Further Topics on Random Variables: Transforms (Moment Generating Functions)

Berlin Chen
Department of Computer Science & Information Engineering
National Taiwan Normal University

Reference:

- D. P. Bertsekas, J. N. Tsitsiklis, *Introduction to Probability* , Section 4.1

Aims of This Chapter

- Introduce methods that are useful in
  - Dealing with the sum of independent random variables, including the case where the number of random variables is itself random
  - Addressing problems of estimation or prediction of an unknown random variable on the basis of observed values of other random variables

Transforms

- Also called **moment generating functions** of random variables

- The **transform** of the distribution of a random variable X is a function $M_X(s)$ of a free parameter s, defined by

$$M_X(s) = \mathbf{E}\left[e^{sX}\right]$$

  – If X is discrete

$$M_X(s) = \sum_x e^{sx} p_X(x)$$

  – If X is continuous

$$M_X(s) = \int_{-\infty}^{\infty} e^{sx} f_X(x) dx$$

Illustrative Examples (1/5)

- **Example 4.1.** Let

$$p_X(x) = \begin{cases} 1/2, & \text{if } x = 2, \\ 1/6, & \text{if } x = 3, \\ 1/3, & \text{if } x = 5. \end{cases}$$

$$\therefore M_X(s) = \mathbf{E}\left[e^{sX}\right] = \sum_x e^{sx} p_X(x)$$

$$= \frac{1}{2}e^{2s} + \frac{1}{6}e^{3s} + \frac{1}{3}e^{5s}$$

Notice that :

$$M_X(0) = \mathbf{E}\left[e^{0X}\right] = \sum_x e^{0x} p_X(x)$$

$$= \sum_x p_X(x) = 1$$

Illustrative Examples (2/5)

- **Example 4.2. The Transform of a Poisson Random Variable.** Consider a Poisson random variable X with parameter λ :

$$p_X(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0,1,2,\ldots$$

$$M_X(s) = \sum_{x=0}^{\infty} e^{sx} \frac{\lambda^x e^{-\lambda}}{x!}$$

$$= e^{-\lambda} \sum_{x=0}^{\infty} \frac{a^x}{x!} \qquad \left(\text{Let } a = e^s \lambda\right)$$

$$= e^{-\lambda} e^a \qquad \left(\because \text{ McLaurin series} \left(1 + a + \frac{a^2}{2!} + \frac{a^3}{3!} + \cdots\right) = e^a\right)$$

$$= e^{a-\lambda}$$

$$= e^{\lambda\left(e^s - 1\right)}$$

Illustrative Examples (3/5)

- **Example 4.3. The Transform of an Exponential Random Variable.** Let X be an exponential random variable with parameter λ:

$$f_X(x) = \lambda e^{-\lambda x}, \quad x \geq 0$$

$$\begin{aligned} M_X(s) &= \int_0^\infty e^{sx} \lambda e^{-\lambda x} dx \\ &= \lambda \int_0^\infty e^{(s-\lambda)x} dx \\ &= \lambda \frac{e^{(s-\lambda)x}}{(s-\lambda)} \Big|_0^\infty \qquad (\text{if } s - \lambda < 0) \\ &= \frac{\lambda}{\lambda - s} \end{aligned}$$

Notice that :
$M_X(s)$ can be calculated only when $s < \lambda$

Illustrative Examples (4/5)

- **Example 4.4. The Transform of a Linear Function of a Random Variable.** Let $M_X(s)$ be the transform associated with a random variable X . Consider a new random variable $Y = aX + b$. We then have

$$M_Y(s) = \mathbf{E}\left[e^{s(aX+b)}\right] = e^{sb}\mathbf{E}\left[e^{saX}\right] = e^{sb}M_X(sa)$$

  – For example, if X is exponential with parameter $\lambda = 1$ and $Y = 2X + 3$, then

$$M_X(s) = \frac{\lambda}{\lambda - s} = \frac{1}{1 - s}$$

$$M_Y(s) = e^{3s}M_X(2s) = e^{3s}\frac{1}{1 - 2s}$$

Illustrative Examples (5/5)

- **Example 4.5. The Transform of a Normal Random Variable.** Let X be normal with mean μ and variance σ^2.

