Lecture 12:
Generating functions and moments

We will see that when working with probability measures over a complicated space $(\Omega, \mathcal{F})$ it will be useful to be able to characterize a probability P on $(\Omega, \mathcal{F})$ by analyzing the value of $\int_\Omega f\,dP$ computed over a range of test functions $f: \Omega \to \overline{\mathbb{R}}$.

Characteristic functions and moment generating functions are an example of this.

MGF's, CF's and Complex generating functions

Definition:

Let $(\Omega, \mathcal{F}, P)$ be a probability space and X be a random variable (taking values in $\mathbb{R}$). For $t \in \mathbb{R}$ and $z \in \mathbb{C}$ define

$$M_X(t) := E(e^{tX})$$
$\leftarrow$ Moment generating function of X (MGF).

$$G_X(z) := E(e^{zX})$$
$\leftarrow$ Complex generating function of X.

$$\phi_X(t) := E(e^{itX})$$
$\leftarrow$ Characteristic function for X (CF).

In general, if μ is a measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ define

$$M_\mu(t) := \int_\mathbb{R} e^{tx}\,d\mu(x)$$

$$G_\mu(z) := \int_\mathbb{R} e^{zx}\,d\mu(x).$$

$$\phi_\mu(t) := \int_\mathbb{R} e^{itx}\,d\mu(x)$$

Since $x \mapsto e^{zx}$ takes values in $\mathbb{C}$ we need to say a word about integrating complex valued functions.

If $f: \Omega \to \mathbb{C}$ then one can decompose it into real and imaginary parts:

$$f(\omega) = \underbrace{\operatorname{Re} f(\omega)} + i\,\underbrace{\operatorname{Im} f(\omega)}$$

functions mapping $\mathbb{R} \to \mathbb{R}$.

If μ is a measure on $(\Omega, \mathcal{F})$ then

$$\int_\Omega f(\omega)\,d\mu(\omega) := \underbrace{\int_\Omega \operatorname{Re} f(\omega)\,d\mu(\omega)} + i\underbrace{\int_\Omega \operatorname{Im} f(\omega)\,d\mu(\omega)}$$

all the properties of $\int_\Omega d\mu$ extent to the complex case with minor changes

when these two are defined i.e. $\operatorname{Re} f, \operatorname{Im} f \in Q(\Omega, \mathcal{F}, \mu)$.

The usefulness of these generating functions come from 3 facts:

1) ϕ_X & G_X (and M_X sometimes) characterizes the distribution of X.

   E.g. if you have two r.v.s X & Y then $X \overset{\mathcal{D}}{=} Y$ iff $\phi_X(t) = \phi_Y(t)$ $\forall t \in \mathbb{R}$.

   Note: This is analogous to c.d.f.s and densities.

2) The generating functions for sums of independent r.v.s is easy to calculate. i.e. If $X_1, \ldots, X_n$ are independent r.v.s all defined on $(\Omega, \mathcal{F}, P)$ then

$$\phi_{X_1 + \cdots + X_n}(t) = \phi_{X_1}(t) \cdots \phi_{X_n}(t).$$

   Note: The corresponding operation for densities is hard, i.e. the density of $X_1 + \cdots + X_n$ is a n-fold convolution of the densities of each X_i.

3) If you know $M_X(t)$, $\phi_X(t)$ or $G_X(z)$ you can compute the moments $E(X^k)$ by differentiating.

Note: if μ is a measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ s.t.

$$d\mu = \delta\, dx$$

then $\phi_\mu(-t)$ is just the Fourier transform of δ. It is actually more natural to think of $\phi_\mu(t)$ as inverse Fourier transform of $\delta(x)$ where x represents frequency.

This explains fact 3) since the FT & inverse FT diagonalizes convolution.

Relating ϕ_X, M_X and G_X

There are two examples of generating functions that are useful to keep in your mind.

