ECE 603
Probability and Random Processes

Lessons 14-16
Chapter 6
Multiple Random Variables

Objectives

- Explore joint distributions and independence.
- Analyze Moment Generating Functions (MGFs).
- Examine random vectors.

Rationale

Until now in this course, you have been working with one and two random variables and how they might be extended to more.

You will now begin considering three or more variables. As the number of random variables increases, you will notice how the functions become computationally intractable. This leads to an exploration of other techniques.

Prior Learning

- Basic Concepts
- Counting Methods
- Random Variables
- Access to the online textbook: https://www.probabilitycourse.com/

Joint Distributions and Independence

Let $X_1, X_2, X_3, \cdots, X_n$ be n discrete random variables.

Joint PMF:

$$P_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n) = P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n).$$

$$P_{XYZ}(1, 2, \sqrt{3}) = P(X = 1, Y = 2, Z = \sqrt{3})$$

Joint Distributions and Independence

Let $X_1, X_2, X_3, \cdots, X_n$ be n continuous random variables.

Joint PDF: $f_{X_1 X_2 \ldots X_n}(x_1, x_2, \ldots, x_n)$

$$P\Big((X_1, X_2, \cdots, X_n) \in A\Big)$$
$$= \int \cdots \underset{A}{\int} \cdots \int f_{X_1 X_2 \cdots X_n}(x_1, x_2, \cdots, x_n) dx_1 dx_2 \cdots dx_n.$$

Joint Distributions and Independence

The joint CDF of n random variables $X_1, X_2, X_3, \cdots, X_n$ (both discrete and continuous) is defined as

$$F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n).$$

Joint Distributions and Independence

Random variables $X_1, X_2, X_3, \cdots, X_n$ **are independent, if**

$$F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = F_{X_1}(x_1)F_{X_2}(x_2)\cdots F_{X_n}(x_n).$$

Discrete:

$$P_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = P_{X_1}(x_1)P_{X_2}(x_2)\cdots P_{X_n}(x_n).$$

Continuous:

$$f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = f_{X_1}(x_1)f_{X_2}(x_2)\cdots f_{X_n}(x_n).$$

Joint Distributions and Independence

If random variables $X_1, X_2, X_3, \cdots, X_n$ are independent, then

$$E[X_1 X_2 \cdots X_n] = E[X_1]E[X_2] \cdots E[X_n].$$

Joint Distributions and Independence

Definition. Random variables $X_1, X_2, X_3, \cdots, X_n$ are said to be **independent and identically distributed (i.i.d.)** if they are *independent*, and they have the *same marginal distributions* :

$$F_{X_1}(x) = F_{X_2}(x) = \ldots = F_{X_n}(x), \text{ for all } x \in \mathbb{R}.$$

Joint Distributions and Independence

Example. if random variables $X_1, X_2, X_3, \cdots, X_n$ are i.i.d., they will have the same means and variances, so we can write

$$
\begin{aligned}
E[X_1 X_2 \cdots X_n] &= E[X_1]E[X_2]\cdots E[X_n] && (\ X_i\text{'s are indepenednt}) \\
&= E[X_1]E[X_1]\cdots E[X_1] && (X_i\text{'s are identically distributed}) \\
&= E[X_1]^n.
\end{aligned}
$$

Joint Distributions and Independence

Example. let $X_1, X_2, X_3, \cdots, X_n$ be i.i.d. with

$$F_{X_1}(x) = F_{X_2}(x) = \cdots = F_{X_n}(x) = F_X(x),$$

$$f_{X_1}(x) = f_{X_2}(x) = \cdots = f_{X_n}(x) = f_X(x).$$

We define $Y = \max\{X_1, X_2, \cdots, X_n\}, \; Z = \min\{X_1, X_2, \cdots, X_n\}.$