We first calculate the transform of a standard normal random variable Y

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$$

$$M_Y(s) = \int_{-\infty}^{\infty} e^{sy} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

$$= e^{s^2/2} \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\left[\left(y^2/2\right) - sy + \left(s^2/2\right)\right]} dy$$

$$= e^{s^2/2} \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(y-s)^2/2} dy$$

1

$$= e^{s^2/2}$$

Since we also know that $Y = \frac{X\text{-}\mu}{\sigma}$,

we can have $X = \sigma Y + \mu$

$$\therefore\ M_X(s) = e^{s\mu} M_Y(s\sigma)$$

$$= e^{s\mu} \cdot e^{s^2\sigma^2/2}$$

$$= e^{s\mu + \left(s^2\sigma^2/2\right)}$$

From Transforms to Moments (1/2)

- Given a random variable X , we have

$$M_X(s) = \mathbf{E}\left[e^{sx}\right] = \int_{-\infty}^{\infty} e^{sx} f_X(x) dx \quad \text{(If } X \text{ is continuous)}$$

Or

$$M_X(s) = \mathbf{E}\left[e^{sx}\right] = \sum_{x} e^{sx} p_X(x) \quad \text{(If } X \text{ is discrete)}$$

- When taking the derivative of the above functions with respect to s (for example, the continuous case)

$$\frac{dM_X(s)}{ds} = \frac{d\int_{-\infty}^{\infty} e^{sx} f_X(x) dx}{ds} = \int_{-\infty}^{\infty} x e^{sx} f_X(x) dx$$

  - If we evaluate it at $s = 0$, we can further have

the first moment of X

$$\frac{dM_X(s)}{ds}\bigg|_{s=0} = \int_{-\infty}^{\infty} x e^{sx} f_X(x) dx \bigg|_{s=0} = \int_{-\infty}^{\infty} x f_X(x) dx = \mathbf{E}[x]$$

From Transforms to Moments (2/2)

- More generally, the differentiation of $M_X(s)$ *n* times with respect to s will yield

$$\frac{d^n M_X(s)}{d^n s}\Big|_{s=0} = \int_{-\infty}^{\infty} x^n e^{sx} f_X(x) dx \Big|_{s=0} = \int_{-\infty}^{\infty} x^n f_X(x) dx = \mathbf{E}[x^n]$$

the *n*-th moment of X

Illustrative Examples (1/2)

- **Example 4.6a.** Given a random variable X with PMF:

$$p_X(x) = \begin{cases} 1/2, & \text{if } x = 2, \\ 1/6, & \text{if } x = 3, \\ 1/3, & \text{if } x = 5. \end{cases}$$

$$M_X(s) = \mathbf{E}\left[e^{sX}\right] = \sum_x e^{sx} p_X(x)$$
$$= \frac{1}{2}e^{2s} + \frac{1}{6}e^{3s} + \frac{1}{3}e^{5s}$$

$$\Rightarrow \mathbf{E}[X] = \frac{dM(s)}{ds}\Big|_{s=0}$$
$$= \frac{1}{2}\cdot 2\cdot e^{2s} + \frac{1}{6}\cdot 3\cdot e^{3s} + \frac{1}{3}\cdot 5\cdot e^{5s}\Big|_{s=0}$$
$$= 1 + \frac{3}{6} + \frac{5}{3} = \frac{19}{6}$$

$$\Rightarrow \mathbf{E}[X^2] = \frac{d^2M(s)}{d^2s}\Big|_{s=0}$$
$$= \frac{1}{2}\cdot 4\cdot e^{2s} + \frac{1}{6}\cdot 9\cdot e^{3s} + \frac{1}{3}\cdot 25\cdot e^{5s}\Big|_{s=0}$$
$$= 2 + \frac{9}{6} + \frac{25}{3} = \frac{71}{6}$$

