9.07 Introduction to Probability and Statistics for Brain and Cognitive Sciences
Emery N. Brown

<div align="center">

Lecture 6: Expectation and Variance, Covariance and Correlation,
And Movement Generating Functions

</div>

I. Objectives

> **Understand how to compute expectations and variances for linear combinations of random variables.**
>
> **Understand how to compute the covariance between linear combinations of random variables.**
>
> **Understand the concept of moments of a probability density.**
>
> **Understand how the moments of a probability density or probability mass function can be derived from the moment generating function.**
>
> **Understand the basic properties of moment generating functions and their use in probability calculations.**

II. Expectations and Covariances

A. Expectation

In **Lectures 2 and 3**, we defined the expected value of a discrete and continuous random variable respectively as

1. Discrete
If
$$X: \ p_x(x) \tag{6.1}$$
then

$$E(X) = \sum x_i p_x(x_i) \tag{6.}$$

provided that $\sum |x_i| p_x(x_i) < \infty.$

2. Continuous
If
$$X: \ f_x(x) \tag{6.3}$$
then

$$E(x) = \int_{-\infty}^{\infty} x f_x(x) dx \tag{6.4}$$

provided that $\int |x| f_x(x) dx < \infty.$

Remark 6.1. The expected value has an important physical interpretation as the center of mass or the first moment.

We also stated that if $Y = g(X)$, we can obtain the expected value of Y directly without first computing $p_y(y)$ or $f_y(y)$. That is, given $Y = g(X)$

1. Discrete

If $X : p_x(x)$ then

$$E(Y) = \sum_x g(x) p_x(x) \tag{6.5}$$

provided that $\sum | g(x) | p(x) < \infty$.

2. Continuous

If $X : f_x(x)$ then

$$E(Y) = \int_{-\infty}^{\infty} g(x) f_x(x) dx \tag{6.5}$$

provided that $\int | g(x) | f_x(x) dx < \infty$.

Proof (Discrete Case)

$$E(Y) = \sum_i y_i p_y(y_i). \tag{6.6}$$

Let A_i be the x's mapped into by y by g. We have $x \in A_i$ if $g(x) = y_i$. Hence,

$$p_y(y_i) = \sum_{x \in A_i} p_x(x)$$

$$E(Y) = \sum_i y_i \sum_{x \in A_i} p_x(x)$$

$$= \sum_i \sum_{x \in A_i} y_i p_x(x) \tag{6.7}$$

$$= \sum_i \sum_{x \in A_i} g(x) p_x(x)$$

$$= \sum_x g(x) p_x(x).$$

Remark 6.2. A very important function $g(x)$ whose expected value we considered is $g(x) = (x - \mu_x)^2$ which is the variance of X. We know that $\sigma_x^2 = E(g(X)) = E(X - \mu_x)^2 = E[X^2 - 2X\mu_x + \mu_x{}^2] = EX^2 - \mu_x^2$.

The **standard deviation** of X is $\sigma_x = [EX^2 - \mu_x^2]^{\frac{1}{2}}$.

In general, if $X_1,..., X_n$ are jointly distributed random variables and $Y = g(X_1,..., X_n)$ then

    1. Discrete

If $X_1, X_2,..., X_n : p_x(X_1, X_2,..., X_n)$ then

$$E(Y) = \sum g(x_1,..., x_n) p_x(x_1,..., x_n) \tag{6.8}$$

provided that $\sum | g(x_1,..., x_n) | p_x(x_1,..., x_n) < \infty$.

    2. Continuous

If $X_1, X_2,..., X_n : f_x(x_1,..., x_n)$ then

$$E(Y) = \int g(x_1,..., x_n) f(x_1,..., x_n) dx_1,..., dx_n \tag{6.9}$$

provided that $E(| Y |) < \infty$.

Remark 6.3. If X and Y are independent and g and h are fixed functions, then $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$ provided that each of the expectations on the right is finite.

B. Expectations of Linear Combinations of Random Variables

In **Lectures 2 and 3** we stated that if $Y = aX + b,$ then $E(Y) = aE(X) + b$.

That is, expectation is a linear operation. We now state a very important generalization of this result.