Find CDF and PDF of $Y \;\&\; Z.$

Sums of Random Variables

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

With regards to the linearity of expectation:

$$EY = EX_1 + EX_2 + \cdots + EX_n.$$

Sums of Random Variables

Variance of a sum of two and three random variables is

$$\begin{aligned}\mathrm{Var}(X_1 + X_2) &= \mathrm{Cov}(X_1 + X_2, X_1 + X_2)\\ &= \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + 2\mathrm{Cov}(X_1, X_2).\end{aligned}$$

$$\begin{aligned}\mathrm{Var}(X_1 + X_2 + X_3) &= \mathrm{Cov}(X_1 + X_2 + X_3, X_1 + X_2 + X_3)\\ &= \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \mathrm{Var}(X_3) + 2\mathrm{Cov}(X_1, X_2)\\ &+ 2\mathrm{Cov}(X_1, X_3) + 2\mathrm{Cov}(X_2, X_3).\end{aligned}$$

Sums of Random Variables

Generally,

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\sum_{i<j} \mathrm{Cov}(X_i, X_j).$$

Sums of Random Variables

If $X_1, X_2, X_3, \cdots, X_n$ are uncorrelated (i.e. $\text{Cov}(X_i, X_j) = 0, \;\; \text{for } i \neq j$), then

$$\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{Var}(X_i).$$

If $X_1, X_2, X_3, \cdots, X_n$ are independent then they are uncorrelated, thus

$$\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{Var}(X_i).$$

Moment Generating Functions

Definition. The nth moment of a random variable X is defined to be $E[X^n]$.

The nth central moment of X is defined to be $E[(X - EX)^n]$.

Moment Generating Functions

The moment generating function (MGF) of a random variable X is a function $M_X(s)$ defined as

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

We say that MGF of X exists, if there exists a positive constant a such that $M_X(s)$ is finite for all $s \in [-a, a]$.

Moment Generating Functions

Example. **For each of the following random variables, find the MGF.**

a) X **is a discrete random variable, with PMF**

$$P_X(k) = \begin{cases} \frac{1}{3} & k = 1 \\ \\ \frac{2}{3} & k = 2 \end{cases}$$

b) Y **is a** $Uniform(0, 1)$ **random variable.**

Moment Generating Functions

Finding Moments from MGF:

Remember

$$e^{x} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots = \sum_{k=0}^{\infty} \frac{x^k}{k!}.$$

$$e^{sX} = \sum_{k=0}^{\infty} \frac{(sX)^k}{k!} = \sum_{k=0}^{\infty} \frac{X^k s^k}{k!}.$$

Moment Generating Functions

Thus, we have

$$M_X(s) = E[e^{sX}] = \sum_{k=0}^{\infty} E[X^k] \frac{s^k}{k!}.$$

Moment Generating Functions

We can obtain all moments of X^k from its MGF:

$$M_X(s) = \sum_{k=0}^{\infty} E[X^k]\frac{s^k}{k!},$$

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

Moment Generating Functions

Example. Let $X \sim Exponential(\lambda)$. Find the MGF of X, $M_X(s)$, and all of its moments, $E[X^k]$.

Moment Generating Functions

Theorem. Consider two random variables X and Y . Suppose that there exists a positive constant c such that MGFs of X and Y are finite and identical for all values of s in $[-c, c]$. Then,

$$F_{\mathbf{X}}(t) = F_{\mathbf{Y}}(t), \text{ for all } t \in \mathbb{R}.$$

MGF determines the distribution.

Moment Generating Functions

Sum of Independent Random Variables:

If X and Y are independent RVs and $Z = X + Y$ then,

$$
\begin{aligned}
M_Z(s) &= E[e^{sZ}] = E[e^{s(X+Y)}] \\
&= E[e^{sX}e^{sY}] = E[e^{sX}]E[e^{sY}] \qquad (\text{Since } X\&Y \text{ independent}) \\
&= M_X(s)M_Y(s).
\end{aligned}
$$

Moment Generating Functions

If $X_1, X_2, \cdots, X_n$ are n independent random variables, then

$$M_{X_1+X_2+\cdots+X_n}(s) = M_{X_1}(s)M_{X_2}(s)\cdots M_{X_n}(s).$$

Moment Generating Functions

Example. If $X \sim Binomial(n, p)$ find the MGF of X .

