Probability and Random Process

Aimin Tang

The University of Michigan- Shanghai Jiao Tong University Joint Institute

Shanghai Jiao Tong University

Sep. 29 2020

Outline

- 2. Random Variables
  - Introduction to Random Variables
  - PMF and Discrete Random Variables
  - PDF and Continuous Random Variables
  - Gaussian CDF
  - Conditional Probability
  - Function of a RV
  - Expectation of a RV
  - Transform Methods and Probability Generating Function

Expectation of a RV

Definition:

Discrete case:
$$E(X) = \sum_i x_i p_X(x_i)$$

General case:
$$E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx$$

E(X) is well-defined if

$$\sum_i |x_i| p_X(x_i) < \infty$$

$$\int_{-\infty}^{+\infty} |x| f_X(x) dx < \infty$$

$E(X)$ is a numerical average of a large number of independent observations of the random variable

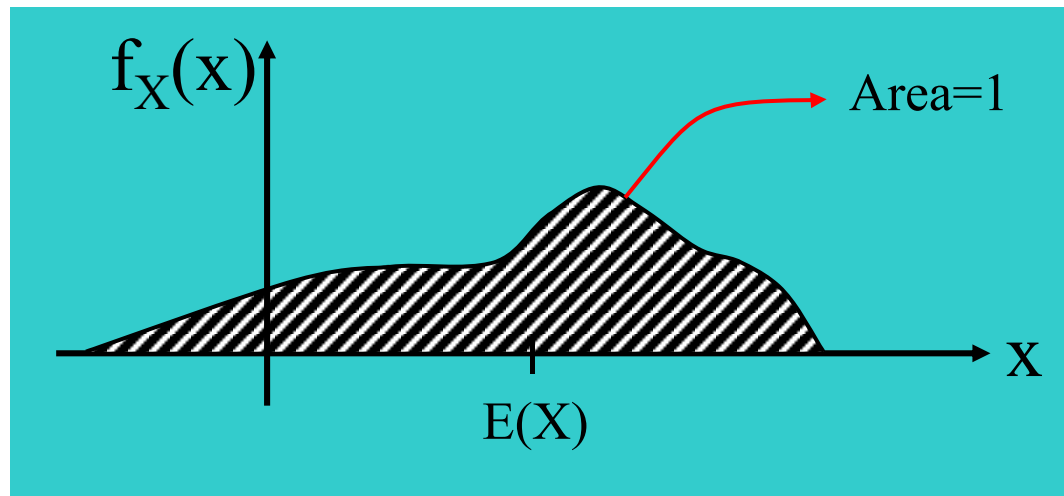$E(X)$ is also known as the:

- first moment
- ensemble average
- mean

$E(X)$ is symbolically expressed:
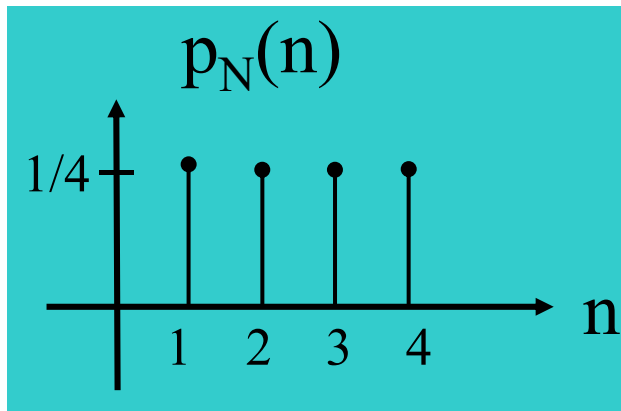
$$\mu_X, m_X, \eta_x, \text{or } \overline{X}$$

or just

$$\mu, m, \text{or } \eta$$

If the probability density is interpreted as a mass density along an axis, then $E(X)$ is the center of mass.

Note that $E(X)$ is not random.

