Sums of Random Variables
Covariance and Correlation
Linear Functions & Bivariate Normal Distributions
A Sum of Two Independent Variables

- $W = X + Y$; $X, Y$ independent

- $f_{W|X}(w \mid x) = f_Y(w - x)$

- $f_{W,X}(w, x) = f_X(x)f_{W|X}(w \mid x) = f_X(x)f_Y(w - x)$

- $f_W(w) = \int_{-\infty}^{\infty} f_X(x)f_Y(w - x) \, dx$
Example: A Sum of Exponentials

The time your TA takes to help each student is exponentially distributed. If there are 2 students before you, how long will it take to help both of them?

\[
f(t, u) = f_T(t) f_U(u) = \lambda e^{-\lambda t} \lambda e^{-\lambda u} = \lambda^2 e^{-\lambda(t+u)} \quad (t, u \geq 0)
\]

\[
S = T + U
\]

\[
f_S(s) = \int_{-\infty}^{\infty} f_T(t) f_U(s - t) dt
\]

\[
= \int_{0}^{s} f_T(t) f_U(s - t) dt \quad \text{since } f_T(t) = 0 \text{ if } t < 0
\]

\[
\text{and } \quad f_U(s - t) = 0 \text{ if } t > s
\]

\[
= \int_{0}^{s} \lambda e^{-\lambda t} \lambda e^{-\lambda(s-t)} dt
\]

\[
= \int_{0}^{s} \lambda^2 e^{-\lambda s} dt
\]

\[
= \lambda^2 s e^{-\lambda s} \quad (s \geq 0)
\]
Example: A Sum of Exponentials

The time your TA takes to help each student is exponentially distributed. If there are 2 students before you, how long will it take to help both of them?

\[ f(t, u) = f_T(t) f_U(u) = \lambda e^{-\lambda t} \lambda e^{-\lambda u} = \lambda^2 e^{-\lambda(t+u)} \quad (t, u \geq 0) \]

\[ S = T + U \]

\[ f_S(s) = \int_{-\infty}^{\infty} f_T(t) f_U(s-t) dt \]

\[ = \int_{0}^{s} f_T(t) f_U(s-t) dt \]

\[ = \int_{0}^{s} \lambda e^{-\lambda t} \lambda e^{-\lambda(s-t)} dt \]

\[ = \int_{0}^{s} \lambda^2 e^{-\lambda s} dt \]

\[ = \lambda^2 se^{-\lambda s} \quad (s \geq 0) \]

Not an exponential distribution! This is a gamma distribution.
Sums of Multiple Exponential Variables

$f_{X_i}(x) = \lambda e^{-\lambda x}, \ x \geq 0. \quad E[X_i] = \frac{1}{\lambda} \quad \text{Var}[X_i] = \frac{1}{\lambda^2}$

$Y = \sum_{i=1}^{n} X_i \quad E[Y] = \frac{n}{\lambda} \quad \text{Var}[Y] = \frac{n}{\lambda^2}$

Gamma PDF: $f_Y(y) = \frac{\lambda^n}{(n-1)!} y^{n-1} e^{-\lambda y}$
Normal Random Variables

\[ f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \left( \frac{x-\mu}{\sigma} \right)^2} \]

\[ E[X] = \mu \]

\[ \text{Var}[X] = E[(X - \mu)^2] = \sigma^2 \]

\[ \sqrt{\text{Var}[X]} = \sigma \] is the standard deviation

**Theorem:** A sum of independent Gaussians is Gaussian!

\[ Y = \sum_{i=1}^{n} X_i \]

\[ f_Y(y) = \frac{1}{\sqrt{2\pi\bar{\sigma}^2}} e^{-\frac{1}{2} \left( \frac{y-\bar{\mu}}{\bar{\sigma}} \right)^2} \]

\[ f_{X_i}(x_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{1}{2} \left( \frac{x_i-\mu_i}{\sigma_i} \right)^2} \]

