CS145: Lecture 14 Outline

- Linear Functions & Bivariate Normal Distributions
- Multivariate Normal Distributions
- Markov and Chebyshev Inequalities
The covariance of random variables $X$ and $Y$ is defined as:
\[
\]

The covariance depends on units of variables $X$ and $Y$

Often convenient to use standardized variables:
\[
\tilde{X} = \frac{X - \mu_x}{\sigma_x}, \quad \tilde{Y} = \frac{Y - \mu_y}{\sigma_y}
\]
\[
\mu_x = E[X], \quad \mu_y = E[Y]
\]
\[
\sigma_x^2 = \text{Var}(X), \quad \sigma_y^2 = \text{Var}(Y)
\]

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 $-1 \leq \rho(X, Y) \leq 1$
Empirical Correlation Coefficients

Correlation coefficient of empirical distribution of $N$ observations:

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

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

$$
\sigma_x^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu_x)^2
$$
Normal Random Variables

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

\[ E[X] = \mu \]

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

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

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

\[ Y = aX + b \]

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

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

\[ \bar{\mu} = a\mu + b, \quad \bar{\sigma} = |a| \sigma \]
Standard Normal Random Variables

- If $X \sim N(\mu, \sigma^2)$ then for any constants $a$ and $b$ the random variable $aX + b$ is distributed $N(a\mu + b, a^2\sigma^2)$.
- If $X \sim N(\mu, \sigma^2)$ then $Z = \frac{X - \mu}{\sigma}$ is distribution $N(0, 1)$.
- $N(0, 1)$ is the standard Normal distribution.

$$
Pr(Z \leq z) = \Phi_Z(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2} dt
$$

$$
\phi_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}
$$
If $X \sim N(\mu, \sigma^2)$ then $Z = \frac{X - \mu}{\sigma}$

$Pr(X \leq x) = Pr(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}) = \Phi(\frac{x - \mu}{\sigma})$

The standard Normal random variable is symmetric around 0.

For $\frac{x - \mu}{\sigma} < 0$,

$\Phi(\frac{x - \mu}{\sigma}) = 1 - \Phi(-\frac{x - \mu}{\sigma})$

With a table of $\Phi(Z)$ for $Z > 0$ we can compute $F(x)$ for any Normal random variable
Modern Computation of Normal CDF

**normcdf**

Normal cumulative distribution function

**Syntax**

```r
p = normcdf(x)
p = normcdf(x, mu, sigma)
[p, plo, pup] = normcdf(x, mu, sigma, pcv, alpha)
[p, plo, pup] = normcdf(__, 'upper')
```

**Description**

$p = \text{normcdf}(x)$ returns the standard normal cdf at each value in $x$. The standard normal distribution has parameters $\mu = 0$ and $\sigma = 1$. $x$ can be a vector, matrix, or multidimensional array.

$p = \text{normcdf}(x, \mu, \sigma)$ returns the normal cdf at each value in $x$ using the specified values for the mean $\mu$ and standard deviation $\sigma$. $x$, $\mu$, and $\sigma$ can be vectors, matrices, or multidimensional arrays that all have the same size. A scalar input is expanded to a constant array with the same dimensions as the other inputs. The parameters in $\sigma$ must be positive.

**norminv**

Normal inverse cumulative distribution function

**Syntax**

```r
X = norminv(P, mu, sigma)
[X, XLO, XUP] = norminv(P, mu, sigma, pcv, alpha)
```

**Description**

$X = \text{norminv}(P, \mu, \sigma)$ computes the inverse of the normal cdf using the corresponding mean $\mu$ and standard deviation $\sigma$ at the corresponding probabilities in $P$. $P$, $\mu$, and $\sigma$ can be vectors, matrices, or multidimensional arrays that all have the same size. A scalar input is expanded to a constant array with the same dimensions as the other inputs. The parameters in $\sigma$ must be positive, and the values in $P$ must lie in the interval $[0, 1]$.
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.
Two Independent Normal Variables

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

\[
\begin{align*}
\text{Var}[X] &= \sigma_x^2 = \sigma_y^2 = \text{Var}[Y] \\
\text{Var}[X] &= \sigma_x^2 < \sigma_y^2 = \text{Var}[Y]
\end{align*}
\]
A \textbf{bivariate normal distribution} is any joint distribution defined as a \textit{linear function of two independent normal distributions}

First consider the following particular linear function:

\begin{align*}
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
\end{align*}

\(-1 \leq \rho \leq 1\)

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

\[
E[X] = E[Y] = 0 \quad \text{Var}(X) = \text{Var}(Y) = 1 \quad \rho(X, Y) = \text{Cov}(X, Y) = \rho
\]
Bivariate Normal Density Functions

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

\( \rho > 0 \) \hspace{1cm} \rho = 0 \hspace{1cm} \rho < 0
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 \textit{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

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

- 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

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

Advanced topic not covered in homeworks or exams!
Requires linear algebra for in-depth study.