$E(X)$ may not be a value that X can take.

$$E(N) = \sum_{n=1}^{4} np_N(n) = 2.5$$

- To calculate $E\{G(X)\}$, there are two options:
  - First, get $f_Y(y)$ for $Y = G(X)$, then calculate $E(Y)$
  - Second, and faster, method: calculate

$$E[Y] = \sum_X G(x)p_X(x)$$

$$E[Y] = \int_{-\infty}^{+\infty} G(x)f_X(x)dx$$

- It is called the law of the unconscious statistician (LOTUS)

$$P_r(g(X) = y) = \sum_{x:g(x)=y} P_r(X = x) = \sum_{x:g(x)=y} p_X(x)$$

$$E[Y] = \sum_y y p_Y(y)$$

$$= \sum_y y \, P_r(Y = y) = \sum_y y \, P_r(g(X) = y)$$

$$= \sum_y y \sum_{x:g(x)=y} p_X(x)$$

$$= \sum_x g(x) p_X(x)$$

1. The expected value of a constant* is that constant.

$$E(c) = c$$

2. The expected value is a linear operator:

$$E\big(cH(X)\big) = cE\big(H(X)\big), \quad c \in C$$
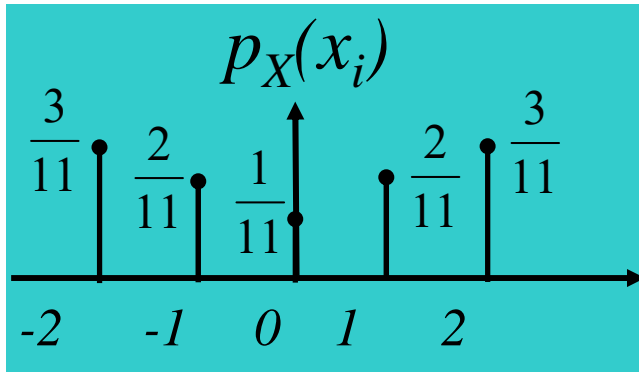
$$E\big(H(X) + G(X)\big) = E\big(H(X)\big) + E\big(G(X)\big)$$

Ex:

$$Y = aX^2 + bX + c$$

$$\Rightarrow E(Y) = aE(X^2) + bE(X) + c$$

* Constant with respect to the random variables

$$p_X(x_i)$$

$$\frac{3}{11} \qquad \frac{2}{11} \qquad \frac{1}{11} \qquad \frac{2}{11} \qquad \frac{3}{11}$$

-2   -1   0   1   2
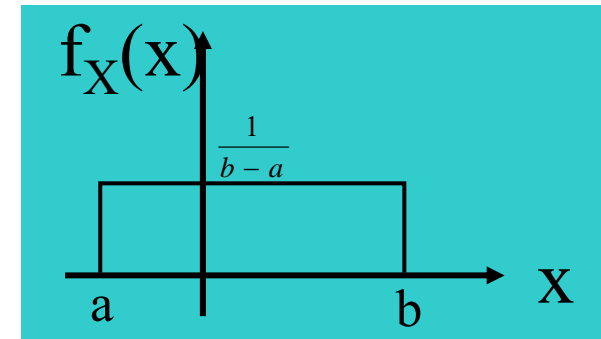
$$R_X : \{0, \pm 1, \pm 2\}$$

$$E(X^2) = \sum_{i=-2}^{2} i^2 \, p_X(x_i)$$

$$= 0 \cdot \frac{1}{11} + 2\left(1^2 \cdot \frac{2}{11} + 2^2 \cdot \frac{3}{11}\right) = \frac{28}{11} = 2.54$$

$E(X)$ is always in the middle of a uniform distribution.