Proposition 6.1. If $X_1,..., X_n$ are jointly distributed random variables with expectations $E(X_i)$ and Y is a linear function of the X_i's,

$$Y = a + \sum_{i=1}^{n} b_i E(X_i). \tag{6.10}$$

Simply stated, the expectation of the linear combination of a set of random variables is the linear combination of the expectations provided that each expectation is finite.

Proof: (see Rice, p. 125).

Example 2.1 (continued). The rat executed 40 trials of a learning experiment with a probability p of a correct response on each trial. What is the expected number of correct responses? Each trial is a Bernoulli experiment with probability p of the animal executing it correctly. Let X_i be the outcome on trial i, then $Y = \sum_{i=1}^{40} X_i$ is the total number of correct responses on each trial. Hence,

$$E(Y) = E(\sum_{i=1}^{40} X_i)$$

$$= \sum_{i=1}^{40} E(X_i) \tag{6.11}$$

$$= \sum_{i=1}^{40} p = 40p.$$

In general, if there were n trials $E(Y) = np$. We showed this result before in **Lecture 2** using the definition of expectation directly and the binomial theorem.

Example 5.4 (continued). Suppose that the number of channel openings in a 500 msec time interval at each of n sites at the frog neuromuscular junction is a Poisson random variable. Each site has an expected number of openings λ_i per millisecond for $i = 1, ..., n$. What is the expected number of openings in 500 msec across all n sites? We have

$$X_i : P(\lambda_i)$$

$$\tag{6.12}$$

$$E(X_i) = \lambda_i$$

and in 500 msec we have

$$Z_i = 500 X_i$$

$$\tag{6.13}$$

$$E(Z_i) = 500 \lambda_i$$

and

$$Y = \sum_{i=1}^{n} Z_i$$

$$\tag{6.14}$$

$$E(Y) = E(\sum_{i=1}^{N} Z_i)$$

$$= \sum_{i=1}^{n} E(Z_i) = 500 \sum_{i=1}^{n} \lambda_i.$$

Example 5.4 (continued). We assumed that the length of time that each of two channels is open is an exponential random variable with parameters λ_1 and λ_2 respectively. What is the expected sum of the opening times at the two sites? We have $X_i : E(\lambda_i)$ $i = 1,2$ and $E(X_i) = \lambda_i^{-1}$. Hence, $Z = X_1 + X_2$ and $E(Z) = E(X_1 + X_2) = \lambda_1^{-1} + \lambda_2^{-1}$. Notice that when $\lambda_1 = \lambda_2 = \lambda$ we have $E(Z) = 2\lambda^{-1}$. This is completely consistent with our observation that when $\lambda_1 = \lambda_2 = \lambda$, $Z : \Gamma(2, \lambda)$, which we showed in **Lecture 5**.

III. Covariance and Correlation

A. Covariance
Example 4.6 (continued). In this example, we saw that positive values on one sensor were associated with positive values on the other sensor and negative values on one sensor were associated with negative values on the other sensor. The values seemed to be related. We formalize the characterization of this relationship in terms of the concept of covariance and correlation. The latter we have already discussed in terms of the correlation coefficient.

Definition 6.1. If X and Y are jointly distributed random variables with $E(X) = \mu_x$ and $E(Y) = \mu_y$, then the covariance of X and Y is

$$\text{Cov}(X,Y) = E(X - \mu_x)(Y - \mu_y)$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (X - \mu_x)(Y - \mu_y) f_{xy}(x,y) dx dy. \tag{6.15}$$

Remark 6.4. The covariance of X and Y measures the strength of the association between X and Y in the sense mentioned above in **Example 4.6**. It is the first cross moment of X and Y. It has units of the random variables squared, assuming X and Y measure the same thing.

Remark 6.5. If $X = Y$, we have

$$\text{Cov}(X,Y) = E(X - \mu_x)(Y - \mu_y)$$
$$= E(X - \mu_x)^2 \tag{6.16}$$
$$= \sigma_x^2.$$

Therefore, the covariance of X with itself is simply the variance of X.

Remark 6.6. Variance is the second central moment, and $E(X^2)$ is the second moment. The variance tells us about how mass is distributed away from the mean. We report the standard deviation because σ_x is on the same scale (units) as X.