Moment Generating Functions

Example. **Using MGFs prove that if $X \sim Binomial(m, p)$ and $Y \sim Binomial(n, p)$ are independent, then**

$$X + Y \sim Binomial(m + n, p).$$

Characteristic Functions

If a random variable does not have a well-defined MGF, we can use the characteristic function defined as

$$\phi_X(\omega) = E[e^{j\omega X}],$$

where $j = \sqrt{-1}$ **and** ω **is a real number.**

Characteristic Functions

$$|\phi_X(\omega)| = |E[e^{j\omega X}]| \leq E[|e^{j\omega X}|] \leq 1.$$

If X and Y are independent, and $Z = X + Y$, then

$$\phi_Z(\omega) = \phi_X(\omega)\phi_Y(\omega).$$

UMASS AMHERST

Random Vectors

Vectors

Column vector:

$$a = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \Rightarrow a = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$$

$$a + b = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} + \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \\ 3 \end{bmatrix}_{1\times 3}$$

Random Vectors

Matrix multiplication

$$A = \begin{bmatrix} 1 & 3 & -1 \\ 2 & 4 & 2 \end{bmatrix}_{2\times 3}$$

$$\longrightarrow \qquad \begin{bmatrix} 1 & 3 & -1 \\ 2 & 4 & 2 \end{bmatrix}_{2\times 3} \times \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}_{3\times 1} = \begin{bmatrix} 0+3+1 \\ 0+4-2 \end{bmatrix}_{2\times 1} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}_{2\times 1}$$

Random Vectors

Matrix multiplication

$$A_{m\times n} \cdot B_{n\times l} = C_{m\times l}$$

$$A = [a_{ij}] \;\Rightarrow\; A^T = [a_{ji}],$$

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}^T = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$$

Random Vectors

When we have n random variables $X_1, X_2, X_3, \cdots, X_n$ we can put them in a (column) vector $\mathbf{X}$.

$$\mathrm{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}_{n \times 1}.$$

Random Vectors

CDF of the random vector $\mathbf{X}$

$$\begin{aligned} F_{\mathbf{X}}(\mathbf{x}) &= F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) \\ &= P(X_1 \le x_1, X_2 \le x_2, \dots, X_n \le x_n). \end{aligned}$$

PDF of the random vector $\mathbf{X}$

$$f_{\mathbf{X}}(\mathbf{x}) = f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n).$$

Random Vectors

Expectation:

The **expected value vector** or the **mean vector** of the random vector X is defined as

$$\boldsymbol{E}\mathbf{X} = \begin{bmatrix} EX_1 \\ EX_2 \\ \vdots \\ EX_n \end{bmatrix}.$$

Random Vectors

random matrix

$$\mathrm{M} = \begin{bmatrix} \boldsymbol{X}_{11} & \boldsymbol{X}_{12} & \cdots & \boldsymbol{X}_{1n} \\ \boldsymbol{X}_{21} & \boldsymbol{X}_{22} & \cdots & \boldsymbol{X}_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{X}_{m1} & \boldsymbol{X}_{m2} & \cdots & \boldsymbol{X}_{mn} \end{bmatrix}.$$

Random Vectors

Mean matrix of $\mathbf{M}$

$$E\mathbf{M} = \begin{bmatrix} EX_{11} & EX_{12} & \dots & EX_{1n} \\ EX_{21} & EX_{22} & \dots & EX_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ EX_{m1} & EX_{m2} & \dots & EX_{mn} \end{bmatrix}.$$

Random Vectors

Linearity of expectation

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b},$$

$$\mathbf{Y}\ \&\ \mathbf{X} : \textbf{Random Vector}$$

$$\mathbf{A}\ \&\ \mathbf{b} : \textbf{fixed (non-random) matrices}$$

$$\mathbf{Y}_{m\times 1} = \underbrace{\mathbf{A}_{m\times n}\ \mathbf{X}_{n\times 1}}_{m\times 1} + \mathbf{b}_{m\times 1}$$