Illustrative Examples (2/2)

- **Example 4.6b.** Given an exponential random variable X with PMF:

$$f_X(x) = \lambda e^{-\lambda x}, \qquad x \geq 0.$$

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

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

$$= \lambda \frac{e^{(s-\lambda)x}}{(s-\lambda)} \Big|_0^\infty \qquad (\text{if } s - \lambda < 0)$$

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

$$\Rightarrow \mathbf{E}[X] = \frac{dM_X(s)}{ds}\Big|_{s=0} = \frac{\lambda}{(\lambda - s)^2}\Big|_{s=0} = \frac{1}{\lambda}$$

$$\Rightarrow \mathbf{E}\left[X^2\right] = \frac{d^2 M_X(s)}{d^2 s}\Big|_{s=0} = \frac{2\lambda}{(\lambda - s)^3}\Big|_{s=0} = \frac{2}{\lambda^2}$$

Two Properties of Transforms

- For any random variable X , we have

$$M_X(0) = \mathbf{E}\left[e^{0X}\right] = \mathbf{E}[1] = 1$$

- If random variable X only takes nonnegative integer values ($x = 0,1,2,\cdots$)

$$\lim_{s \to -\infty} M_X(s) = \mathbf{P}(X = 0)$$

$$\lim_{s \to -\infty} M_X(s) = \lim_{s \to -\infty} \sum_{k=0}^{\infty} \mathbf{P}(X = k) e^{sk} = \mathbf{P}(X = 0)$$

Inversion of Transforms

- Inversion Property
  - The transform $M_X(s)$ associated with a random variable X uniquely determines the probability law of X, assuming that $M_X(s)$ is finite for all s in an interval $[-a, a]$, $a \geq 0$
    - The determination of the probability law of a random variable => The PDF and CDF
    - In particular, if $M_X(s) = M_Y(s)$ for all s in $[-a, a]$, then the random variables X and Y have the same probability law

Illustrative Examples (1/2)

- **Example 4.7.** We are told that the transform associated with a random variable X is

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

If we compare the formula $M_X(s) = \sum_x e^{sx} p_X(x)$, (if X is discrete)

we will have $p_X(-1) = \mathbf{P}(X = -1) = \frac{1}{4}$,

$$p_X(0) = \mathbf{P}(X = 0) = \frac{1}{2},$$

$$p_X(4) = \mathbf{P}(X = 4) = \frac{1}{8},$$

$$p_X(5) = \mathbf{P}(X = 5) = \frac{1}{8}.$$

Illustrative Examples (2/2)

- **Example 4.8. The Transform of a Geometric Random Variable.** We are told that the transform associated with random variable X is of the form

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

  - Where $0 < p \le 1$

If $(1-p)e^s < 1$, we can set $\alpha = (1-p)e^s$.

- Based on the property that

$$\frac{1}{1-\alpha} = 1 + \alpha + \alpha^2 + \ldots, \quad (\alpha < 1)$$

- $M_X(s)$ is then expressed as

$$M_X(s) = pe^s\left(1 + (1-p)e^s + (1-p)^2 e^{2s} + (1-p)^3 e^{3s} + \ldots\right)$$

- It can be infered that X is a discrete random variable with PDF

$$p_X(x) = p(1-p)^{x-1}, \quad x = 1,2,\ldots$$

$\therefore X$ is a geometric random variable

$$\mathbf{E}[X] = \frac{dM_X(s)}{ds}\Big|_{s=0}$$

$$= \frac{d\left(\frac{pe^s}{1-(1-p)e^s}\right)}{ds}\Big|_{s=0}$$

$$= \left[\frac{pe^s}{1-(1-p)e^s} + \frac{(1-p)pe^s}{\left(1-(1-p)e^s\right)^2}\right]\Big|_{s=0}$$

$$= 1 + \frac{(1-p)p}{p^2}$$

$$= \frac{1}{p}$$

Mixture of Distributions of Random Variables (1/3)