Markov and Chebyshev Inequalities
Let $X^T = (X_1, \ldots, X_n)$ be a vector of $n$ independent, standard normal random variables. $E[X_i] = 0$ and $Var[X_i] = 1$.

Let $Y^T = (Y_1, \ldots, Y_m)$ be random variable vector obtained by a linear transformation on the vector $X^T$:

\begin{align*}
Y_1 &= a_{11}X_1 + a_{12}X_2 + \cdots + a_{1n}X_n + \mu_1; \\
Y_2 &= a_{21}X_1 + a_{22}X_2 + \cdots + a_{2n}X_n + \mu_2; \\
\vdots \\
Y_m &= a_{m1}X_1 + a_{m2}X_2 + \cdots + a_{mn}X_n + \mu_m.
\end{align*}

Let $A$ denote the matrix of coefficients $a_{ij}$, and $\bar{\mu}^T = (\mu_1, \ldots, \mu_m)$. Then we can write

$$Y = AX + \bar{\mu}.$$
Mean Vectors & Covariance Matrices

\[ \begin{align*}
Y_1 &= a_{11}X_1 + a_{12}X_2 + \cdots + a_{1n}X_n + \mu_1; \\
Y_2 &= a_{21}X_1 + a_{22}X_2 + \cdots + a_{2n}X_n + \mu_2; \\
\vdots \\
Y_m &= a_{m1}X_1 + a_{m2}X_2 + \cdots + a_{mn}X_n + \mu_m.
\end{align*} \]

\[ Y = AX + \bar{\mu}, \quad E[Y_i] = \mu_i, \quad \text{Var}[Y_i] = \sum_{j=1}^{n} a_{i,j}^2, \quad E[\bar{Y}] = \bar{\mu}, \]

\[ \text{Cov}(Y_i, Y_j) = \sum_{k=1}^{n} a_{i,k}a_{j,k}. \]

The covariance matrix for \( Y \) is given by

\[ \Sigma = AA^T = \begin{pmatrix}
\text{Var}[Y_1] & \text{Cov}(Y_1, Y_i) & \cdots & \text{Cov}(Y_1, Y_n) \\
\text{Cov}(Y_1, Y_2) & \text{Var}[Y_2] & \cdots & \text{Cov}(Y_2, Y_n) \\
\vdots & \vdots & \ddots & \vdots \\
(Y_m, Y_1) & \text{Cov}(Y_m, Y_2) & \cdots & \text{Var}[Y_m]
\end{pmatrix} = E[(Y - \bar{\mu})(Y - \bar{\mu})^T]. \]
If $A$ has a full rank, then $X = A^{-1}(Y - \bar{\mu})$, and we can derive a density function for the joint distribution.

$$
\Pr(Y \leq y) = \Pr(Y - \mu \leq y - \mu) \\
= \Pr(AX \leq y - \mu) \\
= \Pr(X \leq A^{-1}(y - \mu)) \\
= \frac{1}{(2\pi)^{n/2}} \int_{\bar{w} \leq A^{-1}(y - \mu)} e^{-\frac{\bar{w}^T \bar{w}}{2}} \, dw_1 \ldots \, dw_n.
$$

Changing the integration variables to $\bar{z} = A\bar{w} + \bar{\mu}$ we have

$$
\Pr(Y \leq y) = \frac{1}{\sqrt{(2\pi)^n |AA^T|}} \int_{-\infty}^{y_1} \ldots \int_{-\infty}^{y_n} e^{-\frac{1}{2}(\bar{z} - \mu)^T (A^{-1})^T A^{-1} (\bar{z} - \mu)} \, dz_1 \ldots \, dz_n.
$$

Here $|AA^T|$ denotes the determinant of $AA^T$, a term which arises under the multivariate change of variables.
Joint Multivariate Normal Distribution

Applying \((A^{-1})^T A^{-1} = (A^T)^{-1} A^{-1} = (AA^T)^{-1} = \Sigma^{-1}\), we can write the distribution function of \(Y\) as

\[
\Pr(Y \leq \bar{y}) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \int_{-\infty}^{y_1} \cdots \int_{-\infty}^{y_n} e^{-\frac{1}{2}(\bar{z}-\mu)^T \Sigma^{-1} (\bar{z}-\mu)} \, dz_1 \cdots dz_n \tag{1}
\]