Random Vectors

Linearity of expectation

$$\underbrace{E\mathbf{Y}}_{m \times 1} = \underset{m \times n}{\mathbf{A}}\underbrace{E\mathbf{X}}_{n \times 1} + \underset{m \times 1}{\mathbf{b}}.$$

Also, if $\mathbf{X_1}, \mathbf{X_2}, \cdots, \mathbf{X_k}$ are n-dimensional random vectors, then we have

$$\mathbf{E}[\mathbf{X_1} + \mathbf{X_2} + \cdots + \mathbf{X_k}] = \mathbf{E}\mathbf{X_1} + \mathbf{E}\mathbf{X_2} + \cdots + \mathbf{E}\mathbf{X_k}.$$

Random Vectors

Correlation and Covariance Matrix

$$\mathrm{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \qquad \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \times \begin{bmatrix} X_1 & X_2 & \cdots & X_n \end{bmatrix}$$

Random Vectors

Correlation and Covariance Matrix

For a random vector X, we define the correlation matrix, $\mathrm{R_X}$, as

$$\mathrm{R_X} = \mathrm{E}[\mathrm{X}\mathrm{X}^{\mathrm{T}}] = E\begin{bmatrix} X_1^2 & X_1X_2 & \dots & X_1X_n \\ X_2X_1 & X_2^2 & \dots & X_2X_n \\ \vdots & \vdots & \vdots & \vdots \\ X_nX_1 & X_nX_2 & \dots & X_n^2 \end{bmatrix}$$

$$= \begin{bmatrix} EX_1^2 & E[X_1X_2] & \dots & E[X_1X_n] \\ EX_2X_1 & E[X_2^2] & \dots & E[X_2X_n] \\ \vdots & \vdots & \vdots & \vdots \\ E[X_nX_1] & E[X_nX_2] & \dots & E[X_n^2] \end{bmatrix}_{n\times n}$$

$$n = 1 \Rightarrow \mathrm{R_X} = EX^2$$

Random Vectors

Covariance

$$\mathrm{Cov}(X, Y) = E[(X - EX).(Y - EY)],$$

Random Vectors

The **covariance matrix**, $\mathbf{C_X}$, is defined as

$$\mathbf{C_X} = \mathbf{E}[(\mathbf{X} - \mathbf{EX})(\mathbf{X} - \mathbf{EX})^{\mathbf{T}}]$$

$$= E\begin{bmatrix} (X_1 - EX_1)^2 & (X_1 - EX_1)(X_2 - EX_2) & \dots & (X_1 - EX_1)(X_n - EX_n) \\ (X_2 - EX_2)(X_1 - EX_1) & (X_2 - EX_2)^2 & \dots & (X_2 - EX_2)(X_n - EX_n) \\ \vdots & \vdots & \vdots & \vdots \\ (X_n - EX_n)(X_1 - EX_1) & (X_n - EX_n)(X_2 - EX_2) & \dots & (X_n - EX_n)^2 \end{bmatrix}$$

$$= \begin{bmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) & \dots & \mathrm{Cov}(X_1, X_n) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Var}(X_2) & \dots & \mathrm{Cov}(X_2, X_n) \\ \vdots & \vdots & \vdots & \vdots \\ \mathrm{Cov}(X_n, X_1) & \mathrm{Cov}(X_n X_2) & \dots & \mathrm{Var}(X_n) \end{bmatrix}.$$

Random Vectors

$$\mathrm{Var}(X) = \mathrm{Cov}(X, X) = EX^2 - (EX)^2$$

$$\Rightarrow \quad \mathbf{C_X} = \underset{\substack{\big\downarrow \\ E(XX^T)}}{\mathbf{R_X}} - \mathbf{EXEX^T}.$$

Random Vectors

Correlation matrix of $\mathbf{X}$:

$$\mathbf{R_X} = \mathbf{E}[\mathbf{X}\mathbf{X}^{\mathbf{T}}]$$

Covariance matrix of $\mathbf{X}$:

$$\mathbf{C_X} = \mathbf{E}[(\mathbf{X} - \mathbf{EX})(\mathbf{X} - \mathbf{EX})^{\mathbf{T}}] = \mathbf{R_X} - \mathbf{EXEX}^{\mathbf{T}}$$

