ELEC 2600H: Probability and Random Processes in Engineering

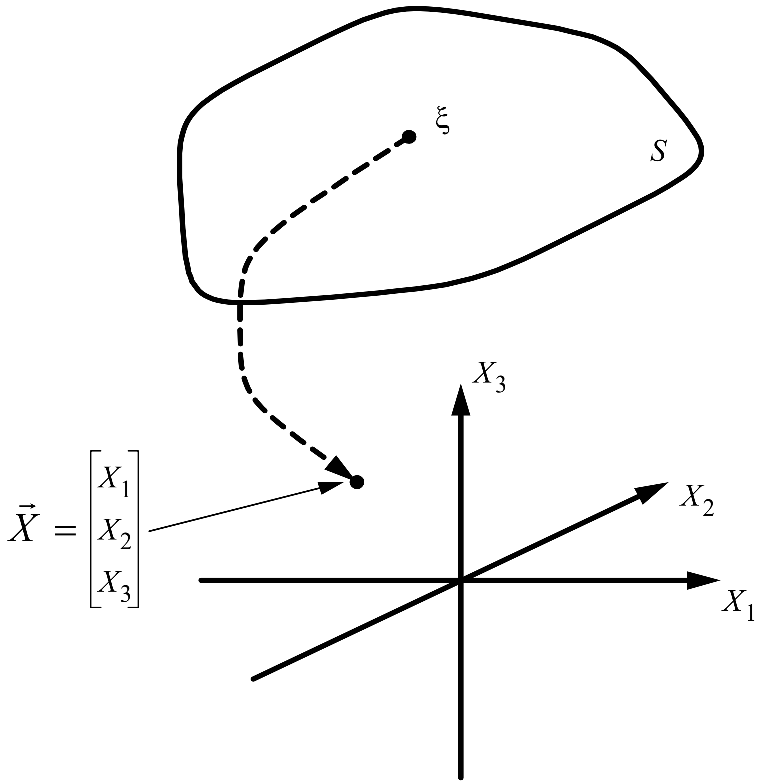
Part III: Multiple Random Variables

Elec2600H: Lecture 10

- **Multiple random variables (RVs)**
- Joint probability mass function of two discrete RVs
- Marginal probability mass function

Vector Random Variables

- A *vector random variable* $\vec{X}$ is a function that assigns a vector of real numbers to every *outcome* of an experiment.



Example: Height, Weight, Age

- Example 5.1 – A student's name is selected randomly from an urn. Let ζ denote the outcome of this experiment, and define:

  $H(\zeta)$ = height of student in inches

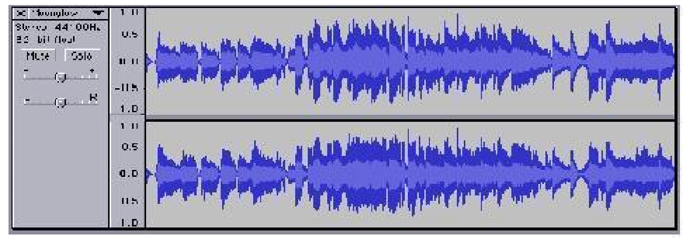
  $W(\zeta)$ = weight of student in pounds,

  $A(\zeta)$ = age of student in years

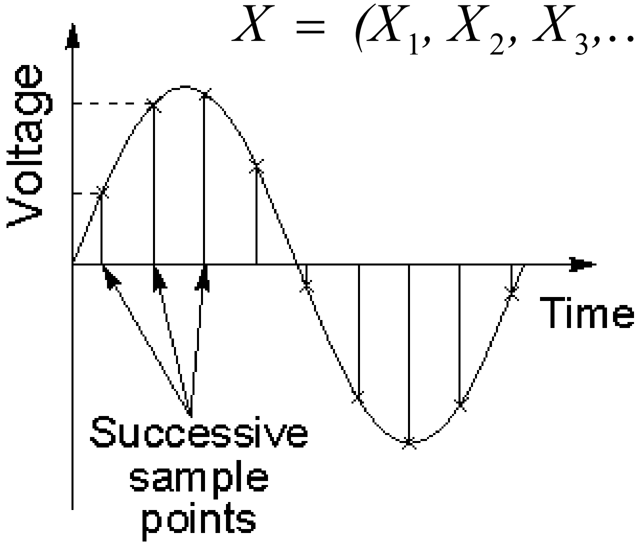
  The vector $[H(\zeta),\ W(\zeta),\ A(\zeta)]$ is a **vector** random variable

Example: Sampled Voltage (ADC)

- Example – Let the outcome ζ of an experiment be a voltage waveform $X(t, \zeta)$. Let the random variable $X_k = X(kT, \zeta)$ be the sampled voltage taken at time $t = kT$. The vector consisting of the first n samples is a vector random variable.

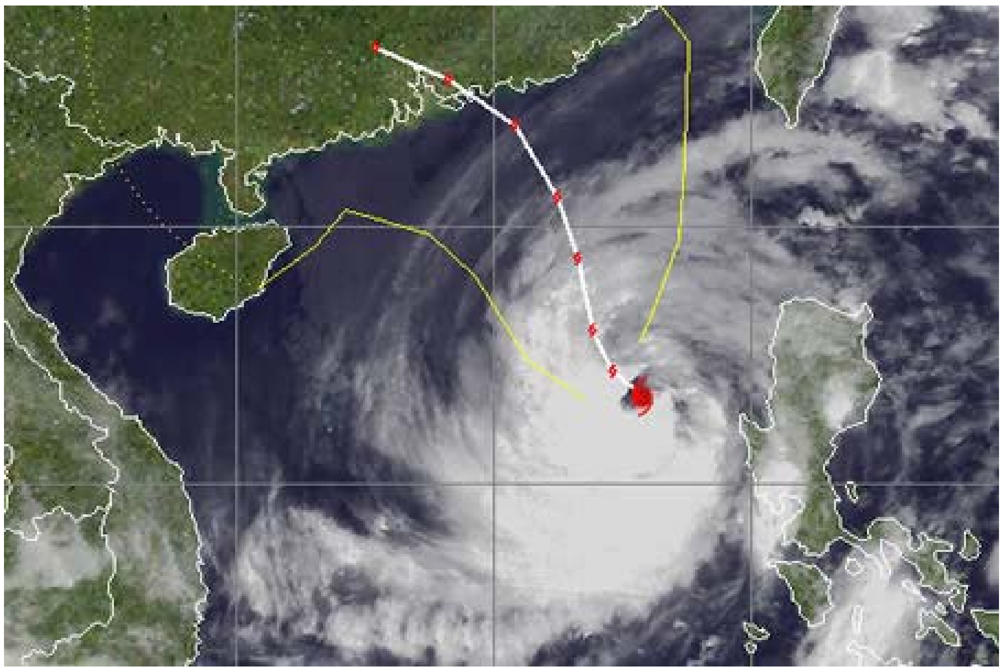


$$\vec{X} = (X_1, X_2, X_3, \ldots, X_n)$$



Example: Typhoon MEGI

- Example – For typhoon MEGI, at a given time, measure the maximum wind speed and the position from land. Each of these is a random variable, and together they form a random vector.



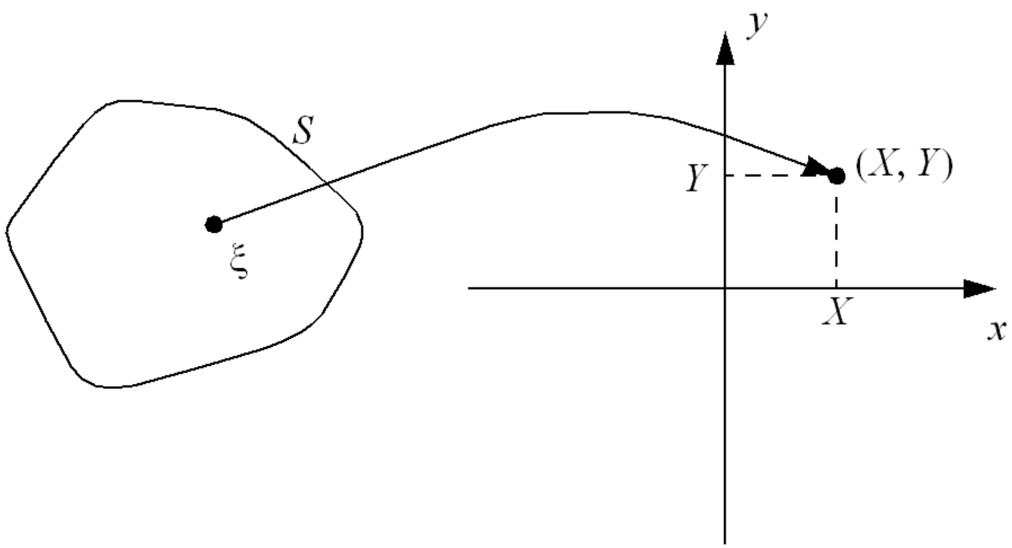
Example: Skin Detection in Digital Images

Question:

What about detecting:
- **Eyes?**
- **Face?**
- **Multiple Faces?**

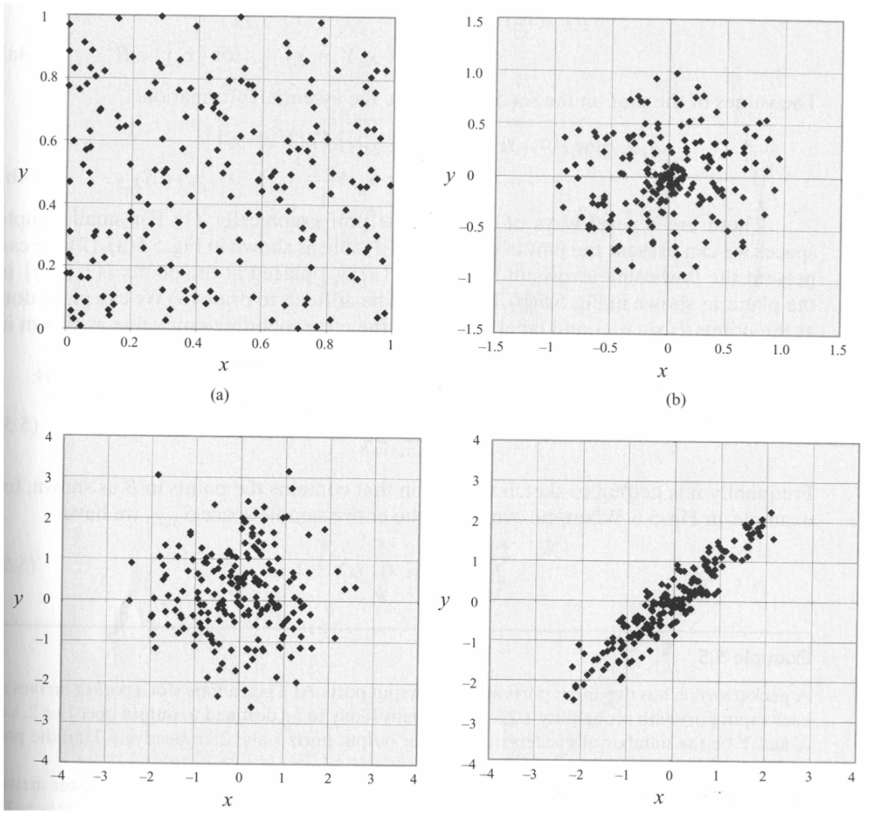
Two Random Variables

- One random variable can be considered as a mapping from the sample space to the real line.
- Two random variables can be considered as a mapping from the sample space to the plane.
- Note that given the outcome of the experiment, both X and Y are determined simultaneously.

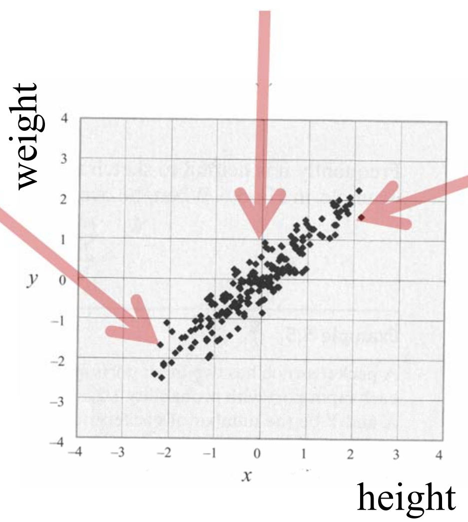


Scattergrams

- Scattergrams are used to visualize the joint behavior of two random variables.
- A set of experiments is performed. For each experiment, we place a dot at every observation pair (x, y).



Scattergrams: Example

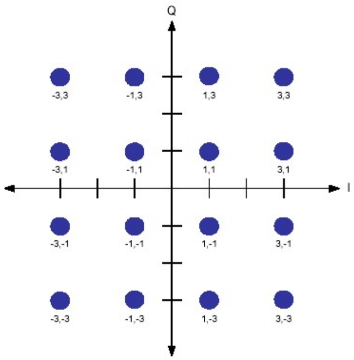


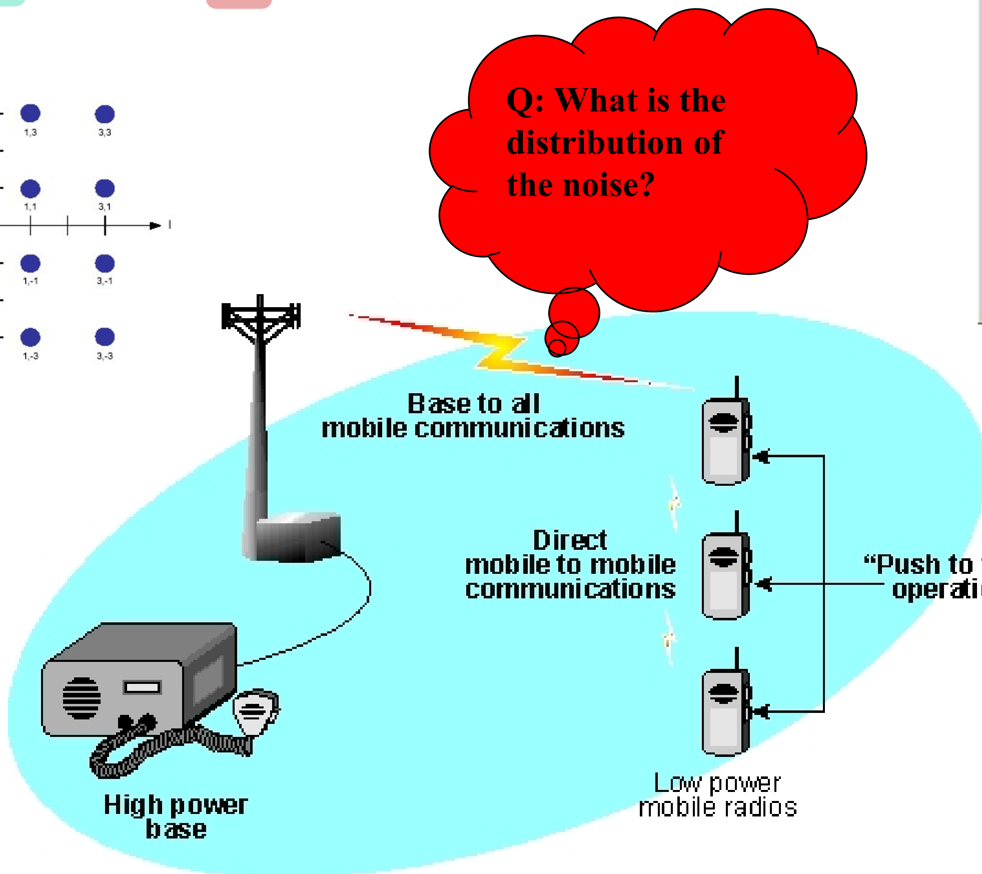
Edward Nino Hernandez (0.7m, 10kg)

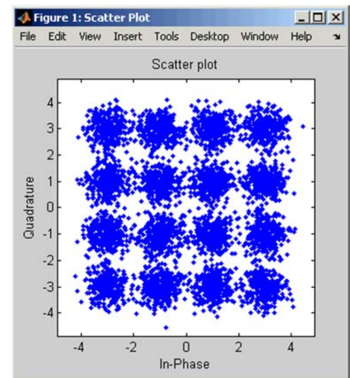
Bao Xishun (2.36m)

Scattergrams: Mobile Communications!

$$s(t) = I(t)\cos(2\pi f_0 t) + Q(t)\sin(2\pi f_0 t)$$

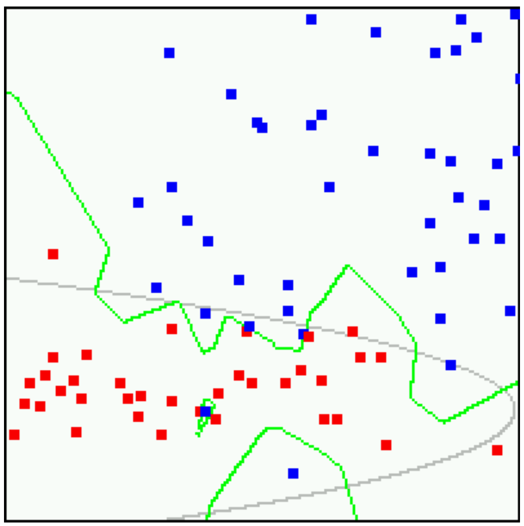






Events involving two RVs

- Events involving two random variables are **regions** in the X-Y plane.
- You can find the region using the following procedure
  - Find the **boundaries** of the event. The boundary splits the plane into one or more regions.
  - Check which regions belong to the event by **testing a point** inside each region.



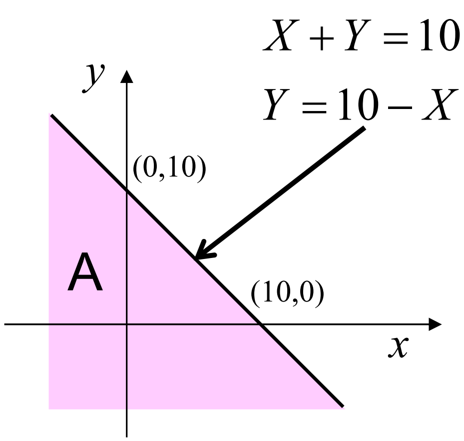
Examples

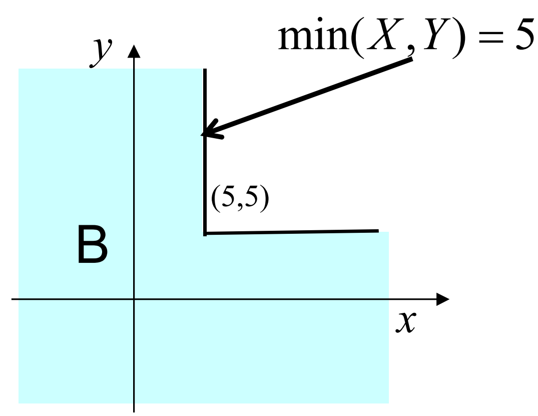
- Consider a 2-dimensional R.V. $\boldsymbol{X} = (X, Y)$.
- Find the **region of the plane** corresponding to the events:

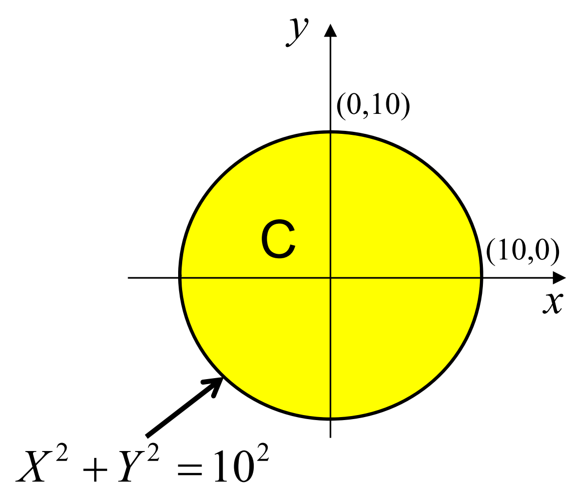
  $A = \{ X + Y \leq 10 \}$ = "the sum is less or equal to 10"

  $B = \{ \min(X, Y) \leq 5 \}$ = "the smaller is less than or equal to 5"

  $C = \{ X^2 + Y^2 \leq 100 \}$ = "the distance from the origin is less than or equal to 10"







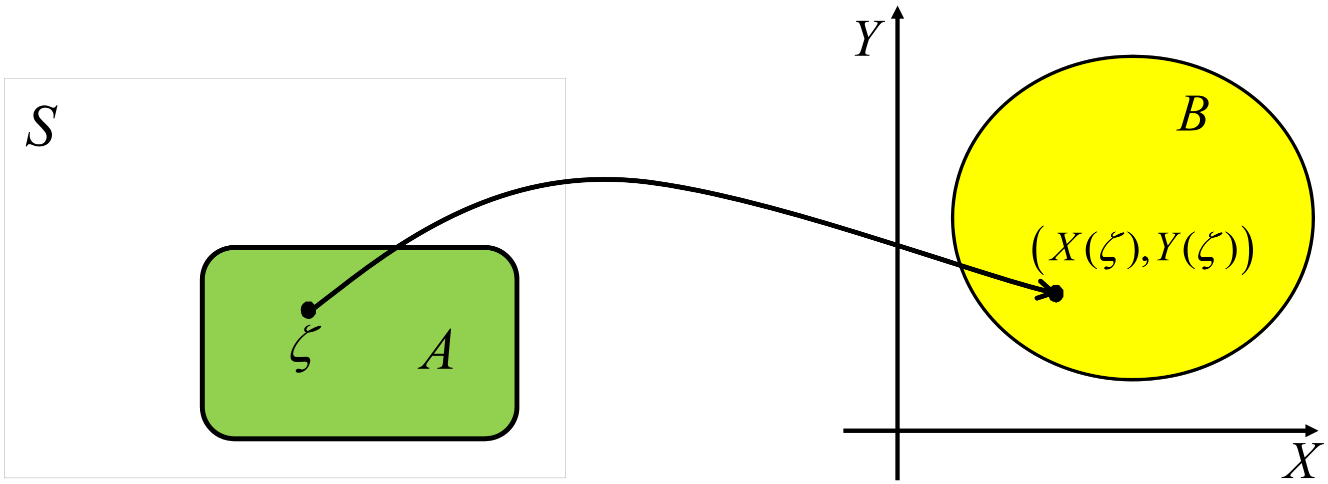
Elec2600H: Lecture 10

- Multiple random variables (RVs)
- **Joint probability mass function of two discrete RVs**
- Marginal probability mass function

Finding the probabilities of events

- Just as in the case for single random variables, we find the probability of an event B defined in terms of X and Y by finding the probability of the equivalent event, A.

$$\boxed{P[(X,Y) \text{ in } B] = P[A] \quad \text{where} \quad A = \{\zeta : (X(\zeta), Y(\zeta)) \text{ in } B\}}$$



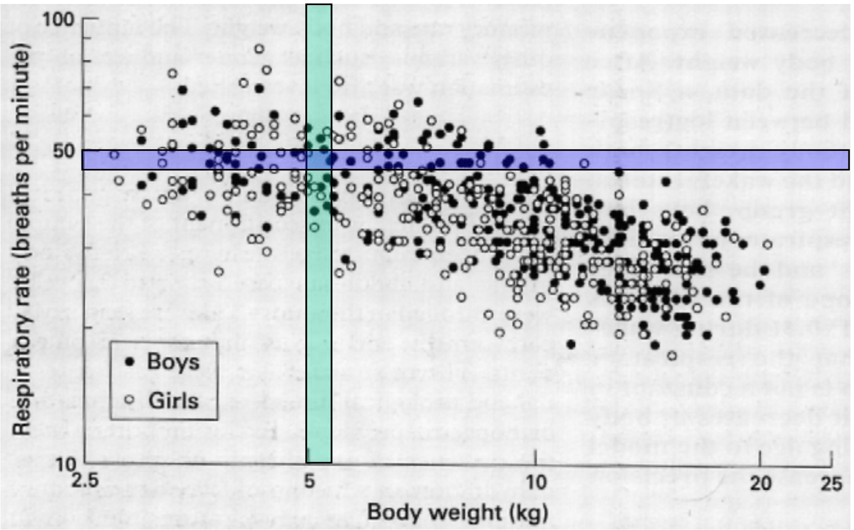
Pairs of discrete random variables

- Let the vector variable random $\vec{X} = (X, Y)$ assume values from a countable set.
- For convenience, we assume X and Y only take integer values.