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x)\,dx = \int_a^b x \frac{1}{b-a}\,dx$$

$$= \frac{1}{b-a}\left(\frac{x^2}{2}\right)\Bigg|_a^b$$

$$= \frac{b^2 - a^2}{2(b-a)} = \frac{b+a}{2} = \text{midpoint of ROS}$$

$$p_N(n) = \binom{m}{n} p^n (1-p)^{m-n}$$

Represent $\quad N = \sum_{i=1}^{m} X_i \quad\quad X_i = $ Independent Bernoulli RV

$$E[N] = E\left[\sum_{i=1}^{m} X_i\right] = \sum_{i=1}^{m} E(X_i) = \sum_{i=1}^{m} p = mp$$

Mean of a sum is the sum of the means

$$E(N) = \sum_{n=0}^{\infty} n \frac{\lambda^n}{n!} e^{-\lambda}$$

$\rightarrow$ poisson

$$= \lambda \sum_{n=1}^{\infty} \frac{\lambda^{n-1}}{(n-1)!} e^{-\lambda}$$

Change variables $i = n - 1$

$$E(N) = \lambda \left(\sum_{i=0}^{\infty} \frac{\lambda^i}{i!} e^{-\lambda} \right) = \lambda$$

1

$$E(X) = \int_{-\infty}^{+\infty} x \left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}} \right) dx$$

Let $y = x - m$. Then $x = y + m$ and $dx = dy$

$$E(X) = \int_{-\infty}^{+\infty} (y + m) \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{y^2}{2\sigma^2}} dy$$

$$= \underbrace{\int_{-\infty}^{+\infty} y \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{y^2}{2\sigma^2}} dy}_{\text{Odd}} + m \underbrace{\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{y^2}{2\sigma^2}} dy}_{\text{Just a PDF}}$$

The mean is m, given that the first term is 0

$$\int_{-\infty}^{+\infty} y \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{y^2}{2\sigma^2}} \, dy$$

$$= \int_{-\infty}^{0} y \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{y^2}{2\sigma^2}} \, dy + \int_{0}^{+\infty} y \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{y^2}{2\sigma^2}} \, dy$$

Change of limits
$$= -\int_{0}^{-\infty} y \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{y^2}{2\sigma^2}} \, dy + \int_{0}^{+\infty} y \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{y^2}{2\sigma^2}} \, dy$$

Change of variable
$$= \int_{0}^{\infty} (-y) \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(-y)^2}{2\sigma^2}} \, dy + \int_{0}^{+\infty} y \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{y^2}{2\sigma^2}} \, dy$$

$$= 0$$

Observe that because $\mathrm{E}(X)$ is not random (a constant),

$$E\big[X + E(X)\big] = 2E(X)$$

Suppose $\qquad H(x) = (x - \mu_x)^2$

$\qquad\qquad\qquad$ = square of distance of X from it's mean

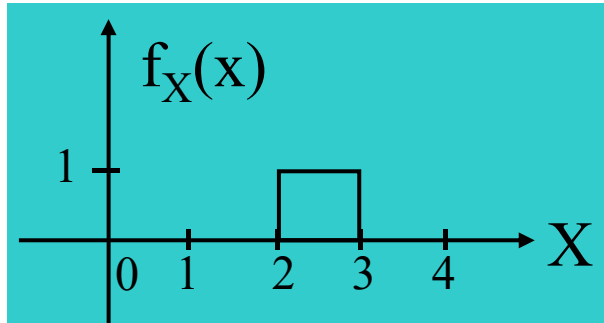Definition for variance:

$$V(X) = E\big[H(X)\big] = E\big[(X - \mu_x)^2\big]$$

Alternative notation: $\mathrm{Var}(X) = \sigma_x^2$

- Observe that since $(X - \mu_x)^2$ is always positive, $V(X)$ must also be positive.

- The standard deviation, $\sqrt{\sigma_x^2} = \sigma_x$ is a measure of the width or spread of the PDF.