Random Vectors

Example. Let X be an n-dimensional random vector and the random vector Y be defined as

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b},$$

where A is a fixed m by n matrix and b is a fixed n-dimensional vector. Show that

$$\mathbf{C}_{\mathbf{Y}} = \mathbf{A}\mathbf{C}_{\mathbf{X}}\mathbf{A}^{\mathrm{T}}.$$

Random Vectors

Note:

$$\mathrm{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \qquad m, n = 1 \quad \mathrm{Y} = \mathrm{AX} + \mathrm{b}$$

$$\Rightarrow \ \mathrm{Var}(Y) = a^2 \mathrm{Var}(X) = a \mathrm{Var}(X) a^T$$

$$(AB)^T = B^T A^T$$

Random Vectors

Proof: Note that by linearity of expectation, we have

$$\mathbf{EY} = \mathbf{AEX} + \mathbf{b}.$$

By definition, we have

$$\begin{aligned}
\mathbf{C_Y} &= \mathbf{E}[(\mathbf{Y} - \mathbf{EY})(\mathbf{Y} - \mathbf{EY})^{\mathbf{T}}] \\
&= \mathbf{E}[(\mathbf{AX} + \mathbf{b} - \mathbf{AEX} - \mathbf{b})(\mathbf{AX} + \mathbf{b} - \mathbf{AEX} - \mathbf{b})^{\mathbf{T}}] \\
&= E[\mathbf{A}(\mathbf{X} - \mathbf{EX})(\mathbf{X} - \mathbf{EX})^{\mathbf{T}}\mathbf{A}^{\mathbf{T}}] \\
&= \mathbf{AE}[(\mathbf{X} - \mathbf{EX})(\mathbf{X} - \mathbf{EX})^{\mathbf{T}}]\mathbf{A}^{\mathbf{T}} \qquad \text{(by linearity of expectation)} \\
&= \mathbf{AC_X A^T}.
\end{aligned}$$

Random Vectors

Normal (Gaussian) Random Vectors:

Random variables $X_1, X_2, \cdots, X_n$ are said to be jointly normal if, for all $a_1, a_2, \cdots, a_n \in \mathbb{R}$, the random variable

$$a_1X_1 + a_2X_2 + \ldots + a_nX_n$$

is a normal random variable.

Random Vectors

A random vector

$$\mathrm{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \qquad (X_1, X_2, \cdots, X_n)$$

is said to be normal or Gaussian if the random variables $X_1, X_2, \cdots, X_n$ are jointly normal.

Random Vectors

Standard Normal Random Variable:

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z^2}{2}\right\}, \qquad \text{for all } z \in \mathbb{R}.$$

$X \sim N(\mu, \sigma^2)$,

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}.$$

Random Vectors

Standard normal random vector:

$$\mathbf{Z} = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{bmatrix}, \qquad Z_1, Z_2, \cdots, Z_n \longrightarrow \text{i.i.d.}$$

$$Z_i \sim N(0,1)$$

Random Vectors

Then,

$$\begin{aligned}
f_{\mathbf{Z}}(\mathbf{z}) &= f_{Z_1,Z_2,\ldots,Z_n}(z_1, z_2, \ldots, z_n) \\
&= \prod_{i=1}^{n} f_{Z_i}(z_i) \\
&= \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} z_i^2\right\} \rightarrow \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z_1^2}{2}\right\} \cdot \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z_2^2}{2}\right\} \cdots \\
&= \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2}\mathbf{z}^T\mathbf{z}\right\}.
\end{aligned}$$

Random Vectors

For a standard normal random vector $\mathbf{Z}$, where Z_i's are i.i.d. and $Z_i \sim N(0,1)$, the PDF is given by

$$f_{\mathbf{Z}}(\mathbf{z}) = \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2}\mathbf{z}^T\mathbf{z}\right\}.$$

