambda$. Each egg hatches independently with probability $p$. Let $K$ denote the number of chicks. We want to find the distribution of $K$.

Clearly,

$$K = \sum_{i=1}^{N} X_i$$

where the $X_i$ are i.i.d. Bernoulli. Hence, using the compounding theorem, we have

$$G_K(s) = G_N(G_X(s)) = e^{(G_X(s)-1)\lambda} = e^{(1-p+sp-1)\lambda} = e^{(s-1)p\lambda}$$

Therefore we conclude that $K$ is Poisson, with parameter $p\lambda$. 
More definitions and facts

- The joint probability generating function of $X$ and $Y$ is given by

$$G_{X,Y}(s, r) = \mathbb{E}[s^X r^Y] = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} p_{X,Y}[i, j] s^i r^j$$

- $X$ and $Y$ are independent if and only if $G_{X,Y}(s, r) = G_X(s) G_Y(r)$ for all $s, r$.

- Defective RVs. It might happen that certain RVs take on the value $+\infty$ with positive probability. For such RVs, we have that $G_X(s)$ converges for $|s| < 1$ and that

$$\lim_{s \uparrow 1} G_X(s) = \sum_i p_X[i] = 1 - \mathbb{P}(X = +\infty)$$

The moments of $X$ are all equal to $+\infty$. 

Applications: matching and occupancy

• We have a collection \( \{A_1, A_2, \ldots, A_n\} \) of events. Let \( X \) be the number of such events that occur in a given random experiment. For example, with \( n \) letters and \( n \) envelopes picked at random, let \( A_i \) denote the event that letter \( i \) is put into its correct envelope, and we are interested in the probability \( \mathbb{P}(X = k) \) for some \( 0 \leq k \leq n \), that is, the probability that exactly \( k \) letters are put into their correct envelopes.

• The problem consists of expressing the pmf of \( X \) in terms of probabilities of the form \( \mathbb{P}(A_{i_1}, A_{i_2}, \ldots, A_{i_m}) \) for \( m = 1, 2, \ldots, n \).

• Let us introduce the notation:

\[
S_m = \sum_{i_1 < i_2 < \ldots < i_m} \mathbb{P}(A_{i_1}, A_{i_2}, \ldots, A_{i_m})
\]

with the convention that \( S_0 = 1 \).
• We have

\[ S_m = \sum_{i_1 < i_2 < \cdots < i_m} \mathbb{E}[I_{A_{i_1}} I_{A_{i_2}} \cdots I_{A_{i_m}}] \]

\[ = \mathbb{E} \left[ \sum_{i_1 < i_2 < \cdots < i_m} I_{A_{i_1}} I_{A_{i_2}} \cdots I_{A_{i_m}} \right] \]

\[ = \mathbb{E} \left[ \mathbb{E} \left[ \sum_{i_1 < i_2 < \cdots < i_m} I_{A_{i_1}} I_{A_{i_2}} \cdots I_{A_{i_m}} \Big| X \right] \right] \]

\[ = \mathbb{E} \left[ \binom{X}{m} \right] \]
• Hence,

\[ S_m = \sum_{i=0}^{n} \binom{i}{m} \mathbb{P}(X = i) \]

• We introduce the generating functions:

\[ G_S(s) = \sum_{m=0}^{n} S_m s^m, \quad G_X(s) = \sum_{i=0}^{n} s^i \mathbb{P}(X = i) \]

• Multiplying by \( s^m \) and summing over \( m \), we obtain

\[ G_S(s) = \sum_{i=0}^{n} \mathbb{P}(X = i) \sum_{m=0}^{n} s^m \binom{i}{m} = \sum_{i=0}^{n} (1 + s)^i \mathbb{P}(X = i) = G_X(s + 1) \]

from which we get \( G_X(s) = G_S(s - 1) \).
Equating the coefficients of $s^i$, we find:

$$\mathbb{P}(X = i) = \sum_{j=i}^{n} (-1)^{j-i}\binom{j}{i} S_j, \quad 0 \leq i \leq n$$

This is the well-known **Waring theorem** formula, that we have already obtained in a Homework in a fairly more complicated way.
Branching processes

- Processes that describe in probabilistic terms the evolution of populations.
- Useful to describe cells, atoms, etc ...
- At each time, each member of the current generation gives birth to a random number of children. The random variable that describes the number of generated children from a given parent is called family size.
- Assumption: the family sizes form a collection of i.i.d. RVs, with common pmf $p$ and probability generating function $G$. 
Generation distribution

- Let $Z_n$ denote the number of elements in generation $n$, where we assume $Z_0 = 1$ with probability 1.

