Topic 9: Moment Generating Function

Lecture Outline

- Moment Generating Function (MGF) of a RV
- Why introducing MGF?
- Notable MGF Properties
  - Uniqueness of MGF
    MGF ←→ PDF or PMF (1-to-1 correspondence)
  - MGF of sum of independent RVs

Reading: Textbook 4.4, 4.5

Definition

Definition

The *moment generating function (MGF)* of a RV X (or, *the transform of the distribution PDF of PMF*) is defined by

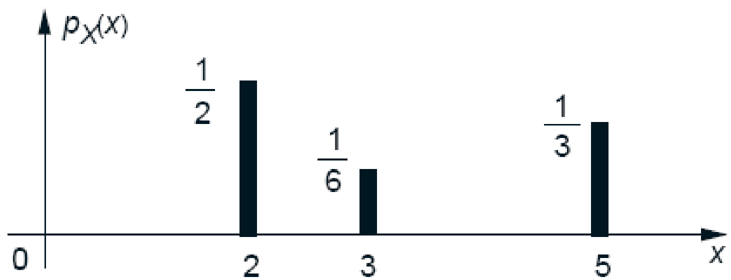
$$M_X(s) \equiv E[e^{sX}]$$
$$= \begin{cases} \sum_x e^{sx} p_X(x) & \text{discrete } X \\ \int e^{sx} f_X(x) dx & \text{continuous } X \end{cases}$$

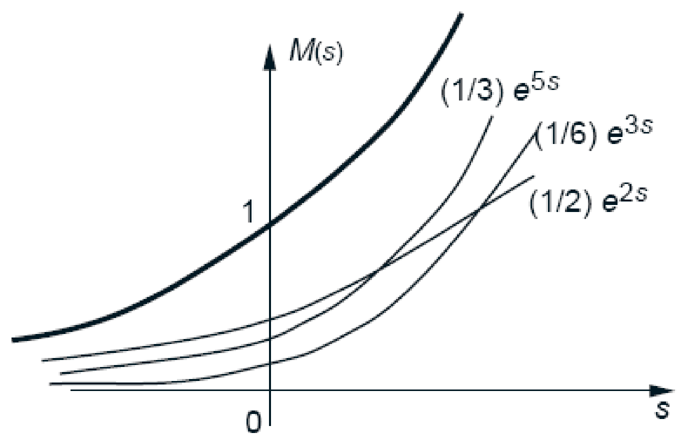
In general s (and hence e^{sX}) can be a complex number.

這類似於「訊號與系統」或「數位訊號處理」裡所介紹的 **Laplace Transform**

Example

- Simple Case:





Examples

- Bernoulli random variable *X* with parameter $p = p_X(1) = 1 - p_X(0)$:

$$M_X(s) = \sum_k e^{sk} p_X(k) = (1-p) + pe^s$$

- Geometric random variable: $p_X(k) = p(1-p)^{k-1}, k = 1,2,...$

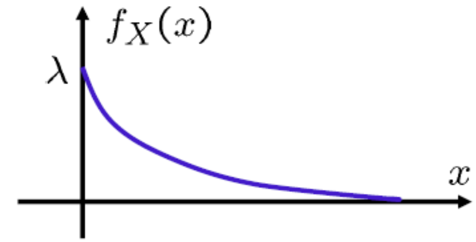
$$M_X(s) = \sum_{k=1}^{\infty} p_X(k) e^{sk} = \sum_{k=1}^{\infty} p(1-p)^{k-1} e^{sk}$$

$$= pe^s \sum_{k=1}^{\infty} (1-p)^{k-1} e^{s(k-1)} = pe^s \sum_{k=0}^{\infty} (1-p)^k e^{sk}$$

$$= \frac{pe^s}{1-(1-p)e^s}$$

MGF of Exponential

- Exponential $f_X(x) = \lambda e^{-\lambda x}, x \geq 0:$
  (p.231)



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

MGF of Gaussian RV

- X is Gaussian with mean m and variance σ^2:

$$M_X(s) \triangleq E[e^{sX}]$$
$$= e^{ms+\frac{1}{2}\sigma^2 s^2}$$

  - *Useful Fact*: If $X=aY+b$, then $M_X(s)= e^{sb} M_Y(as)$

    We can use this fact to obtain the result $M_X(s) = e^{ms+\frac{1}{2}\sigma^2 s^2}$ in the above, starting from the MGF of a standard Gaussian RV.