Let $Z \sim N(0,1)$. Then

$$\phi_Z(t) = e^{-t^2/2}$$

$$M_Z(t) = e^{t^2/2}$$

$$G_Z(z) = e^{z^2/2}$$

Also let Y be a r.v. with density $\delta(x) = \frac{1}{\pi(1+x^2)}$ w.r.t. dx. Then

$$\phi_Y(t) = e^{-|t|}$$

$$M_Y(t) = \begin{cases} 1 & \text{if } t=0 \\ \infty & \text{o.w} \end{cases}$$

$$G_Y(z) = \begin{cases} e^{-|z|} & \text{if } \mathrm{Re}\, z = 0 \\ \infty & \text{if } \mathrm{Re}\, z \neq 0 \ \& \ \mathrm{Im}\, z = 0 \\ \text{Not defined} & \text{o.w} \end{cases}$$

Note: Y is a Cauchy r.v..

Looking at the case of $Z \sim N(0,1)$ we have $M_Z(it) = \phi_Z(t)$. However this can't hold in general since $M_Y(it) \neq \phi_Y(t)$.

To understand the difference we need to analyze G_X.

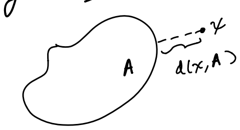
Definition: If Ω is a metric space, with metric d, and $A \subset \Omega$ define

$A^\circ :=$ the open interior of A ← union of all open sets $\subset A$

$\bar{A} :=$ the closure of A ← intersection of all closed sets containing A

$\partial A := \bar{A} - A^\circ$

$d(x, A) := \inf\{d(x,y) : y \in A\}$



Also for any subset of $\mathbb{C}$ let

$$\mathrm{Re}\, A := \{\mathrm{Re}\, z : z \in A\}$$

$$\mathrm{Im}\, A := \{\mathrm{Im}\, z : z \in A\}$$

Definition:

For any r.v. X let

$$\mathcal{D}_X := \{u + iv \in \mathbb{C} : E(e^{uX}) < \infty\}$$

$$= \text{the cylinder in } \mathbb{C} \text{ with base } \{u \in \mathbb{R} : M_X(u) < \infty\}$$

Theorem ($\mathrm{Re}\,\mathcal{D}_X$ is an interval)

If X is a r.v. then $\mathrm{Re}\,\mathcal{D}_X$ is an interval containing 0 (closed, open or half open) and M_X is convex on $\mathrm{Re}\,\mathcal{D}_X$.

Remark: This thm is true for $\mathrm{Re}\,\mathcal{D}_\mu$ & M_μ when μ is a measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ but now the interval can be empty (in which case it will not contain 0).

Proof:

Since $t \mapsto e^{tx}$ is convex we have

$$e^{[\alpha t_1 + (1-\alpha)t_2]x} \le \alpha e^{t_1 x} + (1-\alpha) e^{t_2 x}$$

$\forall\, t_1, t_2 \in \mathbb{R}$ & $\alpha \in [0,1]$.

$$\therefore\ M_X(\alpha t_1 + (1-\alpha)t_2) = E\left(e^{[\alpha t_1 + (1-\alpha)t_2]X}\right)$$

$$\overset{\text{Big 3}}{\le} \alpha E(e^{t_1 X}) + (1-\alpha) E(e^{t_2 X})$$

$$= \alpha M_X(t_1) + (1-\alpha) M_X(t_2)$$

$\therefore$ M_X is convex.

Now suppose $t_1, t_2 \in \operatorname{Re} \mathcal{D}_X$. Then $M_X(t_1) < \infty$, $M_X(t_2) < \infty$ and

$$t_1 \le t \le t_2 \implies M_X(t) \le \alpha M_X(t_1) + (1-\alpha) M_X(t_2) < \infty$$