Example - I

$$\mu_X = 2.5$$

$$V(X)$$

Example - II

$$\mu_W = 2$$

$$V(W)$$
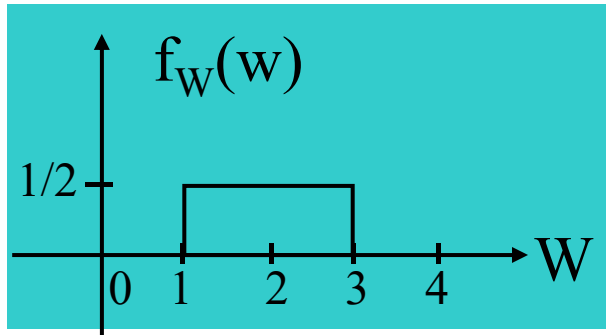
$$V(X) = E\left[(X - \mu_X)^2\right] = E\left(X^2 - 2X\mu_X + \mu_X{}^2\right)$$

$$= E\left(X^2\right) - 2E(X)\mu_X + \mu_X{}^2 = E\left(X^2\right) - \mu_X{}^2$$

or:

$$V(X) = E\left(X^2\right) - E(X)^2$$

Observe that if

$$\mu_x = 0, \quad V(X) = E(X^2)$$

Recall:
$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

$$V(X) = E\left[(X-m)^2\right] = \int_{-\infty}^{+\infty} \frac{(x-m)^2}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}} \, dx$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} y^2 e^{-\frac{y^2}{2}} \, dy \qquad y = \frac{x-m}{\sigma}, \quad dy = \frac{dx}{\sigma}$$

Integration by parts:

$$u = y, \quad dv = ye^{-\frac{y^2}{2}}$$

$$du = dy, \quad v = -e^{-\frac{y^2}{2}}$$

$$V(X) = \frac{\sigma^2}{\sqrt{2\pi}} \left[-ye^{-y^2/2} \Big|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} -e^{-y^2/2} dx \right]$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \left[0 + \sqrt{2\pi} \right] = \sigma^2$$

Almost a Gaussian PDF

Definition: k^{th} moment $= E(X^k)$

$$k^{th} \text{central moment} = E\left[(X - \mu_X)^k \right]$$

$$k^{th} \text{absolute moment} = E\left[|X|^k \right]$$

Observation:

These three PDFs have the same k^{th} central moment
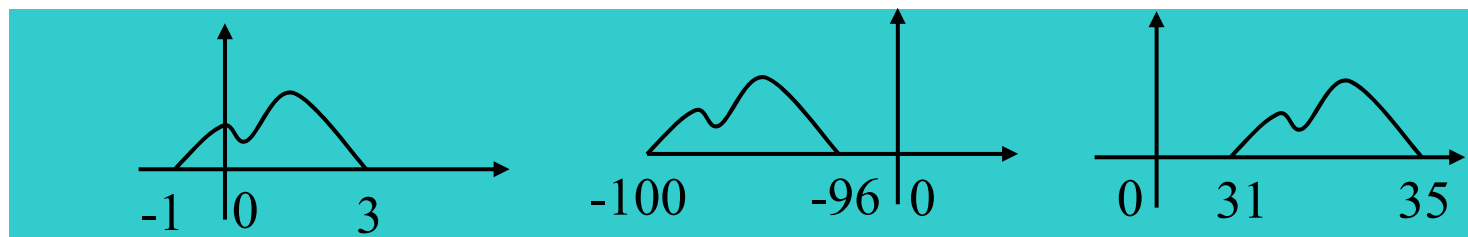
Just shifted versions of the same function.

Short Summary

- Expectation of a RV $\quad E(X) = \int_{-\infty}^{+\infty} x f_X(x)\, dx$

- Variance $\quad V(X) = E\left[(X - \mu_x)^2\right]$ or $E\left(X^2\right) - E\left(X\right)^2$

- Moments
  - kth moment
  - kth central moment
  - kth absolute moment

Transform Methods and Probability Generating Function

Transform Methods

As in linear systems theory, Fourier, Laplace and Z-transforms allow us to avoid integration

Convolution in time-domain is transformed to multiplication in another domain

In probability theory, where can we use them?