  - Consider *zero* mean and *unit variance Gaussian Y*: (p.232)

$$M_Y(s) \triangleq E[e^{sY}] = \int_{-\infty}^{\infty} e^{sy} f_Y(y)dy$$
$$= e^{\frac{s^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-s)^2}{2}} dy$$
$$= e^{\frac{s^2}{2}}.$$

Why Introducing MGF?

- MGF can be used to speed up computation of moments (mean, variance, or more exactly, $E[X^n]$ for any n)

- MGF associated with a RV uniquely determines the PMF/PDF of that RV. In other words, knowing MGF is equivalent to knowing the entire statistical property

  Inversion Theorem: PMF or PDF ←→ MGF

- Using MGF often greatly *simplifies* computations
  - *MGF of the* **sum** *of independent* RVs is the **multiplication** of their corresponding MGFs (see page 13)

- MGF can be used to prove important properties about RVs, such as Central Limit Theorem

Compute Moments with MGF

- Differentiate $M_X(s)$ with respect to s: (consider discrete case)

$$\frac{d}{ds}M_X(s) = \frac{d}{ds}\sum_x p_X(x)e^{sx}$$
$$= \sum_x p_X(x)(xe^{sx})$$

- Evaluate the derivative at $s = 0$ to find

$$M_X'(0) = \frac{d}{ds}M_X(s)\,|_{s=0} = E[X]$$

Thus, in general,

$$E[X] = M_X'(0)$$

The *k*th moment: $$E[X^k] = M^{(k)}(0) = \frac{d^k}{ds^k}M_X(s)\Big|_{s=0}$$

Examples

Proof:

$$\frac{d^k}{ds^k} M_X(s)|_{s=0} = \frac{d^k}{ds^k} E[e^{sX}]|_{s=0}$$

$$= E\left[\frac{d^k}{ds^k} e^{sX}\right]|_{s=0}$$

$$= E[X^k e^{sX}]|_{s=0}$$

$$= E[X^k]$$

Examples

- Bernoulli *X* with parameter $p = p_X(1) = 1 - p_X(0)$ has transform

$$M_X(s) = (1-p) + pe^s$$

Hence

$$M_X{}'(0) = p = E[X] \ , \quad M_X^{(2)}(0) = p = E[X^2]$$

Examples

- Exponential RV

$$M_X(s) = \frac{\lambda}{\lambda - s}$$

$$E[X] = \frac{d}{ds} M_X(s) \,|_{s=0}$$

$$= \frac{\lambda}{(\lambda - s)^2} \,|_{s=0}$$

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

$$E[X^2] = \frac{d^2}{ds^2} M_X(s) \,|_{s=0}$$

$$= \frac{2\lambda}{(\lambda - s)^3} \,|_{s=0}$$

$$= \frac{2}{\lambda^2}$$

Inversion of Transforms

The transform $M_X(s)$ associated with a random variable X uniquely determines the CDF (or equivalently the PDF or PMF) of X.

More specifically, if we know $M_X(s)$, then we can find the PDF of the PMF of X

Example (Ex 4.28, p.235)

A transform of a random variable X is given by

$$M(s) = \frac{1}{4}e^{-s} + \frac{1}{2} + \frac{1}{8}e^{4s} + \frac{1}{8}e^{5s}.$$

Find the probability law of X.

Example – Mixture of Two Random Variables

Let $X_1, \ldots, X_n$ be continuous random variables with PDFs $f_{X_1}, \ldots f_{X_n}$, and let Y be a random variable, which is equal to X_i with probability p_i. Then, by total probability theorem, the PDF of Y is given by

$$f_Y(y) = p_1 f_{X_1}(y) + \cdots + p_n f_{X_n}(y)$$

The moment generating function is

$$M_Y(s) = p_1 M_{X_1}(s) + \cdots + p_n M_{X_n}(s)$$

Example:

If we know the moment generating function of a random variable Y is given by

$$M_Y(s) = \frac{1}{2} \cdot \frac{1}{2-s} + \frac{3}{4} \cdot \frac{1}{1-s}$$

What is the probability law of Y?

