

## 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  $\int f dP$  computed over a range of test functions  $f: \Omega \rightarrow \mathbb{R}$ .

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

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

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

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

In general, if  $\mu$  is a measure on  $(\Omega, \mathcal{B}(\Omega))$  define

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

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

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

①

Since  $x \mapsto e^{zx}$  takes values in  $\mathbb{C}$  we need ② to say a word about integrating complex valued functions.

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

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

functions mapping  $\Omega \rightarrow \mathbb{R}$ .

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

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

↗

all the properties of  $\int_{\Omega} f(w) d\mu(w)$  extend to the complex case with minor changes

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

The usefulness of these generating functions come from 3 facts:

1)  $\phi_X$  &  $G_X$  (and  $M_X$  sometimes) characterizes the distribution of  $X$ .

E.g. if you have two r.v.s  $X$  &  $Y$  then  $X \stackrel{D}{=} Y$  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  $X_1, \dots, X_n$  are independent r.v.s all defined on  $(\Omega, \mathcal{F}, P)$  then

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

Note: The corresponding operation for densities is hard, i.e. the density of  $X_1 + \dots + X_n$  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  $\mu$  is a measure on  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  s.t. (3)

$$d\mu = s dx$$

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

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

### Relating $\phi_X$ , $M_X$ 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  $s(x) = \frac{1}{\pi(1+x^2)}$  w.r.t.  $dx$ . Then

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

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

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

Note:  $Y$  is a Cauchy r.v..

Looking at the case of  $Z \sim N(0,1)$  we have (4)

$$M_Z(it) = \phi_Z(t)$$

$$\text{general since } M_Z(it) \neq \phi_Z(t).$$

To understand the difference we need to analyze  $G_X$ .

Definition: If  $\mathcal{S}$  is a metric space, with

metric  $d$ , and  $A \subset \mathcal{S}$  define

$$A^\circ := \text{the open interior of } A \quad \text{union of all open sets } c \subset A$$

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

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

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



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

$$\operatorname{Re} A := \{ \operatorname{Re} z : z \in A \}$$

$$\operatorname{Im} A := \{ \operatorname{Im} z : z \in A \}$$

Definition:

For any r.v.  $X$  let

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

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

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

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

Remark: This thm is true for  $\operatorname{Re} \mathcal{D}_\mu$  &  $M_\mu$  when  $\mu$  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

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

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

$$\begin{aligned} M_X(\alpha t_1 + (1-\alpha)t_2) &= E(e^{[\alpha t_1 + (1-\alpha)t_2]X}) \\ &\stackrel{\text{Big 3}}{\leq} \alpha E(e^{t_1 X}) + (1-\alpha) E(e^{t_2 X}) \\ &= \alpha M_X(t_1) + (1-\alpha) M_X(t_2) \end{aligned}$$

$\therefore M_X$  is convex.

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

$M_X(t_2) < \infty$  and

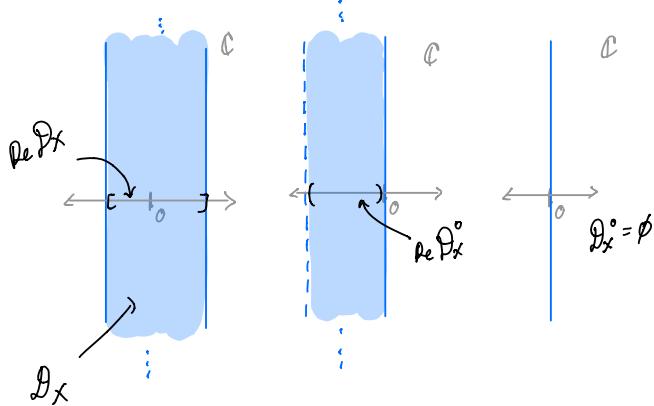
$$\begin{aligned} t_1 \leq t \leq t_2 \implies M_X(t) &\leq \alpha M_X(t_1) + (1-\alpha) M_X(t_2) < \infty \\ &\text{Writing } t = \alpha t_1 + (1-\alpha)t_2 \text{ for some } \alpha \in [0, 1] \\ \implies t &\in \text{Re } \mathcal{D}_X. \end{aligned}$$

$\therefore \text{Re } \mathcal{D}_X$  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:



(5)