Computation of Moments

PDFs of Sums of Independent RVs

Connections

Fourier Transform $\longleftrightarrow$ Characteristic Function

Laplace Transform $\longleftrightarrow$ Moment Generating Function

Z-Transform $\longleftrightarrow$ Probability Generating Function

$$\Phi_X(w) := E\left\{ e^{jwX} \right\}$$

(annotated: $-w$ above w, and $-w$ above jwX)

$$= \int_{-\infty}^{+\infty} f_X(x) e^{jwx} \, dx$$

$$= \left[\int_{-\infty}^{+\infty} f_X(x) e^{-jwx} \, dx \right]^{*}$$

(annotations: * → conjunction; underbrace: fourier transform)

$$= \left[F\left\{ f_X(x) \right\} \right]^{*}$$

(annotation: fourier)

It's not exactly the F.T. of the PDF, which is just the Fourier transform of $f_X(x)$ evaluated at $-\omega$

Characteristic Function

- How about discrete r.v.?

- If X is a integer-valued discrete random variable with PMF $p_X(n)$, then

$$\Phi_X(\omega) = \sum_n e^{j\omega n} p_X(n)$$

- which is just a 2π-periodic **Fourier series**.

Fourier 的逆变换

$$E\left\{X^n\right\} = \int\limits_{-\infty}^{+\infty} x^n f_X(x)\, dx$$

$$= \frac{1}{j^n} \int\limits_{-\infty}^{+\infty} \frac{d^n}{dw^n} e^{jwx} f_X(x) \Bigg|_{w=0}$$

$$e^{jw}\Big|_{w=0} = 1.$$

$$= \frac{1}{j^n}\left(\frac{d^n}{dw^n}\Phi_X(w)\right)\Bigg|_{w=0}$$

Example 1

Suppose X is an exponential RV:

$$f_x(x) = \begin{cases} \alpha e^{-\alpha x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

What is the characteristic function?

By definition, $\phi_x(\omega) = \dfrac{\alpha}{\alpha - j\omega}$ (看 table 更简单)

Example 2

- $X \sim \text{Bernoulli}\{q\}$, i.e. $p_X(1) = q = 1 - p_X(0)$. Find $\Phi_X(\omega)$.

Example 3

1. $\Phi_X(\omega) = \dfrac{3}{3-j\omega}$. Find X's probability distribution.

2. $\Phi_X(\omega) = \dfrac{e^{j\omega}}{5} + \dfrac{4}{5}$. Find X's probability distribution.

1. $\phi_X(\omega) = \int e^{j\omega x} \cdot f(x)$

查 Table!

How to get the First Moment by characteristic function?

$$
E(X) = \frac{1}{j} \frac{d}{dw} \left[\alpha (\alpha - jw)^{-1} \right] \Big|_{w=0}
$$

求导φ, w=0. $\frac{1}{j}$.

$$
= \frac{1}{j} \left[-\alpha (\alpha - jw)^{-2} (-j) \right] \Big|_{w=0} \quad (1)
$$

$$
= \frac{\alpha}{(\alpha - jw)^2} \Big|_{w=0} = \frac{1}{\alpha}
$$

How to get the Second Moment by characteristic function?

$$E(X^2) = \frac{1}{j^2} \frac{d^2}{dw^2} \left[\alpha(\alpha - jw)^{-1} \right] \Big|_{w=0}$$

$$= \frac{1}{j^2} \frac{d}{dw} \left[-\alpha(\alpha - jw)^{-2}(-j) \right] \Big|_{w=0}$$

Substituted first derivative from (1) in previous slide.

$$= \frac{1}{j^2} \left[2\alpha(\alpha - jw)^{-3}(-j)^2 \right] \Big|_{w=0}$$

$$= \frac{2\alpha}{(\alpha - jw)^3} \Big|_{w=0} \qquad = \frac{2}{\alpha^2}$$

$Var[x] = E[x^2] - E[x]^2$

How to get the Variance by characteristic function?

$$Var(X) = E(X^2) - \left(E(X)\right)^2 = \frac{2}{\alpha^2} - \left(\frac{1}{\alpha}\right)^2 = \frac{1}{\alpha^2}$$

- Let X and Y be independent RVs and let $Z=X+Y$.