Remark 6.7. We have on expanding the terms in **Definition 6.1**

$$\text{Cov}(X,Y) = E(X - \mu_x)(Y - \mu_y)$$
$$= E(XY - \mu_x Y - \mu_y X + \mu_x \mu_y)$$
$$= E(XY) - \mu_x E(Y) - \mu_y E(X) + \mu_x \mu_y \tag{6.17}$$
$$= E(XY) - \mu_x \mu_y - \mu_y \mu_x + \mu_x \mu_y$$
$$= E(XY) - E(X)E(Y).$$

Example 6.1 (Rice p. 138). Suppose X and Y have the joint probability density

$$f(x,y) = (2x + 2y - 4xy) \tag{6.18}$$

where $x \in (0,1)$ and $y \in (0,1)$.

Find $\mathrm{Cov}(X,Y)$. First it is easy to show that $E(X) = E(Y) = \dfrac{1}{2}$

$$E(X) = \int_0^1 \int_0^1 x(2x + 2y - 4xy)\,dydx = \frac{1}{2}$$

$$E(Y) = \int_0^1 \int_0^1 y(2x + 2y - 4xy)\,dxdy = \frac{1}{2}$$

(6.19)

We need to compute

$$E(XY) = \int_0^1 \int_0^1 xy(2x + 2y - 4xy)\,dxdy = \frac{2}{9}.$$

(6.20)

Hence,

$$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) = \frac{2}{9} - \left(\frac{1}{2}\right)^2 = -\frac{1}{36}.$$

(6.21)

Here are 4 useful results.

Result 6.1 $\mathrm{Cov}(a + X, Y)$

$$\begin{aligned}
\mathrm{Cov}(a + X, Y) &= E((a + X)Y) - E(a + X)E(Y) \\
&= E(aY) + E(XY) - [E(a) + E(X)]E(Y) \\
&= aE(Y) + E(XY) - E(a)E(Y) - E(X)E(Y) \\
&= E(XY) - E(X)E(Y) + aE(Y) - aE(Y) \\
&= \mathrm{cov}(X,Y)
\end{aligned}$$

(6.22)

Result 6.2 $\mathrm{Cov}(aX, bY)$

$$\begin{aligned}
\mathrm{Cov}(aX, bY) &= E(aXbY) - E(aX)E(bY) \\
&= abE(XY) - abE(X)E(Y) \\
&= ab\,\mathrm{cov}(X,Y)
\end{aligned}$$

(6.23)

Result 6.3 $\mathrm{Cov}(X, Y + Z)$

$$\begin{aligned}
\mathrm{Cov}(X, Y + Z) &= E(XY + Z) - E(X)E(Y + Z) \\
&= E(XY + XZ) - E(X)[E(Y) + E(Z)] \\
&= E(XY) + E(XZ) - E(X)E(Y) - E(X)E(Z) \\
&= E(XY) - E(X)E(Y) + E(XZ) - E(X)E(Z) \\
&= \mathrm{cov}(X,Y) + \mathrm{cov}(X,Z)
\end{aligned}$$

(6.24)

Result 6.4 $\text{Cov}(aW + bX + cY + dZ)$

$$\begin{aligned}
\text{Cov}(aW + bX, cY + dZ) &= \text{Cov}(aW + bX, cY) + \text{Cov}(aW + bX, dZ) \\
&= \text{Cov}(aW, cY) + \text{cov}(bX, cY) + \text{Cov}(aW, dZ) + \text{Cov}(bX, dZ) \\
&= ac\text{Cov}(W, Y) + bc\text{Cov}(X, Y) + ad\text{Cov}(WZ) + bd\text{Cov}(X, Z).
\end{aligned}$$ (6,25)

We can state the following general result

Proposition 6.2. Let $U = a_0 + \sum\limits_{i=1}^{I} a_i X_i$ $\qquad$ $V = b_0 + \sum\limits_{j=1}^{J} b_j X_j$

Then

$$\text{Cov}(U, V) = \sum_{i=1}^{I} \sum_{j=1}^{J} a_i b_j \text{Cov}(X_i, X_j).$$ (6.26)