- ***Joint probability mass function*** **(joint pmf):** $p_{X,Y}(j,k) = P[\{X = j\} \cap \{Y = k\}]$

- The probability of any event A is the sum over all outcomes in A: $P[A] = \sum_{(j,k) \in A} p_{X,Y}(j,k)$
- The sum over all outcomes is 1:

$$\sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} p_{X,Y}(j,k) = 1$$



Example

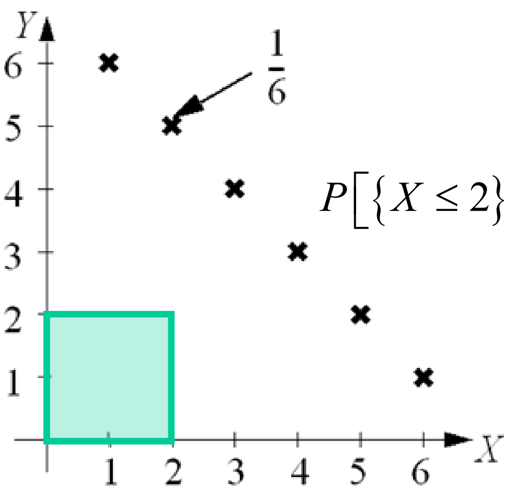
- Toss two fair die, and define

  $X=$ # of dots on first die,

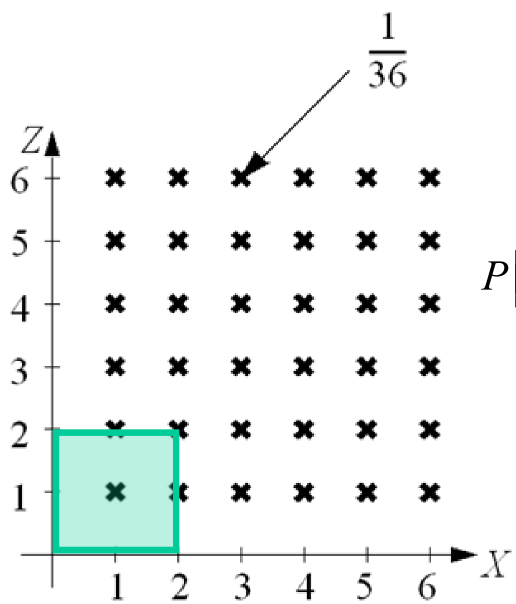
  $Y = 7 - X$

  $Z =$ # of dots on second die

- Consider the **pairs** (X,Y) and (X,Z).
- Find the probabilities that *both numbers* in the pair are less than or equal to 2.
- Solution



$$P\left[\{X \le 2\} \cap \{Y \le 2\}\right] = 0$$



$$P\left[\{X \le 2\} \cap \{Z \le 2\}\right] = \frac{1}{9}$$

Example 5.6

- A random experiment consists of tossing two **"loaded" (unfair) dice**, which have higher probability for pairs.
- Let (X,Y) be the number of dots facing up on the two die.
- The joint pmf $p_{X,Y}(j,k)$ for j=1,…,6 and k=1,…6 is:

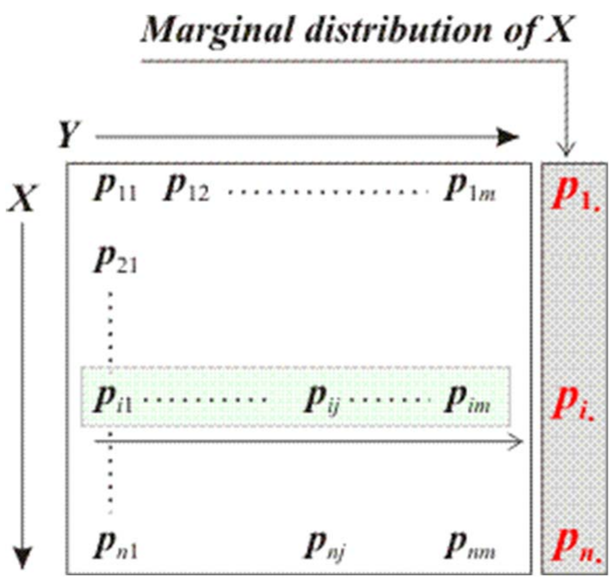
k							
	6	1/42	1/42	1/42	1/42	1/42	**2/42**
	5	1/42	1/42	1/42	1/42	**2/42**	1/42
	4	1/42	1/42	1/42	**2/42**	1/42	1/42
	3	1/42	1/42	**2/42**	1/42	1/42	1/42
	2	1/42	**2/42**	1/42	1/42	1/42	1/42
	1	**2/42**	1/42	1/42	1/42	1/42	1/42
$p_{X,Y}(j,k)$		1	2	3	4	5	6
		j					

- Find $P[\min(X,Y) = 3]$

$$
\begin{aligned}
P[\min(X,Y)=3] &= p_{X,Y}(6,3)+p_{X,Y}(5,3)+p_{X,Y}(4,3) \\
&\quad + p_{X,Y}(3,3) \\
&\quad + p_{X,Y}(3,4)+p_{X,Y}(3,5)+p_{X,Y}(3,6) \\
&= 6\times\frac{1}{42}+\frac{2}{42}=\frac{8}{42}
\end{aligned}
$$

Elec2600H: Lecture 10

- Multiple Random Variables (RVs)
- Joint probability mass function of two discrete RVs
- Marginal probability mass function



Marginal Statistics

- The *marginal probability distribution* of a random variable in a joint distribution is simply the probability of the random variable when **considered in isolation**.

- The *marginal pmf* of X is $p_X(j) = P[X = j] = P[X = j, Y = \text{anything}]$

$$= \sum_{k=-\infty}^{\infty} p_{X,Y}(j,k)$$

- Similarly, $p_Y(k) = \sum_{j=-\infty}^{\infty} p_{X,Y}(j,k)$

be careful when doing these sums, as many of the values may be zero!

Example 5.8

- A random experiment consists of tossing two "loaded" (unfair) dice
- Let (X,Y) be the number of dots facing up on the two die.
- The joint pmf $p_{X,Y}(j,k)$ for $j=1,\ldots,6$ and $k=1,\ldots6$ is:

k	6	1/42	1/42	1/42	1/42	1/42	**2/42**
	5	1/42	1/42	1/42	1/42	**2/42**	1/42
	4	1/42	1/42	1/42	**2/42**	1/42	1/42
	3	1/42	1/42	**2/42**	1/42	1/42	1/42
	2	1/42	**2/42**	1/42	1/42	1/42	1/42
	1	**2/42**	1/42	1/42	1/42	1/42	1/42
$p_{X,Y}(j,k)$		1	2	3	4	5	6
		j					

- Find the marginal pmf's.

Example 5.8 (Solution)

- The marginal pmf for X is obtained by summing along columns:

$$p_X(1) = P[X=1] = \frac{2}{42} + \frac{1}{42} + \frac{1}{42} + \frac{1}{42} + \frac{1}{42} + \frac{1}{42} = \frac{7}{42} = \frac{1}{6}$$

  - More generally, $p_X(j) = \mathrm{P}[X = j] = 1/6$ for j = 1,2,…,6.

k	6	1/42	1/42	1/42	1/42	1/42	**2/42**
	5	1/42	1/42	1/42	1/42	**2/42**	1/42
	4	1/42	1/42	1/42	**2/42**	1/42	1/42
	3	1/42	1/42	**2/42**	1/42	1/42	1/42
	2	1/42	**2/42**	1/42	1/42	1/42	1/42
	1	**2/42**	1/42	1/42	1/42	1/42	1/42
$p_{X,Y}(j,k)$		1	2	3	4	5	6
		j					

- The marginal pmf for Y is obtained by summing along rows: $p_Y(k) = P[Y = k] = \frac{1}{6}$ for $k = 1, 2, \ldots, 6$

Example:

- Suppose X is chosen by counting the number of dots on a fair die and Y is chosen by choosing a number at random between 1 and X.

- It turns out that the joint pmf is:

k	6	0	0	0	0	0	1/36
	5	0	0	0	0	1/30	1/36
	4	0	0	0	1/24	1/30	1/36
	3	0	0	1/18	1/24	1/30	1/36
	2	0	1/12	1/18	1/24	1/30	1/36
	1	1/6	1/12	1/18	1/24	1/30	1/36
$p_{X,Y}(j,k)$		1	2	3	4	5	6
		j					

- Find the marginal pmfs.

Example (Solution; PMF of X)

- The marginal pmf for X is obtained by **summing along columns** of the matrix.

k	6	0	0	0	0	0	1/36
	5	0	0	0	0	1/30	1/36
	4	0	0	0	1/24	1/30	1/36
	3	0	0	1/18	1/24	1/30	1/36
	2	0	1/12	1/18	1/24	1/30	1/36
	1	1/6	1/12	1/18	1/24	1/30	1/36
$p_{X,Y}(j,k)$		1	2	3	4	5	6
		j					

j	1	2	3	4	5	6
$p_X(j)$	1/6	2/12	3/18	4/24	5/30	6/36
$p_X(j)$	1/6	1/6	1/6	1/6	1/6	1/6

Example (Solution; PMF of Y)

- The marginal pmf for Y is obtained by **summing along rows** of the matrix.

k	6	0	0	0	0	0	1/36
	5	0	0	0	0	1/30	1/36
	4	0	0	0	1/24	1/30	1/36
	3	0	0	1/18	1/24	1/30	1/36
	2	0	1/12	1/18	1/24	1/30	1/36
	1	1/6	1/12	1/18	1/24	1/30	1/36
$p_{X,Y}(j,k)$		1	2	3	4	5	6
		j					

k	$p_Y(k)$	$p_Y(k)$
6	10/360	0.028
5	22/360	0.061
4	37/360	0.103
3	57/360	0.158
2	87/360	0.242
1	147/360	0.408

Example 5.9

The number of characters (bytes) N in a message has a geometric distribution: $p_N(n) = (1-q)q^n$ for $n = 0,1,2,\ldots$

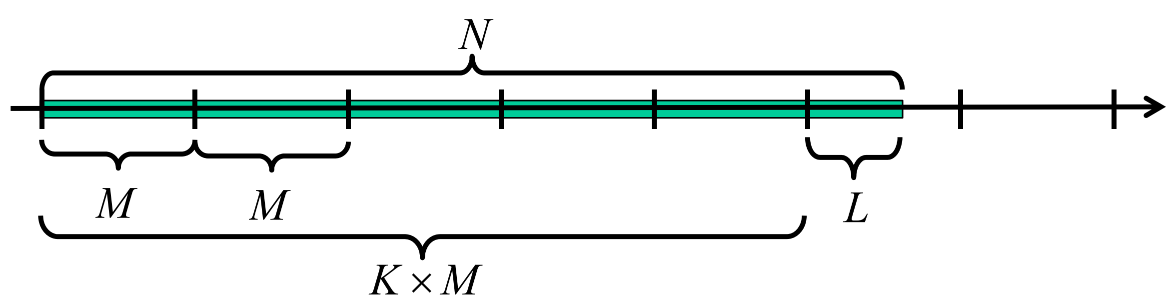
Note that we have defined the geometric distribution here in terms of $q = (1-p)$, where p the probability of success. The usual pmf of the geometric distribution is given by $p_N(n) = p(1-p)^n$ for $n = 0,1,2,\ldots$

In this problem, q can be interpreted as the probability that the message continues beyond the current character, whereas p can be interpreted as the probability that the message stops. The larger q (smaller p) is, the longer the average length of the message: $E[N] = \frac{q}{1-q} = \frac{1-p}{p}$

Suppose that messages are broken into packets of maximum length M bytes.

Let K be the number of full packets

L be the number of bytes left over.



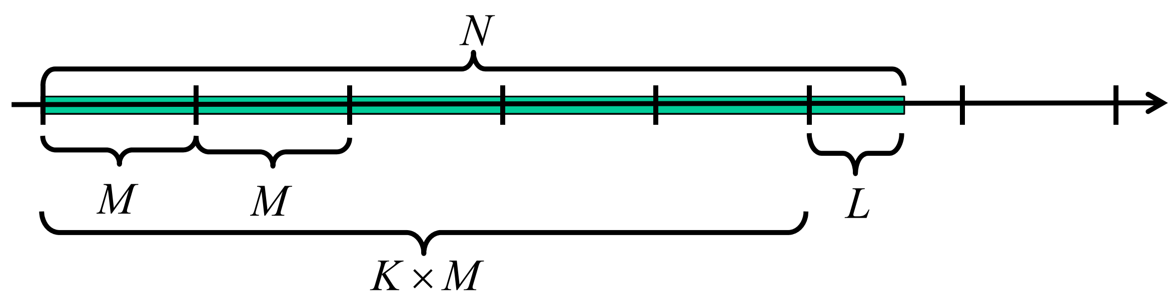
Example 5.9

The number of bytes N in a message has a geometric distribution $p_N(n) = (1-q)q^n$ for $n = 0,1,2,\ldots$

Suppose that messages are broken into packets of maximum length M bytes.

Let K be the number of full packets

L be the number of bytes left over.



Find the joint pmf of K and L.

Solution

The joint pmf is given by

$$\begin{aligned} p_{(K,L)}(k,l) &= P[K = k, L = l] \\ &= P[N = k \times M + l] \\ &= (1-q)q^{kM+l} \end{aligned}$$

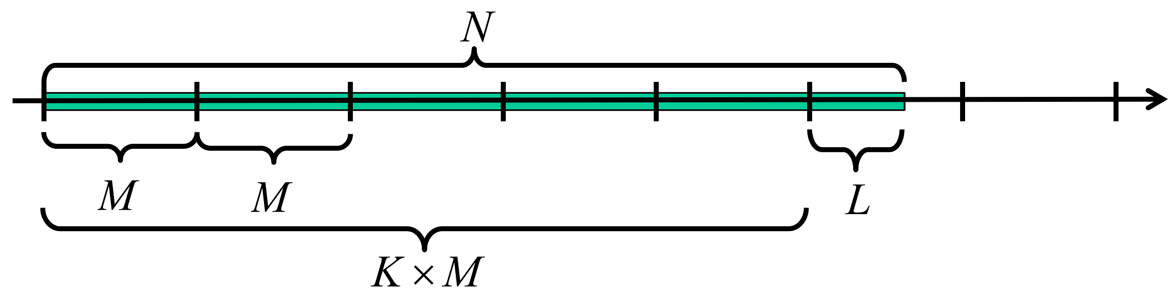
for $k = 0,1,2,\ldots$ and $l = 0,1,2,\ldots, M-1$ and zero otherwise.

Example 5.9 (Cont.)

Find the marginal pmf of K in the previous example, where the joint pmf is

$$p_{K,L}(k,l) = (1-q)q^{kM+l}$$

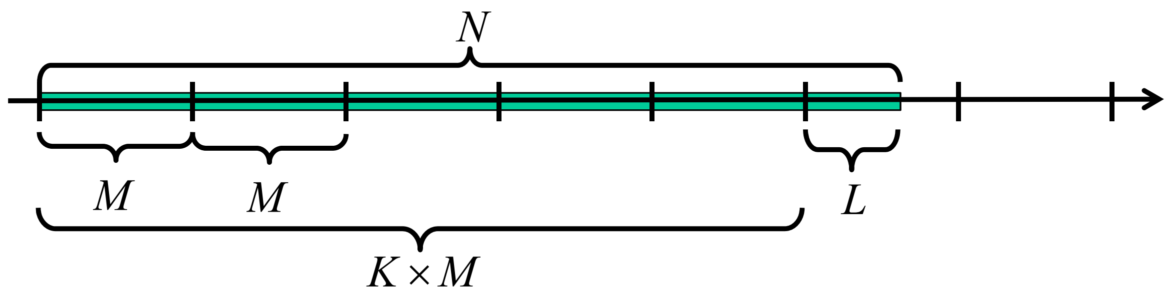
for $k \in \{0,1,2,\dots\}$ and $l \in \{0,1,2,\dots,M-1\}$



Solution

$$\begin{aligned}
p_K(k) &= \sum_{l=0}^{M-1} p_{K,L}(k,l) \\
&= \sum_{l=0}^{M-1} (1-q)q^{kM+l} \\
&= (1-q)q^{kM} \sum_{l=0}^{M-1} q^l \\
&= (1-q)q^{kM} \frac{(1-q^M)}{(1-q)} \\
&= (1-q^M)(q^M)^k \ \text{ for } k \in \{0,1,2,\dots\}
\end{aligned}$$

Example 5.9 (Cont.)



Find the marginal pmfs of L in the previous example, where the joint pmf is

$$p_{K,L}(k,l) = (1-q)q^{kM+l}$$

for $k \in \{0,1,2,\dots\}$ and $l \in \{0,1,2,\dots,M-1\}$

Solution

$$\begin{aligned} p_L(l) &= \sum_{k=0}^{\infty} p_{K,L}(k,l) \\ &= \sum_{k=0}^{\infty} (1-q)q^{kM+l} \\ &= (1-q)q^l \sum_{k=0}^{\infty} (q^M)^k \\ &= (1-q)q^l \frac{1}{(1-q^M)} \\ &= \frac{1-q}{1-q^M} q^l \ \text{ for } l \in \{0,1,2,\dots(M-1)\} \end{aligned}$$

Major Points from this Lecture:

- **Joint PMF contains all information about discrete random vectors**
- **Obtain the marginal distributions from joint PMF**

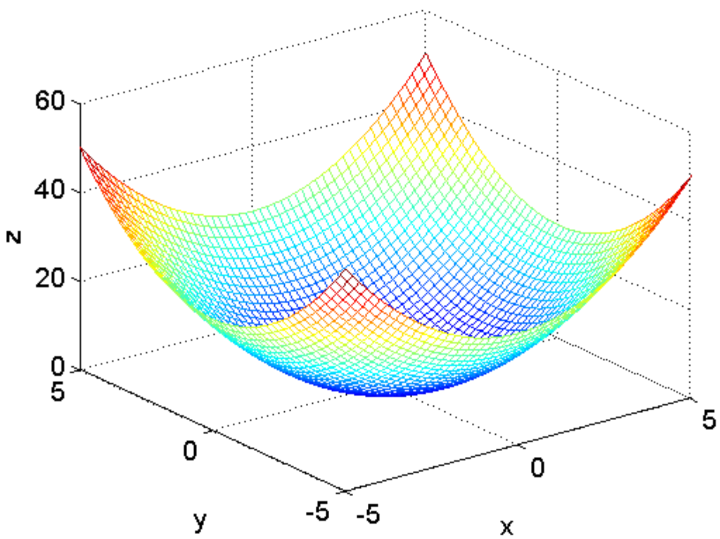
ELEC 2600H: Probability and Random Processes in Engineering

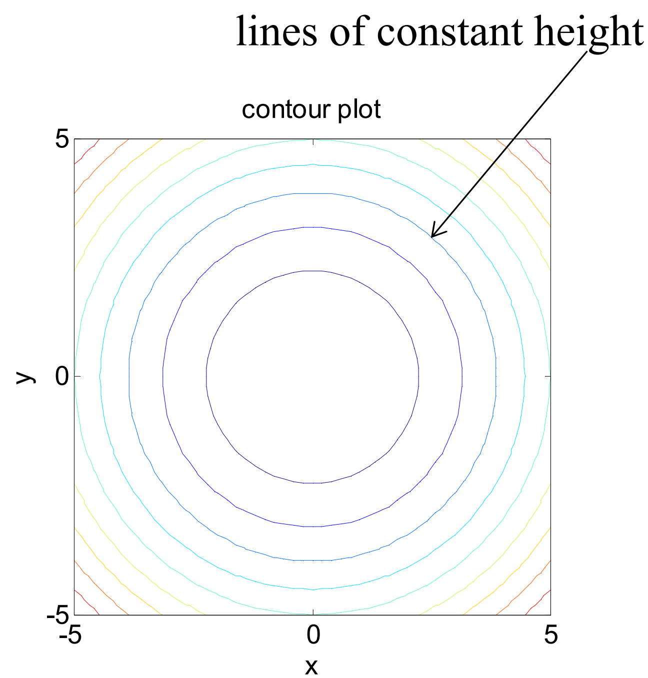
Part III: Multiple Random Variables

- **Lecture 10: Pairs of Discrete Random Variable**
  - **Out-of-Class Reading: 2D Calculus**
- Lecture 11: Pairs of Continuous Random Variable
- Lecture 12: Independence, Joint Moments
- Lecture 13: Correlation Coefficients and Their Properties
- Lecture 14: Conditional PDF, Conditional Expectation
- Lecture 15: Sum of Two Random Variables
- Lecture 16: Pairs of Jointly Gaussian Random Variables
- Lecture 17: More than Two Random Variables
- Lecture 18: Laws of Large Numbers
- Lecture 19: Central Limit Theorem and Characteristic Function

Review: 2D Functions

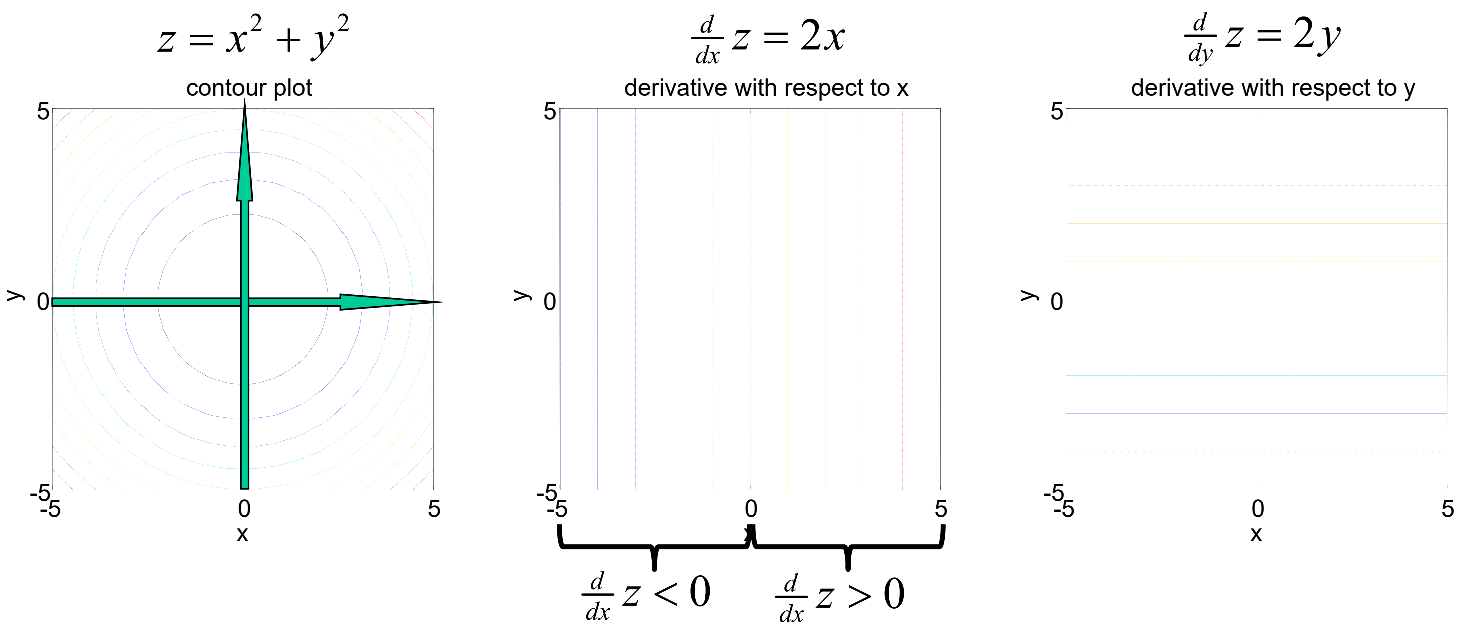
- A real-valued 2D function, $z = f(x,y)$, assigns a real value to each point on the (x,y) plane.
- Think of z as being the height of the ground and regard x and y as your east/west and north/south coordinates.
- Example: $z = x^2 + y^2$





Differentiating 2D functions

- The derivative of a 2D function with respect to x indicates how the function changes as you move from **west to east**. It is calculated using the normal rules of differentiation, assuming that y **is constant.**
- The derivative of a 2D function with respect to y indicates how the function changes as you move from **south to north.** It is calculated using the normal rules of differentiation, assuming that x **is constant.**



Multiple Differentiations

- Multiple differentiations are performed in sequence.
- The sequence of the differentiation does **not matter**.

- Example:

$$z = (x^2 + y)^2 = x^4 + 2x^2 y + y^2$$

$$\begin{aligned}\frac{d}{dx}\frac{d}{dy} z &= \frac{d}{dx}\frac{d}{dy}\left(x^4 + 2x^2 y + y^2\right) \\ &= \frac{d}{dx}\left(0 + 2x^2 + 2y\right) \\ &= 4x\end{aligned}$$

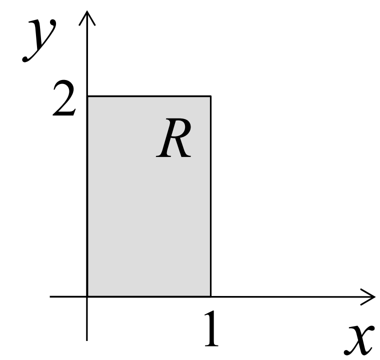
$$\begin{aligned}\frac{d}{dy}\frac{d}{dx} z &= \frac{d}{dy}\frac{d}{dx}\left(x^4 + 2x^2 y + y^2\right) \\ &= \frac{d}{dy}\left(4x^3 + 4xy + 0\right) \\ &= 4x\end{aligned}$$

2D Integration

- Recall that the integral of a 1D function gives us the **area** under the curve between two endpoints.
- The integral of a 2D function gives us the **volume** under the surface within a region.
- In order to evaluate a 2D integral numerically, we need **two** things:
  - The function to be integrated $f(x,y)$.
  - A region in the 2D plane over which the function will be integrated. The region determines the limits of the integration.
- A 2D integral is evaluated as a sequence of two 1D integrals.
- For each 1D integral (e.g. over x), the other variable (e.g. y) is assumed to be constant.
- The order of integration can usually be switched. However, the limits of the integral may change.

Example

Integrate the function x^2y over the region R.