- Let $X_1, \ldots, X_n$ be continuous random variables with PDFs $f_{X_1}, \cdots, f_{X_n}$, and let Y be a random variable, which is equal to X_i with probability p_i ($\sum_{i=1}^{n} p_i = 1$). Then,

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

$$(\text{Note that this is quite different from}: Y = p_1 X_1 + \cdots + p_n X_n)$$

and

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

Mixture of Distributions of Random Variables (2/3)

$$f_Y(y) = p_1 f_{X_1}(y) + \cdots + p_n f_{X_n}(y), \quad \sum_{i=1}^{n} p_i = 1$$

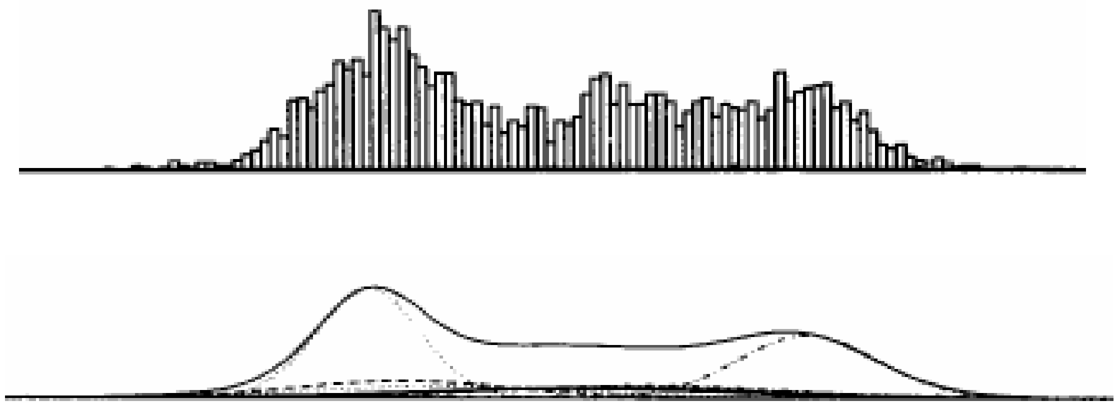
$$\begin{aligned}
M_Y(s) = \mathbf{E}\left[e^{sY}\right] &= \int_{-\infty}^{\infty} e^{sy} f_Y(y)\,dy \\
&= \int_{-\infty}^{\infty} e^{sy}\left(p_1 f_{X_1}(y) + \cdots + p_n f_{X_n}(y)\right)dy \\
&= \left[\int_{-\infty}^{\infty} e^{sy} p_1 f_{X_1}(y)\,dy\right] + \cdots + \left[\int_{-\infty}^{\infty} e^{sy} p_n f_{X_n}(y)\,dy\right] \\
&= \left[p_1 \int_{-\infty}^{\infty} e^{sx_1} f_{X_1}(x_1)\,dx_1\right] + \cdots + \left[p_n \int_{-\infty}^{\infty} e^{sx_n} f_{X_n}(x_n)\,dx_n\right] \\
&= p_1 M_{X_1}(s) + \cdots + p_n M_{X_n}(s)
\end{aligned}$$

Mixture of Distributions of Random Variables (3/3)

- Mixture of Gaussian Distributions
  - More complex distributions with multiple local maxima can be approximated by Gaussian (a unimodal distribution) mixture