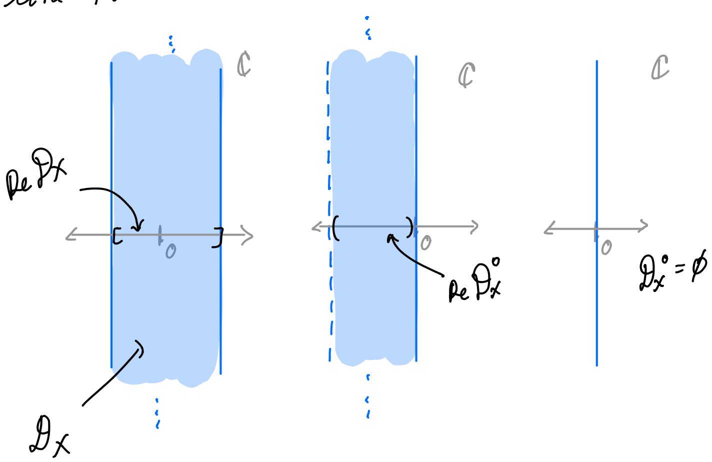
writing $t = \alpha t_1 + (1-\alpha)t_2$ for some $\alpha \in [0,1]$

$$\implies t \in \operatorname{Re} \mathcal{D}_X.$$

$\therefore$ $\operatorname{Re} \mathcal{D}_X$ is an interval containing 0, since clearly $M_X(0) = 1$.

QED

So for any r.v. X $\mathcal{D}_X$ could look something like this:



Theorem: (The Analyticity of G_X over $\mathcal{D}_X^\circ$)

Let X be a r.v. (mapping into $\mathbb{R}$) such that $\mathcal{D}_X^\circ \ne \emptyset$. Then $\forall z \in \mathcal{D}_X^\circ$

i) $E|X^n e^{zX}| < \infty$ for $n = 0, 1, 2, \dots$

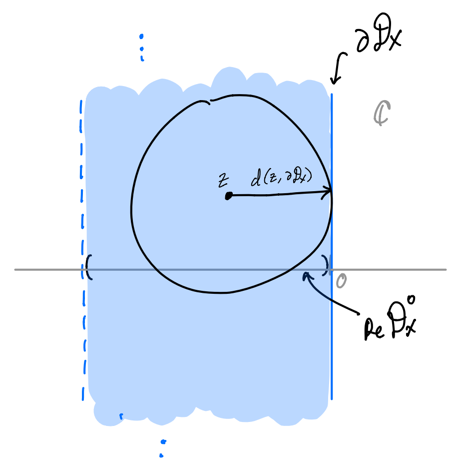
ii) $$G_X(\zeta) = \sum_{n=0}^{\infty} E(X^n e^{zX}) \frac{(\zeta - z)^n}{n!}$$

for all ζ in the open ball of $\mathbb{C}$ centered at z with radius $d(z, \partial \mathcal{D}_X)$.

iii) G_X is infinitely differentiable on $\mathcal{D}_X^\circ$ with complex derivative

$$\frac{d^n}{dz^n} G_X(z) = E(X^n e^{zX}).$$

e.g.



Proof:

First note that for any $z, \zeta \in \mathbb{C}$ we have

$$e^{\zeta X} = e^{zX} e^{(\zeta - z)X} = \sum_{n=0}^{\infty} e^{zX} \frac{X^n (\zeta - z)^n}{n!}$$

and

(*) $$\sum_{n=0}^{\infty} \left| e^{zX} \frac{X^n (\zeta - z)^n}{n!} \right| \le |e^{zX}| \sum_{n=0}^{\infty} \frac{|X(\zeta - z)|^n}{n!}$$

$$= e^{uX} e^{r|X|}$$

where $u := \operatorname{Re}(z)$, $r := |\zeta - z|$

$$\le e^{(u-r)X} + e^{(u+r)X}$$

since $e^{r|X|} = \begin{cases} e^{-rX} \\ e^{rX} \end{cases}$ or

Notice that when $z, \zeta \in \mathcal{D}_X^\circ$ and $r := |z-\zeta| < d(z, \partial \mathcal{D}_X)$ then

$$u \pm r := \operatorname{Re} z \pm |z-\zeta| \in \operatorname{Re} \mathcal{D}_X^\circ$$

so that

$$E(e^{(u-r)X}) = M_X(u-r) < \infty$$

$$E(e^{(u+r)X}) = M_X(u+r) < \infty$$

$$\therefore \quad E\left(\sum_{n=0}^{\infty} \left| e^{zX} \frac{X^n(\zeta-z)^n}{n!} \right| \right) < \infty \qquad (**)$$