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

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

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

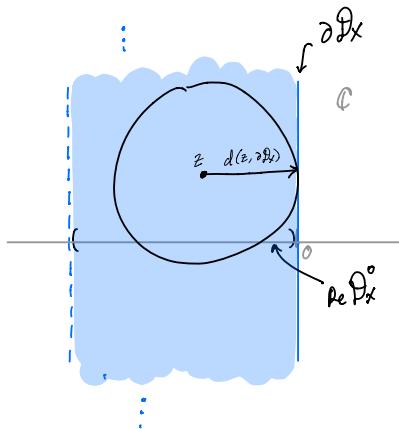
$$\text{ii) } G_X(z) = \sum_{n=0}^{\infty} E(X^n e^{zx}) \frac{(z-z)^n}{n!}$$

for all  $z$  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  $\mathcal{D}_X^o$  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

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

and

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

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

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

$$\text{where } u := \text{Re}(z), r := |\zeta-z| \quad \text{since } e^{r|X|} = \begin{cases} e^{-rX} & \text{if } r < 0 \\ e^{rX} & \text{otherwise} \end{cases}$$

Notice that when  $z, \zeta \in \mathcal{D}_X$  and  
 $r := |z - \zeta| < d(z, \partial \mathcal{D}_X)$  then  
 $u \pm r := Re z \pm |z - \zeta| \in Re \mathcal{D}_X^0$

so that

$$E(e^{(u-r)x}) = M_x(u-r) < \infty$$

$$E(e^{(u+r)x}) = M_x(u+r) < \infty$$

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

|| ← By monotone convergence in Big 3

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

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

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

To show ii) notice that

$$\sum_{n=0}^{\infty} e^{zx} \frac{x^n (\zeta-z)^n}{n!} \xrightarrow{N \rightarrow \infty} \sum_{n=0}^{\infty} e^{zx} \frac{x^n (\zeta-z)^n}{n!} \quad P\text{-q.e.}$$

and DCT applies with upper bound given

by the LHS of (\*\*) which is integrable by (\*\*).

$$\therefore G_x(\zeta) = E(e^{\zeta x})$$

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

$$\stackrel{\text{DCT}}{=} \lim_N \sum_{n=0}^N E\left(e^{zx} x^n\right) \frac{(\zeta-z)^n}{n!}$$

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

This gives iii).

Finally iv) follows directly from iii).

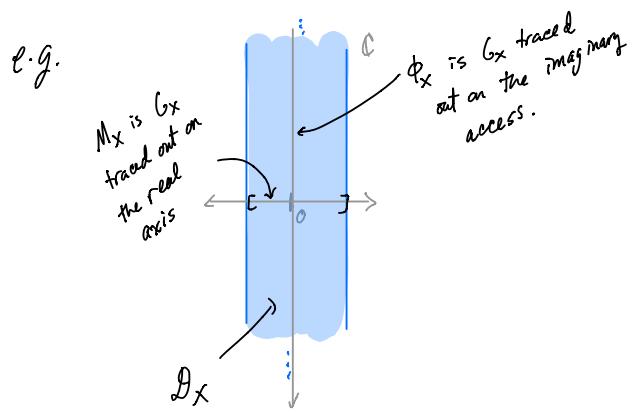
QED

Now we can understand the relationship btwn  $\phi_x$  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  $0 \in Re \mathcal{D}_X$

- $\phi_x(it) = G_x(it)$   $\forall t \in \mathbb{R}$  and  
 $M_x(it) = G_x(it)$   $\forall t \in Re \mathcal{D}_X$



- $G_x(z) =$  the unique analytic extension of  $\phi_x$  on  $\{it : t \in \mathbb{R}\}$  to  $\mathcal{D}_X$ .  
 only when  $Re \mathcal{D}_X \neq \emptyset$  ↗ = the unique analytic extension of  $M_x$  on  $Re \mathcal{D}_X$  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  $z_1, z_2, \dots \in \mathcal{D}$  s.t.  $\lim_{n \rightarrow \infty} z_n \in \mathcal{D}$  then

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

Now suppose we have a formula for (9)

$M_X(t)$  s.t.  $\operatorname{Re} \mathcal{D}_X \neq \{0\}$  and a extension  $H(z)$  defined on  $\mathcal{D}_X$  s.t.

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

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

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

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

↗ defined by continuity to  $\partial \mathcal{D}_X$   
in case  $\operatorname{Re} \mathcal{D}_X = [0, a)$  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 neighborhood of 0 (use similar arguments for the thm on  $G_X$ ).