Sum of Independent Random Variables

- Suppose that X and Y are independent random variables.
  Recall that for any functions g and h,

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$$

In particular, suppose that $W = X + Y$. Then, the MGF $M_W(s)$ of W is

$$\boxed{M_W(s) = M_X(s)M_Y(s)}$$

i.e., *adding independent RV's produces a new RV whose transform is the product of the original transforms*

- More generally, if $X_1, \ldots, X_n$ is a collection of independent random variables, and

$$W = X_1 + X_2 + \cdots + X_n$$

Then,

$$\boxed{M_W(s) = M_{X_1}(s) \cdots M_{X_n}(s)}$$

Sum of Two Independent RVs

- The probability distribution of $W=X+Y$ for independent X and Y can by obtained by
  - Inverting the transform $M_W(s)=M_X(s)M_Y(s)$

我們由原本PDF之間convolution的運算<u>轉換</u>成了MGF之間乘法的運算

- *Cf.* Direct Evaluation ➔ convolution (see topic 8)

<u>Discrete Case</u>

Let X and Y are independent discrete RVs with PMFs $p_X(x)$ and $p_Y(y)$. Then,

$$p_W(w) = \sum_x p_X(x) p_Y(w-x)$$

<u>Continuous Case</u>

Let X and Y be indep. continuous RVs with PDFs $f_X(x)$ and $f_Y(y)$.

$$f_W(w) = \int_{-\infty}^{\infty} f_X(x) f_W(w-x) dx$$

Examples

Assume that X is Bernoulli with parameter $p = p_X(1) = 1 - p_X(0)$
and

$$M_X(s) = \sum_{k=0}^{1} e^{sk} p_X(k) = (1-p) + pe^s$$

If $\{ X_i ; i = 1, \ldots, n \}$ are independent Bernoulli random variables with identical distributions and

$$Y_n = X_1 + X_2 + \cdots + X_n$$

Then, we know

$$M_{Y_n}(s) = [(1-p) + pe^s]^n$$

Sum of Independent Normal RVs

- Let X and Y be independent **normal** random variables with means μ_x and μ_y and variances $\sigma^2{}_x$ and $\sigma^2{}_y$, respectively.

  What is the PDF of $W = X + Y$?

  - General normal X: $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
  - Transform: $M_X(s) = e^{(s^2\sigma^2/2)+s\mu}$
  - **Sum of independent normals**:

$$X \sim N(\mu_x, \sigma_x^2) \qquad Y \sim N(\mu_y, \sigma_y^2) \qquad W = X + Y$$

$$\begin{aligned} M_W(s) &= M_X(x) M_Y(s) \\ &= e^{(s^2\sigma_x^2/2)+s\mu_x} \cdot e^{(s^2\sigma_y^2/2)+s\mu_y} \\ &= e^{[s^2(\sigma_x^2+\sigma_y^2)/2+s(\mu_x+\mu_y)]} \end{aligned}$$

  - Conclude: $W \sim N(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$

Sum of A Random Number of Independent RVs

- We consider the sum

$$Y = X_1 + X_2 + \cdots + X_N$$

where N is a random variable that takes nonnegative integer values, and $X_1, X_2, \ldots X_N$ are i.i.d. random variables with $E[X_i]=\mu$, $\text{var}(X_i)=\sigma^2$

Then,

$$E[Y] = \mu E[N]$$

$$\text{var}(Y) = E[N]\sigma^2 + \mu^2 \text{var}(N)$$

- **Mean:** $\mathbf{E}[Y] = \mathbf{E}[\mathbf{E}[Y \mid N]] = \mathbf{E}[N\mathbf{E}[X]] = \underline{\mathbf{E}[N]\mathbf{E}[X]}$
- **Variance:**

$$\begin{aligned}\text{Var}(Y) &= \mathbf{E}[\text{Var}(Y \mid N)] + \text{Var}(\mathbf{E}[Y \mid N]) \\ &= \underline{\mathbf{E}[N]\text{Var}(X) + (\mathbf{E}[X])^2 \text{Var}(N)}\end{aligned}$$

Transform of Sum of A Random Number of RVs

- For the sum of a random number of independent RVs

$$Y = X_1 + X_2 + \cdots + X_N$$

Then, $$\boxed{M_Y(s) = M_N\left(\log M_X(s)\right)}$$

Thus, to get $M_Y(s)$, we start with $M_N(s)$ and replace each occurence e^s of by $M_X(s)$.