We will be more interested in the following special cases

i) $\begin{aligned}
\text{Cov}(X + Y, X + Y) &= \text{Cov}(X, X) + \text{Cov}(Y, Y) + 2\text{Cov}(X, Y) \\
&= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)
\end{aligned}$ (6.27)

ii) $\text{Cov}(X - Y, X - Y) = \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X, Y)$ (6.28)

iii) If $U = a + \sum\limits_{i=1}^{I} a_i X_i$ and X_i are independent then

$$\begin{aligned}
\text{Var}(U) = \text{Cov}(U, U) &= \text{Cov}(a + \sum_{i=1}^{I} a_i X_i, a + \sum_{i=1}^{I} a_j X_j) \\
&= \sum_{i=1}^{I} \sum_{j=1}^{I} a_i a_j \, \text{cov}(X_i X_j).
\end{aligned}$$ (6.29)

If $i \neq j$ $\text{Cov}(X_i, X_j) = 0$ because the X_i are independent. If $i = j$ $\text{Cov}(X_i, X_j) = \text{Var}(X_i)$.

Thus,

$$\text{Var}(U) = \sum_{i=1}^{I} a_i^2 \text{Var}(X_i) = \sum_{i=1}^{I} a_i^2 \sigma_i^2.$$ (6.30)

This result is to some extent the analog for variances of the result for expectations established in **Proposition 6.1**. We required independence in the variance calculations.

Example 2.1 (continued). Here $X : B(n, p)$ where $n = 40$. What is the variance of X? We can write

$$X = \sum_{i=1}^{40} Y_i \tag{6.31}$$

where Y_i : Bernoulli with probability of a correct response p.

If we assume (as we do) that the Y_i's are independent, then by **Proposition 6.2 iii**

$$\text{Var}(X) = \sum_{i=1}^{40} \text{Var}(Y_i) = \sum_{i=1}^{40} p(1-p) = 40p(1-p). \tag{6.32}$$

In general, for n we have $\text{Var}(X) = np(1-p)$. We did not show this before because it is very tedious to do directly from the definition of the variance.

Example 6.2. Integrate and Fire Model as a Random Walk + Drift
As we discussed in **Lecture 3**, an elementary stochastic model for the subthreshold membrane voltage of a neuron is the random walk with drift model proposed by Gerstein and Mandelbrot (1964). In continuous time, it is

$$V(t) = \mu t + \sigma \int_{t_0}^{t} dW(u) \tag{6.33}$$

where $W(t)$ is a Weiner process, i.e. a continuous time Gaussian random variable with mean 0 and variance 1. We assume $E(W(s)W(t)) = 0$ for $s \neq t$. To simulate this on the computer we rewrite it in discrete time as

$$V(t_i) = \mu t_i + \sigma \sum_{j=0}^{i} W_i \tag{6.34}$$

where $W_i : N(0,1)$. Find the variance of $V(t_i)$. Applying **Proposition 6.2 iii** we get

$$\text{Var}(V(t_i)) = \text{Var}(\mu t_i + \sigma \sum_{j=0}^{i} W_i)$$

$$= \sigma^2 \sum_{j=0}^{i} \text{Var}(W_i) \tag{6.35}$$

$$= \sigma^2 (i+1).$$

This is a restatement of the well-known classic result that the variance of a random walk is proportional to the length of the walk. Uncertainty increases with time. Notice however, that

$$E(V(t_i)) = E(\mu t_i + \sigma \sum_{j=0}^{i} W_i) = E(\mu t_i) + \sigma \sum_{j=0}^{i} E(W_i) = \mu t_i + 0 \tag{6.36}$$
$$= \mu t_i.$$

On average the random walk with drift will be on the course of linear increase in membrane voltage. (see Gerstein GL, Mandelbrot M. Random walk models for the spiking activity of a single neuron. *Biophysical Journal* 4:41-68, 1964).

B. Correlation

Definition 6.2. If X and Y are jointly distributed random variables and the variances and covariances of X and Y are non-zero, then the correlation of X and Y is

$$\rho = \frac{\text{Cov}(X,Y)}{[\text{Var}(X)\text{Var}(Y)]^{\frac{1}{2}}} = \frac{\text{Cov}(X,Y)}{\sigma_x \sigma_y}. \tag{6.37}$$

Example 6.1 (continued). Find ρ. We have $E(X) = E(Y) = \frac{1}{2}$ and we computed $\text{Cov}(X,Y) = -\frac{1}{36}$.