$\parallel \longleftarrow$ By monotone convergence in Big 3

$$\sum_{n=0}^{\infty} E\left| e^{zX} \frac{X^n(\zeta-z)^n}{n!} \right|$$

$$\therefore \quad E\left| e^{zX} \frac{X^n(\zeta-z)^n}{n!} \right| = \frac{|\zeta-z|^n}{n!} E(|e^{zX} X^n|) < \infty$$

for all $n = 1, 2, \ldots$ so i) holds

To show ii) notice that

$$\sum_{n=0}^{N} e^{zX} \frac{X^n(\zeta-z)^n}{n!} \xrightarrow{N\to\infty} \sum_{n=0}^{\infty} e^{zX} \frac{X^n(\zeta-z)^n}{n!} \quad P\text{-a.e.}$$

and DCT applies with upper bound given by the LHS of (*) which is integrable by (**).

$$\therefore \quad G_X(\zeta) = E(e^{\zeta X})$$

$$= E\left(\lim_N \sum_{n=0}^{N} e^{zX} \frac{X^n(\zeta-z)^n}{n!}\right)$$

$$\overset{DCT}{=} \lim_N \sum_{n=0}^{N} E(e^{zX} X^n) \frac{(\zeta-z)^n}{n!}$$

$$= \sum_{n=0}^{\infty} E(e^{zX} X^n) \frac{(\zeta-z)^n}{n!}$$

This gives ii).

Finally iii) follows directly from ii).

QED

Now we can understand the relationship btwn ϕ_X and M_X:

- $G_X(z)$ always exist and is finite on $\mathcal{D}_X$

  since $u+iv \in \mathcal{D}_X \Rightarrow M_X(u) = E(e^{uX}) < \infty$

  $$\Rightarrow E(|e^{(u+iv)X}|) = E(e^{uX}) < \infty$$

  $$\Rightarrow |G_X(u+iv)| < \infty.$$

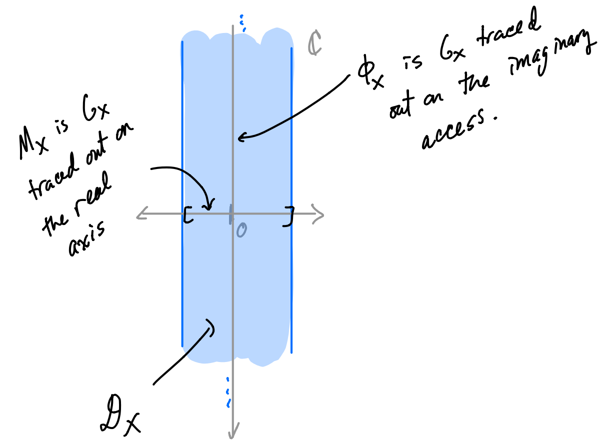
- $\mathcal{D}_X$ always contains $\{it : t \in \mathbb{R}\}$

  since $0 \in \operatorname{Re} \mathcal{D}_X$

- $\phi_X(t) = G_X(it) \;\; \forall t \in \mathbb{R}$ and

  $M_X(t) = G_X(t) \;\; \forall t \in \operatorname{Re} \mathcal{D}_X$

e.g.



- $G_X(z)$ = the unique analytic extension of ϕ_X on $\{it : t \in \mathbb{R}\}$ to $\mathcal{D}_X$.

  only when $\operatorname{Re} \mathcal{D}_X \neq \{0\}$ $\;=$ the unique analytic extension of M_X on $\operatorname{Re} \mathcal{D}_X$ to $\mathcal{D}_X$

This follows by a complex analysis result:

Thm: Suppose $\mathcal{D} \subset \mathbb{C}$ is open and connected. If f and g are differentiable complex-valued functions defined on $\mathcal{D}$ which agree on distinct $z_1, z_2, \ldots \in \mathcal{D}$ s.t. $\lim_{n\to\infty} z_n \in \mathcal{D}$ then

$$f(z) = g(z) \;\; \forall z \in \mathcal{D}.$$

- Now suppose we have a formula for $M_X(t)$ s.t. $\mathrm{Re}\,\mathcal{D}_X \neq \{0\}$ and a extension $H(z)$ defined on $\mathcal{D}_X$ s.t.

$$H(t) = M_X(t) \quad \forall t \in \mathrm{Re}\,\mathcal{D}_X.$$

Then if H is complex differentiable on $\mathcal{D}_X^\circ$ must be that

$$H(z) = G_X(z), \ \forall z \in \mathcal{D}_X^\circ \ \text{ and}$$

$$H(it) = \phi_X(t), \ \forall t \in \mathbb{R}.$$

defined by continuity to $\partial \mathcal{D}_X$ in case $\mathrm{Re}\,\mathcal{R}_X = [0,a)$ for e.g.

- This also works if you can compute

$$\alpha_n := \mathbb{E} X^n$$

when $\beta_n := \mathbb{E}|X|^n$ decay fast enough so $\sum_{n=0}^\infty \beta_n \frac{t^n}{n!}$ has a non-zero radius of convergence. In which case

$$M_X(t) = \sum_{n=0}^\infty \alpha_n \frac{t^n}{n!}$$

for all t in an open neighbarhood of 0 (use similar arguments for the thm on G_X).

This completely determines G_X, and thus ϕ_X, by analytic extension to $\mathcal{D}_X$.

<u>Note</u>: once we show that $\phi_X(t)$ completely characterizes the distribution of X we will have:

The moments $\{\mathbb{E}|X|^n\}_{n\geq 1}$ characterize the distribution of X only when $\sum_{n=0}^\infty \mathbb{E}|X|^n \frac{t^n}{n!}$ has a non-zero radius of convergence.

<u>e.g.</u> If $X \sim N(0,1)$ then one can derive that $M_X(t) = e^{t^2/2}$ & $\mathrm{Re}\,\mathcal{D}_X = \mathbb{R}$.
Here are two extensions defined on $\mathcal{D}_X = \mathbb{C}$:

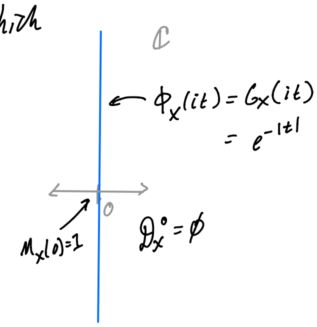
$$H_1(z) = e^{|z|^2/2} \quad \leftarrow \text{not analytic}$$

$$H_2(z) = e^{z^2/2} \quad \leftarrow \text{analytic}$$

$\therefore$ $G_X(z) = H_2(z)$ but not $H_1(z)$ and

$\phi_X(t) = H_2(it) = e^{-t^2/2}$ but not $H_1(it) = e^{t^2/2}$

<u>e.g.</u> If X is a cauchy R.V. then $\mathrm{Re}\,\mathcal{D}_X^\circ = \phi$ which explains why we can't get ϕ_X or G_X from M_X.



<u>e.g.</u> Suppose $X \geq 0$ which satisfies

$$\mathbb{E}X^n = \mathbb{E}|X|^n = n!$$

Can we infere what ϕ_X is?

Since $\sum_{n=0}^\infty n! \frac{t^n}{n!} < \infty \quad \forall t \in (-1,1)$,

$M_X(t)$ is finite on $(-1,1) = \mathrm{Re}\,\mathcal{D}_X$

$$\therefore \quad M_X(t) = G_X(t) = \sum_{n=0}^\infty E(X^n)\frac{t^n}{n!} = \sum_{n=0}^\infty t^n = \frac{1}{1-t}$$

when $t < 1$

Since $\frac{1}{1-z}$ is an analytic extension of M_X to $\mathcal{D}_X$ we must have

$$\phi_X(t) = G_X(it) = \frac{1}{1-it} \quad \forall t \in \mathbb{R}.$$

This r.v. is an exponential r.v. where

$$dPX^{-1} = e^{-x} I_{[0,\infty)}(x)\,dx$$

Another consequence of our analyticity theorem for G_X is that both ϕ_X and M_X give the moments of X.