Solution:

x first, then y:

$$\int_0^2\int_0^1 x^2 y\,dx\,dy = \int_0^2 y\left(\int_0^1 x^2 dx\right)dy$$

$$= \int_0^2 y\left(\left.\frac{x^3}{3}\right|_0^1\right)dy$$

$$= \int_0^2 y\left(\frac{1}{3}-0\right)dy$$

$$= \frac{1}{3}\int_0^2 y\,dy = \frac{1}{3}\left(\left.\frac{y^2}{2}\right|_0^2\right) = \frac{2}{3}$$

y first, then x:

$$\int_0^1\int_0^2 x^2 y\,dy\,dx = \int_0^1 x^2\left(\int_0^2 y\,dy\right)dx$$

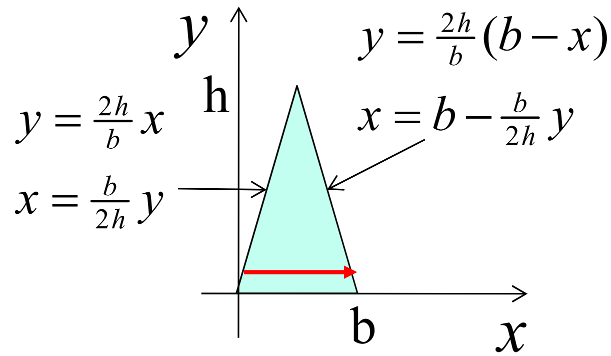
$$= \int_0^1 y\left(\left.\frac{y^2}{2}\right|_0^2\right)dx$$

$$= \int_0^1 x^2\left(\frac{4}{2}-0\right)dx$$

$$= 2\int_0^1 x^2 dx = 2\left(\left.\frac{x^3}{3}\right|_0^1\right) = \frac{2}{3}$$

Example (x first, then y)

Integrate the function c (a constant) over the shaded region.



Solution:

$$\int_0^h \int_{\frac{b}{2h}y}^{b-\frac{b}{2h}y} c\,dx\,dy = c \times \int_0^h \left(\int_{\frac{b}{2h}y}^{b-\frac{b}{2h}y} 1\,dx \right) dy$$

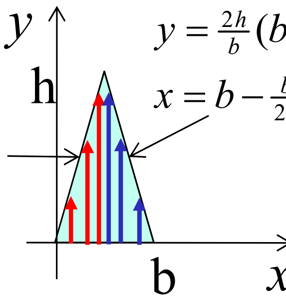
$$= c \times \int_0^h \left(x \Big|_{\frac{b}{2h}y}^{b-\frac{b}{2h}y} \right) dy = c \times \int_0^h \left(b - \tfrac{b}{h} y \right) dy$$

$$= c \times \left(by - \frac{b}{h}\frac{y^2}{2} \right) \Bigg|_0^h$$

$$= c \times \left(bh - \frac{b}{h}\frac{h^2}{2} \right) = c \times \left(\tfrac{1}{2} bh \right)$$

Note:
The 2D integral of a constant over a region is **equal to the constant times the area of the region.**

Example (y first, then x)



Integrate the function c (a constant) over the shaded region.

Solution:

$$\int_0^{\frac{b}{2}}\int_0^{\frac{2h}{b}x} c\,dy\,dx+\int_{\frac{b}{2}}^{b}\int_0^{\frac{2h}{b}(b-x)} c\,dy\,dx = c\times\int_0^{\frac{b}{2}}\left(\int_0^{\frac{2h}{b}x}1\,dy\right)dx + c\times\int_{\frac{b}{2}}^{b}\left(\int_0^{\frac{2h}{b}(b-x)}1\,dy\right)dx$$

$$= c\times\int_0^{\frac{b}{2}}\left(y\Big|_0^{\frac{2h}{b}x}\right)dx + c\times\int_{\frac{b}{2}}^{b}\left(y\Big|_0^{\frac{2h}{b}(b-x)}\right)dx$$

$$= c\times\int_0^{\frac{b}{2}}\frac{2h}{b}x\,dx + c\times\int_{\frac{b}{2}}^{b}\frac{2h}{b}(b-x)\,dx$$

$$= c\times\frac{2h}{b}\left(\frac{x^2}{2}\Big|_0^{\frac{b}{2}}\right) + c\times\frac{2h}{b}\left(bx-\frac{x^2}{2}\Big|_{\frac{b}{2}}^{b}\right)$$

$$= c\times\frac{2h}{b}\left(\frac{b^2}{8}\right) + c\times\frac{2h}{b}\left(b^2-\frac{b^2}{2}-\frac{b^2}{2}+\frac{b^2}{8}\right)$$

$$= c\times\left(\frac{1}{2}bh\right)$$

ELEC 2600H: Probability and Random Processes in Engineering

Part III: Multiple Random Variables

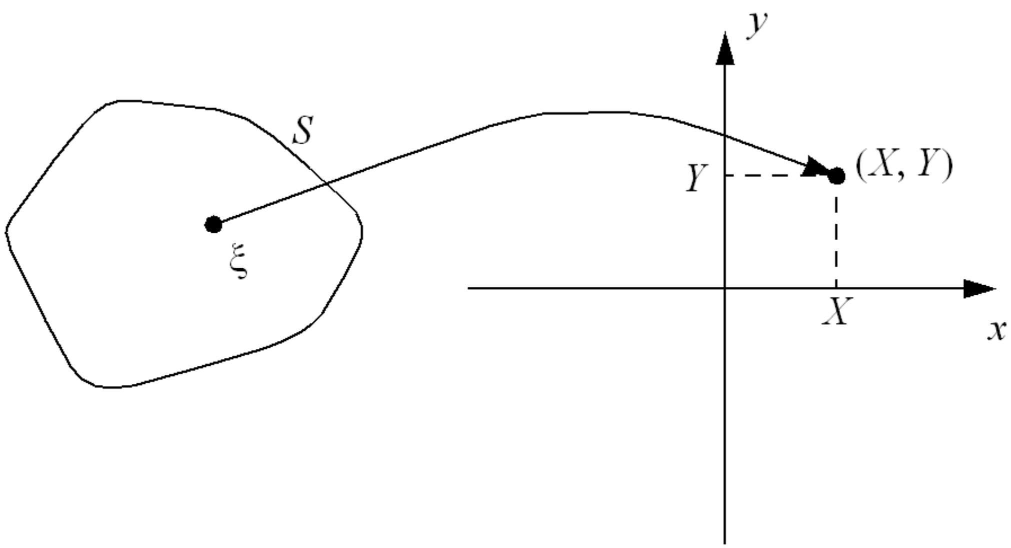
- Lecture 10: Pairs of Discrete Random Variable
  - Out-of-Class Reading: 2D Calculus
- **Lecture 11: Pairs of Continuous Random Variable**
- Lecture 12: Independence, Joint Moments
- Lecture 13: Correlation Coefficients and Their Properties
- Lecture 14: Conditional PDF, Conditional Expectation
- Lecture 15: Sum of Two Random Variables
- Lecture 16: Pairs of Jointly Gaussian Random Variables
- Lecture 17: More than Two Random Variables
- Lecture 18: Laws of Large Numbers
- Lecture 19: Central Limit Theorem and Characteristic Function

Elec2600H: Lecture 11

- **Pairs of continuous random variables**
  - **Joint cumulative distribution function**
  - Joint density function
- Pairs of continuous and discrete random variables

Two Random Variables

- One random variable can be considered as a mapping from the sample space to the real line.
- Two random variables can be considered as a mapping from the **sample space to the plane.**
- Note that given the outcome of the experiment, **both X and Y are determined simultaneously**.

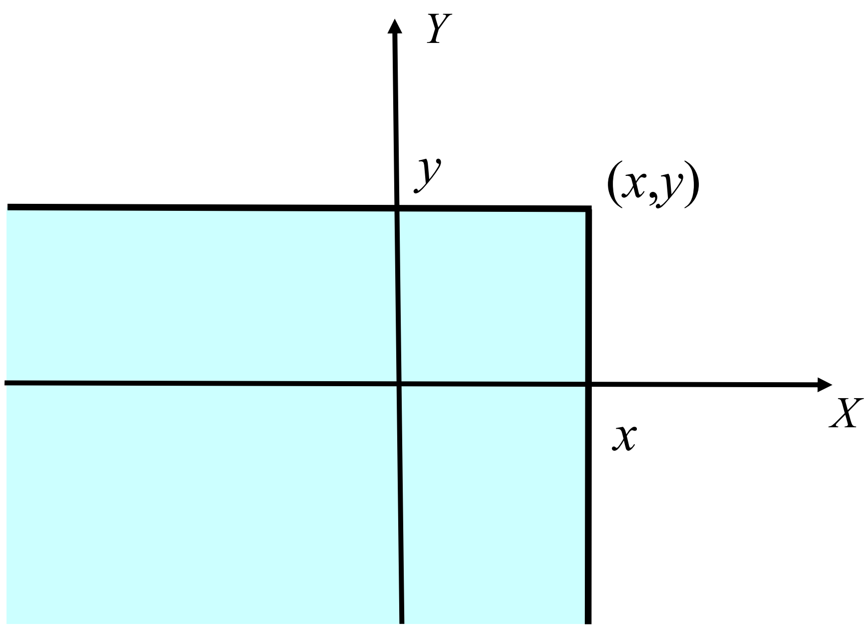


Joint Cumulative Distribution Function

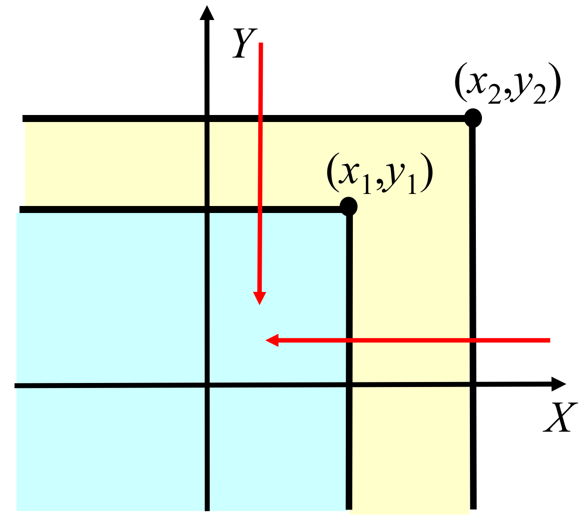
- The ***joint cumulative distribution function*** of two random variables X and Y, $F_{X,Y}(x,y)$ is defined as

$$\boxed{F_{X,Y}(x,y) = P\left[\{X \le x\} \cap \{Y \le y\}\right] \quad \text{where } x, y \in \Re}$$

- The subscript indicates the variables and the order they are referred to in the function.



Properties of the Joint CDF



- $F_{X,Y}(x_1, y_1) \le F_{X,Y}(x_2, y_2)$ if $x_1 \le x_2$ and $y_1 \le y_2$

- $F_{X,Y}(-\infty, -\infty) = 0$

  $F_{X,Y}(-\infty, y) = 0$

  $F_{X,Y}(x, -\infty) = 0$

  $F_{X,Y}(\infty, \infty) = 1$

- The joint cdf is continuous from the "north" and "east."

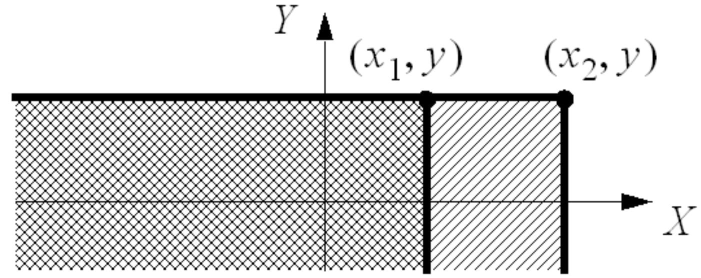
$$\lim_{x \to a^+} F_{X,Y}(x, y) = F_{X,Y}(a, y)$$

$$\lim_{y \to b^+} F_{X,Y}(x, y) = F_{X,Y}(x, b)$$

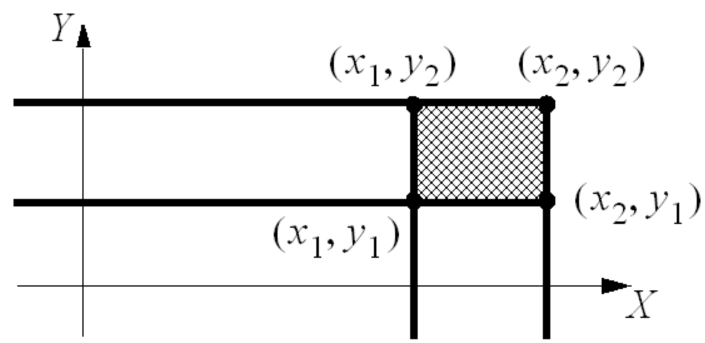
Properties of the Joint CDF

- $P\left[\{x_1 < X \le x_2\} \cap \{Y \le y\}\right] = F_{X,Y}\left(x_2, y\right) - F_{X,Y}\left(x_1, y\right)$

  $P\left[\{X \le x\} \cap \{y_1 < Y \le y_2\}\right] = F_{X,Y}\left(x, y_2\right) - F_{X,Y}\left(x, y_1\right)$



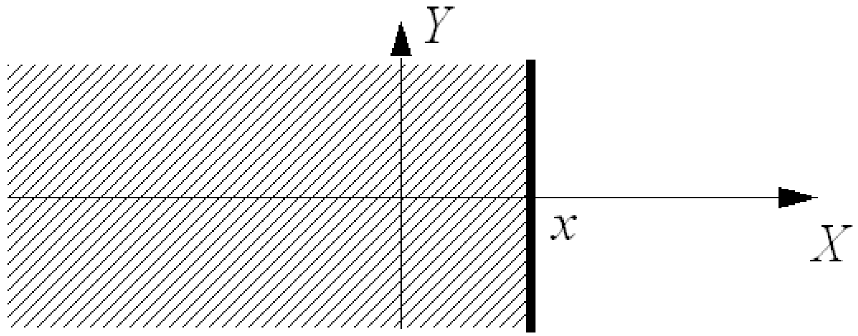
- $P\left[\{x_1 < X \le x_2\} \cap \{y_1 < Y \le y_2\}\right] = F_{X,Y}\left(x_2, y_2\right) - F_{X,Y}\left(x_2, y_1\right) - F_{X,Y}\left(x_1, y_2\right) + F_{X,Y}\left(x_1, y_1\right)$



Properties of the Joint CDF

- $F_X(x) = P[X \le x] = P[X \le x, Y \le \infty] = F_{X,Y}(x, \infty)$

  $F_Y(y) = F_{X,Y}(\infty, y)$



- Thus, the **marginal statistics can be derived from the joint** (but not in general vice versa).

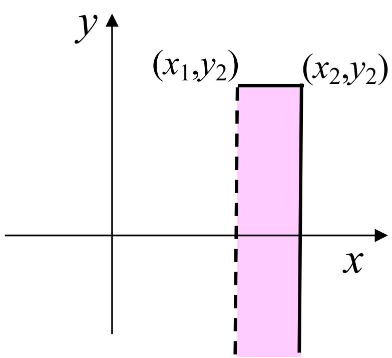
Product Form Events…

Product form events

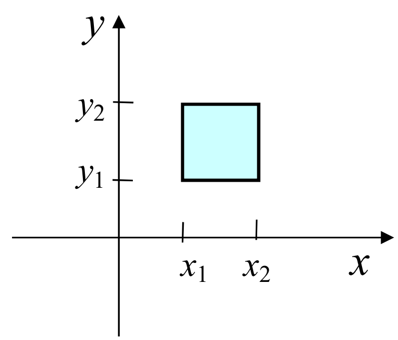
- A ***product form*** *event* is an event that is the intersection of events that involve only one random variable.

$$A = \{X \in A_1\} \bigcap \{Y \in A_2\}$$

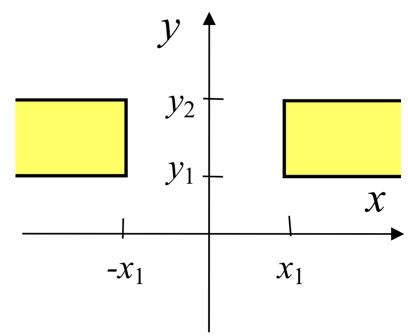
- Product form events have a rectangular shape.



$\{x_1 < X \le x_2\} \bigcap \{Y \le y_2\}$



$\{x_1 \le X \le x_2\} \bigcap \{y_1 \le Y \le y_2\}$

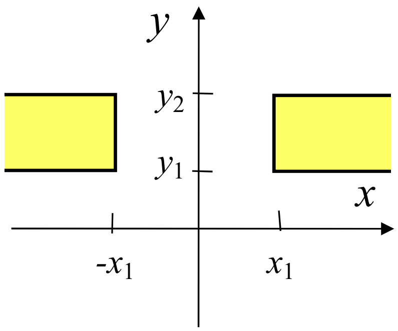


$\{|X| \ge x_1\} \bigcap \{y_1 < Y \le y_2\}$

- Their probability can be computed by adding/subtracting the cdf at different points.

Example 5.11

- Compute the probability of the product form event shown in terms of the cumulative distribution function. Assume that the cdf is continuous.



$$\{|X| > x_1\} \cap \{y_1 < Y \le y_2\}$$

Solution

$$P\left[\{|X| > x_1\} \cap \{y_1 < Y \le y_2\}\right] = \overbrace{P\left[\{y_1 < Y \le y_2\}\right]} - \overbrace{P\left[\{-x_1 < X \le x_1\} \cap \{y_1 < Y \le y_2\}\right]}$$

$$= \left(F_{X,Y}(\infty, y_2) - F_{X,Y}(\infty, y_1)\right) - \left(F_{X,Y}(x_1, y_2) - F_{X,Y}(x_1, y_1) - F_{X,Y}(-x_1, y_2) + F_{X,Y}(-x_1, y_1)\right)$$

Example 5.12

The joint cdf for the vector random variable **X**=(X, Y) is given by

$$F_{X,Y}(x,y) = \begin{cases} (1-e^{-\alpha x})(1-e^{-\beta y}) & x \geq 0, y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Find the marginal cdf's.

Solution – We let one argument go to infinity

$$F_X(x) = F_{X,Y}(x,\infty) = 1-e^{-\alpha x} \quad x \geq 0$$

$$F_Y(y) = F_{X,Y}(\infty,y) = 1-e^{-\beta y} \quad y \geq 0$$

Thus, X and Y individually have exponential distributions with parameters α and β.

Example 5.13

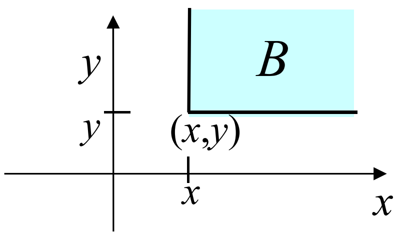
For the cdf in Example 5.12, find the probability of the events

$B = \{X > x,\ Y > y\}$ where $x > 0$ and $y > 0$, and

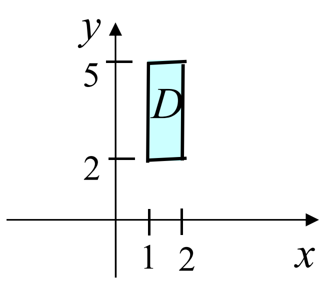
$D = \{1 < X \leq 2,\ 2 < Y \leq 5\}$

Solution:

$$
\begin{aligned}
P[B] &= 1 - \left(P[X \leq x] + P[Y \leq y] - P[X \leq x \text{ and } Y \leq y]\right) \\
&= 1 - \left(F_{X,Y}(x,\infty) + F_{X,Y}(\infty, y) - F_{X,Y}(x,y)\right) \\
&= 1 - ((1-e^{-\alpha x}) + (1-e^{-\beta y}) - (1-e^{-\alpha x})(1-e^{-\beta y})) \\
&= 1 - (1 - e^{-\alpha x}e^{-\beta y}) = e^{-\alpha x}e^{-\beta y}
\end{aligned}
$$

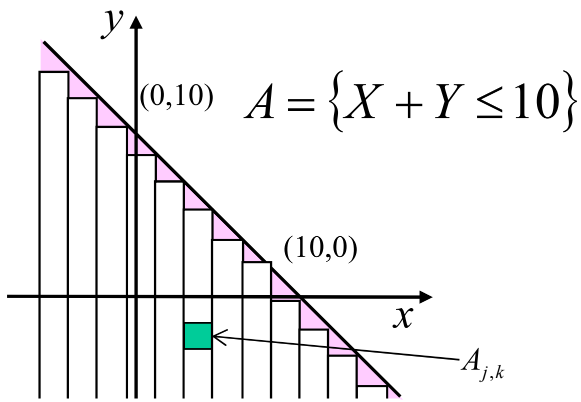


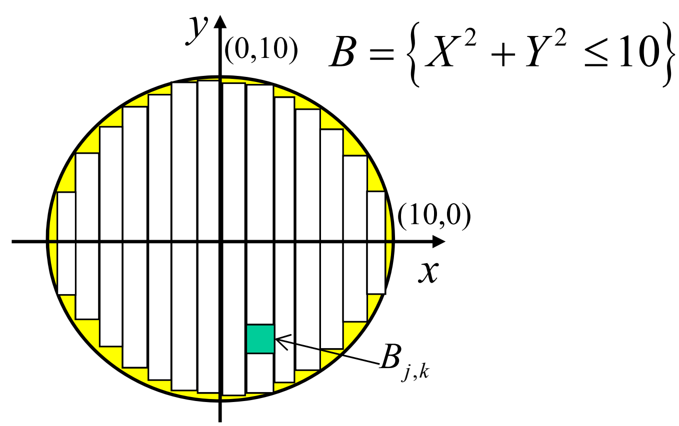
$$
\begin{aligned}
P[D] &= \left(F_{X,Y}(2,5) - F_{X,Y}(2,2) - F_{X,Y}(1,5) + F_{X,Y}(1,2)\right) \\
&= (1-e^{-2\alpha})(1-e^{-5\beta}) - (1-e^{-2\alpha})(1-e^{-2\beta}) \\
&\quad - (1-e^{-\alpha})(1-e^{-5\beta}) + (1-e^{-\alpha})(1-e^{-2\beta})
\end{aligned}
$$



Approximations by Product Form Events

- Many events cannot be expressed as the union of a finite number of product form events .
- However, they can be **approximated arbitrarily well** by rectangles of infinitesimal width.





- If the cdf is sufficiently smooth, probabilities can be computed by integrating a density function

$$P[A] = \sum_j \sum_k P[A_{j,k}] = \sum_j \sum_k f_{X,Y}(x_j, y_k)\Delta x \Delta y \to \int_y \int_x f_{X,Y}(x,y)dxdy$$

Elec2600H: Lecture 11

- **Pairs of continuous random variables**
  - Joint cumulative distribution function
  - **Joint density function**
- Pairs of continuous and discrete random variables

Jointly continuous random variables

- We say that two random variables are **jointly continuous** if the joint cumulative distribution function is continuous and differentiable.
- We define the **joint probability density function** to be the second derivative of the joint distribution:

$$\boxed{f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y}}$$

- The joint probability density function is also referred to as the **joint pdf** or the **joint density**.
- Note that the probability density function is not necessarily continuous.
- Just as in the case of a single random variable, the probability that a pair of continuous random variables achieves any particular combination of values is zero: $P\left[\{X = x\} \cap \{Y = y\}\right] = 0$

Properties of the joint probability density function

- The joint pdf is always non-negative: $f_{X,Y}(x,y) \geq 0 \quad \text{for all } x, y$

- The joint cdf can be computed by integrating the joint pdf: $F_{X,Y}(x,y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(\alpha,\beta)\, d\alpha\, d\beta$

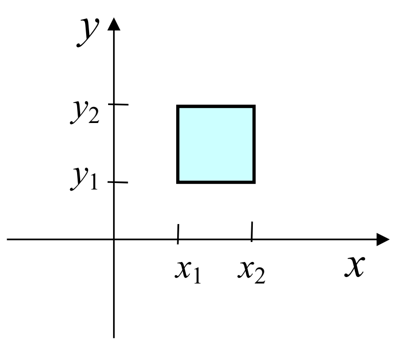
- The total area under the joint pdf is 1: $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx dy = 1$

- The probability of any event A is the integral of the joint pdf over the corresponding region in the X-Y plane:

$$P[(X,Y) \in A] = \iint_A f_{X,Y}(x,y)\, dx dy$$

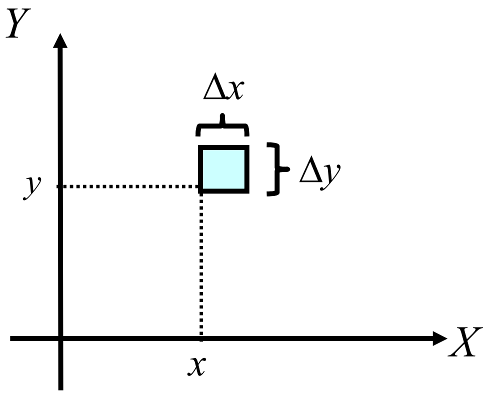
For example,

$$P[\{x_1 < X \leq x_2\} \cap \{y_1 < Y \leq y_2\}] = \int_{y_1}^{y_2} \int_{x_1}^{x_2} f_{X,Y}(x,y)\, dx dy$$



The joint pdf is **NOT** a probability

- Note that the value of the density function is not a probability.
- However, it is large in regions of the plane with high probability and small in regions of the plane with low probability.



$$P\left[\{x < X \le x + \Delta x\} \cap \{y < Y \le y + \Delta y\}\right] = \int_{y}^{y+\Delta y} \int_{x}^{x+\Delta x} f_{X,Y}(\alpha, \beta) d\alpha d\beta$$

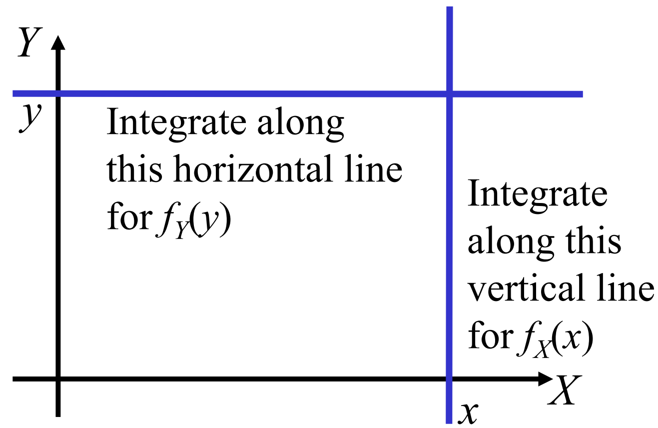
$$\approx f_{X,Y}(x, y) \Delta x \Delta y$$

- It is true that $f_{X,Y}(x,y) \ge 0$ for all x and y.
- However, it is not necessarily true that $f_{X,Y}(x,y) \le 1$.