- The Convolution Theorem says:

$$f_Z(u) = f_X(u) * f_Y(u)$$

- We know

$$F\left\{f_Z(u)\right\} = F\left\{f_X(u)\right\} F\left\{f_Y(u)\right\}$$

- Conjugating both sides yields

$$\Phi_Z(w) = \Phi_X(w)\Phi_Y(w)$$

- The **moment generating functions (MGF)** of a random variable X is defined by

$$\varphi_X(t) = E\left\{e^{tX}\right\}$$

$$= \int_{-\infty}^{+\infty} f_X(x)e^{tx}\, dx$$

$$= \left(\int_{-\infty}^{+\infty} f_X(x)e^{-sx}\, dx\right)\Bigg|_{s=-t}$$

<u>Laplace Transform</u>

$$= \left(\mathcal{L}\left\{f_X(x)\right\}\right)\Bigg|_{s=-t}$$

Sometimes we also use $M_X(s) = E[e^{sX}]$ to represent moment generating function

- Note that $M_X(-s)$ is the Laplace transform of $f_X(x)$

  - Note the RoC of Laplace transform

    $\hookrightarrow$ range of convergence

- Since Laplace transform pairs are unique, given $f_X(x)$ one can obtain $M_X(s)$, and given $M_X(s)$ one can obtain $f_X(x)$

MGF

$$E\left\{X^n\right\} = \left(\frac{d^n}{dt^n}\phi_X(t)\right)\bigg|_{t=0}$$

Don't have to worry about j's

Can you derive that?

"Easy", omitted

$Z=X+Y$, X and Y independent RVs:

$$\phi_Z(t) = \phi_X(t)\phi_Y(t)$$

- If X is an exponential random variable with parameter $\lambda > 0$, find its moment generating function.

$$M_X(s) = \mathsf{E}\left[e^{sX}\right] = \int_0^\infty e^{sx} \overbrace{\lambda e^{-\lambda x}}^{f_X(x)} dx$$

$$= \lambda \int_0^\infty e^{x(s-\lambda)} dx \longrightarrow \frac{1}{s-\lambda} e^{x(s-\lambda)}\Big|_0^\infty$$

$$= 0 - \frac{1}{s-\lambda}$$

$$= \frac{1}{\lambda-s}$$

$$= \frac{\lambda}{\lambda - s}, \quad (\text{real}\{S\} < \lambda)$$

ROC → range of convergence

Hence mgf of an $X \sim \exp(\lambda)$ is defined only for $\text{real}\{S\} < \lambda$

If the random variable N is discrete and takes non-negative integer values, e.g. Poisson or Binomial, then we can define the Probability Generating Function (PGF) as:

$$G_N(z) = E\left\{z^N\right\} = \sum_{k=0}^{\infty} p_N(k) z^k$$

The PGF is not quite the Z-Transform of the PMF

Relating to Z-Transform

$$G_N(z^{-1}) = \sum_{k=0}^{\infty} p_N(k) z^{-k} = Z\left\{p_N(k)\right\}$$

Given the PGF, we can recover the PMF:

$$p_N(n) = \frac{1}{n!}\frac{d^n}{dz^n}G_N(z)\bigg|_{z=0}$$

Proof for *n=2*. Recall that

$$G_N(z) = p_N(0) + p_N(1)z^1 + p_N(2)z^2 + p_N(3)z^3 + \ldots$$

$$\frac{d^2}{dz^2}G_N(z) = 2p_N(2) + 3\cdot 2 p_N(3)z + 4\cdot 3 p_N(4)z^2 + \ldots$$

