Probability and Random Process

Aimin Tang

The University of Michigan- Shanghai Jiao Tong University Joint Institute
Shanghai Jiao Tong University

Sep. 29 2020

Ve501 2020-2021 Fall

Based on Lecture Notes by Prof. Chong Han

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

Expectation of a Random Variable

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

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

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

Interpretation and Notation

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:

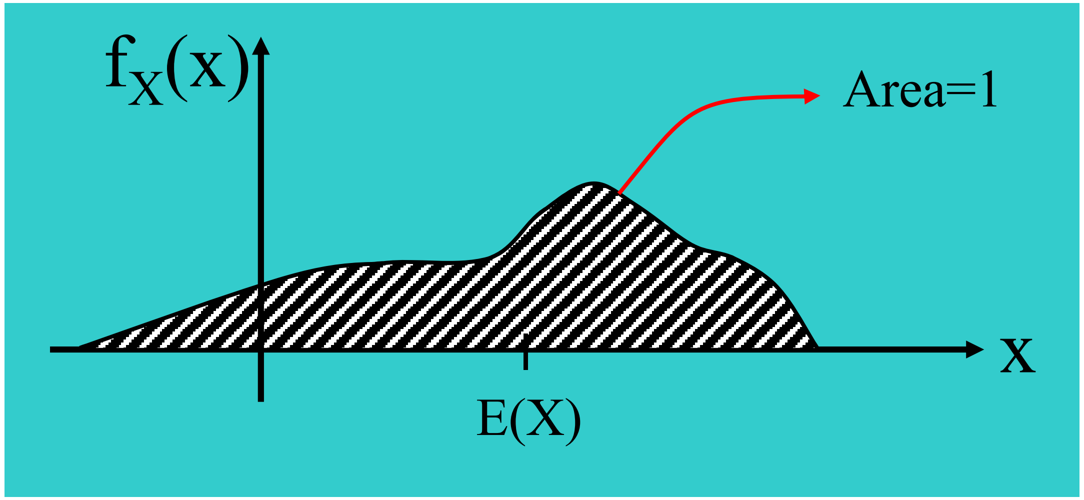
$$\mu_X, m_X, \eta_x, \text{or } \overline{X}$$

or just

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

Interpretation

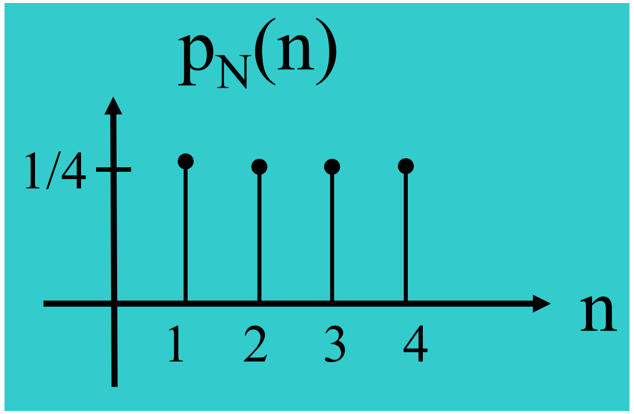
If the probability density is interpreted as a mass density along an axis, then $E(X)$ is the center of mass.



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

Discrete Example

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



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

Mean of a Function of a RV

- 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)

Proof of LOTUS

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

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

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

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

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

Properties of Expected Value

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

$$E(c) = c$$

2. The expected value is a linear operator:

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

$$E\big(H(X) + G(X)\big) = E\big(H(X)\big) + E\big(G(X)\big)$$

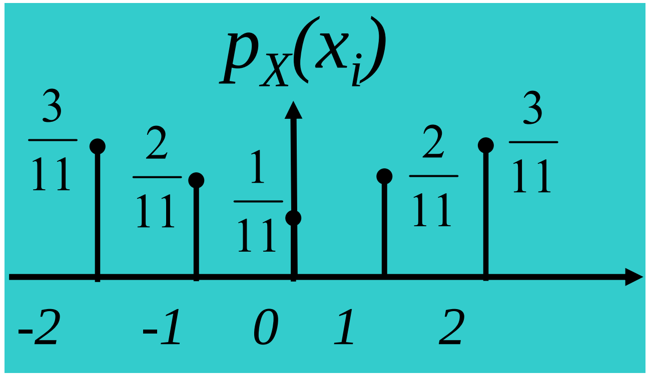
Ex:

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

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

* Constant with respect to the random variables

Example Calculations of Expectation



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

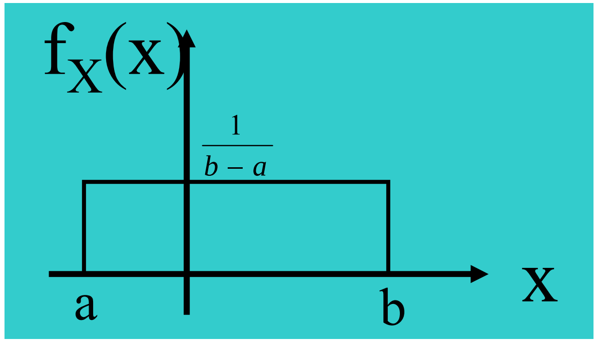
$$E(X^2) = \sum_{i=-2}^{2} i^2 p_X(x_i)$$

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

Mean of a Uniform RV

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

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



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

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

Expected Value of a Binomial RV

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

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

Expected Value of a Poisson RV

$$E(N)=\sum_{n=0}^{\infty} n\frac{\lambda^n}{n!}e^{-\lambda}=\sum_{n=1}^{\infty} n\frac{\lambda^n}{n!}e^{-\lambda}=\lambda\sum_{n=1}^{\infty}\frac{\lambda^{n-1}}{(n-1)!}e^{-\lambda}$$

Dropped $n = 0$

Change variables $i = n - 1$

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

Mean of a Gaussian RV

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

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

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

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

Proof for the Mean

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

$$= 0$$

Variance

Observe that because E(X) is not random (a constant),

$$E\left[X + E(X)\right] = 2E(X)$$

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

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

Definition for variance:

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

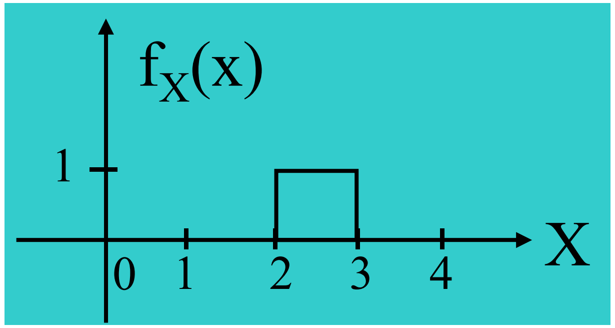
Alternative notation: Var(X)=σ_x^2

Interpretation

- 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



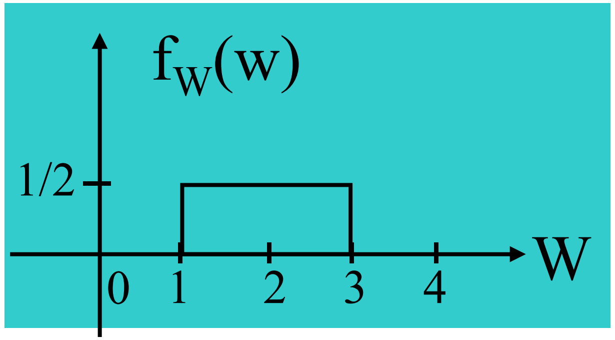
$$\mu_X = 2.5$$

$$V(X) = \int_2^3 (x-2.5)^2 \cdot 1dx = \int_2^3 \left(x^2 - 5x + (2.5)^2\right)dx$$

$$= \left(\frac{x^3}{3} - \frac{5x^2}{2} + (2.5)^2 x\right)\Bigg|_2^3 = \frac{27-8}{3} - \frac{5(9-4)}{2} + (2.5)^2(3-2)$$

$$= \frac{19}{3} - \frac{25}{2} + \frac{25}{4} = \frac{76-150+75}{12} = \frac{1}{12}$$

Example - II



$$\mu_W = 2$$

$$V(W) = \int_1^3 (w-2)^2 \cdot \frac{1}{2} dw = \frac{1}{2}\left(\frac{w^3}{3} - 2w^2 + 4w\right)\Bigg|_1^3$$

$$= \frac{1}{2}\left[\frac{27-1}{3} - 2(9-1) + 4 + (3-1)\right] = \frac{1}{2}\left[\frac{26}{3} - 16 + 8\right] = \frac{1}{3}$$

Alternative Formula

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

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

or:

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

Observe that if

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

Variance of a Gaussian RV

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

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

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

Variance of a Gaussian RV, Concluded

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

Moments

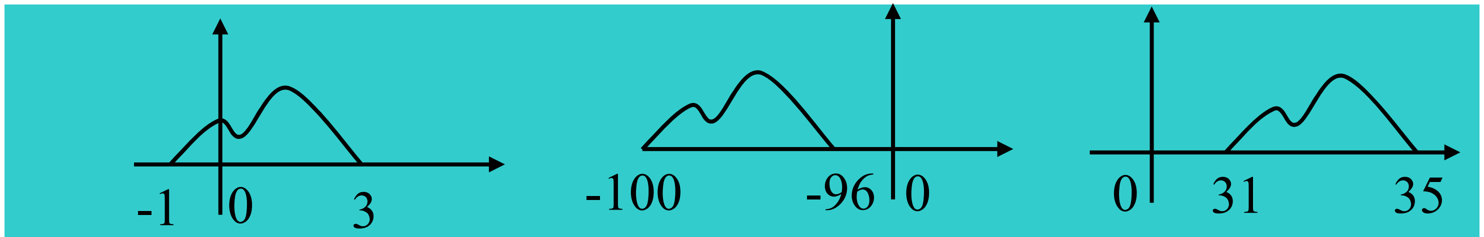
Definition: $$\text{k}^{\text{th}}\ \text{moment} = E(X^k)$$

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

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

Observation:

These three PDFs have the same k^{th} central moment



Just shifted versions of the same function.