Because X and Y are both uniform on $(0,1)$ it is easy to show that $\text{Var}(X) = \text{Var}(Y) = \frac{1}{12}$. Now, we have

$$\rho = \frac{-\frac{1}{36}}{(\frac{1}{12} \times \frac{1}{12})^{\frac{1}{2}}} = \frac{-\frac{1}{36}}{\frac{1}{12}} = -\frac{1}{3}. \tag{6.38}$$

We now establish an important property of the correlation coefficient.

Proposition 6.3. $-1 \le \rho \le 1$. $\rho = \pm 1$ if and only if $\Pr(Y = a + bX) = 1$ for some constants a and b.

Proof:

$$0 \le \text{Var}(\frac{X}{\sigma_x} + \frac{Y}{\sigma_y})$$
$$= \frac{\text{Var}(X)}{\sigma_x} + \frac{\text{Var}(Y)}{\sigma_y} + 2\text{Cov}(\frac{X}{\sigma_x}, \frac{Y}{\sigma_x})$$
$$= \frac{\text{Var}(X)}{\sigma_x^2} + \frac{\text{Var}(Y)}{\sigma_y^2} + 2\frac{\text{Cov}(X,Y)}{\sigma_x \sigma_y} \tag{6.39}$$
$$\quad 1 \quad + \quad 1 \quad + \quad 2\rho$$
$$= 2(1 + \rho)$$

Or $2(1 + \rho) \ge 0 \Rightarrow 1 + \rho \ge 0$ or $\rho \ge -1$. Also

$$0 \le \text{Var}(\frac{X}{\sigma_x} - \frac{X}{\sigma_y}) = 2(1 - \rho) \tag{6.40}$$

Or $2(1-\rho) \geq 0 \Rightarrow (1-\rho) \geq 0$ $1 \geq \rho$. To prove the last statement of the proposition requires Chebyshev's inequality which we will prove in **Lecture 7**.

Remark 6.8. Correlation provides a normalized (scale-free) measure of covariance.

Example 4.6 (continued). We previously stated in **Lecture 4** that ρ was the correlation coefficient for a bivariate Gaussian probability density. Now we demonstrate this by showing that $\text{cov}(X,Y) = \rho \sigma_x \sigma_y$.

$$\text{Cov}(X,Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)(y - \mu_y) f(x,y) \, dxdy. \tag{6.41}$$

Make the change of variables $u = (x - \mu_x)/\sigma_x$, $v = (y - \mu_y)/\sigma_y$ and we have

$$= \frac{\sigma_x \sigma_y}{2\pi(1-\rho^2)^{\frac{1}{2}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} uv \exp\left\{ -\frac{1}{2} \frac{1}{(1-\rho^2)}(u^2 + v^2 - 2\rho uv) \right\} dudv \tag{6.42}$$

completing the square

$$= \frac{\sigma_x \sigma_y}{2\pi(1-\rho^2)^{\frac{1}{2}}} \int_{-\infty}^{\infty} v \exp\left\{ \frac{v^2}{2} \right\} \left(\int_{-\infty}^{\infty} u \exp\left\{ -\frac{1}{2(1-\rho^2)}(u - \rho v)^2 \right\} du \right) du, \tag{6.43}$$

The inner integral is $E(u) \times [2\pi(1-\rho^2)]^{\frac{1}{2}} = \rho v [2\pi(1-\rho^2)]^{\frac{1}{2}}$

$$= \frac{\rho \sigma_x \sigma_y [2\pi(1-\rho^2)]^{\frac{1}{2}}}{2\pi(1-\rho^2)^{\frac{1}{2}}} \times \int_{-\infty}^{\infty} v^2 \exp\left\{ \frac{v^2}{2} \right\} dv$$

$$= \frac{\rho \sigma_x \sigma_y (2\pi)^{\frac{1}{2}}}{2\pi} (2\pi)^{\frac{1}{2}} E(v) \tag{6.44}$$

$$= \rho \sigma_x \sigma_y.$$

IV. Moment Generating Functions

"*Learning to use moment generating functions in probability calculations is like going from using a PC to using a supercomputer.*" -- E.N. Brown

As we will see, the moment generating function simplifies many of the calculations we have already done and will enable many more, the most important of which is the proof of the **Central Limit Theorem**.