Random Vectors

Generally,

For a normal random vector X with mean m and covariance matrix C , the PDF is given by

$$f_{\mathrm{X}}(\mathrm{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{\det \mathrm{C}}} \exp\left\{-\frac{1}{2}(\mathrm{x}-\mathrm{m})^T \mathrm{C}^{-1}(\mathrm{x}-\mathrm{m})\right\}.$$

$$\mathrm{C} \rightarrow \mathrm{Var}(X) = \sigma^2 f_X(x) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2}} \exp\left\{-\frac{1}{2}(\mathrm{x}-\mu)^T . \frac{1}{\sigma^2}(\mathrm{x}-\mu)\right\}$$

Random Vectors

If $\mathbf{X} = [X_1, X_2, ..., X_n]^T$ is a normal random vector, and we know $\mathrm{Cov}(X_i, X_j) = 0$ for all $i \neq j$, then $X_1, X_2, ..., X_n$ are independent.

Another important result is that if

$$X \sim N(\mu_X, \sigma_X^2) \rightarrow Y = aX + b \Rightarrow Y \sim N(a\mu_X + b, a^2\sigma_X^2)$$

$$\mathbf{C_Y} = \mathbf{A}\mathbf{C_X}\mathbf{A}^\mathbf{T}$$

UMASS AMHERST

Random Vectors

If $\mathrm{X} = [X_1, X_2, \ldots, X_n]^T$ is a normal random vector, $\mathrm{X} \sim N(\mathrm{m}, \mathrm{C})$, A is an m by n fixed matrix, and b is an m-dimensional fixed vector, then the random vector $\mathrm{Y} = \mathrm{AX} + \mathrm{b}$ is a normal random vector with mean $\mathrm{A}E\mathrm{X} + \mathrm{b}$ and covariance matrix ACA^T.

$$\mathrm{Y} \sim N(\mathrm{A}E\mathrm{X} + \mathrm{b}, \mathrm{ACA}^T).$$

Random Vectors

Example. Let $\mathrm{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ be a normal random vector with the following mean vector and covariance matrix

$$\mathrm{m} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \mathrm{C} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}.$$

Let also

$$\mathrm{A} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 1 & 1 \end{bmatrix}, \mathrm{b} = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}, \mathrm{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} = \underset{3\times 2}{\mathrm{A}}\,\underset{2\times 1}{\mathrm{X}} + \underset{3\times 1}{\mathrm{b}}.$$

Random Vectors

a) Find $P(0 \leq X_2 \leq 1)$.

b) Find the expected value vector of $\mathbf{Y}$, $\mathbf{m_Y} = E\mathbf{Y}$.

c) Find the covariance matrix of $\mathbf{Y}$, $\mathbf{C_Y}$.

d) Find $P(Y_3 \leq 1)$.

Probability Bounds

1) Union Bound and its Extensions
2) Markov and Chebyshev's Inequalities
3) Chernoff Bounds
4) Cauchy-Schwarz Inequality
5) Jenson's Inequality

Probability Bounds

Usefulness:

1) Generality
2) When exact computation is not possible.

Probability Bounds

Remember:

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &\leq P(A) + P(B). \end{aligned}$$

More generally:

$$P(A \cup B \cup C \cup \cdots) \leq P(A) + P(B) + P(C) + \cdots$$

Probability Bounds

The Union Bound

For any events $A_1, A_2, \ldots, A_n$, we have

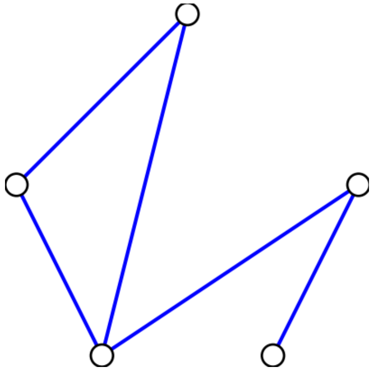
$$P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i).$$

Probability Bounds

Example. **Random Graphs:** $G(n, p)$.

n : The number of nodes

p : Probability of connection between two nodes (independent from other edges).