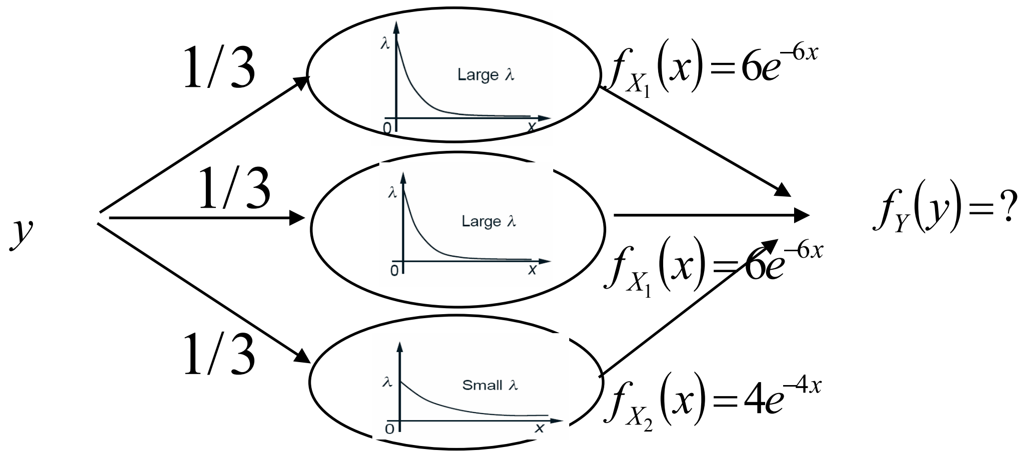
$$f_Y(y) = \sum_{i=1}^{n} p_i N_i\left(y; \mu_i, \sigma_i^2\right), \qquad \sum_{i=1}^{n} p_i = 1$$

  - Gaussian mixtures with enough mixture components can approximate any distribution



An Illustrative Example (1/2)

- **Example 4.9. The Transform of a Mixture of Two Distributions.** The neighborhood bank has three tellers, two of them fast, one slow. The time to assist a customer is exponentially distributed with parameter λ = 6 at the fast tellers, and λ = 4 at the slow teller. Jane enters the bank and chooses a teller at random, each one with probability 1/3. Find the PDF of the time it takes to assist Jane and the associated transform



An Illustrative Example (2/2)

– The service time of each teller is exponentially distributed

$$f_{X_1}(x) = 6e^{-6x}, \qquad x \geq 0.$$ the faster teller

$$f_{X_2}(x) = 4e^{-4x}, \qquad x \geq 0.$$ the slower teller

– The distribution of the time that a customer spends in the bank

$$f_Y(y) = \frac{2}{3} \cdot 6e^{-6y} + \frac{1}{3} \cdot 4e^{-4y}, \qquad y \geq 0.$$

- The associated transform

$$M_Y(s) = \mathbf{E}\left[e^{sy}\right] = \int_0^\infty e^{sy} \left(\frac{2}{3} \cdot 6e^{-6y} + \frac{1}{3} \cdot 4e^{-4y} \right) dy$$

$$= \frac{2}{3} \int_0^\infty e^{sy} \cdot 6e^{-6y} dy + \frac{1}{3} \int_0^\infty e^{sy} \cdot 4e^{-4y} dy$$

$$= \frac{2}{3} \cdot \frac{6}{6-s} + \frac{1}{3} \cdot \frac{4}{4-s} \qquad (\text{for } s < 4)$$

cf. p.12

Sum of Independent Random Variables

- Addition of **independent** random variables corresponds to multiplication of their transforms
  - Let X and Y be independent random variables, and let $W = X + Y$. The transform associated with W is,

$$M_W(s) = \mathbf{E}\left[e^{sW}\right] = \mathbf{E}\left[e^{s(X+Y)}\right] = \mathbf{E}\left[e^{sX}e^{sY}\right] = \mathbf{E}\left[e^{sX}\right]\mathbf{E}\left[e^{sY}\right] = M_X(s)M_Y(s)$$

    - Since X and Y are independent, and e^{sX} and e^{sY} are functions of X and Y, respectively

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

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

Illustrative Examples (1/3)

- **Example 4.10. The Transform of the Binomial.**
  Let $X_1, \ldots, X_n$ be independent Bernoulli random variables with a common parameter p . Then,

$$M_{X_i}(s) = (1 - p)e^{s \cdot 0} + pe^{s \cdot 1} = 1 - p + pe^s, \quad \text{for } i = 1, \ldots, n$$