Since ϕ_X is always defined and not every r.v.s has finite moments of all orders it suggests that

The larger n is s.t. $E|X|^n<\infty$ $\iff$ The smoother $\phi_X(t)$ is at $t=0$

Corollary: (Moments from M_X or ϕ_X)

If X is a r.v. s.t. $0\in \operatorname{Re} \mathcal{D}_X^\circ$ then $E|X|^n<\infty$ for all n and

$$M_X^{(n)}(0) = (-i)^n \phi_X^{(n)}(0) = E(X^n).$$

Proof:
Note that $G_X(t)=M_X(t)$ when $t\in\mathbb{R}$ and the complex derivatives of an analytic function equal the "directional derivatives".

$$\therefore \quad \frac{d^n}{dt^n} M_X(t)\Big|_{t=0} = \frac{d^n}{dz^n} G_X(z)\Big|_{z=0} = E(X^n).$$

A similar argument holds for ϕ_X. QED

Getting Moments from ϕ_X

What about the case when $0\notin \operatorname{Re}\mathcal{D}_X^\circ$

Now either $E|X|^n$ doesn't decay fast enough to be summable, or they are ∞ for all large n.

We need a more fine tuned argument.

Studying Taylor's thm one gets

$$(*) \quad \left| e^{itx} - \sum_{n=0}^{N} \frac{(itx)^n}{n!} \right| \le \frac{|tx|^{N+1}}{(N+1)!}$$

$\forall t\in\mathbb{R}$ & $N\ge 0$.

Note: this bound gives a slightly sub-optimal bound on the regularity of $\phi_X(t)$ near $t=0$ when $E|X|^{N+1}=\infty$. See Billingsley p. 343 for the more fine tuned result.

Now $(*)$ already gives

$$\left| \phi_X(t) - \sum_{n=0}^{N} E(X^n)\frac{(it)^n}{n!} \right| \le \frac{|t|^{N+1}}{(N+1)!} E(|X|^{N+1})$$

Theorem: (Moments from ϕ_X)

If X is a r.v. that satisfies $E|X|^{n+1}<\infty$ for some $n\in\{1,2,\dots\}$ then ϕ_X is n times differentiable and

$$\phi_X^{(m)}(t) = E\left((iX)^m e^{itX}\right).$$

$\forall\, m\le n$.

Proof: start with $m=1$.

$$\frac{\phi_X(t+\varepsilon)-\phi_X(t)}{\varepsilon} = E\left(\frac{e^{itX}e^{i\varepsilon X} - e^{itX}}{\varepsilon}\right)$$

$$= E\left(e^{itX}\frac{e^{i\varepsilon X}-1}{\varepsilon}\right)$$

Therefore

$$\frac{\phi_X(t+\varepsilon)-\phi_X(t)}{\varepsilon} - E(iXe^{itX}) = E\Bigg(e^{itX}\underbrace{\frac{e^{i\varepsilon X}-1-i\varepsilon X}{\varepsilon}}_{\text{bdd in magnitude by } \frac{|\varepsilon X|^2}{\varepsilon 2!} \text{ from } (*)}\Bigg)$$

$\therefore \lim_{\varepsilon \to 0}$ RHS $= 0$ by DCT.

$\therefore \quad \phi_X'(t) = E(iXe^{itX})$.

Repeating the argument gives the result.

QED.

- In the lecture on separating classes we will show that if $\phi_X(t) = \phi_Y(t)$ $\forall t \in \mathbb{R}$ then $X \overset{\mathcal{D}}{=} Y$
- In the lecture on convergence in distribution we will show $\phi_{X_n}(t) \to \phi_X(t)$ $\forall t$ then $X_n \xrightarrow{\mathcal{D}} X$