$n = 5$

UMASS AMHERST

Probability Bounds

Example. Let B_n be the event that a graph randomly generated according to $G(n, p)$ model has at least one isolated node (a node that is not connected to any other nodes). Show that

$$P(B_n) \leq n(1-p)^{n-1}.$$

b) And conclude that for any $\epsilon > 0$, if $p = p_n = (1+\epsilon)\frac{\ln(n)}{n}$ then

$$\lim_{n\to\infty} P(B_n) = 0.$$

Probability Bounds

It is an interesting exercise to calculate $P(B_n)$ exactly using the inclusion-exclusion principle:

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) - \cdots + (-1)^{n-1} P\left(\bigcap_{i=1}^{n} A_i\right).$$

Probability Bounds

Generalization of the Union Bound: Bonferroni Inequalities

For any events $A_1, A_2, \ldots, A_n$, we have

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i);$$

$$P\left(\bigcup_{i=1}^{n} A_i\right) \geq \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j);$$

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k).$$

$$\vdots$$

Probability Bounds

Markov's Inequality

If X is any nonnegative random variable, then

$$P(X \geq a) \leq \frac{EX}{a}, \quad \text{for any } a > 0.$$

Probability Bounds

Let X be any positive continuous random variable, we can write

$$\begin{aligned} EX = \int_0^\infty x f_X(x)dx &\geq \int_a^\infty x f_X(x)dx \qquad \text{(for any } a > 0 \text{ and } x > a\text{)} \\ &\geq \int_a^\infty a f_X(x)dx = a \underbrace{\int_a^\infty f_X(x)dx}_{P(X \geq a)} \end{aligned}$$

$$\Rightarrow \quad P(X \geq a) \leq \frac{EX}{a}, \qquad \text{for any } a > 0.$$

Probability Bounds

Chebyshev's Inequality

If X is any random variable, then for any $b > 0$, we have

$$P(|X - EX| \geq b) \leq \frac{Var(X)}{b^2}.$$

Probability Bounds

Let X be any random variable. If $Y = (X - EX)^2$, then Y is a nonnegative random variable.

$$P(Y \geq b^2) \leq \frac{EY}{b^2} = \frac{\mathrm{Var}(X)}{b^2},$$

$$P\big((X - EX)^2 \geq b^2\big) = P\big(|X - EX| \geq b\big).$$

Probability Bounds

Chernoff Bounds

$$P(X \geq a) \leq e^{-sa} M_X(s), \quad \text{for all } s > 0$$

$$P(X \leq a) \leq e^{-sa} M_X(s), \quad \text{for all } s < 0$$

Probability Bounds

For $s > 0$, we can write

$$\begin{aligned} P(X \geq a) &= P(e^{sX} \geq e^{sa}) \\ &\leq \frac{E[e^{sX}]}{e^{sa}}, \qquad \text{by Markov's inequality.} \\ &= \frac{M_X(s)}{e^{sa}} \end{aligned}$$

Probability Bounds

Cauchy-Schwarz Inequality

For any two random variables X and Y, we have

$$|EXY| \leq \sqrt{E[X^2]E[Y^2]},$$

where equality holds if and only if $X = \alpha Y$, for some constant $\alpha \in \mathbb{R}$.

Probability Bounds

$|\rho(X,Y)| \leq 1:$

assuming $EX = EY = 0$

$$|\rho(X,Y)| = \left|\frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}\right| = \left|\frac{EXY}{\sqrt{E[X^2]E[Y^2]}}\right| \leq 1$$