where, again,

\[
\Sigma = AA^T = E[(Y - \mu)(Y - \mu)^T].
\]
A vector $Y^T = (Y_1, \ldots, Y_n)$ has a multivariate normal distribution, denoted $Y \sim N(\bar{\mu}, \Sigma)$, if and only if there is an $n \times k$ matrix $A$, a vector $X^T = (X_1, \ldots, X_k)$ of $k$ independent standard normal random variables, and a vector $\bar{\mu}^T = (\mu_1, \ldots, \mu_n)$, such that

$$Y = AX + \bar{\mu}.$$ 

If $\Sigma = AA^T = E[(Y - \bar{\mu})(Y - \bar{\mu})^T]$ has full rank, then the density of $Y$ is

$$\frac{1}{\sqrt{(2\pi)^n|\Sigma|}} e^{-\frac{1}{2} (Y-\bar{\mu})^T \Sigma^{-1} (Y-\bar{\mu})}.$$ 

If $\Sigma$ is not invertible then the joint distribution has no density function.
Multivariate Normal Probability Density

\[ \mathcal{N}(x|\mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp \left\{ -\frac{1}{2\sigma^2}(x - \mu)^2 \right\} \]

\[ \mathcal{N}(x|\mu, \Sigma) = \frac{1}{(2\pi)^{D/2} |\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu) \right\} \]

\[ \mu = E[X] \]
\[ \Sigma = E[(X - \mu)(X - \mu)^T] \]

D-dimensional ellipsoids parameterized by mean vector & covariance matrix
Linear Functions & Bivariate Normal Distributions
Multivariate Normal Distributions
Markov and Chebyshev Inequalities
Expectation is Not Everything…

Which Game Would You Prefer?

1. With probability $\frac{1}{2}$ win $1$, with probability $\frac{1}{2}$ pay $1$.
2. With probability $\frac{1}{2}$ win $100,000$, with probability $\frac{1}{2}$ pay $100,000$.
3. With probability $\frac{1}{1,000,000}$ win $1,000,000$, with probability $\frac{1}{2}$ pay $5$, else $0$.

Which Job Would You Prefer?

- A job that pays $1000$ a week.
- A job that pays $1$ a week plus a bonus of $1,000,000$ with probability $\frac{1}{1000}$. 
Suppose we care about some summary of random variable $X$:

$$Y = g(X)$$

- **Expected Value:** Value of a “typical” sample
  $$E[Y] = E[g(X)] = \int g(x) f_X(x) \, dx$$

- **Variance:** Expected squared distance from mean
  $$\text{Var}[Y] = E[(g(X) - E[g(X)])^2]$$

- **Cumulative distribution:** A complete characterization of $g(X)$
  $$F_Y(y) = P(Y \leq y) = P(g(X) \leq y)$$
  $$f_Y(y) = \frac{dF_Y(y)}{dy}$$

*If $g(X)$ is complex or high-dimensional, may be hard to compute!*
Markov’s Inequality

**Theorem**

[Markov Inequality] For any non-negative random variable, and for all $a > 0$,

$$\Pr(X \geq a) \leq \frac{E[X]}{a}.$$  

Fix some constant $a > 0$, and define

$$Y_a = \begin{cases} 
0, & \text{if } X < a, \\
 a, & \text{if } X \geq a. 
\end{cases}$$

$$a \Pr(X \geq a) = E[Y_a] \leq E[X]$$
Markov’s Inequality

Theorem

[Markov Inequality] For any non-negative random variable, and for all \( a > 0 \),

\[
Pr(X \geq a) \leq \frac{E[X]}{a}.
\]

Fix some constant \( a > 0 \), and define

\[
Y_a = \begin{cases} 
0, & \text{if } X < a, \\
 a, & \text{if } X \geq a.
\end{cases}
\]

- No such inequality would hold if \( X \) could take negative values. Why?
- If \( a < E[X] \), Markov’s inequality is vacuous, but no better bound is possible. Why?
Markov’s Inequality

**Theorem**

[Markov Inequality] For any non-negative random variable, and for all \(a > 0\),

\[
Pr(X \geq a) \leq \frac{E[X]}{a}.
\]

Another proof of Markov’s inequality:

\[
E[X] = \int_0^\infty x f_X(x) \, dx \geq \int_a^\infty x f_X(x) \, dx \geq \int_a^\infty a f_X(x) \, dx = a P(X \geq a)
\]
Chebyshev’s Inequality

**Theorem**

For any random variable $X$, and any $a > 0$,

$$Pr(|X - E[X]| \geq a) \leq \frac{Var[X]}{a^2}.$$  

**Proof.**

$$Pr(|X - E[X]| \geq a) = Pr((X - E[X])^2 \geq a^2)$$

By Markov inequality

$$Pr((X - E[X])^2 \geq a^2) \leq \frac{E[(X - E[X])^2]}{a^2}$$

$$= \frac{Var[X]}{a^2}$$