- If $Y = X_1 + \cdots + X_N$, we have:

$$\begin{aligned} M_Y(s) &= \mathbf{E}[e^{sY}] \\ &= \mathbf{E}\left[\mathbf{E}[e^{sY} \mid N]\right] \\ &= \mathbf{E}\left[\mathbf{E}[e^{s(X_1+\cdots+X_N)} \mid N]\right] \\ &= \mathbf{E}\left[M_X(s)^N\right] = \sum_{n=0}^{\infty} (M_X(s))^n p_N(n) \end{aligned}$$

- Compare with: $M_N(s) = \mathbf{E}[(e^s)^N]$

Bookstore Example (1)

- Jane visits a number of bookstores, looking for "Great Expectations" (a novel by Charles Dickens) (Example *4.35*)

  A bookstore caries such a book with probability p=1/3.

  The time Jane spends in each bookstore is exponentially distributed with λ=3.

  Jane will visit bookstores until she finds the book.

  We wish to find the mean, variance, and PDF of the total time she spent in bookstores.

Solution:

Let N be the number of book stores, X_i be the time spent at bookstore i.

Total time: $Y=X_1+X_2+ \ldots +X_N$

Bookstore Example (2)

- Number of bookstores, N :
  - **PMF** $p_N(n) = \frac{1}{3}\left(\frac{2}{3}\right)^{n-1}$ (geometric, from *n*=1)
  - **Mean** $\mathbf{E}[N] = \frac{1}{\frac{1}{3}} = 3$
  - **Variance** $\mathrm{Var}(N) = \frac{1-\frac{1}{3}}{(\frac{1}{3})^2} = 6$
- Time in each bookstore, X (i.i.d., indep of N):
  - **PDF** $f_X(x) = 3e^{-3x}$ $x \geq 0$
  - **Mean** $\mathbf{E}[X] = \frac{1}{3}$
  - **Variance** $\mathrm{Var}(X) = \frac{1}{9}$
- Total time, Y :
  - **Mean** $\mathbf{E}[Y] = \mathbf{E}[N]\mathbf{E}[X] = 1$
  - **Variance** $\mathrm{Var}(Y) = \mathbf{E}[N]\mathrm{Var}(X) + (\mathbf{E}[X])^2\mathrm{Var}(N)$
    $= 1$

Bookstore Example (3)

- Number of bookstores:
  - **Transform** $M_N(s) = \dfrac{e^s/3}{1 - 2e^s/3} = \mathbf{E}\left[(e^s)^N\right]$
- Time in each bookstore:
  - **Transform** $M_X(s) = \dfrac{3}{3-s}$
- Total time:
  - **Transform** $M_Y(s) = \mathbf{E}\left[M_X(s)^N\right]$

$$= \frac{\left(\frac{3}{3-s}\right)/3}{1 - 2\left(\frac{3}{3-s}\right)/3} = \frac{1}{1-s}$$

  - **PDF:** $f_Y(y) = e^{-y} \quad y \geq 0$ (exponential, with λ = 1)

Bookstore Example (4)

In the general case,

$$M_X(s) = \frac{\lambda}{\lambda - s}$$

$$M_N(s) = \frac{pe^s}{1-(1-p)e^s}$$

$$M_Y(s) = \frac{pM_X(s)}{1-(1-p)M_X(s)} = \frac{\dfrac{p\lambda}{\lambda - s}}{1-(1-p)\dfrac{\lambda}{\lambda - s}},$$

$$f_Y(y) = p\lambda e^{-p\lambda y}, \qquad y \geq 0.$$