Marginal densities

- The marginal densities $f_X(x)$ or $f_Y(y)$ can always be recovered from the joint density $f_{X,Y}(x,y)$:

$$
\begin{aligned}
f_X(x) &= \frac{dF_X(x)}{dx} = \frac{dF_{X,Y}(x,\infty)}{dx} \\
&= \frac{d}{dx}\int_{-\infty}^{\infty}\int_{-\infty}^{x} f_{X,Y}(\alpha,\beta)\,d\alpha\, d\beta \\
&= \int_{-\infty}^{\infty}\left\{\frac{d}{dx}\int_{-\infty}^{x} f_{X,Y}(\alpha,\beta)\,d\alpha\right\}d\beta \\
&= \int_{-\infty}^{\infty} f_{X,Y}(x,\beta)\,d\beta
\end{aligned}
$$

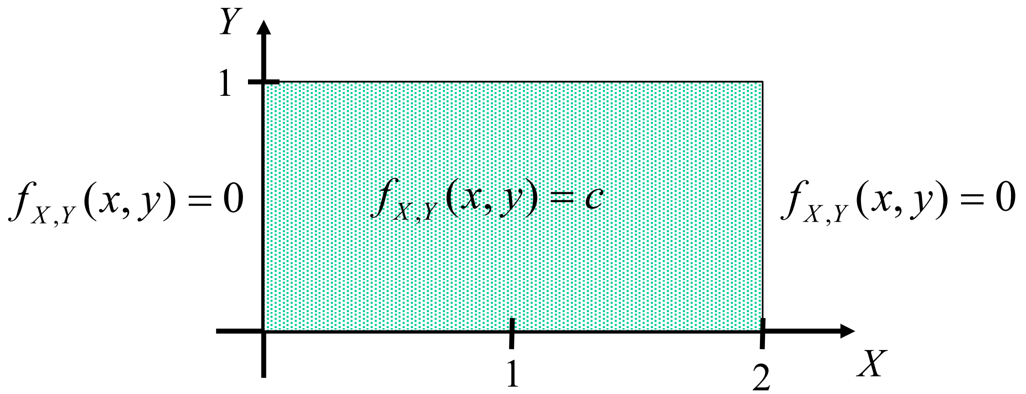


$$
f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(\alpha,y)\,d\alpha
$$

- It is **not** generally possible to determine the joint density from the marginal densities.

Example: Uniform Distribution

- Suppose that the joint pdf of X and Y is uniform (equal to a constant c) in the region shown, and zero otherwise.



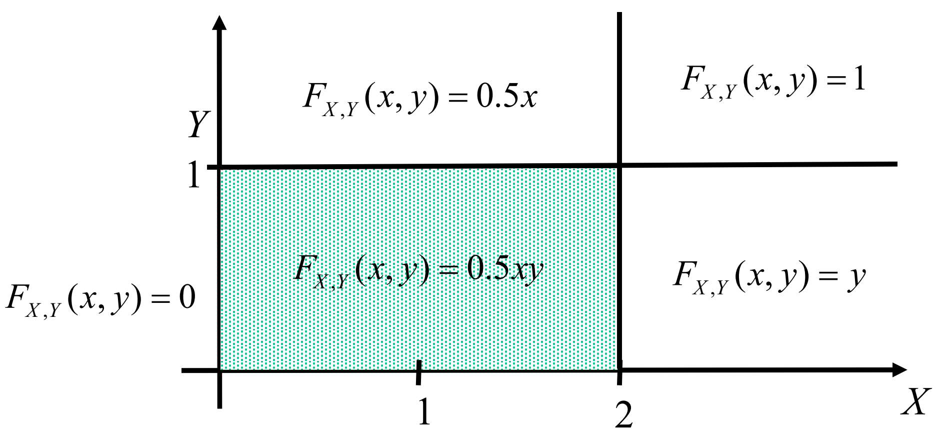
- Find
  - The constant c
  - The cdf
  - The marginal distributions of X and Y.
  - The following probabilities: $P[X+Y<0.5]$, $P[X+Y<1.5]$, $P[X+Y<2.5]$

Answers (1)

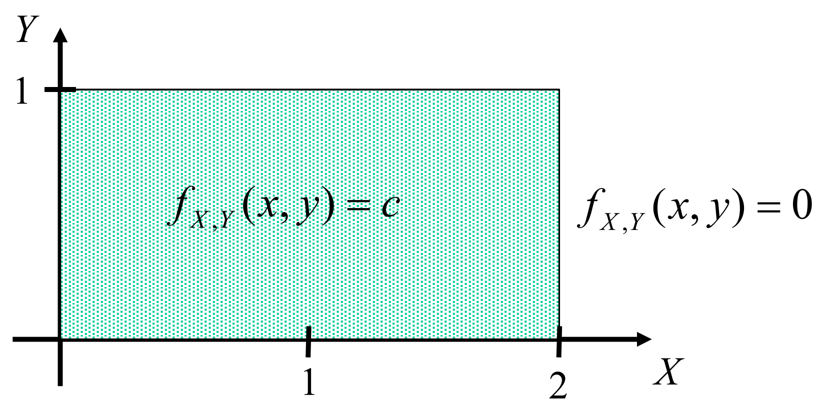
Note that for this problem, double integrals reduce to finding the area of the regions where the integrand is non-zero and multiplying by c.

The constant c = 0.5 is the inverse of the area (2).

Cumulative distribution function:



Marginal Densities:
X is uniform over [0,2]
Y is uniform over [0,1]

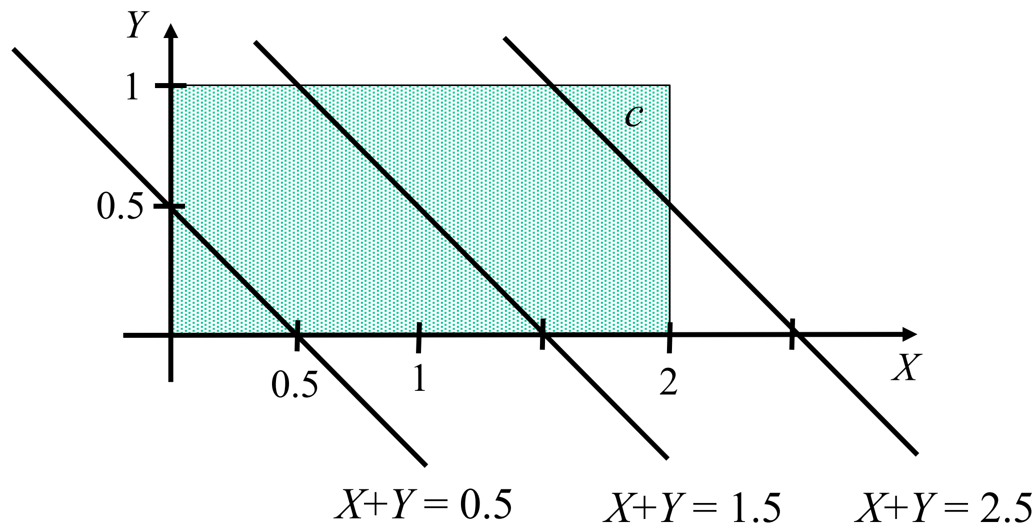


Answers (2)

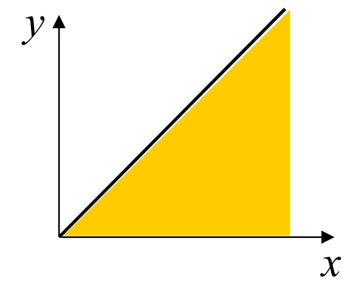
P[X+Y < 0.5] = 1/16

P[X+Y < 1.5] = 1/2

P[X+Y < 2.5] = 15/16



Example 5.16



Suppose that random variables X and Y are jointly distributed according to

$$f_{X,Y}(x,y) = \begin{cases} ce^{-x}e^{-y} & 0 \le y \le x \le \infty \\ 0 & \text{elsewhere.} \end{cases}$$

Find the <u>constant c</u> and the <u>marginal densities</u> of X and Y.

<u>Solution</u>:

Integrate along y first in yellow region, then along x.

$$1 = \int_0^\infty \int_0^x ce^{-x}e^{-y}\,dy\,dx = c\int_0^\infty e^{-x}\left(\int_0^x e^{-y}\,dy\right)dx \qquad \boxed{\int e^{-\lambda y}\,dy = -\frac{1}{\lambda}e^{-\lambda y}}$$

$$= c\int_0^\infty e^{-x}\left(-e^{-y}\right)\Big|_0^x dx = c\int_0^\infty e^{-x}\left(1 - e^{-x}\right)dx$$

$$= c\int_0^\infty \left(e^{-x} - e^{-2x}\right)dx = c\left(1 - \frac{1}{2}\right) = \frac{c}{2}$$

$$\boxed{\int_0^\infty e^{-\lambda x}\,dx = \frac{1}{\lambda}} \qquad \Rightarrow \qquad \boxed{c = 2}$$

Example 5.16 (marginal densities)

If $0 \le y \le \infty$,

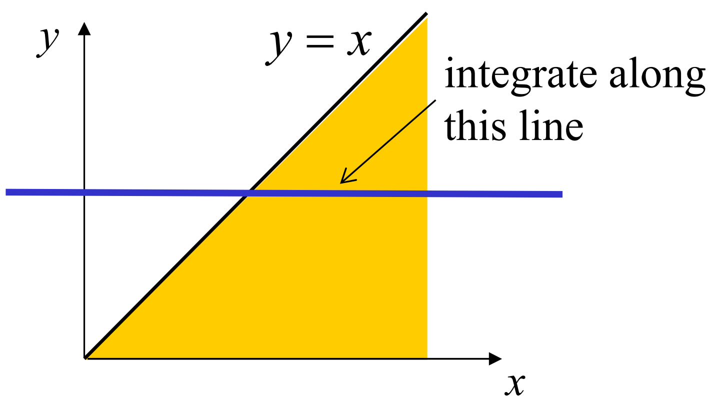
$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dx$$

$$= \int_{y}^{\infty} 2e^{-x}e^{-y}dx$$

$$= 2e^{-y}\int_{y}^{\infty} e^{-x}dx$$

$$= 2e^{-y}(e^{-y}) = 2e^{-2y}$$

Otherwise, $f_Y(y) = 0$.



If $0 \le x \le \infty$,

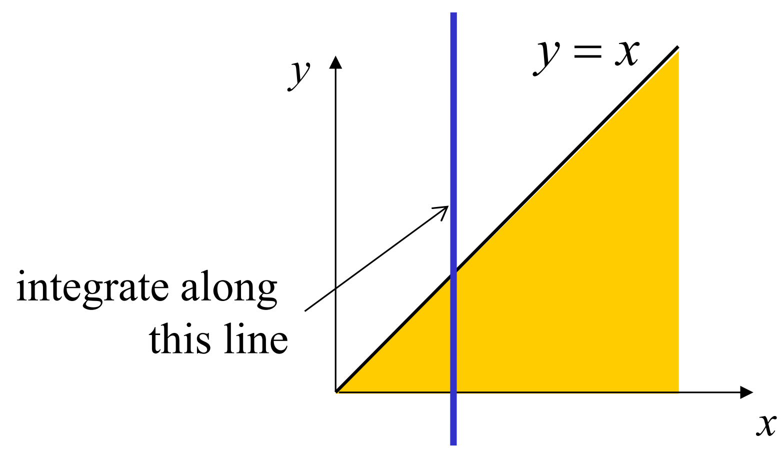
$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dy$$

$$= \int_{0}^{x} 2e^{-x}e^{-y}dy$$

$$= 2e^{-x}\int_{0}^{x} e^{-y}dy$$

$$= 2e^{-x}\left(1 - e^{-x}\right)$$

Otherwise, $f_X(x) = 0$.



Example 5.17

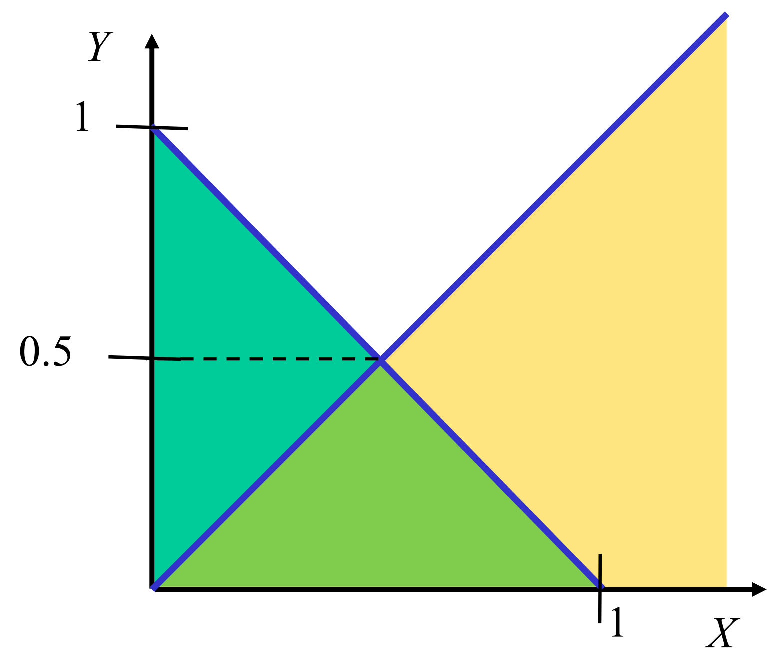
Find $P[X+Y \leq 1]$ in Example 5.16

Solution:

The probability is computed by integrating over the triangular region which is the intersection of the area where the pdf is non-zero (yellow), and the region of the plane where $X + Y \leq 1$ (green).

We integrate over x first, then y:

$$P[\{X+Y \leq 1\}] = \int_0^{0.5} \int_y^{1-y} 2e^{-x} e^{-y} \, dx dy$$
$$= 2\int_0^{0.5} e^{-y} \left(\int_y^{1-y} e^{-x} dx \right) dy$$
$$= 2\int_0^{0.5} e^{-y} \left(e^{-y} - e^{-(1-y)} \right) dy$$
$$= 1 - 2e^{-1}$$



Elec2600H: Lecture 11

- Pairs of continuous random variables
  - Review of 2D functions, differentiation and integration
  - Joint cumulative distribution function
  - Joint density function
- **Pairs of continuous and discrete random variables**

Random variables that **differ** in type

- In problems where it is necessary to deal with random variables of different types (e.g. X is discrete and Y is continuous), it is usually preferable to work with probabilities that have the form of a mixture of a pmf and a cdf.

$$P[\{X = k\} \cap \{Y \le y\}]$$

or

$$P[\{X = k\} \cap \{y_1 < Y \le y_2\}]$$

rather than the joint CDF.

- Events of this form are easier because the discreteness of X is explicit. However, we can always compute the joint CDF from events of this form if necessary.

Pairs of continuous and discrete random variables

- Pairs of discrete and continuous random variables are becoming increasingly important, with the widespread development and use of pattern recognition and pattern classification.
- Typically,
  - X is a discrete random variable, which assume one of N possible values (e.g. N classes)
  - Y is a continuous random variable corresponding to a measurement
- Examples
  - X = sex of person (male or female), Y = height of person
  - X = bit input to a communication channel, Y = voltage output of channel
  - X = whether or not it rains, Y = percentage of cloud cover
  - X = whether or not a flash is used to take a picture, Y = amount of light in the room
  - X = final course grade (A, B, C, D, F), Y = amount of time spent studying
- Typically, we are interested in assigning probabilities to outcomes in X based on measurements of Y (later)

Joint distributions (X discrete, Y continuous)

- Joint CDF: $F_{X,Y}(x, y) = P[X \leq x, Y \leq y]$

- PMF in X and CDF in Y: $p_{X,Y}(x, y) = P[X = x, Y \leq y]$

- PMF in X and PDF in Y: $f_{X,Y}(x, y) = \frac{d}{dy} P[X = x, Y \leq y]$

- Note that the symbols p and f can change, i.e. p might be used to indicate the PMF/PDF. There is no strict convention. You must read the definition of the symbols given in each context.

Example

- Let X be a Bernoulli random variable indicating the sex of person (0 = male, 1 = female) and Y be his/her height. Assume the joint pmf/pdf is

$$f_{X,Y}(x,y) = \begin{cases} 0.6\mathcal{N}(y|175{,}64) & \text{if } x = 0 \\ 0.4\mathcal{N}(y|160{,}40) & \text{if } x = 1 \end{cases}$$

where $\mathcal{N}(y|m,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-m)^2}{2\sigma^2}}$

- Find the marginal pmf of X.
- **Solution:**

$$P[X = x] = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$$

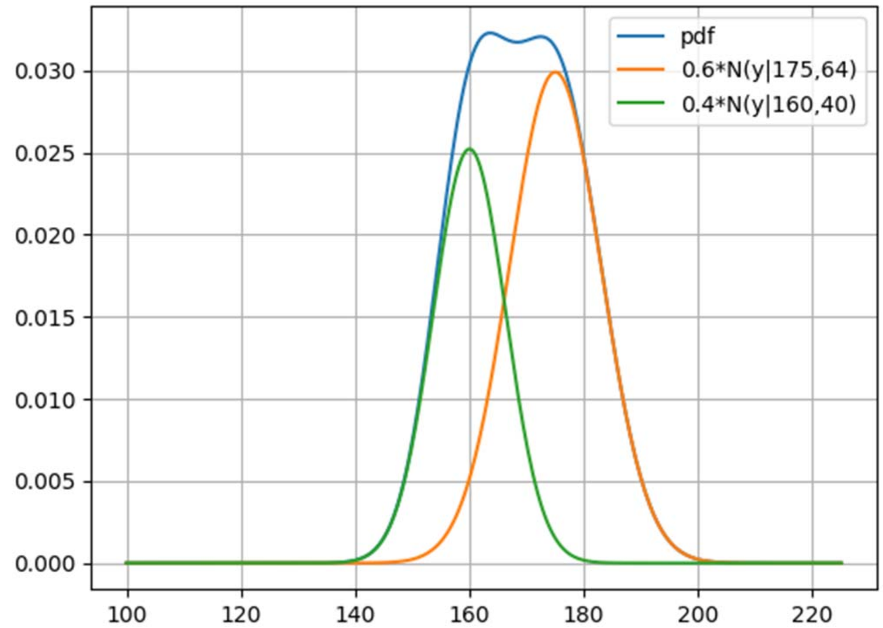
$$= \begin{cases} 0.6 & \text{if } x = 0 \\ 0.4 & \text{if } x = 1 \end{cases} \qquad \text{since } \int_{-\infty}^{\infty} \mathcal{N}(y|m,\sigma^2)dy \ = 1$$

Example

- Let X be a Bernoulli random variable indicating the sex of person (0 = male, 1 = female) and Y be his/her height. Assume the joint pmf/pdf is

$$f_{X,Y}(x,y) = \begin{cases} 0.6\mathcal{N}(y|175{,}64) & \text{if } x = 0 \\ 0.4\mathcal{N}(y|160{,}40) & \text{if } x = 1 \end{cases}$$

- Find the marginal pdf of Y.
- **Solution:**



$$f_Y(Y) = \sum_{x=0}^{1} f_{X,Y}(x,y)$$
$$= 0.6\mathcal{N}(y|175{,}64) + 0.4\mathcal{N}(y|160{,}40)$$

Major Points from this Lecture:

- **Joint CDF and PDF of pairs of random variables**
- **Product form events**
- **Marginal densities (and how to compute them)**
- **Combinations of discrete and continuous random variables**

ELEC 2600H: Probability and Random Processes in Engineering

Part III: Multiple Random Variables

- Lecture 10: Pairs of Discrete Random Variable
  - Out-of-Class Reading: 2D Calculus
- Lecture 11: Pairs of Continuous Random Variable
- **Lecture 12: Independence, Joint Moments**
- Lecture 13: Correlation Coefficients and Their Properties
- Lecture 14: Conditional PDF, Conditional Expectation
- Lecture 15: Sum of Two Random Variables
- Lecture 16: Pairs of Jointly Gaussian Random Variables
- Lecture 17: More than Two Random Variables
- Lecture 18: Laws of Large Numbers
- Lecture 19: Central Limit Theorem and Characteristic Function

Elec2600H: Lecture 12

- **Independence**
- Moments and Central Moments

Independence

- **Definition:** Two random variables X and Y are said to be *independent* or *statistically independent* if for any events, A_X and A_Y, defined in terms of X and Y respectively,

$$P[\{X \in A_X\} \cap \{Y \in A_Y\}] = P[X \in A_X]P[Y \in A_Y]$$

- The probability of any product form event can be expressed as the product of the probabilities of the events in X and Y.

- Intuitively, when two random variables are independent, they do *not influence each other*.
  - In other words, knowing one does not tell us anything about the other one.

Independent Discrete Random Variables

- Suppose that X and Y are a pair of independent discrete random variables with joint and marginal probability mass functions given by

$$p_{X,Y}(j,k) = P[X = j, Y = k]$$
$$p_X(j) = P[X = j]$$
$$p_Y(k) = P[Y = k]$$

- Since the random variables are independent, for any j and k,

$$\begin{aligned} p_{X,Y}(j,k) &= P\left[\{X = j\} \cap \{Y = k\}\right] \\ &= P\left[\{X = j\}\right] P\left[\{Y = k\}\right] \\ &= p_X(j) p_Y(k) \end{aligned}$$

- Thus, **if X and Y are independent, then the joint pmf is the product of the marginal pmfs.**
- **Question: Is the converse also true?**

Independent Discrete Random Variables

- Suppose that X and Y are discrete random variables that satisfy $p_{X,Y}(j,k) = p_X(j)p_Y(k)$ for all j,k

- Let A_X and A_Y be events defined in terms of only X and Y respectively

- Consider the product form event $A = A_X \cap A_Y$

$$
\begin{aligned}
P[A] &= \sum_{j \in A_X} \sum_{k \in A_Y} p_{X,Y}(j,k) \\
&= \sum_{j \in A_X} \sum_{k \in A_Y} p_X(j) p_Y(k) \\
&= \sum_{j \in A_X} p_X(j) \sum_{k \in A_Y} p_Y(k) = P[A_X]P[A_Y]
\end{aligned}
$$

- Thus, X and Y are independent.
- Combined with previous result**:**

  A pair of discrete RVs are independent if and only if the joint pmf is the product of the marginal pmfs!

Conditional Probability and Independence

- If X and Y are *independent*, then the conditional probability is the same as the marginal probability.

$$p_{Y|X}(k \mid j) = \frac{p_{X,Y}(j,k)}{p_X(j)} = \frac{p_X(j)p_Y(k)}{p_X(j)} = p_Y(k)$$

$$p_{X|Y}(j \mid k) = \frac{p_{X,Y}(j,k)}{p_Y(k)} = \frac{p_X(j)p_Y(k)}{p_Y(k)} = p_X(j)$$

- In other words, knowing X does not tell us anything about Y, and vice versa.

Example 5.19

- Consider the outcomes (X,Y) of the toss of a pair of loaded dice with joint pmf shown.

$p_{X,Y}(j,k)$						
k = 6	1/42	1/42	1/42	1/42	1/42	**2/42**
5	1/42	1/42	1/42	1/42	**2/42**	1/42
4	1/42	1/42	1/42	**2/42**	1/42	1/42
3	1/42	1/42	**2/42**	1/42	1/42	1/42
2	1/42	**2/42**	1/42	1/42	1/42	1/42
1	**2/42**	1/42	1/42	1/42	1/42	1/42
j	1	2	3	4	5	6

- We have already seen that the marginal pmfs are: $p_X(j) = \frac{1}{6}$ for all $j \in \{1,\ldots,6\}$

$p_Y(k) = \frac{1}{6}$ for all $k \in \{1,\ldots,6\}$

- The outcomes are not independent, since if they were, all of the joint probabilities in the table would satisfy

$$p_{X,Y}(j,k) = p_X(j)p_Y(k) = \tfrac{1}{36} \quad \text{for all } j \in \{1,\ldots,6\} \text{ and } k \in \{1,\ldots,6\}$$

Example 5.20

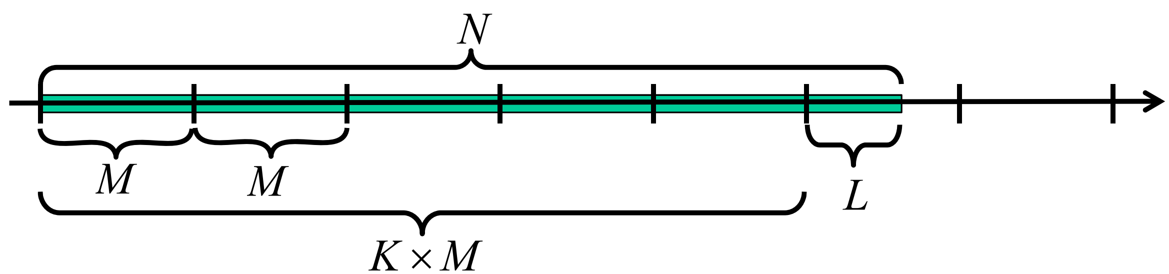
The number of bytes N in a message has a geometric distribution:

Suppose that messages are broken into packets of maximum length M bytes.