Definition 6.3. The rth moment of a random variable X is

$$E(X^r) = \int_{-\infty}^{\infty} x^r f_x(x)dx \tag{6.45}$$

provided the expectation exists. The rth central moment of a random variable X is

$$E\{[X - E(X)]^r\} = \int_{-\infty}^{\infty} (x - E(x))^r f_x(x)dx \tag{6.46}$$

provided that the expectation exists.

Remark 6.9. Some key factoids about moments

$$E(X) = \mu \qquad 1^{\text{st}} \text{ moment}$$

$$E(X^2) \qquad 2^{\text{nd}} \text{ moment}$$

$$E(X - \mu)^2 = \sigma^2 \quad 2^{\text{nd}} \text{ central moment (variance)}$$

$$E(X - \mu)^3 \qquad 3^{\text{rd}} \text{ central moment}$$

$$E(X - \mu)^4 \qquad 4^{\text{th}} \text{ central moment}$$
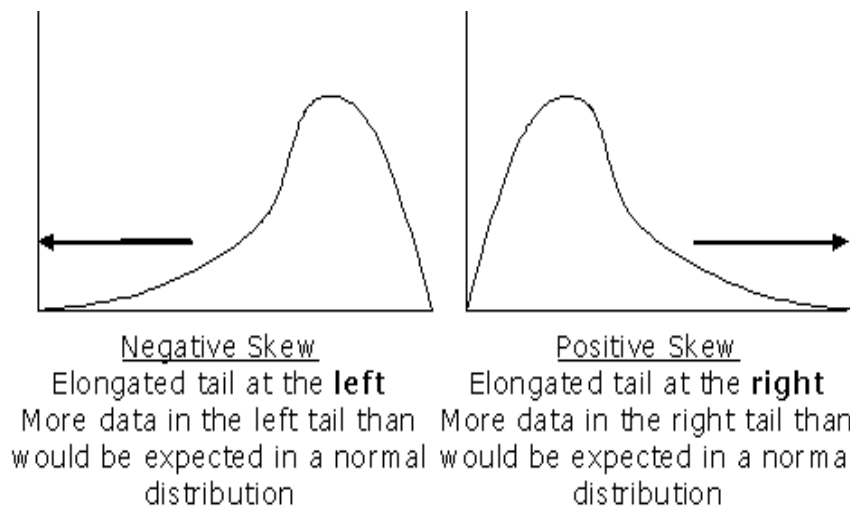
The skewness measures the asymmetry of the distribution

Negative Skew
Elongated tail at the **left**
More data in the left tail than
would be expected in a normal
distribution

Positive Skew
Elongated tail at the **right**
More data in the right tail than
would be expected in a normal
distribution

Figure 6.A. Examples of Negative (left panel) and positive (right) skew in a probability density.

It is defined is

$$\frac{E(X - \mu)^3}{\sigma^3} \tag{6.47}$$

The skewness of standard Gaussian distribution is 0.

The kurtosis measures how "fat" the tails of a distribution is often reported as

$$\frac{E(X - \mu)^4}{\sigma^4} \tag{6.48}$$

or to make a comparison relative to a standard Gaussian distribution as

$$\frac{E(X - \mu)^4}{\sigma^4} - 3 \tag{6.49}$$

because for kurtosis for the standard Gaussian distribution is 3.

Definition 6.4. The moment generating function (mgf) of a random variable X is $E(e^{tx})$ which is in the

Discrete Case

$$M_x(t) = \sum_x e^{tx} p_x(x) \tag{6.50}$$

and is in the

Continuous Case

$$M_x(t) = \int_{-\infty}^{\infty} e^{tx} f_x(x) dx \tag{6.51}$$

provided the expectations exist. This depends critically on how fast the tails in the probability density "die out".

Result 6.5. If $M_x(t)$ exists in an open interval containing zero it uniquely determines $f_x(x)$.