$$\Rightarrow \; |EXY| \leq \sqrt{E[X^2]E[Y^2]}.$$

Probability Bounds

Jensen's Inequality

Remember

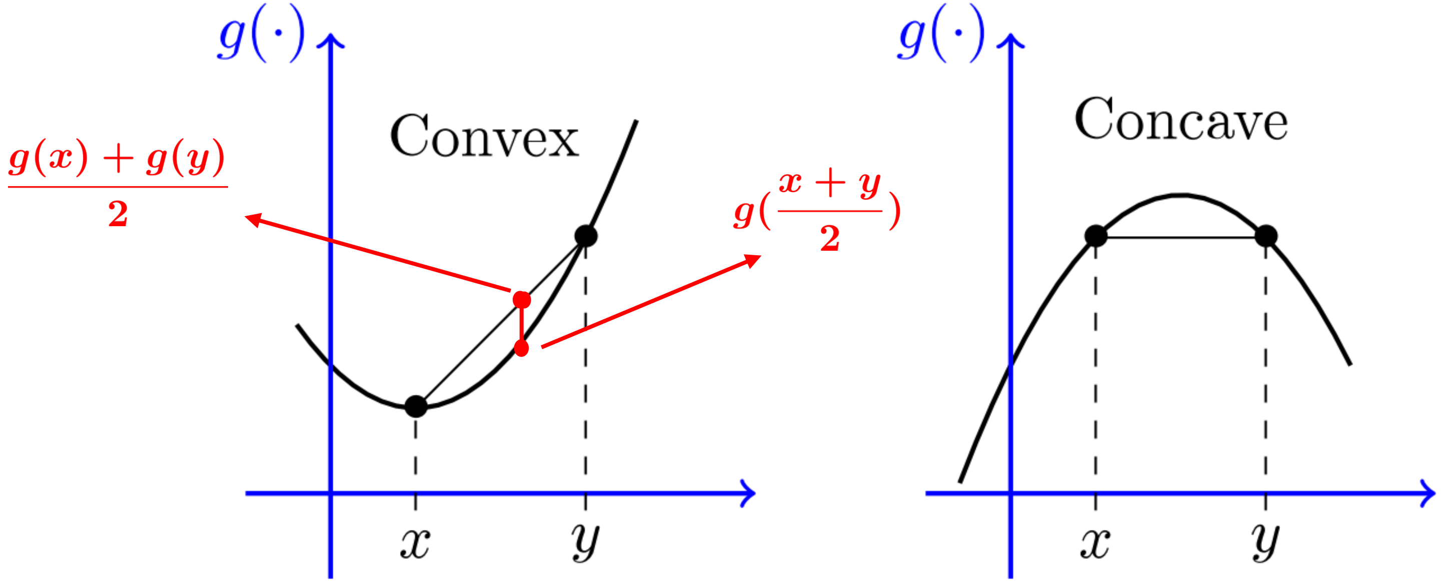
$$Var(X) = EX^2 - (EX)^2 \geq 0 \;\Rightarrow\; EX^2 \geq (EX)^2.$$

If $g(x) = x^2$,

$$E[g(X)] \geq g(E[X]).$$

Probability Bounds

Jensen's Inequality



$$\Rightarrow \text{ Convex : } \frac{g(x)+g(y)}{2} \geq g(\frac{x+y}{2})$$

Probability Bounds

Definition. Consider a function $g : I \to \mathbb{R}$, where I is an interval in $\mathbb{R}$. We say that g is a **convex** function if, for any two points x and y in I and any $\alpha \in [0, 1]$, we have

$$g(\alpha x + (1-\alpha)y) \leq \alpha g(x) + (1-\alpha)g(y).$$

We say that g is **concave** if

$$g(\alpha x + (1-\alpha)y) \geq \alpha g(x) + (1-\alpha)g(y).$$

Probability Bounds

Jensen's Inequality

If $g(x)$ is a convex function on R_X, and $E[g(X)]$ and $g(E[X])$ are finite, then

$$E[g(X)] \geq g(E[X]).$$

Probability Bounds

A twice-differentiable function $g : I \to \mathbb{R}$ is convex if and only if $g''(x) \geq 0$ for all $x \in I$.

Post-work for Lesson

- **Complete homework assignment for Lessons 14-16:**

HW#8

Go to the online classroom for details.

To Prepare for the Next Lesson

- Read Chapter 7 in your online textbook:

  https://www.probabilitycourse.com/chapter6/6_0_0_intro.php

- Complete the Pre-work for Lesson 17.

Visit the online classroom for details.