  - If $Y = X_1 + \cdots + X_n$, Y can be thought of as a binomial random variable with parameters n and p , and its corresponding transform is given by

$$M_Y(s) = \prod_{i=1}^{n} M_{X_i}(s) = \left(1 - p + pe^s\right)^n$$

Illustrative Examples (2/3)

- **Example 4.11. The Sum of Independent Poisson Random Variables is Poisson.**
  - Let X and Y be independent Poisson random variables with means λ and μ , respectively
    - The transforms of X and Y will be the following, respectively

$$M_X(s) = e^{\lambda(e^s - 1)}, \; M_Y(s) = e^{\mu(e^s - 1)}$$

cf. p.5

  - If $W = X + Y$, then the transform of the random variable W is

$$\begin{aligned} M_W(s) &= M_X(s) M_Y(s) \\ &= e^{\lambda(e^s - 1)} e^{\mu(e^s - 1)} \\ &= e^{(\lambda + \mu)(e^s - 1)} \end{aligned}$$

    - From the transform of W , we can conclude that W is also a Poisson random variable with mean $\lambda + \mu$

Illustrative Examples (3/3)

- **Example 4.12. The Sum of Independent Normal Random Variables is Normal.**
  - Let X and Y be independent normal random variables with means μ_x, μ_y , and variances σ_x^2, σ_y^2 , respectively
    - The transforms of X and Y will be the following, respectively

$$M_X(s) = e^{\frac{\sigma_x^2 s^2}{2} + \mu_x s}, \quad M_Y(s) = e^{\frac{\sigma_y^2 s^2}{2} + \mu_y s}$$

cf. p.8

  - If $W = X + Y$, then the transform of the random variable W is

$$M_W(s) = M_X(s) M_Y(s)$$

$$= e^{\frac{(\sigma_x^2 + \sigma_y^2) s^2}{2} + (\mu_x + \mu_y) s}$$

    - From the transform of W , we can conclude that W also is normal with mean $\mu_x + \mu_y$ and variance $\sigma_x^2 + \sigma_y^2$

Tables of Transforms (1/2)

Transforms for Common Discrete Random Variables

Bernoulli(p)

$$p_X(k) = \begin{cases} p, & \text{if } k = 1, \\ 1 - p, & \text{if } k = 0. \end{cases} \qquad M_X(s) = 1 - p + pe^s.$$

Binomial(n, p)

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n.$$

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

Geometric(p)

$$p_X(k) = p(1-p)^{k-1}, \quad k = 1, 2, \ldots \qquad M_X(s) = \frac{pe^s}{1 - (1-p)e^s}.$$

Poisson(λ)

$$p_X(k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0, 1, \ldots \qquad M_X(s) = e^{\lambda(e^s - 1)}.$$

Uniform(a, b)

$$p_X(k) = \frac{1}{b - a + 1}, \quad k = a, a+1, \ldots, b.$$

$$M_X(s) = \frac{e^{as}}{b - a + 1} \frac{e^{(b-a+1)s} - 1}{e^s - 1}.$$

Tables of Transforms (2/2)

Transforms for Common Continuous Random Variables

Uniform(a, b)

$$f_X(x) = \frac{1}{b-a}, \quad a \le x \le b. \qquad M_X(s) = \frac{1}{b-a}\frac{e^{sb} - e^{sa}}{s}.$$

Exponential(λ)

$$f_X(x) = \lambda e^{-\lambda x}, \quad x \ge 0. \qquad M_X(s) = \frac{\lambda}{\lambda - s}, \qquad (s > \lambda).$$

Normal(μ, σ^2)

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2}, \quad -\infty < x < \infty. \qquad M_X(s) = e^{\frac{\sigma^2 s^2}{2} + \mu s}.$$

Recitation

- SECTION 4.1 Transforms
  - Problems 2, 4, 5, 7, 8