The 1/n! will get rid of the "2"　　　　*z=0 gets rid of all other terms.*

What is the PGF of the Poisson PMF?

$$G_N(z) = \sum_{k=0}^{\infty} p_N(k) z^k$$

$$= e^{-a} \sum_{k=0}^{\infty} \frac{(az)^k}{k!} = e^{-a} e^{az} = e^{a(z-1)}$$

Taylor Expansion

- $G_X(z) : \mathbb{C} \to \mathbb{C}$ (set of all complex numbers.) It is well defined when $|z| \leq 1$.

$$
\begin{aligned}
|G_X(z)| &= |\sum_{n=0}^{\infty} z^n p_X(n)| \leq \sum_{n=0}^{\infty} |z^n p_X(n)| \quad \text{(triangular inequality)} \\
&= \sum_{n=0}^{\infty} |z^n| p_X(n) \leq \sum_{n=0}^{\infty} p_X(n) = 1
\end{aligned}
$$

Take derivatives and evaluate at 1 to get various moments

$$\frac{d}{dz}G_N(z)\bigg|_{z=1} = E\left(\frac{d}{dz}z^N\right)\bigg|_{z=1} = E\left(Nz^{N-1}\right)\bigg|_{z=1} = E(N)$$

$$\frac{d^2}{dz^2}G_N(z)\bigg|_{z=1} = E\left(N(N-1)z^{N-2}\right)\bigg|_{z=1} = E(N^2 - N)$$

Not quite the variance

Since $\quad \dfrac{d^2}{dz^2} G_N(z)\bigg|_{z=1} = E(N^2 - N),$

$$\mathrm{Var}(N) = \dfrac{d^2}{dz^2} G_N(z)\bigg|_{z=1} + E(N) - \big[E(N)\big]^2$$

Use moment theorem to get Poisson moments

$$\frac{d}{dz}e^{a(z-1)}\bigg|_{z=1} = ae^{a(z-1)}\bigg|_{z=1} = a = E(N)$$

$$\frac{d^2}{dz^2}e^{a(z-1)}\bigg|_{z=1} = \frac{d}{dz}ae^{a(z-1)}\bigg|_{z=1} = a^2 e^{a(z-1)}\bigg|_{z=1} = a^2$$

$$\text{Var}(N) = \frac{d^2}{dz^2} G_N(z) \bigg|_{z=1} + E(N) - \left[E(N) \right]^2$$

$$= a^2 + a - a^2 = a$$

Why so many transforms?

- We have now discussed
  - characteristic function
  - moment generating functions
  - probability generating function

- Why do we need them all?
  - After all, the characteristic functions exists for all random variable, and we can use it to recover PMF and PDF and to find expectations.

Why so many transforms?

In the case of nonnegative, integer-valued random variables
- the formula of pgf is simpler to derive and to remember.

$$G_X(z) = E\left[z^X\right], \quad \varphi_X(\omega) = E\left[e^{j\omega X}\right]$$

$$\varphi_X(\omega) = G_X(e^{j\omega}), z = e^{j\omega}$$

- It is easier to compute the pmf

$$p_X(n) = \frac{G_X^{(n)}(0)}{n!}$$

If $M_x(s)$ exists,

- using mgf

$$M_X^{(k)}(s)\Big|_{s=0} = E\left[X^k\right]$$

is simpler than the characteristic function

$$\varphi_X^{(k)}(\omega)\Big|_{\omega=0} = j^k\, E\left[X^k\right]$$

- Transform Methods
    - Computation of Moments
    - PDFs of Sums of Independent RVs

Fourier Transform $\longleftrightarrow$ Characteristic Function

Laplace Transform $\longleftrightarrow$ Moment Generating Function

- The probability generating function is for discrete RVs
- The Moment Theorem is slightly different from the other transforms
- The Convolution Theorem would also hold

Thank You!