\[ \bar{\mu} = \sum_{i=1}^{n} \mu_i, \quad \bar{\sigma}^2 = \sum_{i=1}^{n} \sigma_i^2 \]
Normal Random Variables

\[ f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \left( \frac{x-\mu}{\sigma} \right)^2} \]

\[ E[X] = \mu \]

\[ \text{Var}[X] = E[(X - \mu)^2] = \sigma^2 \]

\[ \sqrt{\text{Var}[X]} = \sigma \] is the standard deviation

**Theorem:** A sum of independent Gaussians is Gaussian!

- Let \( W = X + Y \)

\[ f_W(w) = \int_{-\infty}^{\infty} f_X(x) f_Y(w - x) \, dx \]

\[ = \frac{1}{2\pi\sigma_x\sigma_y} \int_{-\infty}^{\infty} e^{-x^2/2\sigma_x^2} e^{-(w-x)^2/2\sigma_y^2} \, dx \]

(algebra) \[ = ce^{-\gamma w^2} \]

\[ E[X] = 0, E[Y] = 0 \]

\[ \text{Var}[X] = \sigma_x^2, \text{Var}[Y] = \sigma_y^2 \]
CS145: Lecture 13 Outline

➢ Sums of Random Variables
➢ Covariance and Correlation
➢ Linear Functions & Bivariate Normal Distributions
Application: Body Shape Modeling
Variance of Sums of Random Variables

- If $Z = X + Y$ for (possibly dependent) random variables $X$ and $Y$:
  $$E[Z] = E[X] + E[Y]$$

- To simplify analysis, define “centered” random variables:
  $$\tilde{X} = X - E[X], \quad \tilde{Y} = Y - E[Y]$$

- The variance of $Z$ is then equal to:
  $$\text{Var}[Z] = E[(Z - E[Z])^2] = E[(\tilde{X} + \tilde{Y})^2]$$
  $$\text{Var}[Z] = \text{Var}[X] + \text{Var}[Y] + 2\text{Cov}[X, Y]$$

- The covariance of $X$ and $Y$ is defined as:
Covariance

By definition, the covariance of random variables $X$ and $Y$ equals:


Intuition via “centered” random variables:

$$\tilde{X} = X - E[X], \quad E[\tilde{X}] = 0.$$  
$$\tilde{Y} = Y - E[Y], \quad E[\tilde{Y}] = 0.$$  

$$\text{Cov}(X, Y) = E[\tilde{X}\tilde{Y}]$$

Independent variables have zero covariance:

$$\text{Cov}(X, Y) = E[\tilde{X}\tilde{Y}] = E[\tilde{X}]E[\tilde{Y}] = 0$$

if $f_{XY}(x, y) = f_X(x)f_Y(y)$
The covariance depends on units of variables X and Y

Often convenient to use standardized variables:

\[
\tilde{X} = \frac{X - \mu_x}{\sigma_x} \quad \tilde{Y} = \frac{Y - \mu_y}{\sigma_y}
\]

\[
\mu_x = E[X], \quad \mu_y = E[Y] \\
\sigma_x^2 = \text{Var}(X), \quad \sigma_y^2 = \text{Var}(Y)
\]

For these standardized variables, we have changed the “coordinate system” or “units” so that:

\[
E[\tilde{X}] = 0, \quad \text{Var}(\tilde{X}) = 1 \\
E[\tilde{Y}] = 0, \quad \text{Var}(\tilde{Y}) = 1
\]
Correlation Coefficient

- The covariance depends on units of variables $X$ and $Y$
- Often convenient to use **standardized variables**:

\[
\tilde{X} = \frac{X - \mu_x}{\sigma_x} \quad \tilde{Y} = \frac{Y - \mu_y}{\sigma_y}
\]

\[
\mu_x = E[X], \mu_y = E[Y]
\]
\[
\sigma_x^2 = \text{Var}(X), \sigma_y^2 = \text{Var}(Y)
\]