This is an important, deep result which allows us to perform computations on $M_x(t)$ to establish properties of X and $f_x(x)$. For example, if we differentiate $M(t)$ we obtain

$$
\begin{aligned}
M'(t) &= \frac{d}{dt} \int_{-\infty}^{\infty} e^{tx} f_x(x) dx \\
&= \int_{-\infty}^{\infty} \frac{d}{dt} e^{tx} f_x(x) dx \\
&= \int_{-\infty}^{\infty} x e^{tx} f_x(x) dx
\end{aligned}
\tag{6.52}
$$

and

$$M'(0) = \int_{-\infty}^{\infty} x f_x(x) dx = E(X) \tag{6.53}$$

Differentiating r times and evaluating the derivative at zero gives

$$M^{(r)}(0) = E(X^r). \tag{6.54}$$

Result 6.6. If $M(t)$ exists in an open interval containing 0, then $M^{(r)}(0) = E(X^r)$. In this sense $M(t)$ can be used to "generate" the moments of X.

Example 6.3. Poisson MGF

$$M(t) = \sum_{k=0}^{\infty} \frac{e^{tk} \lambda^k}{k!} e^{-\lambda}$$

$$= \sum_{k=0}^{\infty} \frac{(\lambda e^t)^t}{k!} e^{-\lambda} \tag{6.55}$$

$$= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^t)^k}{k!} = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)}.$$

Then sum averages for all t. Differentiating, we have

$$M'(t) = \lambda e^t e^{\lambda(e^t - 1)}$$

$$M''(t) = \lambda e^t e^{\lambda(e^{t-1})} + \lambda^2 e^{2t} (e^{\lambda(e^{t-1})}) \tag{6.56}$$

$$M'(0) \text{ and } \lambda \Rightarrow E(k) = \lambda$$

$$M''(0) = \lambda + \lambda^2 \text{ and } E(k^2) = \lambda^2 + \lambda \text{ or } E(k^2) - [E(k)]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Example 6.4 Gamma MGF

$$M(t) = \int_0^{\infty} \frac{e^{tx} \beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx$$

$$= \frac{\beta^{\alpha}}{\Gamma(\alpha)} \int_0^{\infty} x^{\alpha-1} e^{(t-\beta)x} dx \tag{6.57}$$

$$= \frac{\beta^{\alpha}}{\Gamma(\alpha)} \frac{\Gamma(\alpha)}{(\beta-t)^{\alpha}} = \frac{\beta^{\alpha}}{(\beta-t)^{\alpha}}$$

where we require $t < \beta$ for the integral to exist

$$M'(t) = \alpha\beta^{\alpha}(\beta - t)^{-(\alpha+1)} \Rightarrow M'(0) = E(X) = \frac{\alpha}{\beta}$$

(6.58)

$$M''(t) = \alpha(\alpha+1)\beta^{\alpha}(\beta - t)^{-(\alpha+2)} \Rightarrow M''(0) = E(X^2) = \frac{\alpha(\alpha+1)}{\beta^2}$$

and $\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = \frac{\alpha(\alpha+1)}{\beta^2} - \left(\frac{\alpha}{\beta}\right)^2 = \frac{\alpha}{\beta^2}.$

Example 6.5. Standard Gaussian MGF

$$M(t) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} e^{tx} e^{-\frac{x^2}{2}} dx$$

(6.59)

Complete the square to rewrite

$$\frac{x^2}{2} - tx = \frac{1}{2}(x^2 - 2tx + t^2) - \frac{1}{2}t^2$$

(6.60)

and hence,

$$M(t) = \frac{e^{\frac{t^2}{2}}}{(2\pi)^{\frac{1}{2}}} \int_{-\infty}^{\infty} e^{-\frac{(x-t)^2}{2}} dx.$$

(6.61)

Let $u = x - t$ then $dx = du$

$$= \frac{e^{\frac{t^2}{2}}}{(2\pi)^{\frac{1}{2}}} \int_{-\infty}^{\infty} e^{\frac{u^2}{2}} du = \frac{e^{\frac{t^2}{2}}}{(2\pi)^{\frac{1}{2}}} (2\pi)^{\frac{1}{2}} = e^{\frac{t^2}{2}}.$$

(6.62)

Result 6.7 (Linear Transformation). If X has mgf $M_x(t)$ and $Y = a + bX$, then Y has mgf $M_y(t) = e^{at} M_x(bt)$.
Proof:

$$\begin{aligned} M_Y(t) &= E(e^{tY}) \\ &= E(e^{at+btX}) \\ &= E(e^{at} e^{btX}) = e^{at} E(e^{btX}) = e^{at} M_x(bt) \end{aligned}$$

(6.63)

Example 6.5 (continued). Let $Y = \mu + \sigma X$ where $X : N(0,1)$ then $M_Y(t) = e^{\mu t} M_x(\sigma t) = e^{\mu t} e^{\sigma^2 \frac{t^2}{2}}$.