Short Summary

- Expectation of a RV $E(X) = \int_{-\infty}^{+\infty} x f_X(x)dx$
- Variance $V(X) = E\left[(X-\mu_x)^2\right]$ 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

Ve501 2020-2021 Fall

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

Characteristic Function

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

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

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

conjugate

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

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**.

Getting Moments

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

$(jx)^n e^{jwx} \Big|_{w=0}$

$j^n \cdot x^n,$

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

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

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?

Integration

For the exponential RV, X,

$$\Phi_X(w) = \int_0^\infty \alpha e^{-\alpha x} e^{jwx} dx = \alpha \int_0^\infty e^{(-\alpha + jw)x} dx$$

$$= \alpha \left. \frac{e^{(-\alpha + jw)x}}{-\alpha + jw} \right|_0^{+\infty}$$

$$= 0 - \frac{\alpha}{(-\alpha + jw)}$$

$$= \frac{\alpha}{\alpha - jw}$$

Actually, this could be obtained easily from a Fourier Transform table.

Example 2

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

- Solution:

$$\Phi_X(\omega) = \sum_n e^{j\omega n} p_X(n) = 1 - q + e^{j\omega} q$$

Example 3

1. $\Phi_X(\omega) = \frac{3}{3-j\omega}$. Find X's probability distribution.
2. $\Phi_X(\omega) = \frac{e^{j\omega}}{5} + \frac{4}{5}$. Find X's probability distribution.

1. $f_X(x) = 3\exp(-3x), x \geq 0$
2. $X \sim$ Bernoulli$\{\frac{1}{5}\}$, *i.e.* $p_X(1) = \frac{1}{5} = 1 - p_X(0)$.

$\frac{4}{5}\delta(t-0) + \frac{1}{5}\delta(t-1)$

$= [\frac{4}{5} + \frac{1}{5}e^{-j\omega}]^*$

$= \frac{4}{5} + \frac{1}{5}e^{j\omega}$

First Moment

How to get the First Moment by characteristic function?

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

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

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

Second Moment and Variance

How to get the Second Moment by characteristic function?

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

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

Substituted first derivative from (1) in previous slide.

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

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

Second Moment and Variance

How to get the Variance by characteristic function?

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

Convolution Theorem

- 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

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

Moment Generating Function

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

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

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

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

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

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

Moment generating function & Laplace transform

- Note that $M_X(-s)$ is the Laplace transform of $f_X(x)$
  - Note the RoC of Laplace transform
- Since Laplace transform pairs are unique, given $f_X(x)$ one can obtain $M_X(s)$, and given $M_X(s)$ one can obtain $f_X(x)$

Same Uses

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

Don't have to worry about j's

Can you derive that?

$x^n = \frac{d^n}{dt^n} e^{tx}\Big|_{t=0}$

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

$$\varphi_Z(t) = \varphi_X(t)\varphi_Y(t)$$

MGF example

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

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

$$= \lambda \int_0^{\infty} e^{x(s-\lambda)} dx$$

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

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

Probability Generating Function

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

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

Recovering the PMF

Given the PGF, we can recover the PMF:

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

Proof for $n=2$. Recall that

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

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

The $1/n!$ will get rid of the "2"

$z=0$ gets rid of all other terms.

Example: Poisson r.v.

What is the PGF of the Poisson PMF?

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

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

Two Viewpoints

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

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

Moment Theorem

Take derivatives and evaluate at 1 to get various moments

$\frac{d}{dz}(p_N(0)+p_N(1)z^1+p_N(2)z^2+\cdots)\Big|_{z=1} = p_N(1)+2p_N(2)+3p_N(3)+\cdots = \sum_{k\ge1}^{N} k\,p_N(k)$

$= E(N)$

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

$\sum_{k=0}^{N} \frac{d}{dz} z^k \cdot p(k)$

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

Not quite the variance

Variance

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

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

Poisson, Cont'd

Use moment theorem to get Poisson moments

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

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

Poisson Variance

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

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

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.*

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

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

- *It is easier to compute the pmf*

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

Why so many transforms?

If $M_X(s)$ exists,

- *using mgf*

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

is simpler than the characteristic function

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

Short Summary

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

Fourier Transform ⟷ Characteristic Function

Laplace Transform ⟷ 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!