- The probability generating function of $Z_n$ is denoted by $G_n(s) = \mathbb{E}[s^{Z_n}]$.

**Theorem 27.** $G_{n+m}(s) = G_m(G_n(s)) = G_n(G_m(s))$, therefore,

$$G_n(s) = G(G(G(\cdots G(s)))\cdots)_{n \text{ times}}$$
Lemma 26. Let $Z_1$ denote the family size RV, with moments $\mathbb{E}[Z_1] = \mu$ and $\text{var}(Z_1) = \sigma^2$. Then,

$$\mathbb{E}[Z_n] = \mu^n, \quad \text{var}(Z_n) = \begin{cases} n\sigma^2 & \text{if } \mu = 1 \\ \frac{\sigma^2(\mu^n - 1)\mu^{n-1}}{\mu - 1} & \text{if } \mu \neq 1 \end{cases}$$
Example: geometric branching process

• Let $Z_1$ be geometric, such that

$$p_{Z_1}[k] = qp^k, \quad \text{with } q = 1 - p$$

• We have

$$G(s) = \sum_{k=0}^{\infty} qp^k s^k = \frac{q}{1 - sp}$$

• We can show using induction that

$$G_n(s) = \begin{cases} 
\frac{n-(n-1)s}{n+1-ns} & \text{if } p = q = 1/2 \\
\frac{q(p^n - q^n - ps(p^{n-1} - q^{n-1}))}{p^{n+1} - q^{n+1} - ps(p^n - q^n)} & \text{if } p \neq q
\end{cases}$$
• A typical question that we ask is the probability of ultimate extinction: notice that if at a certain $n$ we have $Z_n = 0$, then the population is extinct, it will remain equal to zero for ever.

• The probability of ultimate extinction is defined as

$$\mathbb{P}(\text{ultimate extinction}) = \mathbb{P}\left( \bigcup_{n} \{Z_n = 0\} \right)$$

• We have that $\{Z_n = 0\} \subseteq \{Z_{n+1} = 0\}$. Hence, the sequence of events $\{Z_n = 0\}$ is nested (increasing).

• Using continuity of the probability measure, we have

$$\mathbb{P}\left( \bigcup_{n} \{Z_n = 0\} \right) = \mathbb{P}\left( \lim_{n \to \infty} \{Z_n = 0\} \right) = \lim_{n \to \infty} \mathbb{P}(Z_n = 0)$$

Copyright G. Caire (Sample Lectures)
• For the geometric branch process, we have explicitly $G_n(s)$, hence,

$$P(Z_n = 0) = G_n(0) = \begin{cases} \frac{n}{q(p^n - q^n)} & \text{if } p = q \\ \frac{n+1}{p^{n+1} - q^{n+1}} & \text{if } p \neq q \end{cases}$$

• Hence, it follows that $P(\text{ultimate extinction}) = 1$ if $p \leq q$, and that

$$P(\text{ultimate extinction}) = \frac{q}{p}, \quad \text{if } p > q$$

• This result is intuitive: notice that $E[Z_1] = p/q$, then, if $E[Z_1] \leq 1$ the process will sooner or later die, almost surely. On the contrary, if $E[Z_1] > 1$, the process will last forever with positive probability $1 - q/p$. 
Theorem 28. For a branching process with family size distribution with mean $\mathbb{E}[Z_1] = \mu$, the probability of ultimate extinction

$$P(\text{ultimate extinction}) = \lim_{n \to \infty} P(Z_n = 0)$$

is equal to the value $\eta$ of the smallest non-negative solution of the equation $s = G(s)$. Also, $\eta = 1$ if $\mu < 1$ and $\eta < 1$ if $\mu > 1$. If $\mu = 1$, then $\eta = 1$ if $Z_1$ has strictly positive variance.
End of Lecture 7