Result 6.8 (Convolution of Independent Random Variables)
If X and Y are independent with mgf's $M_x(t)$ and $M_y(t)$ respectively, and $Z = X + Y$, then $M_z(t) = M_x(t)M_y(t)$ on the common interval where both mgf's exist.

$$M_z(t) = E(e^{tZ})$$
$$= E(e^{t(X+Y)})$$
$$= E(e^{tX}e^{tY}) \qquad (6.64)$$
$$= E(e^{tX})E(e^{tY})$$
$$= M_x(t)M_y(t)$$

by independence.

Example 5.4 (continued). We had $X:$ *Exponential*(λ_1) $Y:$ *Exponential*(λ_2) and $Z = X + Y$. We showed that if $\lambda_1 = \lambda_2 = \lambda$ and X and Y are independent, then $Z: \Gamma(\alpha = 2, \beta = \lambda)$. We now establish the same result using mgf's (**Example 6.4** and **Result 6.8**).

$$M_z(t) = M_x(t)M_y(t)$$
$$= (\frac{\lambda}{\lambda - t})(\frac{\lambda}{\lambda - t}) \qquad (6.65)$$
$$= (\frac{\lambda}{\lambda - t})^2$$

which is the mgf of a gamma random variable with $\alpha = 2$ and $\beta = \lambda$.

Example 6.6. $X: N(\mu_x, \sigma_x^2)$ $Y: N(\mu_y, \sigma_y^2)$ and X and Y are independent. Find the distribution of $Z = X + Y$. By **Example 6.5**,

$$M_z(t) = M_x(t)M_y(t)$$
$$= e^{\mu_x}e^{\frac{\sigma_x^2 t^2}{2}}e^{\mu_y}e^{\frac{\sigma_y^2 t^2}{2}} \qquad (6.66)$$
$$= e^{(\mu_x + \mu_y)}e^{(\sigma_x^2 + \sigma_y^2)\frac{t^2}{2}}$$

Hence, $Z: N(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$.

Remark 6.10. There is an obvious generalization of **Result 6.8**, namely if $X_1,...,X_n$ are independent random variables with mgf's $M_{x_i}(t)$ $i = 1,...,n$, and $Z = \sum_{i=1}^{n} X_i$, then

$$M_z(t) = \prod_{i=1}^{n} M_{x_i}(t). \qquad (6.67)$$

The X_i's do not have to come from the same distribution.

Remark 6.11. In general, if two random variables follow the same distribution the sum will not necessarily follow the same distribution. See **Example 5.4**, **5.5** and consider the case of two independent gamma random variables with $\beta_1 = \beta_2$.

Remark 6.12. Result 6.8 is the special case of the well-known mathematical fact that the convolution of two functions in this case probability density functions in the probability (time)

domain corresponds to multiplying the transforms in the t (frequency) domain. This result is exploited a lot in linear systems theory and Fourier analysis.

Remark 6.13. There is a multivariate generalization of the mgf which we will not consider.

Remark 6.14. The limitation that the mgf may not exist is rectified by working instead with the characteristic function

$$\phi_x(t) = E(e^{itX})$$
$$= \int_{-\infty}^{\infty} e^{itx} f_x(x)dx \tag{6.68}$$

where $i = (-1)^{\frac{1}{2}}$. The integral converges for all t because $|e^{itx}| \leq 1$. The characteristic function has the same properties of the mgf without the existence limitation. Its use requires familiarity with complex numbers and their manipulations.

V. Summary
We can now perform basic probability calculations on linear combinations of random variables. These results have now established the essential groundwork for proving the **Central Limit Theorem** in **Lecture 7**.

Acknowledgments
I am grateful to Julie Scott for technical assistance in preparing this lecture.

Reference
Rice JA. *Mathematical Statistics and Data Analysis*, 3rd edition. Boston, MA, 2007.

9.07 Statistics for Brain and Cognitive Science
Fall 2016