Let K be the number of full packets

L be the number of bytes left over

Are K and L independent?



Solution

We have already seen that $p_{K,L}(k,l) = (1-q)q^{kM+l}$ $\quad p_K(k) = (1-q^M)(q^M)^k$ $\quad p_L(l) = \frac{1-q}{1-q^M}q^l$

for $k \in \{0,1,2,\ldots\}$ and $l \in \{0,1,\ldots,M\text{-}1\}$.

The random variables K and L are independent, since

$$p_K(k) \times p_L(l) = (1-q^M)(q^M)^k \times \frac{1-q}{1-q^M}q^l = (1-q)q^{kM+l} = p_{K,L}(k,l)$$

Theorem (continuous random variables)

The following three statements are equivalent:

1. X and Y are independent.
2. $F_{X,Y}(x,y) = F_X(x)F_Y(y)$ for all x and y.
3. $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ for all x and y.

Proof:

(1) → (2) Since X and Y are independent, for any A_X and A_Y, $P\left[\{X \in A_X\} \cap \{Y \in A_Y\}\right] = P\left[X \in A_X\right] \cdot P\left[Y \in A_Y\right]$

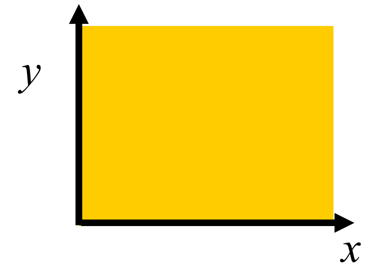
Set $A_X = \{X \leq x\}$ and $A_Y = \{Y \leq y\}$. Thus, $F_{X,Y}(x,y) = F_X(x)F_Y(y)$

(2) → (3) Differentiating, we obtain $f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y} = \frac{\partial F_X(x)}{\partial x} \times \frac{\partial F_Y(y)}{\partial y} = f_X(x)f_Y(y)$

(3) → (1) For any A_X and A_Y, $P\left[\{X \in A_X\} \cap \{Y \in A_Y\}\right] = \int_{A_Y}\int_{A_X} f_{X,Y}(x,y)\,dxdy = \int_{A_X} f_X(x)\,dx \times \int_{A_Y} f_Y(y)\,dy$

$$= P\left[\{X \in A_X\}\right] \cdot P\left[\{Y \in A_Y\}\right]$$

Example



Suppose that random variables X and Y are jointly distributed according to

$$f_{X,Y}(x,y) = \begin{cases} \alpha\beta e^{-\alpha x} e^{-\beta y} & 0 \le x \text{ and } 0 \le y \\ 0 & \text{otherwise} \end{cases}$$

Are X and Y independent?

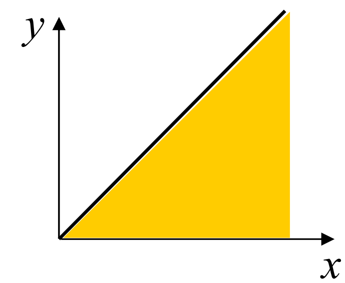
Solution:

The marginal density of X is $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy = \begin{cases} \alpha e^{-\alpha x} \int_0^{\infty} \beta e^{-\beta y} dy = \alpha e^{-\alpha x} & \text{if } 0 \le x \\ 0 & \text{otherwise} \end{cases}$

Similarly, $f_Y(y) = \begin{cases} \beta e^{-\beta y} & 0 \le y \\ 0 & \text{otherwise} \end{cases}$

Thus, the random variables are **independent**, since $f_X(x) f_Y(y) = \begin{cases} \alpha\beta e^{-\alpha x} e^{-\beta x} & 0 \le x \text{ and } 0 \le y \\ 0 & \text{otherwise} \end{cases}$

Example 5.12



Suppose that random variables X and Y are jointly distributed according to

$$f_{X,Y}(x,y)=\begin{cases} 2e^{-x}e^{-y} & 0 \le y \le x \le \infty \\ 0 & \text{elsewhere.} \end{cases}$$

Are X and Y independent?

Solution:

We have already seen (Lecture 14) that the marginal densities are

$$f_X(x)=\begin{cases} 2e^{-x}\left(1-e^{-x}\right) & 0 \le x \le \infty \\ 0 & \text{otherwise} \end{cases} \qquad f_Y(y)=\begin{cases} 2e^{-2y} & 0 \le \text{y} < \infty \\ 0 & \text{otherwise} \end{cases}$$

The random variables are **dependent** (not independent) since

$$f_X(x)f_Y(y)=\begin{cases} 4e^{-2y}e^{-x}\left(1-e^{-x}\right) & 0 \le x, y \le \infty \\ 0 & \text{otherwise} \end{cases}$$

Elec2600H: Lecture 12

- Independence
- **Moments and Central Moments**

Moments and Central Moments

Joint moment: $E[X^j Y^k] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^j y^k f(x,y)\,dx\,dy$ (the jkth moment)

Central moment: $E[(X - E[X])^j (Y - E[Y])^k]$

The first and second order moments are most important

$$\text{means} \begin{cases} E[X] \\ E[Y] \end{cases}$$

$$\text{variances} \begin{cases} E[(X - E[X])^2] = \mathrm{VAR}[X] \\ E[(Y - E[Y])^2] = \mathrm{VAR}[Y] \end{cases}$$

$$\text{covariance} \quad E[(X - E[X])(Y - E[Y])] = \mathrm{COV}(X, Y)$$

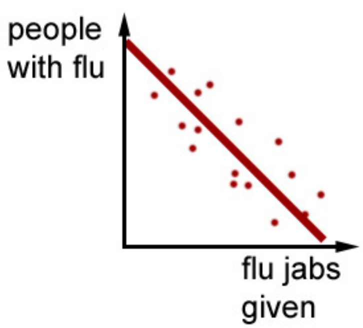
$$\text{correlation} \quad E[XY]$$

Interpretation of Covariance

- The covariance indicates how X and Y **vary together**. $\mathrm{COV}(X,Y) = E[(X - E[X])(Y - E[Y])]$

- When the covariance is *positive*:
  - If X is greater than its mean, Y is also usually greater than its mean.
- When the covariance is *negative:*
  - If X is greater than its mean, then it is likely that Y is less than its mean.

POSITIVE CORRELATION
- people who do more revision get higher exam results.
- revising increases success.



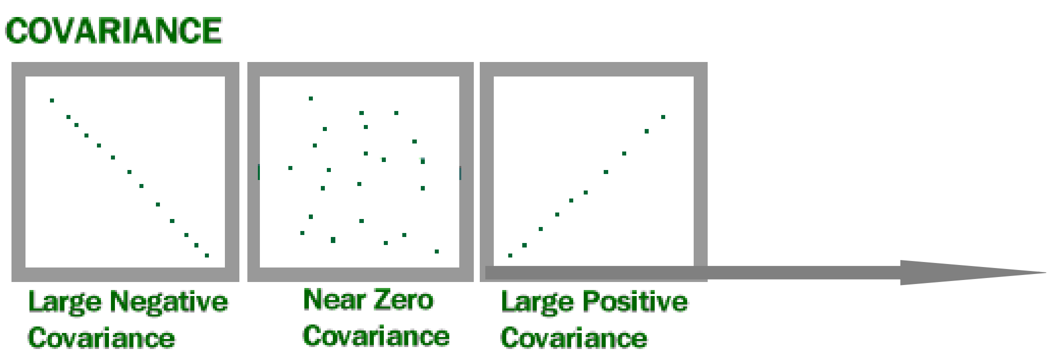
NEGATIVE CORRELATION
- when more jabs are given the number of people with flu falls.
- flu jabs prevent flu.

Interpretation of Covariance

- However, the magnitude of the covariance is not a good measure of how much the two random variables depend upon each other, since I can make the covariance arbitrarily large simply by scaling one of the RVs:

$$\mathrm{COV}[X,Y] = E[XY] - E[X]\cdot E[Y]$$

$$\mathrm{COV}(aX,Y) = a\mathrm{COV}(X,Y)$$



Computing the Covariance

The following formula is often useful in computing the covariance:

$$\mathrm{COV}[X,Y] = E[XY] - E[X]\cdot E[Y]$$

Proof:

$$\begin{aligned}
\mathrm{COV}[X,Y] &= E[(X - E[X])(Y - E[Y])] \\
&= E\big[XY - X\cdot E[Y] - E[X]\cdot Y + E[X]\cdot E[Y]\big] \\
&= E[XY] - E[X]\cdot E[Y] - E[X]\cdot E[Y] + E[X]\cdot E[Y] \\
&= E[XY] - E[X]\cdot E[Y]
\end{aligned}$$

If either X or Y is zero mean, then the covariance and ***correlation*** are identical:

$$\mathrm{COV}[X,Y] = E[XY]$$

Uncorrelated Random Variables

Definition: X and Y are *uncorrelated* if $\mathrm{COV}[X,Y]=0$.

Theorem: X and Y are uncorrelated if and only if $E[XY]=E[X]E[Y]$

Proof

From the previous page, $\mathrm{COV}[X,Y]=E[XY]-E[X]\cdot E[Y]$

Thus, $\mathrm{COV}[X,Y]=0 \Leftrightarrow E[XY]=E[X]\cdot E[Y]$

Uncorrelatedness and Independence

Theorem: If X and Y are independent, then they are uncorrelated.

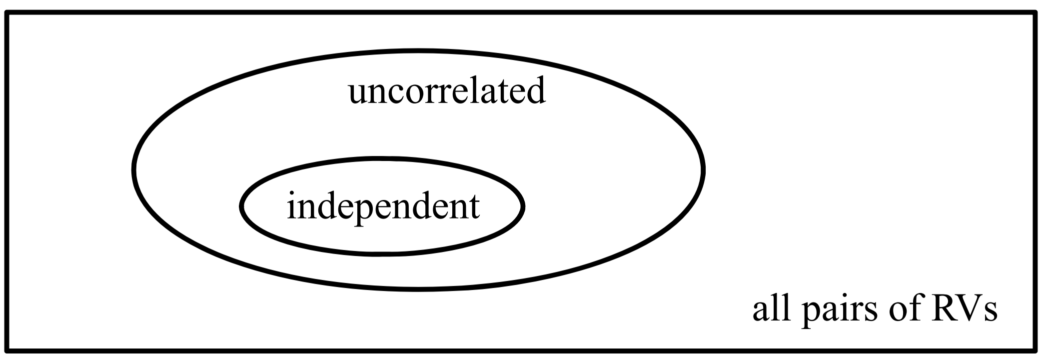
Proof:

Let $g_1(X)=X$ and $g_2(Y)=Y$ in Example 5.25.

Since X and Y are independent, $E[XY]=E[X]E[Y]$.

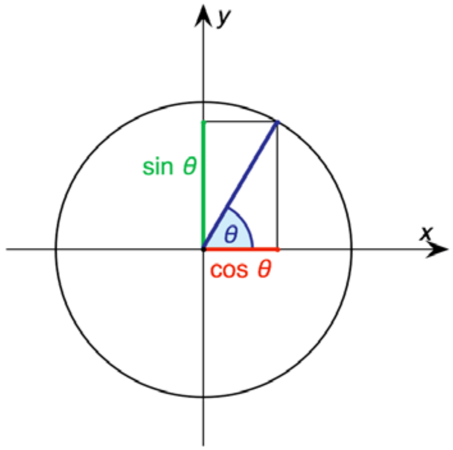
Thus, by the result on the previous page, X and Y are uncorrelated.

- Uncorrelatedness is a weak version of independence:
  - Two independent random variables are uncorrelated.
  - However, uncorrelated random variables are not necessarily independent.



Example 5.27: Uncorrelated but dependent RVs

Let $X = \cos(\theta)$ and $Y = \sin(\theta)$, where θ is uniform on $[0, 2\pi]$
Are X and Y uncorrelated? Are they independent?



Solution

X and Y are uncorrelated, since

$$\left.\begin{aligned} E[X] &= \frac{1}{2\pi}\int_0^{2\pi}\cos\theta \cdot d\theta = 0 \\ E[Y] &= \frac{1}{2\pi}\int_0^{2\pi}\sin\theta \cdot d\theta = 0 \\ E[XY] &= \frac{1}{2\pi}\int_0^{2\pi}\cos\theta\sin\theta \cdot d\theta \\ &= \frac{1}{2\pi}\int_0^{2\pi}\frac{\sin 2\theta}{2} \cdot d\theta = 0 \end{aligned}\right\} E[XY] = E[X]\cdot E[Y]$$

However, they are not independent. Given X, Y can assume only two possible values: $Y = \pm\sqrt{1 - X^2}$

Magnitude of covariance depends on measurement units

- The actual value of the covariance does not directly tell us how much X depends upon Y, since its value depends upon how much X and Y vary. In particular, the covariance changes with the units that we use to measure X and Y.
- For example, consider a student chosen at random.
  - Let H be his/her height in meters
  - Let W be his/her weight in kg
  - Let $1000W$ be his/her weight in grams.

$$
\begin{aligned}
\mathrm{COV}(H, 1000W) &= E[(H - E[H])(1000W - E[1000W])] \\
&= E[(H - E[H])(1000W - 1000E[W])] \\
&= E[(H - E[H])1000(W - E[W])] \\
&= 1000E[(H - E[H])(W - E[W])] \\
&= 1000 \times \mathrm{COV}(H, W)
\end{aligned}
$$

  - The covariance between height and weight in grams is 1000 times the covariance between height and weight in kg.
- More generally, for any a and b,
  - $\mathrm{COV}(aH, \mathrm{W}) = a\mathrm{COV}(H, W)$
  - $\mathrm{COV}(H, b\mathrm{W}) = b\mathrm{COV}(H, W)$
  - $\mathrm{COV}(aH, b\mathrm{W}) = ab\mathrm{COV}(H, W)$

Major Points from this Lecture:

- Independence
- Joint moments

ELEC 2600H: Probability and Random Processes in Engineering

Part III: Multiple Random Variables

- Lecture 10: Pairs of Discrete Random Variable
  - Out-of-Class Reading: 2D Calculus
- Lecture 11: Pairs of Continuous Random Variable
- Lecture 12: Independence, Joint Moments
- **Lecture 13: Correlation Coefficients and Their Properties**
- Lecture 14: Conditional PDF, Conditional Expectation
- Lecture 15: Sum of Two Random Variables
- Lecture 16: Pairs of Jointly Gaussian Random Variables
- Lecture 17: More than Two Random Variables
- Lecture 18: Laws of Large Numbers
- Lecture 19: Central Limit Theorem and Characteristic Function

Correlation Coefficient

- The **correlation coefficient** gives us a measure of how much X depends on Y (or Y depends on X) that is *independent of the units* used to measure them

Definition: The ***correlation coefficient*** between X and Y is

$$\rho_{X,Y} = \frac{\mathrm{COV}[X,Y]}{\sigma_X \sigma_Y} = \rho_{Y,X} \qquad \begin{aligned} \sigma_X{}^2 &= \mathrm{VAR}(X) \\ \sigma_Y{}^2 &= \mathrm{VAR}(Y) \end{aligned}$$

Lemma

The magnitude of the correlation coefficient is ≤ 1.

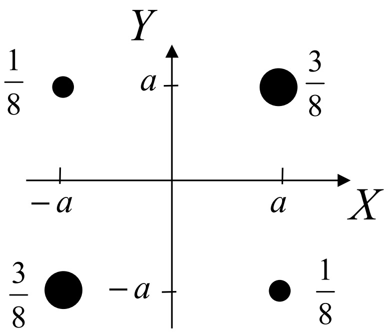
Proof:

$$
\begin{aligned}
0 &\leq E\left[\left(\frac{X-E[X]}{\sigma_X} \pm \frac{Y-E[Y]}{\sigma_Y}\right)^2\right] \\
&= E\left[\left(\frac{X-E[X]}{\sigma_X}\right)^2 \pm 2\left(\frac{X-E[X]}{\sigma_X}\right)\left(\frac{Y-E[Y]}{\sigma_Y}\right) + \left(\frac{Y-E[Y]}{\sigma_Y}\right)^2\right] \\
&= \frac{\mathrm{VAR}(X)}{\sigma_X{}^2} \pm 2\frac{\mathrm{COV}(\mathrm{X},\mathrm{Y})}{\sigma_X\sigma_Y} + \frac{\mathrm{VAR}(Y)}{\sigma_Y{}^2} \\
&= 2 \pm 2\rho_{X,Y}
\end{aligned}
$$

Thus, $1 \pm \rho_{X,Y} \geq 0 \quad \Longleftrightarrow \quad \left|\rho_{X,Y}\right| \leq 1$

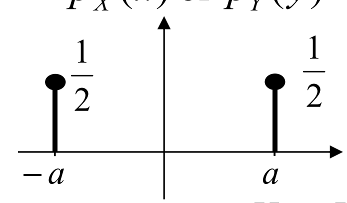
Example

Suppose that X and Y are discrete random variables that can be either $+a$ or $-a$, with joint pmf shown. Find ρ_{XY}.



Solution

By summing horizontally or vertically we see that the marginal pmfs assign equal probability $(1/2)$ to $+a$ and $-a$.



Thus, $\mathrm{E}[X] = 0$ and $\mathrm{Var}[X] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2 = \mathrm{E}[X^2] = \frac{1}{2}(-a)^2 + \frac{1}{2}(a)^2 = a^2$

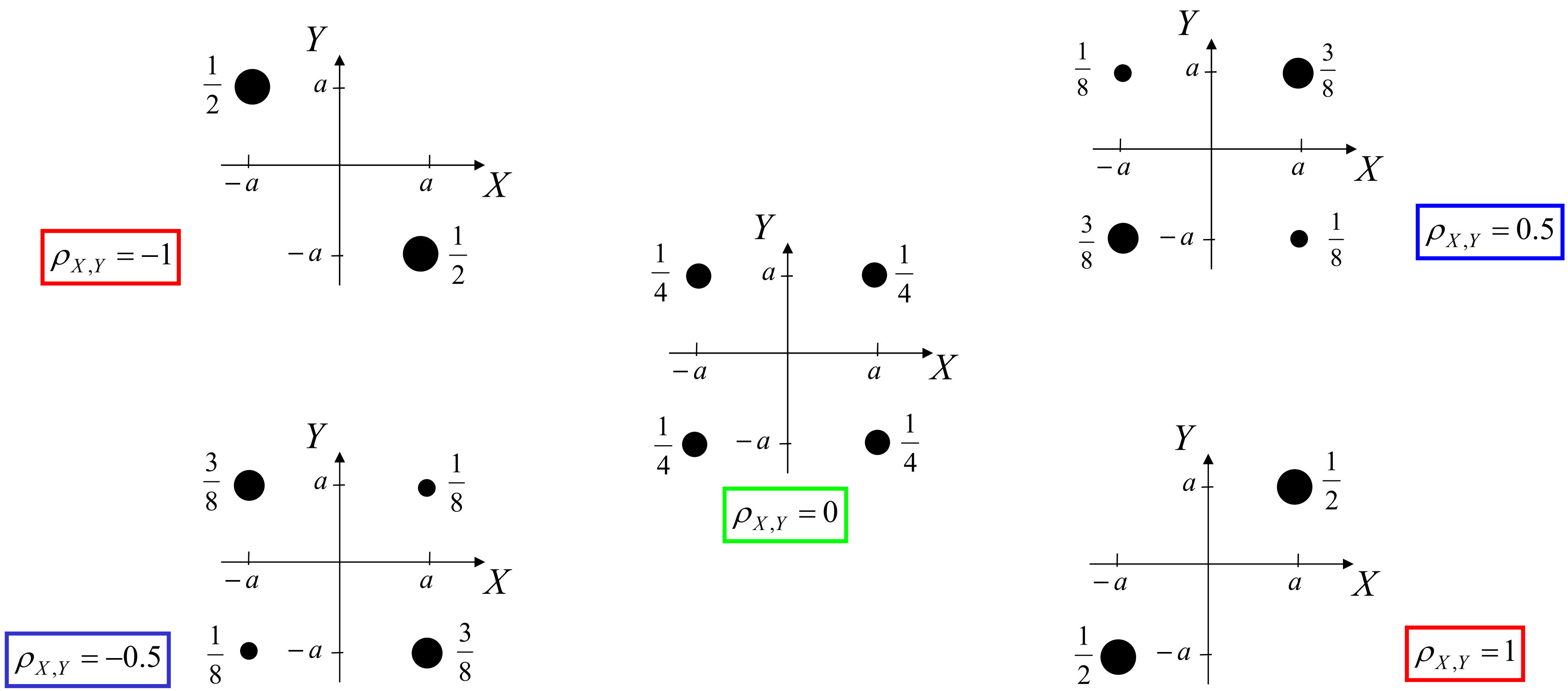
Similarly, $\mathrm{E}[Y] = 0$ and $\mathrm{Var}[Y] = a^2$.

$$
\begin{aligned}
\mathrm{COV}[X,Y] &= \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y] = \mathrm{E}[XY] \\
&= \tfrac{1}{8}(-a)(a) + \tfrac{3}{8}(a)(a) + \tfrac{3}{8}(-a)(-a) + \tfrac{1}{8}(a)(-a) \\
&= \tfrac{4}{8}a^2 = \tfrac{1}{2}a^2
\end{aligned}
$$

$$
\text{Thus, } \rho_{X,Y} = \frac{\mathrm{COV}[X,Y]}{\sqrt{\mathrm{Var}[X]}\sqrt{\mathrm{Var}[Y]}} = \frac{\frac{1}{2}a^2}{\sqrt{a^2}\sqrt{a^2}} = \frac{1}{2}
$$

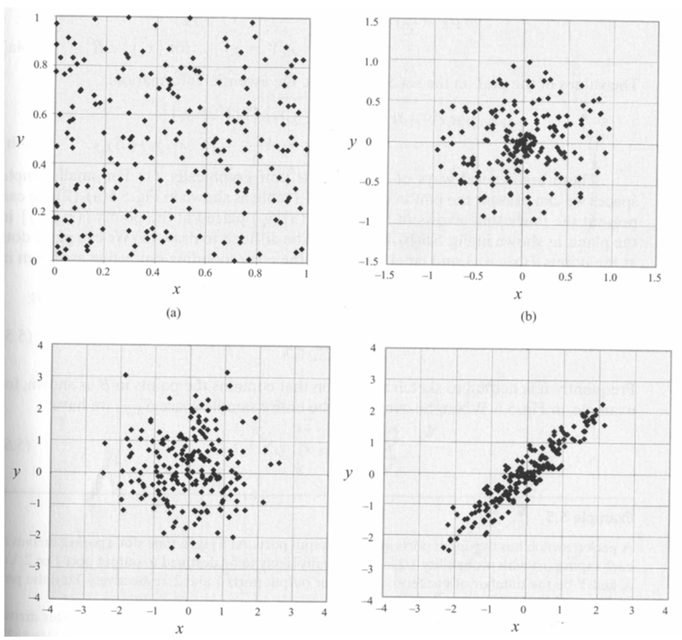
Example (cont)

For the joint pmf's below, X and Y have the same marginals, but different $\rho_{X,Y}$.

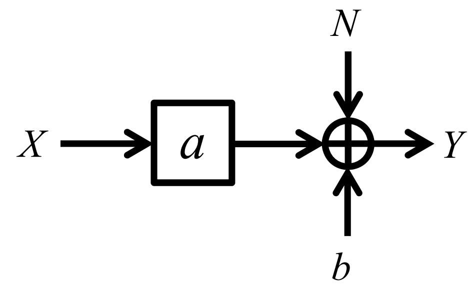


Interpretation of correlation coefficient

- When the correlation coefficient is *positive*,
  - If X is greater than its mean, then it is likely that Y is also greater than its mean.
- When the correlation coefficient is *negative*
  - If X is greater than its mean, then it is likely that Y is less than its mean.
- The closer the absolute value of the correlation is to one, the stronger the relationship between X and Y.



Example: A communication system



Let X be a RV with mean 0 and variance σ_X^2.

Let $Y = aX + b + N$ be the output of a communication channel, where a is a constant gain parameter, b is a constant offset and N is random noise with mean 0 and standard deviation σ_N^2. Assume X and N are independent.

Find $\rho_{X,Y}$.