- The **correlation coefficient** “rho” is defined to equal:

\[
\rho(X, Y) = E[\tilde{X}\tilde{Y}] = E\left[\left(\frac{X - \mu_x}{\sigma_x}\right)\cdot\left(\frac{Y - \mu_y}{\sigma_y}\right)\right] = \frac{\text{Cov}(X, Y)}{\sigma_x\sigma_y}
\]

- For any joint distribution, we have

\[
-1 \leq \rho(X, Y) \leq 1
\]

\[
0 \leq E[(\tilde{X} - \tilde{Y})^2] =
\]
\[
0 \leq E[(\tilde{X} + \tilde{Y})^2] =
\]
Empirical Correlation Coefficients

Correlation coefficient of empirical distribution of \( N \) observations:

\[
\rho = \frac{1}{N} \sum_{i=1}^{N} \frac{(x_i - \mu_x)(y_i - \mu_y)}{\sigma_x \sigma_y}
\]

\[
\mu_x = \frac{1}{N} \sum_{i=1}^{N} x_i
\]

\[
\sigma_x^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu_x)^2
\]
Correlation coefficient of empirical distribution of $N$ observations:

- Dependence grows stronger as $\rho$ approaches $-1$ or $+1$:

  If $\rho = +1$ then $(X - \mu_x) = c(Y - \mu_y)$ for some $c > 0$.
  If $\rho = -1$ then $(X - \mu_x) = c(Y - \mu_y)$ for some $c < 0$. 
Empirical Correlation Coefficients

Correlation coefficient of empirical distribution of $N$ observations:

<table>
<thead>
<tr>
<th>1.0</th>
<th>0.8</th>
<th>0.4</th>
<th>0.0</th>
<th>-0.4</th>
<th>-0.8</th>
<th>-1.0</th>
</tr>
</thead>
</table>

Example empirical statistics of real data:

<table>
<thead>
<tr>
<th>Fathers:</th>
<th>mean height: 5'9''</th>
<th>SD: 2''</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sons:</td>
<td>mean height: 5'10''</td>
<td>SD: 2''</td>
</tr>
<tr>
<td></td>
<td>correlation: 0.5</td>
<td></td>
</tr>
</tbody>
</table>

Karl Pearson’s study of 1078 father, son pairs (~1900).
Data from Pitman Sec. 6.5.
### Empirical Correlation Coefficients

**Correlation coefficient of empirical distribution** of $N$ observations:

<table>
<thead>
<tr>
<th></th>
<th>1.0</th>
<th>0.8</th>
<th>0.4</th>
<th>0.0</th>
<th>-0.4</th>
<th>-0.8</th>
<th>-1.0</th>
</tr>
</thead>
</table>

> **Example empirical statistics of real data:**

<table>
<thead>
<tr>
<th>PSAT Score</th>
<th>Correlation Coefficient</th>
<th>SAT Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Verbal</td>
</tr>
<tr>
<td>PSAT</td>
<td>Verbal</td>
<td>0.817</td>
</tr>
<tr>
<td>Score</td>
<td>Math</td>
<td></td>
</tr>
<tr>
<td>PSAT</td>
<td>Writing</td>
<td></td>
</tr>
<tr>
<td>Score</td>
<td>Verbal + Math</td>
<td></td>
</tr>
<tr>
<td>PSAT</td>
<td>Verbal + Math + Writing</td>
<td></td>
</tr>
</tbody>
</table>

Empirical Correlation Coefficients

Correlation coefficient of empirical distribution of $N$ observations:

1.0  0.8  0.4  0.0  −0.4  −0.8  −1.0

1.0  1.0  1.0  0.0  −1.0  −1.0  −1.0

0.0  0.0  0.0  0.0  0.0  0.0  0.0

WARNING: Zero correlation does not imply independence!!!
Sums of Random Variables
Covariance and Correlation
Linear Functions & Bivariate Normal Distributions
Normal Random Variables

\[ f_X(x) = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{1}{2} \left( \frac{x-\mu}{\sigma} \right)^2} \]

\[ E[X] = \mu \]

\[ \text{Var}[X] = E[(X - \mu)^2] = \sigma^2 \]

\[ \sqrt{\text{Var}[X]} = \sigma \] is the standard deviation

**Theorem:** A linear function of a Gaussian variable is Gaussian!

\[ Y = aX + b \]

\[ f_Y(y) = \frac{1}{\sqrt{2\pi}\bar{\sigma}^2} e^{-\frac{1}{2} \left( \frac{y-\bar{\mu}}{\bar{\sigma}} \right)^2} \]

\[ f_Y(y) = \frac{1}{|a|} f_X \left( \frac{y-b}{a} \right) \]

\[ \bar{\mu} = a\mu + b, \quad \bar{\sigma} = |a|\sigma \]

How to find them

- **The continuous case**
  - **Discrete case**
  - Two-step procedure:
    - Obtain probability mass for each
    - Get CDF of \( Y \) for each possible value of \( Y \)
    - Diﬀerentiate to get \( f_Y(y) \)

**Example.**

- \( X \) is uniform on \([a, b]\).
  - Find PDF of \( V = aX + b \).
  - Solution:
    - \( f_Y(y) = \frac{1}{|a|} f_X \left( \frac{y-b}{a} \right) \)
    - \( \bar{\mu} = a\mu + b, \quad \bar{\sigma} = |a|\sigma \)
Two Independent Normal Variables

\[ f_{XY}(x, y) = f_X(x) f_Y(y) \]
\[ = \frac{1}{2\pi\sigma^2} \exp \left\{ - \frac{(x - \mu_x)^2 + (y - \mu_y)^2}{2\sigma^2} \right\} \]

The set of points where \( f_{XY}(x, y) = c \), for any constant \( c \), is a circle centered at the mean.

\[ \text{Var}[X] = \text{Var}[Y] = \sigma^2 \]
Two Independent Normal Variables

\[ f_{XY}(x, y) = f_X(x)f_Y(y) \]

\[ = \frac{1}{2\pi\sigma_x\sigma_y} \exp \left\{ -\frac{(x - \mu_x)^2}{2\sigma_x^2} - \frac{(y - \mu_y)^2}{2\sigma_y^2} \right\} \]

The set of points where \( f_{XY}(x, y) = c \), for any constant \( c \), is an ellipse centered at the mean.

\[ \text{Var}[X] = \sigma_x^2 = \sigma_y^2 = \text{Var}[Y] \]

\[ \text{Var}[X] = \sigma_x^2 < \sigma_y^2 = \text{Var}[Y] \]
Bivariate Normal Distribution

\[ f_U(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \quad f_V(v) = \frac{1}{\sqrt{2\pi}} e^{-v^2/2} \]

A **bivariate normal distribution** is any joint distribution defined as a **linear function of two independent normal distributions**

First consider the following particular linear function:

\[
X = \sqrt{\frac{1+\rho}{2}} U + \sqrt{\frac{1-\rho}{2}} V
\]

\[
Y = \sqrt{\frac{1+\rho}{2}} U - \sqrt{\frac{1-\rho}{2}} V
\]

The variables \(X\) and \(Y\) are Gaussian with statistics:

\[
E[X] = E[Y] = 0 \quad \text{Var}(X) = \text{Var}(Y) = 1
\]

\[
\rho(X, Y) = \text{Cov}(X, Y) = \rho
\]
Bivariate Normal Density Functions

\[ \rho = 1.0 \quad 0.8 \quad 0.4 \quad 0.0 \quad -0.4 \quad -0.8 \quad -1.0 \]

\[ \rho > 0 \quad \rho = 0 \quad \rho < 0 \]
A bivariate normal distribution is any joint distribution defined as a linear function of two independent normal distributions.