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

To summarize:

Corollary:

i)  $M_X$  characterizes  $\phi_X$  when  $M_X(t)$  finite over some non-empty open interval of  $\mathbb{R}$ .

ii) The moments  $\{\mathbb{E}|X|^n\}_{n \geq 1}$  characterize  $\phi_X$  only when

$$\sum_{n=0}^{\infty} \mathbb{E}|X|^n \frac{t^n}{n!}$$

has a non-zero radius of convergence about  $t=0$ .

(10)

e.g. If  $X \sim N(0, 1)$  then one can derive

$$\text{That } M_X(t) = e^{t^2/2} \quad \& \quad \operatorname{Re} \mathcal{D}_X = \mathbb{R}.$$

Here are two extensions defined on  $\mathcal{D}_X = \mathbb{C}$ :

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

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

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

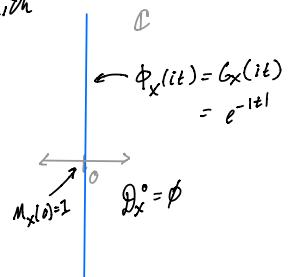
$$\phi_X(t) = H_2(it) = e^{-t^2/2} \text{ but not } H_1(it) = e^{t^2/2}$$

e.g. If  $X$  is a Cauchy R.V.

then  $\operatorname{Re} \mathcal{D}_X^o = \emptyset$  which

explains why we can't get

$\phi_X$  or  $G_X$  from  $M_X$ .



e.g. Suppose  $X \geq 0$  which satisfies

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

Can we infer what  $\phi_X$  is?

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

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

$$\begin{aligned} \therefore M_X(t) &= G_X(t) = \sum_{n=0}^{\infty} \mathbb{E}(X^n) \frac{t^n}{n!} \\ &= \sum_{n=0}^{\infty} t^n = \frac{1}{1-t} \end{aligned}$$

when  $t < 1$

Since  $\frac{1}{1-t}$  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

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

Another consequence of our analyticity theorem for  $G_X$  is that both  $\phi_X$  and  $M_X$  give the moments of  $X$ . (11)

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

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

Corollary : (Moments from  $M_X$  or  $\phi_X$ )

If  $X$  is a r.v. s.t.  $\sigma \in \text{Re } \phi_X^0$  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".

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

A similar argument holds for  $\phi_X$ . GEP

Getting Moments from  $\phi_X$  (12)

What about the case when  $\sigma \notin \text{Re } \phi_X^0$ ? Now either  $E|X|^n$  doesn't decay fast enough to be summable, or they are  $\infty$  for all large  $n$ .

We need a more fine tuned argument.

Studying Taylor's thm one gets

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

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

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

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

Theorem: (Moments from  $\phi_X$ )

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

$$\phi_X^{(n)}(t) = E((iX)^n e^{itX}).$$

If  $m \leq n$ .

Proof: Start with  $m=1$ .

$$\begin{aligned} \frac{\phi_X(t+\varepsilon) - \phi_X(t)}{\varepsilon} &= E\left(\frac{e^{itX} e^{i\varepsilon X} - e^{itX}}{\varepsilon}\right) \\ &= E\left(e^{itX} \frac{e^{i\varepsilon X} - 1}{\varepsilon}\right) \end{aligned}$$

Therefore

(13)

$$\frac{\phi_X(t+\epsilon) - \phi_X(t)}{\epsilon} - E(iX e^{itX}) = E\left(e^{itX} \underbrace{\frac{e^{i\epsilon X} - 1 - i\epsilon X}{\epsilon}}_{\text{bdd in magnitude}}$$

by  $\frac{|i\epsilon X|^2}{\epsilon^2}$  from (\*)

$$\therefore \lim_{\epsilon \rightarrow 0} \text{RHS} = 0 \text{ by DCT.}$$

Repeating the argument gives the result.

QED.

(14)

- In the lecture on Separating classes we will show that if  $\phi_X(t) > \phi_Y(t) \ \forall t \in \mathbb{R}$  then  $X \stackrel{d}{=} Y$

- In the lecture on Convergence in distribution we will show  $\phi_{X_n}(t) \rightarrow \phi_X(t) \ \forall t$  then  $X_n \xrightarrow{d} X$