Solution: First note that $E[Y] = b$ and $\sigma_Y^2 = a^2\sigma_X^2 + \sigma_N^2$ (next page)

The fraction of the standard deviation of Y due to aX.

$$\rho_{X,Y} = \frac{COV(X,Y)}{\sigma_X\sigma_Y} = \frac{E[(X - E[X])(aX + b + N - E[aX + b + N])]}{\sigma_X\sigma_Y}$$

$$= \frac{E[X(aX + N)]}{\sigma_X\sigma_Y} = \frac{E[aX^2 + XN]}{\sigma_X\sigma_Y} = \frac{aE[X^2] + E[X]E[N]}{\sigma_X\sigma_Y} = \frac{a\sigma_X{}^2}{\sigma_X\sigma_Y} = \frac{a\sigma_X}{\sigma_Y} = \frac{a\sigma_X}{\sqrt{a^2\sigma_X^2 + \sigma_N^2}}$$

- Note that if $\sigma_N^2 = 0$, then $\rho_{X,Y} = \pm 1$ depending upon the sign of a.
- The correlation coefficient $\rho_{X,Y}$ (also called "Pearson" correlation coefficient) can be seen as giving a measure of how linearly related X and Y are. It may be seen as a statistical measure describing how well Y may be predicted as a linear function of X.

Variance of Sum

Let $Y = X + N$, where X and N are random variables.
The variance of Y is

$$\mathrm{VAR}\left[Y\right] = \mathrm{VAR}\left[X\right] + 2\mathrm{COV}(X,N) + \mathrm{VAR}\left[N\right]$$

Proof

$$\begin{aligned}
\mathrm{VAR}\left[Y\right] &= E[(Y - E[Y])^2] \\
&= E[(X + N - E[X] - E[N])^2] \\
&= E[\left((X - E[X]) + (N - E[N])\right)^2] \\
&= E[(X - E[X])^2] + 2E[(X - E[X])(N - E[N])] + E[(N - E[N])^2] \\
&= \mathrm{VAR}\left[X\right] + 2\mathrm{COV}(X,N) + \mathrm{VAR}\left[N\right]
\end{aligned}$$

Note: If X and N are uncorrelated, then $\mathrm{VAR}\left[Y\right] = \mathrm{VAR}\left[X\right] + \mathrm{VAR}\left[N\right]$
i.e. the variance of the sum is the sum of the variances.

Invariance under linear transformations

Consider $aX + b$ and $cY + d$, where $a > 0$ and $c > 0$.

$$\begin{aligned} COV(aX+b, cY+d) &= E[(aX+b-E[aX+b])(cY+d-E[cY+d])] \\ &= E[(aX+b-aE[X]-b)(cY+d-cE[Y]-d)] \\ &= acE[(X-E[X])(Y-E[Y])] \\ &= acCOV(X,Y) \end{aligned}$$

We also have that

$$\sigma_{aX+b}^2 = a^2\sigma_X^2 \Rightarrow \sigma_{aX+b} = a\sigma_X$$
$$\sigma_{cY+d}^2 = c^2\sigma_Y^2 \Rightarrow \sigma_{cY+d} = c\sigma_Y$$

Combining

$$\rho_{aX+b,cY+d} = \frac{acCOV(X,Y)}{a\sigma_X c\sigma_Y} = \rho_{X,Y}$$

Hence, the correlation coefficient is unchanged under a linear transformation of X and Y.

Example: Nonlinear relationship between X and Y

Consider $Y = X^2$ where $E[X] = 0$. Then,

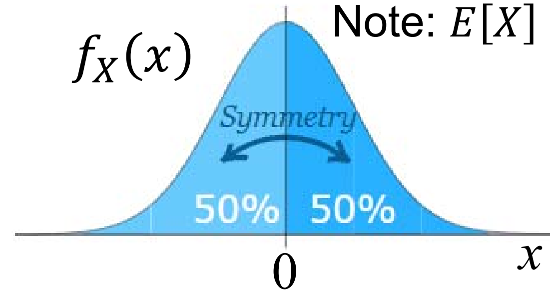
$$COV(X, Y) = E[XY] - E[X]E[Y] = E[XY] = E[X(X^2)] = E[X^3]$$

If $X \sim \mathcal{N}(0,1)$ (Gaussian RV with mean 0 and variance 1), then $E[X^3] = 0$ (see next slide).

In this case, $COV(X, Y) = 0$. Thus, X and Y are uncorrelated, but not independent.

The quadratic dependence is not captured by the correlation coefficient, unlike linear dependencies.

Odd moments of a symmetric PDF are zero



Note: $E[X] = 0$

- Consider a random variable X whose PDF is symmetric around 0.

$$E[X^n] = \int_{-\infty}^{\infty} x^n f_X(x)\,dx = \int_{-\infty}^{0} x^n f_X(x)\,dx + \int_{0}^{\infty} x^n f_X(x)\,dx$$

$$\int_{-\infty}^{0} x^n f_X(x)\,dx = \int_{\infty}^{0} (-u)^n f_X(-u)(-du)$$

$$= (-1)^n \int_{0}^{\infty} u^n f_X(-u) du$$

$$= (-1)^n \int_{0}^{\infty} u^n f_X(u) du$$

same!

change of variables: $x = -u$
$dx = -du$

symmetry: $f_X(u)$=$f_X(-u)$

Hence, $E[X^n] = \begin{cases} 2\int_0^{\infty} x^n f_X(x)\,dx & \text{if } n \text{ is even} \\ 0 & \text{if } n \text{ is odd} \end{cases}$

- If $E[X] \neq 0$ but pdf symmetric around $E[X]$, then odd central moments are zero, i.e., $E[(X - E[X])^n] = 0$

ELEC 2600H: Probability and Random Processes in Engineering

Part III: Multiple Random Variables

- Lecture 10: Pairs of Discrete Random Variable
  - Out-of-Class Reading: 2D Calculus
- Lecture 11: Pairs of Continuous Random Variable
- Lecture 12: Independence, Joint Moments
- Lecture 13: Correlation Coefficients and Their Properties
- **Lecture 14: Conditional PDF, Conditional Expectation**
- Lecture 15: Sum of Two Random Variables
- Lecture 16: Pairs of Jointly Gaussian Random Variables
- Lecture 17: More than Two Random Variables
- Lecture 18: Laws of Large Numbers
- Lecture 19: Central Limit Theorem and Characteristic Function

Elec2600H: Lecture 14

- **Conditional Probability**
  - **Product Rule**
  - **Total Probability Theorem**
  - **Bayes Theorem and Pattern Classification**
- Conditional Expectation

Conditional Probability Mass Functions

- Suppose that X and Y are discrete RVs assuming integer values.
- The conditional pmf of Y given X is $p_{Y|X}(k|j) = \frac{P[Y=k,X=j]}{P[X=j]} = \frac{p_{X,Y}(j,k)}{p_X(j)}$ if $p_X(j) > 0$. It is undefined if $p_X(j) = 0$ (since you condition on an "impossible" event).

- Interpretation: The probability that $Y = k$ assuming $X = j$.

- The conditional pmf of Y given X satisfies all the properties of a regular pmf in Y, but not X. For example,

$$\sum_{k=-\infty}^{\infty} p_{Y|X}(k|j) = 1 \qquad \text{Always true}$$

$$\sum_{j=-\infty}^{\infty} p_{Y|X}(k|j) \neq 1 \qquad \text{Usually}$$

Example 5.29: Loaded Dice

Consider the pmf for tossing two "loaded" die.

Compare the

- marginal pmf for X, $p_X(j)$
- conditional pmf of X if $Y = 5$, $p_{X|Y}(j|5)$

k	6	1/42	1/42	1/42	1/42	1/42	**2/42**
	5	1/42	1/42	1/42	1/42	**2/42**	1/42
	4	1/42	1/42	1/42	**2/42**	1/42	1/42
	3	1/42	1/42	**2/42**	1/42	1/42	1/42
	2	1/42	**2/42**	1/42	1/42	1/42	1/42
	1	**2/42**	1/42	1/42	1/42	1/42	1/42
$p_{X,Y}(j,k)$		1	2	3	4	5	6
		j					

Solution

We find the marginal by summing along the columns:

j	1	2	3	4	5	6
$p_X(j)$	1/6	1/6	1/6	1/6	1/6	1/6

We find the conditional by rescaling row 5:

j	1	2	3	4	5	6
$p_{X\|Y}(j\|5)$	1/7	1/7	1/7	1/7	2/7	1/7

$$p_{X|Y}(j|5) = \frac{p_{X,Y}(j,5)}{p_Y(5)}$$

$$p_Y(5) = 1/6$$

The Product Rule

- Starting from the definitions of the conditional probability

$$p_{Y|X}(k \mid j) = \frac{p_{X,Y}(j,k)}{p_X(j)} \qquad p_{X|Y}(j \mid k) = \frac{p_{X,Y}(j,k)}{p_Y(k)}$$

- We can obtain the product rules: $p_{X,Y}(j,k) = p_{Y|X}(k \mid j)p_X(j)$ $\quad p_{X,Y}(j,k) = p_{X|Y}(j \mid k)p_Y(k)$

- This suggests a way to generate X and Y sequentially:
  - Generate X first according to $p_X(j)$, then generate Y according to $p_{Y|X}(k|j)$.
  - Alternatively, generate Y first according to $p_Y(k)$, then generate X according to $p_{X|Y}(j|k)$.

Total Probability Theorem

- Combining the sum and product rules, we can find a "divide and conquer" strategy for computing the probability that $Y=k$.

$$P[Y=k] = \sum_j p_{X,Y}(j,k) = \sum_j p_{Y|X}(k \mid j)\, p_X(j)$$

- In words, we first compute $P[Y=k|X=j]$ for all j, then average the results over all possible values of j.

- Similarly, it can be shown that $P[Y \text{ in } A] = \sum_j P\left[Y \text{ in } A \mid X=j\right] p_X(j)$

Example 5.30: Defects on a Chip

The total number of defects X on a chip is a **Poisson** random variable with mean $\alpha > 0$.

If there are X of defects on a chip, the number of defects in a specific region R follows a Binomial distribution with parameters X and p, where p is a constant in $[0,1]$.

Find the pmf of the number of defects Y in the region R.

Solution

By the total probability theorem,

$$p_Y(k) = \sum_n p_{Y|X}(k \mid n) \cdot p_X(n) \text{ for } k \geq 0 \quad \text{where } p_{Y|X}(k \mid n) = \binom{n}{k} p^k (1-p)^{n-k} \text{ for } 0 \leq k \leq n$$

Substituting, $$p_Y(k) = \sum_{n=k}^{\infty} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \cdot \frac{\alpha^n}{n!} e^{-\alpha}$$

$$= \frac{(\alpha p)^k}{k!} e^{-\alpha} \times \sum_{n=k}^{\infty} \frac{(\alpha(1-p))^{n-k}}{(n-k)!}$$

$$= \frac{(\alpha p)^k}{k!} e^{-\alpha} \times e^{\alpha(1-p)} = \frac{(\alpha p)^k}{k!} e^{-\alpha p}$$

Poisson with parameter αp!

Conditional pdf (Continuous X and Y)

- If X is continuous, $P[X = x] = 0$ for all x, so we **cannot** define $\cancel{F_{Y|X}(y \mid x) = \dfrac{P[Y \le y, X = x]}{P[X = x]}}$

- Instead, we define the conditional cdf of Y given X by

$$F_{Y|X}(y|x) = \lim_{h \to 0} \frac{P\left[\{Y \le y\} \cap \{x \le X \le x+h\}\right]}{P\left[\{x \le X \le x+h\}\right]}$$

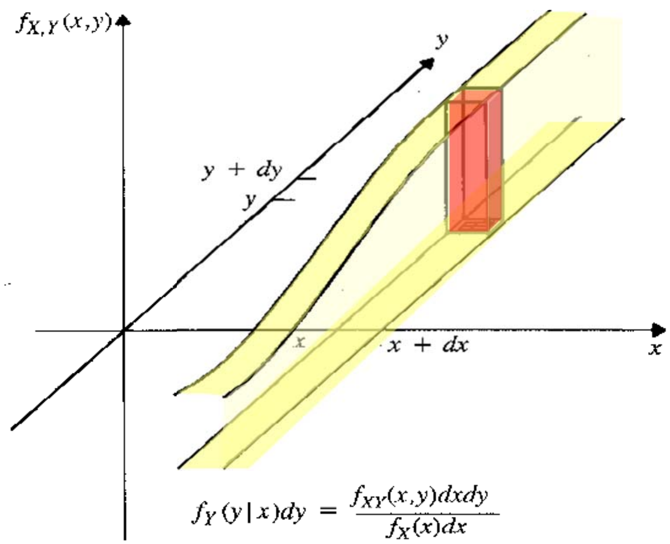
$$= \lim_{h \to 0} \frac{\int_{-\infty}^{y} \int_{x}^{x+h} f_{X,Y}(\alpha, \beta)\, d\alpha\, d\beta}{\int_{x}^{x+h} f_X(\alpha)\, d\alpha} \approx \lim_{h \to 0} \frac{h \int_{-\infty}^{y} f_{X,Y}(x, \beta)\, d\beta}{f_X(x)\, h}$$

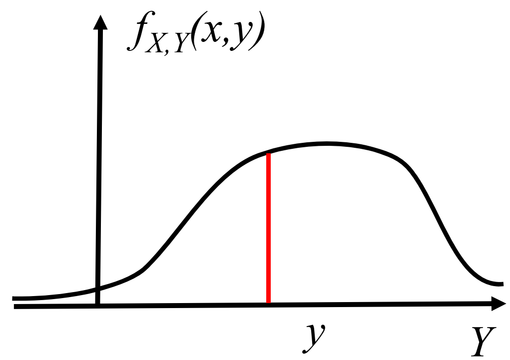
$$= \frac{\int_{-\infty}^{y} f_{X,Y}(x, \beta)\, d\beta}{f_X(x)} \qquad \text{if } f_X(x) > 0 \text{ and undefined otherwise}$$

- Differentiating with respect to y, we obtain the conditional pdf of Y given X, $f_{Y|X}(y|x) = \dfrac{f_{X,Y}(x, y)}{f_X(x)}$

Graphical Interpretation

- The conditional pdf specifies the probability that Y is in the infinitesimal strip defined by $(y, y+dy)$ given that X is in the infinitesimal strip defined by $(x, x+dx)$.
- The conditional pdf is the mass of the red column divided by the mass of the entire yellow strip.
- It can also be interpreted as a slice of the joint density at $X = x$ normalized to unit area.





$$f_{Y|X}(y\,|\,x) = \frac{f_{X,Y}(x,y)}{\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy}$$

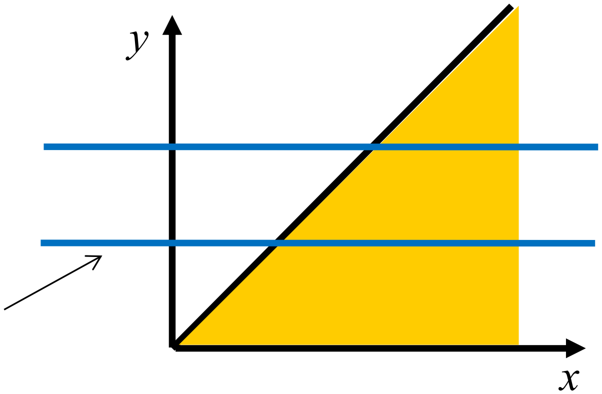
Example 5.32(a)

Suppose that random variables X and Y are jointly distributed according to

$$f_{X,Y}(x,y)=\begin{cases}2e^{-x}e^{-y} & 0\le y\le x\le\infty\\ 0 & \text{elsewhere.}\end{cases}$$

Find $f_{X|Y}(x|y)$.

$f_{X|Y}(x|1)\propto f_{X,Y}(x,1)$



Solution:

Recall that $f_Y(y)=2e^{-2y}u(y)$

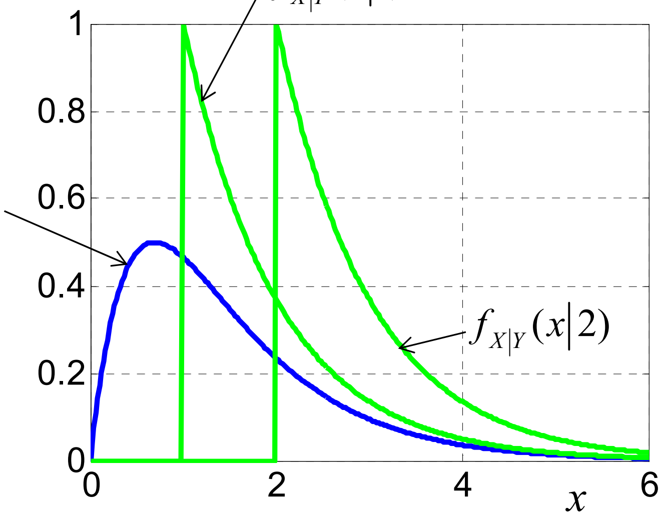
If $y\ge 0$, $f_{X|Y}(x|y)=\dfrac{f_{X,Y}(x,y)}{f_Y(y)}$

$$=\begin{cases}\dfrac{2e^{-x}e^{-y}}{2e^{-2y}}=e^{-(x-y)} & x\ge y\\ 0 & x<y\end{cases}$$

$f_{X|Y}(x|1)$

$f_X(x)=2e^{-x}\left(1-e^{-x}\right)u(x)$

$f_{X|Y}(x|2)$

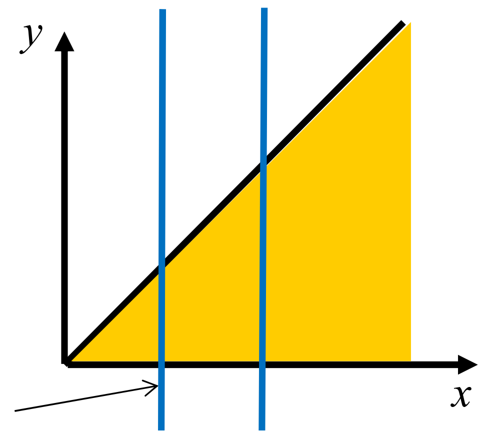


Example 5.32(b)

Suppose that random variables X and Y are jointly distributed according to

$$f_{X,Y}(x,y) = \begin{cases} 2e^{-x}e^{-y} & 0 \le y \le x \le \infty \\ 0 & \text{elsewhere.} \end{cases}$$

Find $f_{Y|X}(y|x)$

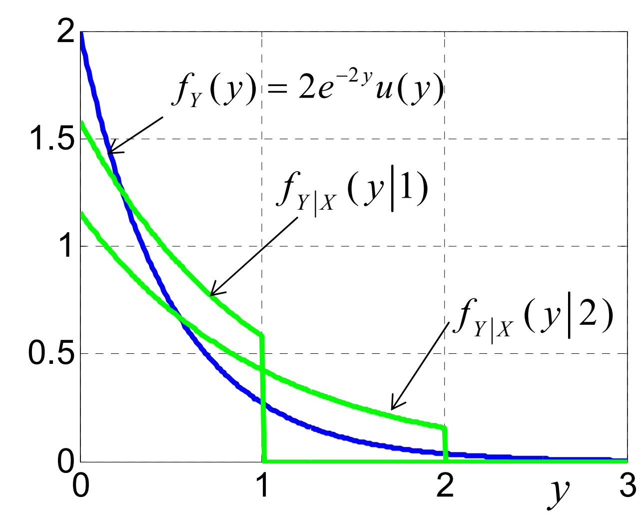


Solution:

Recall that $f_X(x) = 2e^{-x}\left(1-e^{-x}\right)u(x)$

If $x \ge 0$, $f_{Y|X}(y|x) = \dfrac{f_{X,Y}(x,y)}{f_X(x)}$

$$= \begin{cases} \dfrac{e^{-y}}{1-e^{-x}} & 0 \le y \le x \\ 0 & \text{otherwise} \end{cases}$$



Properties of the Conditional Density

- Product rules:

$$f_{X,Y}(x,y) = f_{X|Y}(x|y)f_Y(y)$$
$$f_{X,Y}(x,y) = f_{Y|X}(y|x)f_X(x)$$

- Total Probability Theorem:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|x)f_X(x)dx$$

Proof: Use $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx$ and product rule.

- Bayes Theorem: $f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)f_X(x)}{\int_{-\infty}^{\infty} f_{Y|X}(y|x)f_X(x)dx}$

- Similar equations hold for pmfs (e.g. $p_{X,Y}(x,y)$).
- If X is discrete, replace integral by sum.

Example 5.34 (Total probability theorem)

The random variable X is selected at random from the unit interval. The random variable Y is then selected at random from $[0, X]$. Find the pdf of Y.

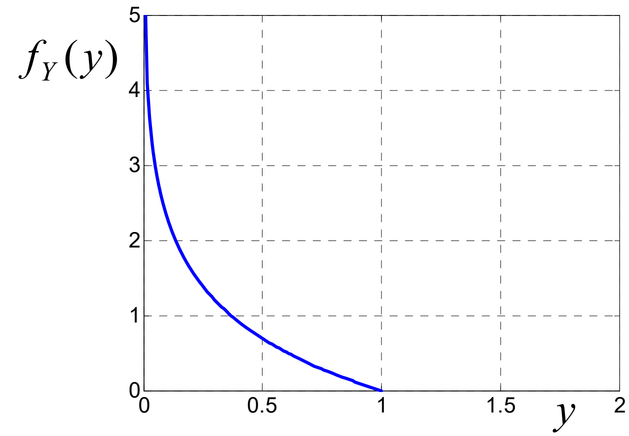
Solution

By the total probability theorem, $f_Y(y) = \int f_{Y|X}(y \mid x) \cdot f_X(x)dx$

where $f_{Y|X}(y|x) = \begin{cases} \frac{1}{x} & 0 \le y \le x \\ 0 & \text{otherwise} \end{cases}$ $\quad f_X(x) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$

Thus, $f_Y(y) = \int_y^1 \frac{1}{x} dx = \ln x \Big|_y^1$

$$= -\ln y \quad \text{for } 0 \le y \le 1$$



Random variables that **differ** in type

- In problems where it is necessary to deal with random variables of different types (e.g. X is discrete and Y is continuous), it is often easiest to work with conditional probabilities

Conditional pmf or pdfs (X discrete, Y continuous)

- Let $f_{X,Y}(x,y) = \frac{d}{dy} P[X = x, Y \leq y]$ (joint pmf in X and pdf in Y)
- Marginal pmf/pdf:

$$p_X(x) = P[X = x]$$
$$f_Y(y) = \frac{d}{dy} P[Y \leq y]$$

- Conditional pmf of X given Y: $p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p(y)}$ if $p_Y(y) > 0$, else undefined.

- Conditional pdf of Y given X: $f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$ if $f_X(x) > 0$, else undefined.

Example 4.14

Let X be the input to a communication channel and let Y be the output.
The input X is either –1 volt or +1 volt with equal probability.
The output $Y = X + N$, where N is a continuous noise voltage uniformly distributed in [-2, 2] volts.
Find $P[X = +1,\ Y \leq 0]$.

Solution:

According to the definition of the conditional probability: $P[X = +1, Y \leq 0] = P[Y \leq 0 \mid X = +1]P[X = +1]$

When the $X = +1$, the Y is uniformly distributed in [-1,3]. Thus, $P[Y \leq 0 \mid X = +1] = \frac{1}{4}$
(Alternatively, the noise N must be less than or equal to -1.)

Therefore, $P[X = +1, Y \leq 0] = \frac{1}{4} \times \frac{1}{2} = \frac{1}{8}$

Example 5.33: Bank Queue

The number of customers that arrive at a bank during a time t is a Poisson RV with parameter βt.
The time required to service each customer is an exponential RV with parameter α.
Find the pmf of the number of customers N that arrive during the service time T of a specific customer.
Assume that customer arrivals and service times are independent.

Solution

By the total probability theorem, $p_N(k) = \int_{-\infty}^{\infty} p_{N|T}(k|t) \cdot f_T(t)dt$

Substituting $p_{N|T}(k|t) = \frac{(\beta t)^k}{k!}e^{-\beta t}$ and $f_T(t) = \alpha e^{-at}u(t)$

we obtain
$$p_N(k) = \int_0^{\infty} \frac{(\beta t)^k}{k!} e^{-\beta t} \cdot \alpha e^{-\alpha t} dt$$
$$= \frac{\alpha \beta^k}{k!} \int_0^{\infty} t^k e^{-(\alpha+\beta)t} dt \quad \text{for } k \in \{0,1,2,\dots\}$$

Example 5.33: Integral Evaluation

$$p_N(k) = \frac{\alpha\beta^k}{k!}\int_0^\infty t^k e^{-(\alpha+\beta)t}dt$$
$$= \frac{\alpha\beta^k}{(\alpha+\beta)^{k+1}}\frac{1}{k!}\int_0^\infty r^k e^{-r}dr$$
$$= \frac{\alpha\beta^k}{(\alpha+\beta)^{k+1}}\frac{1}{k!}\Gamma(k+1)$$
$$= \frac{\alpha\beta^k}{(\alpha+\beta)^{k+1}}$$
$$= \left(\frac{\alpha}{\alpha+\beta}\right)\left(\frac{\beta}{\alpha+\beta}\right)^k$$
$$= p(1-p)^k \text{ for } k \in \{0,1,\dots\}$$

Change of variables:
$r = (\alpha+\beta)t$
$dr = (\alpha+\beta)dt$

Gamma function: $\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}dx$ for $z > 0$

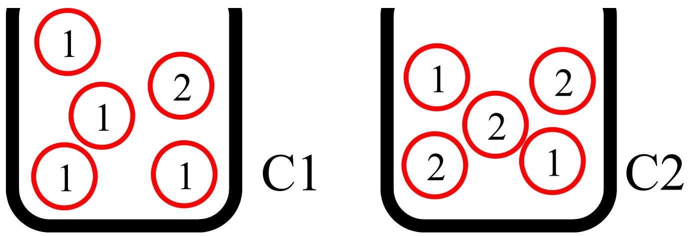
For integer $k \geq 0$,
$\Gamma(k+1) = k!$

Geometric RV! with $p = \dfrac{\alpha}{\alpha+\beta} = \dfrac{1}{1+\frac{\beta}{\alpha}}$

Total probability example

- We are presented with one of two urns, labeled C1 and C2, with equal probability, but we don't know which urn is presented.
- In both urns, there are balls labeled 1 and 2, but in different proportions as shown.

$$P["1" \mid \mathrm{C2}] = 0.4 \qquad P["1" \mid \mathrm{C1}] = 0.8 \qquad P[\mathrm{C1}] = 0.5$$
$$P["2" \mid \mathrm{C2}] = 0.6 \qquad P["2" \mid \mathrm{C1}] = 0.2 \qquad P[\mathrm{C2}] = 0.5$$



- Suppose that we draw a ball at random from the urn.