First consider the following particular linear function:

\[ X = \sqrt{\frac{1 + \rho}{2}} U + \sqrt{\frac{1 - \rho}{2}} V \quad Y = \sqrt{\frac{1 + \rho}{2}} U - \sqrt{\frac{1 - \rho}{2}} V \]

The joint probability density function of \( X \) and \( Y \) equals:

\[ f_{XY}(x, y) = \frac{1}{2\pi \sqrt{1 - \rho^2}} \exp \left\{ -\frac{x^2}{2(1 - \rho^2)} - \frac{y^2}{2(1 - \rho^2)} + \frac{\rho xy}{1 - \rho^2} \right\} \]

\( \rho = 0 \Rightarrow f_{XY}(x, y) = \frac{1}{2\pi} \exp \left\{ -\frac{x^2}{2} - \frac{y^2}{2} \right\} = f_X(x) f_Y(y) \Rightarrow \text{Independence!} \]
Consider two independent “standard” normal variables

\[ f_U(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \quad f_V(v) = \frac{1}{\sqrt{2\pi}} e^{-v^2/2} \quad E[U] = E[V] = 0 \quad \text{Var}(U) = \text{Var}(V) = 1 \]

We construct two new Gaussian variables via a linear function:

\[ X = aU + bV + c \quad Y = dU + eV + f \]

The joint probability density of \( X, Y \) is then bivariate normal:

\[ f_{XY}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp \left\{ -\frac{(x - \mu_x)^2}{2\sigma_x^2(1-\rho^2)} - \frac{(y - \mu_y)^2}{2\sigma_y^2(1-\rho^2)} + \frac{\rho(x - \mu_x)(y - \mu_y)}{\sigma_x\sigma_y(1-\rho^2)} \right\} \]

\[ \mu_x = E[X], \mu_y = E[Y] \quad \sigma_x^2 = \text{Var}(X), \sigma_y^2 = \text{Var}(Y) \quad \rho = \frac{\text{Cov}(X, Y)}{\sigma_x\sigma_y} \]
Interpreting Normal Parameters

\[ f_{XY}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1 - \rho^2}} \exp \left\{ -\frac{(x - \mu_x)^2}{2\sigma_x^2(1 - \rho^2)} - \frac{(y - \mu_y)^2}{2\sigma_y^2(1 - \rho^2)} + \frac{\rho(x - \mu_x)(y - \mu_y)}{\sigma_x\sigma_y(1 - \rho^2)} \right\} \]

- Coordinate system and units for random variable \( X \):
  - Mean: \( \mu_x = E[X] \)
  - Standard deviation: \( \sigma_x = \sqrt{\text{Var}(X)} \)

- Coordinate system and units for random variable \( Y \):
  - Mean: \( \mu_y = E[Y] \)
  - Standard deviation: \( \sigma_y = \sqrt{\text{Var}(Y)} \)

- Dependence between \( X, Y \) measured by correlation coefficient:
  \[
  \rho = \frac{\text{Cov}(X, Y)}{\sigma_x\sigma_y}, \quad -1 \leq \rho \leq 1
  \]

For normal variables: \( X \) and \( Y \) independent if and only if \( \rho = 0 \)
Two Correlated Normal Variables

\[ f_{XY}(x, y) = \frac{1}{2\pi \sigma_x \sigma_y \sqrt{1 - \rho^2}} \exp \left\{ -\frac{(x - \mu_x)^2}{2\sigma_x^2 (1 - \rho^2)} - \frac{(y - \mu_y)^2}{2\sigma_y^2 (1 - \rho^2)} + \frac{\rho(x - \mu_x)(y - \mu_y)}{\sigma_x \sigma_y (1 - \rho^2)} \right\} \]