- What are the probabilities of observing a "1" or of observing at "2"?

$$P["1"] = P["1" \mid \mathrm{C1}]P[\mathrm{C1}] + P["1" \mid \mathrm{C2}]P[\mathrm{C2}]$$
$$= 0.8 \times 0.5 + 0.4 \times 0.5 = 0.6$$
$$P["2"] = P["2" \mid \mathrm{C1}]P[\mathrm{C1}] + P["2" \mid \mathrm{C2}]P[\mathrm{C2}]$$
$$= 0.2 \times 0.5 + 0.6 \times 0.5 = 0.4$$

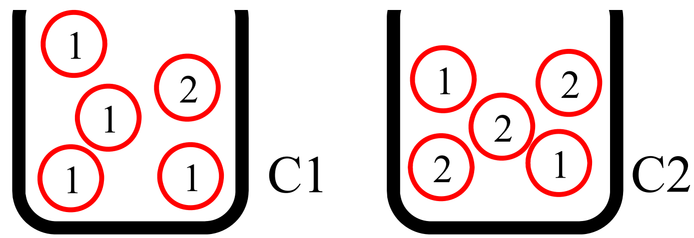
Note that $P["1"] + P["2"] = 1$

Bayes classification example

- We are presented with one of two urns, labeled C1 and C2 with equal probability. In both urns, there are balls labeled 1 and 2, but in different proportions as shown.

$$P["1" \mid C1] = 0.8 \qquad P["1" \mid C2] = 0.4 \qquad P[C1] = 0.5$$

$$P["2" \mid C1] = 0.2 \qquad P["2" \mid C2] = 0.6 \qquad P[C2] = 0.5$$



- Suppose that we draw a ball at random from the urn and observe a "2". What is the probability that C2 was presented?

$$P[C2 \mid "2"] = \frac{P["2" \mid C2]P[C2]}{P["2" \mid C1]P[C1] + P["2" \mid C2]P[C2]}$$

$$= \frac{0.6 \times 0.5}{0.2 \times 0.5 + 0.6 \times 0.5} = 0.75$$

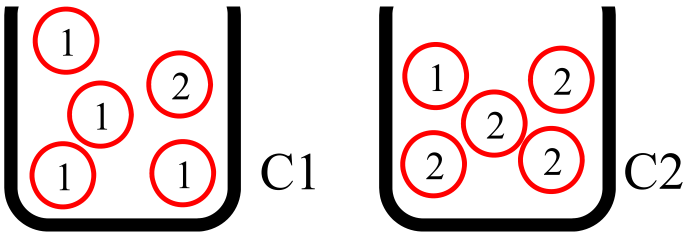
- It is more likely that C2 was presented, but we are only 75% sure.

Classification example (cont.)

- Note that the probability that C2 was presented depends upon the number of balls labelled “2” in C2. Suppose one of the “1” balls in C2 is replaced by a “2” ball.

$$P[\text{"1"}\,|\,\text{C1}] = 0.8 \qquad P[\text{"1"}\,|\,\text{C2}] = 0.2 \qquad P[\text{C1}] = 0.5$$

$$P[\text{"2"}\,|\,\text{C1}] = 0.2 \qquad P[\text{"2"}\,|\,\text{C2}] = 0.8 \qquad P[\text{C2}] = 0.5$$



$$P[\text{C2}|\text{"2"}] = \frac{P[\text{"2"}\,|\,\text{C2}]P[\text{C2}]}{P[\text{"2"}\,|\,\text{C1}]P[\text{C1}] + P[\text{"2"}\,|\,\text{C2}]P[\text{C2}]}$$

$$= \frac{0.8\times 0.5}{0.2\times 0.5 + 0.8\times 0.5} = 0.8 > 0.75$$

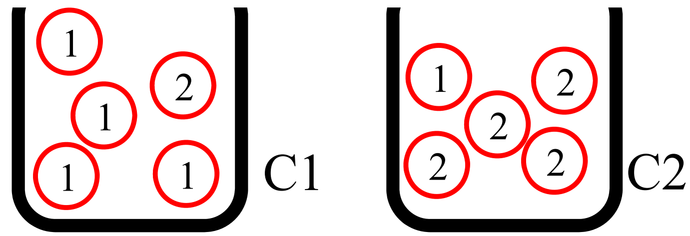
- In this case, we become more certain that C2 was presented.

Classification example (cont.)

- Note that the probability that C2 was presented also depends upon the prior class probabilities. For example, suppose that C1 and C2 are not equally likely to be presented, but rather C1 is 9 times more likely that C2.

$$P["1" \mid \mathrm{C1}] = 0.8 \qquad P["1" \mid \mathrm{C2}] = 0.2 \qquad P[\mathrm{C1}] = 0.9$$

$$P["2" \mid \mathrm{C1}] = 0.2 \qquad P["2" \mid \mathrm{C2}] = 0.8 \qquad P[\mathrm{C2}] = 0.1$$



$$P[\mathrm{C2} \mid "2"] = \frac{P["2" \mid \mathrm{C2}]P[\mathrm{C2}]}{P["2" \mid \mathrm{C1}]P[\mathrm{C1}] + P["2" \mid \mathrm{C2}]P[\mathrm{C2}]}$$

$$= \frac{0.8 \times 0.1}{0.2 \times 0.9 + 0.8 \times 0.1} \approx 0.3077 < 0.8$$

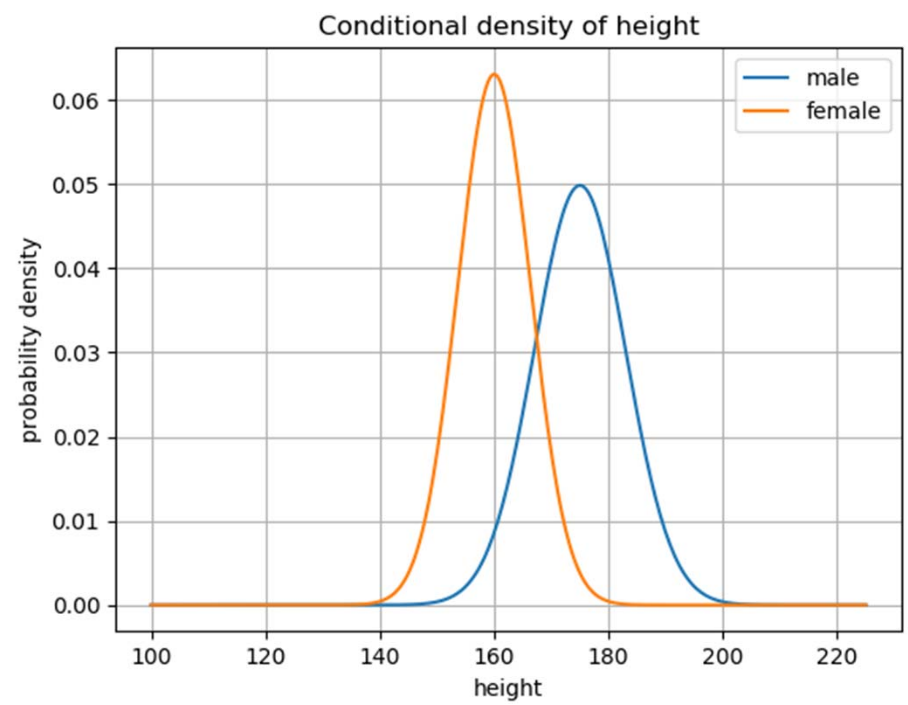
- Although C1 has fewer “2” balls than C1, it is still more likely that C1 was presented!

Example: Classification by height (Discrete X, Continuous Y)

- Suppose we pick someone randomly from the room. Let X be the sex of person (M or F) and Y be his/her height.
- Assume the conditional densities of the height given sex are

$$f_{Y|X}(y|x) = \begin{cases} \mathcal{N}(y|175,64) & \text{if } x = \mathrm{M} \\ \mathcal{N}(y|160,40) & \text{if } x = \mathrm{F} \end{cases}$$

where $\mathcal{N}(y|m,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-m)^2}{2\sigma^2}}$



- Suppose that the prior probabilities of sex are

$$p_X(x) = \begin{cases} 0.7 & \text{if } x = \mathrm{M} \\ 0.3 & \text{if } x = \mathrm{F} \end{cases}$$

- Suppose we observe the height Y, find the probability that the person is a male or female:

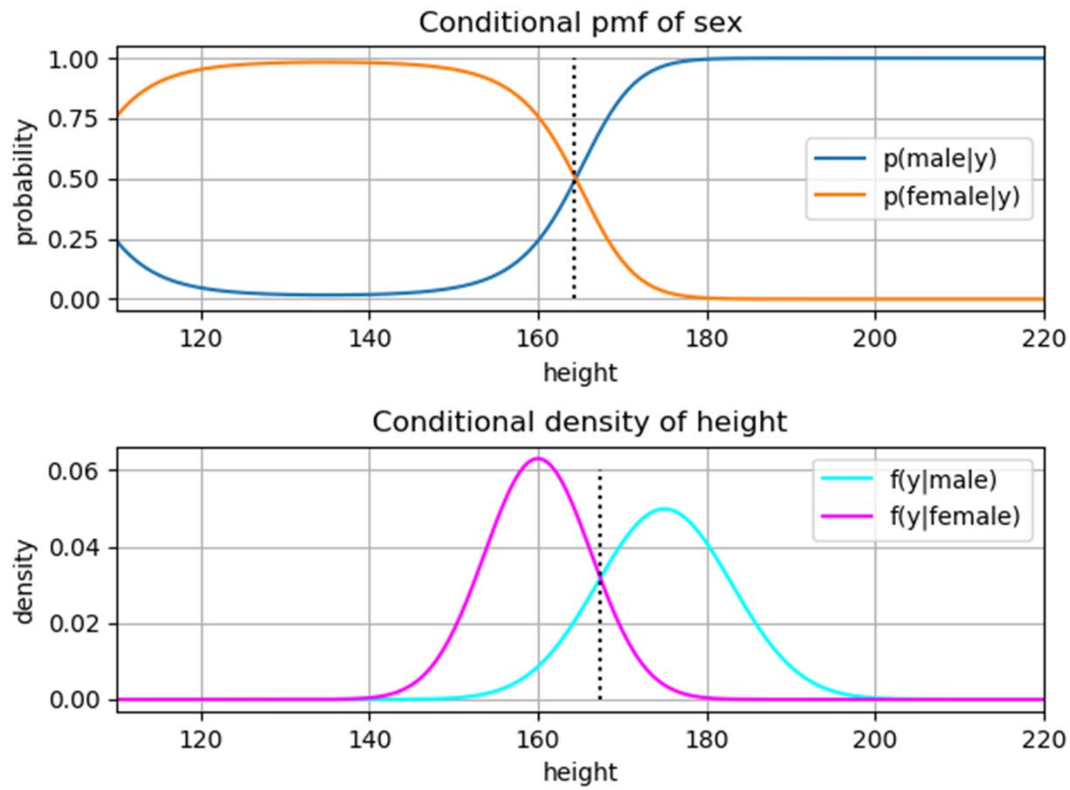
$$p_{X|Y}(x|y)$$

Solution: Classification by Height

- By Bayes Theorem: $p_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)p_X(x)}{\sum_\xi f_{Y|X}(y|\xi)p_X(\xi)}$
- If $x = M$, then

$$p_{X|Y}(\mathrm{M}|y) = \frac{f_{Y|X}(y|\mathrm{M})p_X(\mathrm{M})}{f_{Y|X}(y|\mathrm{M})p_X(\mathrm{M}) + f_{Y|X}(y|\mathrm{F})p_X(\mathrm{F})}$$
$$= \frac{0.7\mathcal{N}(y|175{,}64)}{0.7\mathcal{N}(y|175{,}64) + 0.3\mathcal{N}(y|160{,}40)}$$

- Note: $p_{X|Y}(\mathrm{F}|y) = 1 - p_{X|Y}(\mathrm{M}|y)$



Elec2600H: Lecture 14

- **Conditional Probability**
  - **Product Rule**
  - **Total Probability Theorem**
  - **Bayes Theorem and Pattern Classification**
- **Conditional Expectation**

Conditional Expectation

- **Definition:** The *conditional expectation* of Y given $X = x$ is defined as

$$E[Y \mid x] = \int_{-\infty}^{\infty} y f_{Y|X}\left(y \mid x\right) dy \quad \text{if } Y \text{ is continuous}$$

$$E[Y \mid x] = \sum_{j} j \cdot p_{Y|X}(j \mid x) \quad \text{if } Y \text{ is discrete and integer valued}$$

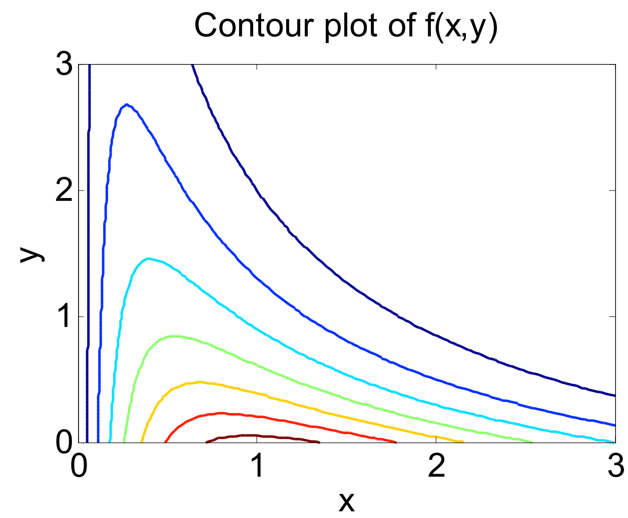
- **Intuition:** Our best guess (in a minimum mean squared error sense) of the value of Y if we observe $X = x$.
- Thus, the conditional expected value is a function of x.

- More generally, if $g(Y)$ is a random variable, then $E[g(Y) \mid x] = \int_{-\infty}^{\infty} g(y) f_{Y|X}(y \mid x) dy$

Example

For the density below, find $\mathrm{E}[Y|x]$.

$$f(x, y) = \begin{cases} xe^{-x(y+1)} & x \ge 0, y \ge 0 \\ 0 & \text{otherwise} \end{cases}$$



Solution

Find the marginal density of X:

$$f_X(x) = \int_0^\infty xe^{-x(y+1)}dy = e^{-x}\int_0^\infty xe^{-xy}dy = e^{-x},$$

Find the conditional density: $f_{Y|X}(y|x) = \dfrac{f(x, y)}{f_X(x)} = xe^{-xy}$ for $y \ge 0$.

Since Y given $X = x$ is exponential with parameter x,

$$E[Y|X] = \int_{-\infty}^{\infty} yf_{Y|X}(y|x)dy$$

$$= \int_0^\infty yxe^{-xy}dy = \frac{1}{x}$$

Removing Conditioning via Expectation

Since $E[Y|X]$ is a function of X, it is also a random variable.
Taking the expectation over X, the conditioning disappears!

over X

over Y

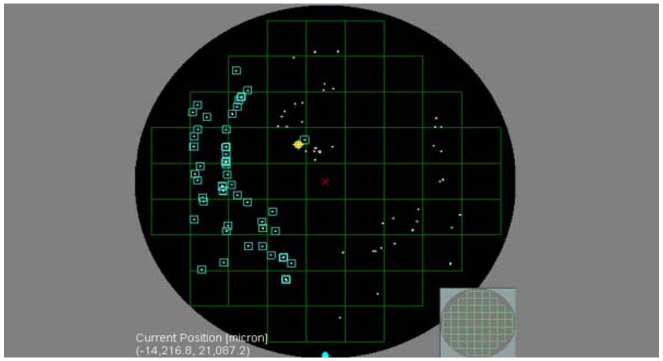
$$
\begin{aligned}
E[E[Y|X]] &= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} y f_{Y|X}(y|x)\,dy\right] f_X(x)\,dx \\
&= \int_{-\infty}^{\infty} y \int_{-\infty}^{\infty} f_{Y|X}(y|x) f_X(x)\,dx\,dy \\
&= \int_{-\infty}^{\infty} y f_Y(y)\,dy = E[Y]
\end{aligned}
$$

This is an "expectation" version of the total probability theorem.

More generally, $E[g(Y)] = E[E[g(Y)|X]]$

Example 5.36: Defects on a Chip

The total number of defects X on a chip is a Poisson RV with mean α. Each defect has a probability p of falling into a region R. Assume the defect locations are independent. Let Y be the number of defects in R. Find $E[Y]$.



Solution

Condition on the total number of defects, X: $E[Y] = E\big[E[Y \mid X]\big]$

If X is fixed, Y has a binomial distribution with parameter X and p. Thus, $E[Y \mid X] = pX$

Taking the expectation over X, $E[Y] = E\big[E[Y \mid X]\big] = E[pX] = pE[X] = p\alpha$

Intuitively, this result makes. The average number of defects in R is the average number of defects on the chip times the probability that each defect lies in R.

Example 5.38(a): Service Station

The time required to service each customer at a service station is an exponential RV with parameter α.
The number of customer arriving in a time interval of t minutes follows a Poisson distribution with mean βt.
Assume that the service times and arrivals are independent.
Find the **mean** of the number of customer arrivals, N, during the service time, T, of a particular customer.

Example 5.38(a): Service Station

The time required to service each customer at a service station is an exponential RV with parameter α.
The number of customer arriving in a time interval of t minutes follows a Poisson distribution with mean βt.
Assume that the service times and arrivals are independent.
Find the **mean** of the number of customer arrivals, N, during the service time, T, of a particular customer.

Solution

Condition on the service time T : $\mathrm{E}[N] = \mathrm{E}[\mathrm{E}[N|T]]$.

Since the number of customer arrivals is Poisson, $\mathrm{E}[N|T] = \beta T$.

Since the service time T is exponential, $\mathrm{E}[T] = 1/\alpha$ and $\mathrm{Var}[T] = 1/\alpha^2$.

Thus, $\mathrm{E}[N] = \beta \mathrm{E}[T] = \beta/\alpha$.

Intuitively, the average number of customer arrivals is the average service time $1/\alpha$ multiplied by the rate at which customers arrive, β (customers/min).

Example 5.38 (cont.)

Find the **variance** of the number of customer arrivals, N, during the service time, T, of a particular customer. Assume that the service times and arrivals are independent.

Solution

$$\begin{aligned}\mathrm{Var}[N] &= \mathrm{E}[N^2] - \mathrm{E}[N]^2 \\ &= \beta \mathrm{E}[T] + \beta^2 \mathrm{E}[T^2] - \beta^2 (\mathrm{E}[T])^2 \\ &= \underbrace{\beta \mathrm{E}[T]}_{\text{variance if service time constant at E}[T]} + \underbrace{\beta^2 \mathrm{Var}[T]}_{\text{extra term due to randomness of }T}\end{aligned}$$

$E[N] = \beta E[T]$ (previous example)

$$\begin{aligned}E[N^2] &= E[E[N^2|T]] \\ &= E[\mathrm{Var}[N|\mathrm{T}] + (E[N|T])^2] \\ &= E[\beta T + \beta^2 T^2] \\ &= \beta E[T] + \beta^2 E[T^2]\end{aligned}$$

Since N is Poisson,

$\mathrm{Var}[N|\mathrm{T}] = \beta T$

$E[N|T] = \beta T$.

Since T is exponential, $\mathrm{E}[T] = 1/\alpha$ and $\mathrm{Var}[T] = 1/\alpha^2$. Thus, $\mathrm{Var}[N] = \dfrac{\beta}{\alpha} + \dfrac{\beta^2}{\alpha^2}$

Major Points from this Lecture:

- **Conditional PDF**
- **Conditional Expected Value**

ELEC 2600H: Probability and Random Processes in Engineering

Part III: Multiple Random Variables

Elec2600H: Lecture 15

- **Sums of two random variables**
  - **Discrete random variables**
  - Continuous random variables
  - Expected Value and Variance

PMF of a function of two discrete RV's

Suppose X takes integer values i.
Suppose Y takes integer values j.
Suppose that $Z = X+Y$.

For each k, we can find the probability mass function of Z by

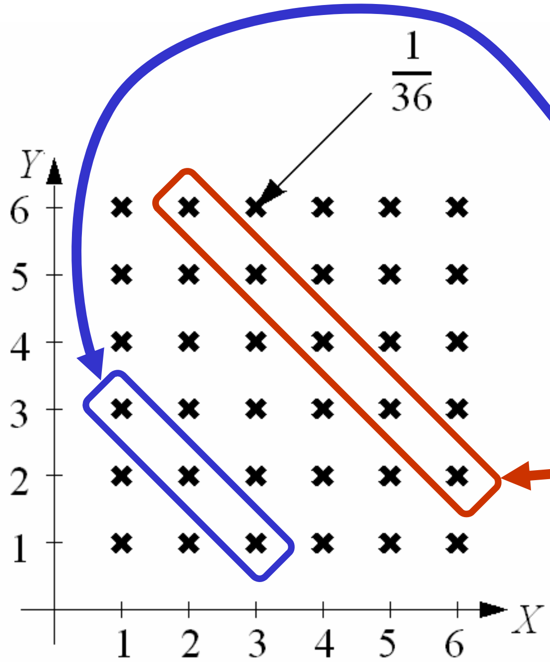
$$p_Z(k) = P[Z = k] = \sum_{i+j=k} p_{X,Y}(i, j)$$

the sum over all pairs (i, j) such that $i+j = k$

Example (Rolling Dice)

Toss two fair dice and let X and Y be the numbers of dots observed. Find the probability mass function of $Z = X + Y$

Solution:



z	P[Z = z]
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36

Example (Sum of independent discrete RVs)

Say X and Y are independent discrete RV's taking values $\{0, 1, 2, \ldots\}$ where

$$\begin{aligned} P(\{X = n\}) &= a_n \\ P(\{Y = n\}) &= b_n \end{aligned} \qquad \forall(n \in \{0, 1, 2, \ldots\})$$

Find $P(\{Z = n\})$ if $Z = X + Y$.

Solution:

$$\begin{aligned} P(\{Z = n\}) &= \sum_{k=0}^{n} P(\{X = k\} \cap \{Y = n-k\}) \\ &= \sum_{k=0}^{n} P(\{X = k\})P(\{Y = n-k\}) \\ &= \sum_{k=0}^{n} a_k b_{n-k} \quad \text{for } n \geq 0 \quad \text{(and 0 otherwise)} \end{aligned}$$

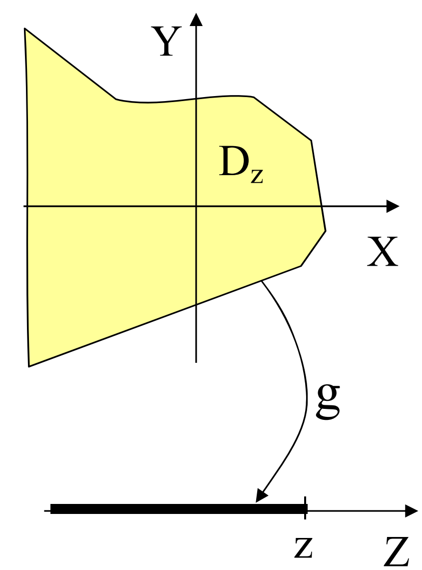
Elec2600H: Lecture 15

- **Sums of two random variables**
  - Discrete random variables
  - **Continuous random variables**
  - Expected Value and Variance

One function of two continuous random variables

Given two RV's X and Y with density $f_{X,Y}(x, y)$, find the density $f_Z(z)$ where $Z = g(X, Y)$

Solution Approach



Step 1. Pick a real number z

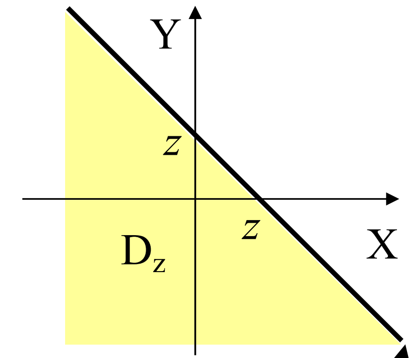
Step 2. Find the region $D_z = \{(x, y) \mid g(x, y) \le z\}$

Step 3. Evaluate $F_z(z) = \int\int_{D_z} f(x, y)dxdy$

Step 4. Differentiate: $f_z(z) = \dfrac{d}{dz} F_z(z)$

Example 5.39

$$Z = X + Y$$



integrate along this line

Step 1. Pick a real number z

Step 2. Find the region $D_z = \{(x, y) : x + y \le z\}$

Step 3. Integrate over D_z : $F_z(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{(z-x)} f(x, y) dy dx$

Step 4. Differentiate: $f_z(z) = \frac{d}{dz} \int_{-\infty}^{\infty} \int_{-\infty}^{(z-x)} f(x, y) dy dx = \int_{-\infty}^{\infty} \frac{d}{dz} \left[\int_{-\infty}^{(z-x)} f(x, y) dy \right] dx$

$$= \int_{-\infty}^{\infty} f(x, z - x) \frac{d(z - x)}{dz} dx$$

chain rule!

$$= \int_{-\infty}^{\infty} f(x, z - x) dx$$

Example

Suppose two random variables X and Y are independent.
Find the pdf of $Z = X + Y$.

Solution:

Since X and Y are independent, $f_{X,Y}(x, y) = f_X(x) f_Y(y)$

Substituting into the result from the previous example,

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(x, z - x)dx$$
$$= \int_{-\infty}^{\infty} f_X(x) f_Y(z - x)dx$$

Example 5.41

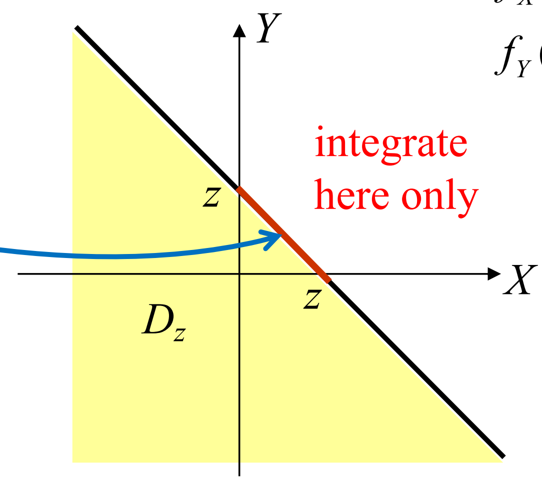
A system with standby redundancy has a single key component in operation and a duplicate of that component in standby mode. When the first component fails, the second is put into operation.
Find the pdf of the lifetime of the system, Z, if the two components have independent exponentially distributed lifetimes X and Y with the same parameter λ,

$$f_X(x) = \lambda e^{-\lambda x} u(x)$$

$$f_Y(y) = \lambda e^{-\lambda y} u(y)$$

Solution:

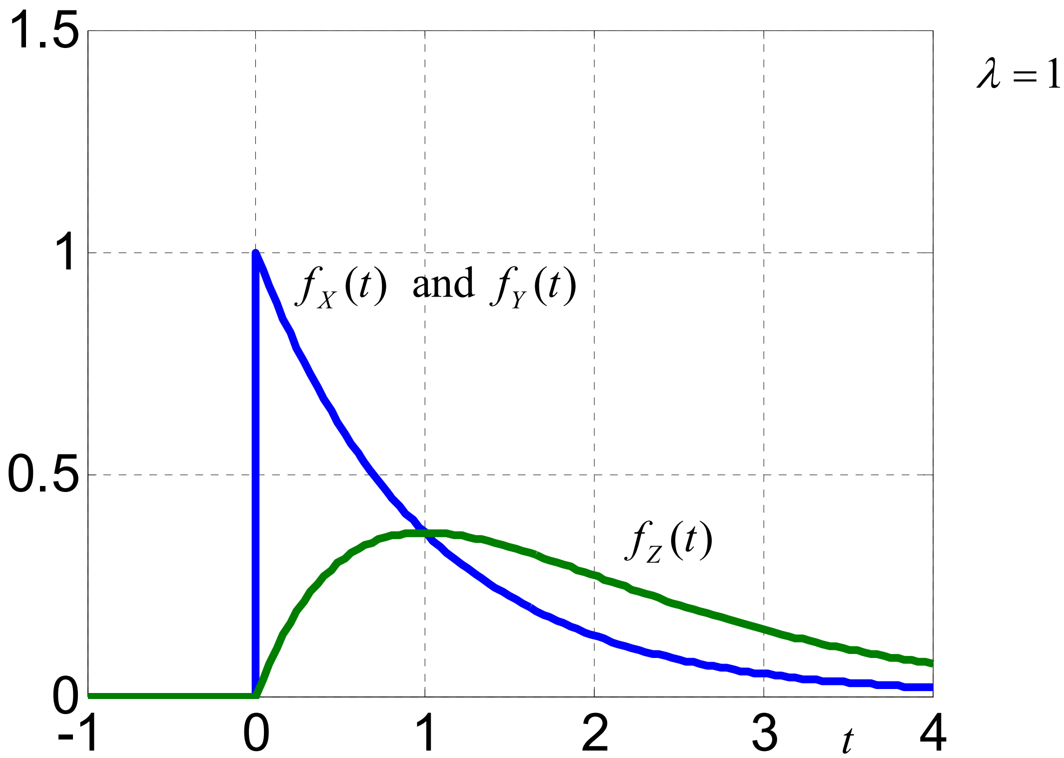
For z ≥ 0,
$$\begin{aligned} f_z(z) &= \int_{-\infty}^{\infty} f_X(x) \cdot f_Y(z-x)dx \\ &= \int_0^z \lambda e^{-\lambda x} \cdot \lambda e^{-\lambda(z-x)}dx \\ &= \lambda^2 e^{-\lambda z} \int_0^z e^{-\lambda x} \cdot e^{\lambda x} dx \\ &= \lambda^2 e^{-\lambda z} \int_0^z 1 \cdot dx \\ &= \lambda^2 z e^{-\lambda z} \end{aligned}$$



For z < 0, $f_z(z) = 0$

This is an example of an Erlang distribution,
which is a special case of the Gamma distribution

Example 5.41



Elec2600H: Lecture 15

- **Sums of two random variables**
  - Discrete random variables
  - Continuous random variables
  - **Expected Value and Variance**

Example 5.24: Sum of two RVs

Let $Z = X + Y$. Find $E[Z]$.

$$E[Z] = E[X+Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+y)f(x,y)dxdy$$

$$= \int_{-\infty}^{\infty} x\int_{-\infty}^{\infty} f(x,y)dydx + \int_{-\infty}^{\infty} y\int_{-\infty}^{\infty} f(x,y)dxdy$$

$$= \int_{-\infty}^{\infty} xf_X(x)dx + \int_{-\infty}^{\infty} yf_Y(y)dy = E[X] + E[Y]$$

Thus, the mean of the sum is the sum of the means.

This result also holds for more than two random variables.

$$E\left[X_1 + X_2 + \ldots + X_n\right] = E[X_1] + E[X_2] + \ldots + E[X_n]$$

Variance of Sum

Let $Z = X + Y$, where X and Y are random variables.
The variance of Z is

$$\mathrm{VAR}[Z] = \mathrm{VAR}[X] + 2\mathrm{COV}(X,Y) + \mathrm{VAR}[Y]$$

Proof

$$\begin{aligned}
\mathrm{VAR}[Z] &= E[(Z - E[Z])^2] \\
&= E[(X + Y - E[X] - E[Y])^2] \\
&= E[\left((X - E[X]) + (Y - E[Y])\right)^2] \\
&= E[(X - E[X])^2] + 2E[(X - E[X])(Y - E[Y])] + E[(Y - E[Y])^2] \\
&= \mathrm{VAR}[X] + 2\mathrm{COV}(X,Y) + \mathrm{VAR}[Y]
\end{aligned}$$

Note: If X and Y are uncorrelated, then $\mathrm{VAR}[Z] = \mathrm{VAR}[X] + \mathrm{VAR}[Y]$
i.e. the variance of the sum is the sum of the variances.

Example: EV of affine combination of two RVs

Let $Z = 5X + 2Y+1$. Find $E[Z]$.

$$E[Z] = E[5X + 2Y + 1] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(5x + 2y + 1)f(x, y)dxdy$$

$$= 5\int_{-\infty}^{\infty} x\int_{-\infty}^{\infty} f(x, y)dydx + 2\int_{-\infty}^{\infty} y\int_{-\infty}^{\infty} f(x, y)dxdy + 1\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)dxdy$$

$$= 5\int_{-\infty}^{\infty} xf_X(x)dx + 2\int_{-\infty}^{\infty} yf_Y(y)dy + 1 = 5E[X] + 2E[Y] + 1$$

More generally, for any a, b and c, $E\left[aX + bY + c\right] = aE[X] + bE[Y] + c$

This property is called "linearity of the expectation"

Example: Variance of affine combination

Let $Z = 5X + 2Y + 1$, where X and Y are random variables. Find $Var[Z]$.

Solution

$$
\begin{aligned}
\mathrm{VAR}\left[Z\right] &= E[(Z - E[Z])^2] \\
&= E[(5X + 2Y + 1 - (5E[X] + 2E[Y] + 1))^2] \\
&= E[\left(5(X - E[X]) + 2(Y - E[Y])\right)^2] \\
&= E[25(X - E[X])^2 + 20(X - E[X])(Y - E[Y]) + 4(Y - E[Y])^2] \\
&= 25E[(X - E[X])^2] + 20E[(X - E[X])(Y - E[Y])] + 4E[(Y - E[Y])^2] \\
&= 25\mathrm{VAR}\left[X\right] + 20\mathrm{COV}(X, Y) + 4\mathrm{VAR}\left[Y\right]
\end{aligned}
$$

More generally, for any a, b and c, $\mathrm{VAR}\left[aX + bY + c\right] = a^2\mathrm{VAR}\left[X\right] + 2ab\mathrm{COV}[X, Y] + b^2\mathrm{VAR}\left[Y\right]$

Note: If X and Y are uncorrelated, then $\mathrm{VAR}\left[aX + bY + c\right] = a^2\mathrm{VAR}\left[X\right] + b^2\mathrm{VAR}\left[Y\right]$

Major Points from this Lecture:

- Computing the distribution of the sum of 2 random variables

- Adding independent RVs -> Convolution of distributions!
  - Be careful about regions where distributions are zero in your integral!
  - Draw figures as much as possible.

- Mean and Variance of Affine (linear plus a constant) Combinations of Pairs of Random Variables

ELEC 2600H: Probability and Random Processes in Engineering

Part III: Multiple Random Variables

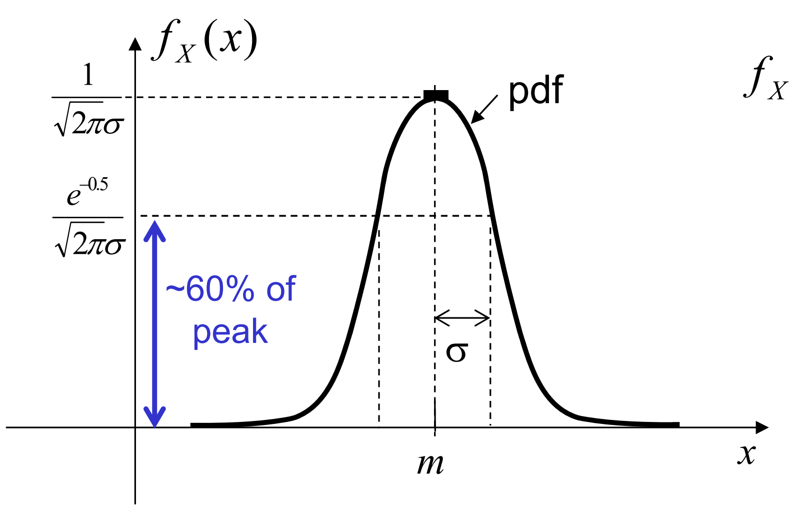
- Lecture 10: Pairs of Discrete Random Variable
  - Out-of-Class Reading: 2D Calculus
- Lecture 11: Pairs of Continuous Random Variable
- Lecture 12: Independence, Joint Moments
- Lecture 13: Correlation Coefficients and Their Properties
- Lecture 14: Conditional PDF, Conditional Expectation
- Lecture 15: Sum of Two Random Variables
- **Lecture 16: Pairs of Jointly Gaussian Random Variables**
- Lecture 17: More than Two Random Variables
- Lecture 18: Laws of Large Numbers
- Lecture 19: Central Limit Theorem and Characteristic Function

Elec2600H: Lecture 16

- Single Gaussian Random Variable
- Pairs of Gaussian Random Variables

Gaussian Random Variable

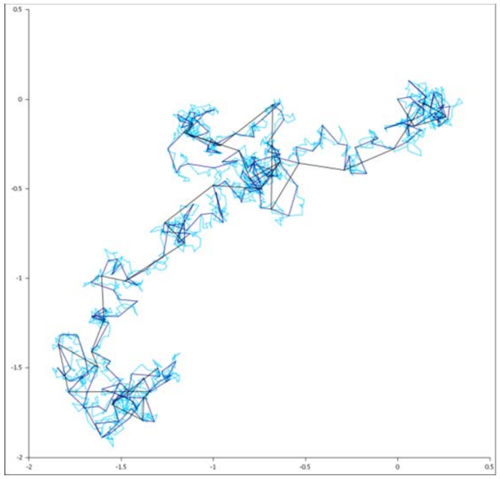
- The Gaussian random variable is used to model variables that tend to occur around a certain value, m, called the mean.
- This random variable is so common that it is also called the "normal" random variable.
- It can assume any value on the real line between -∞ and ∞.
- Applications
  - Noise voltages that corrupt transmission signals.
  - Voltage across a resistor

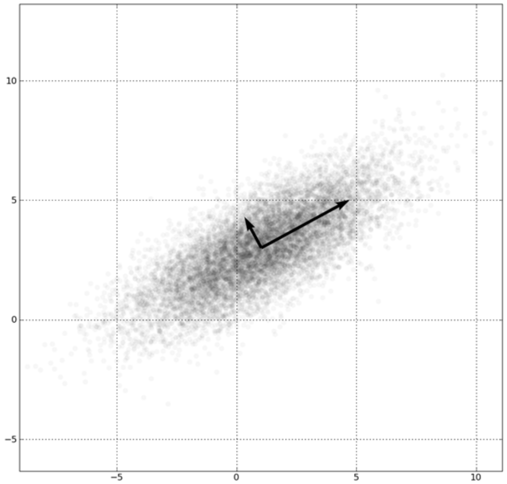


$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-m)^2}{2\sigma^2}}$$

Elec2600H: Lecture 16

- Single Gaussian Random Variable
- Pairs of Gaussian Random Variables
  - o Applications
    - Position of particle undergoing Brownian motion
    - Joint distribution of height and weight





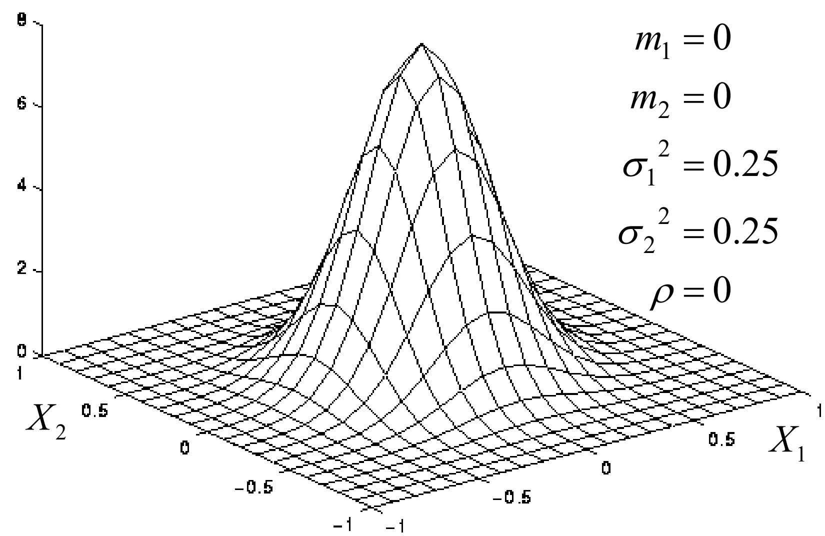
Example 6.20 Two Jointly Gaussian RVs

If X_1 and X_2 are jointly Gaussian, their pdf is given by

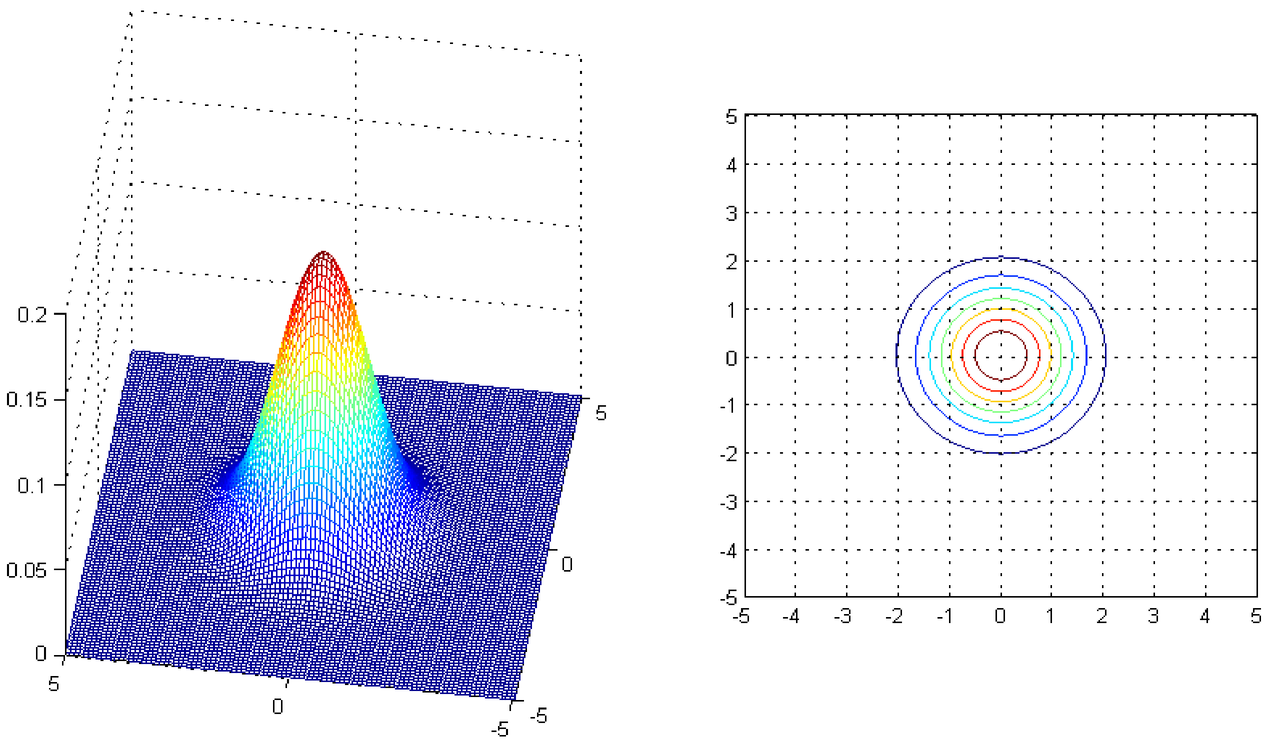
$$f_{X_1X_2}(x_1,x_2)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{\frac{(x_1-m_1)^2}{\sigma_1^{\ 2}}-2\rho\frac{x_1-m_1}{\sigma_1}\frac{x_2-m_2}{\sigma_2}+\frac{(x_2-m_2)^2}{\sigma_2^{\ 2}}}{2(1-\rho^2)}\right\}$$

where $m_i=E[X_i]$, $\sigma_i^{\ 2}=\mathrm{Var}\left[X_i\right]$ and ρ is the correlation coefficient.

The pdf is a bell-shaped hill centered at (m_1, m_2).



PDF of Two Independent Jointly Gaussian RVs

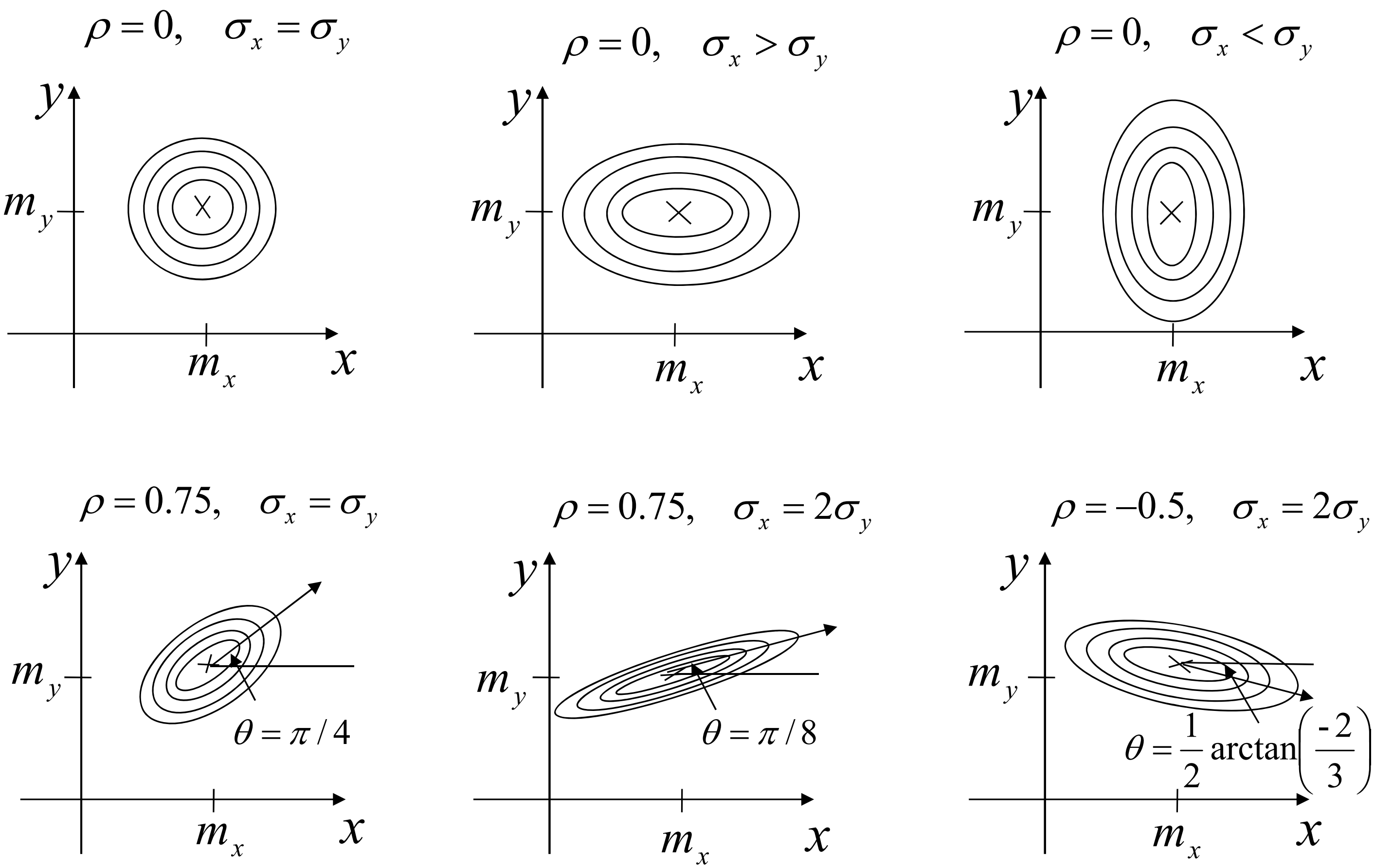


$m_1 = 0 \qquad m_2 = 0$

$\sigma_1{}^2 = 1 \qquad \sigma_2{}^2 = 1$

$\rho = 0$

Contour diagrams of the pdf



Vector notation

- Define the random 2D vector $\vec{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$

- The expectation of a vector or matrix is defined element-wise:
  - $E[\vec{X}] = E\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} E[X_1] \\ E[X_2] \end{bmatrix}$
  - $E\begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix} = \begin{bmatrix} E[X_{11}] & E[X_{12}] \\ E[X_{21}] & E[X_{22}] \end{bmatrix}$

- Mean vector: $E[\vec{X}] = \begin{bmatrix} E[X_1] \\ E[X_2] \end{bmatrix}$
  - Also denoted by $\vec{m} = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix}$ or $\vec{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$

Covariance Matrix

- The **covariance matrix**, $\boldsymbol{C}$, $\boldsymbol{K}$, or $\boldsymbol{\Sigma}$ contains all pairwise covariances between the elements of a random vector

$$
\begin{aligned}
\mathbf{C} &= E[(\vec{X} - E[\vec{X}])(\vec{X} - E[\vec{X}])^T] \\
&= E\left[\begin{bmatrix} X_1 - E[X_1] \\ X_2 - E[X_2] \end{bmatrix} \begin{bmatrix} X_1 - E[X_1] & X_2 - E[X_2] \end{bmatrix}\right] \\
&= \begin{bmatrix} \mathrm{E}\left[(X_1 - E[X_1])^2\right] & \mathrm{E}\left[(X_1 - E[X_1])(X_2 - E[X_2])\right] \\ \mathrm{E}\left[(X_2 - E[X_2])(X_1 - E[X_1])\right] & \mathrm{E}\left[(X_2 - E[X_2])^2\right] \end{bmatrix} \\
&= \begin{bmatrix} \mathrm{VAR}(X_1) & \mathrm{COV}(X_1, X_2) \\ \mathrm{COV}(X_2, X_1) & \mathrm{VAR}(X_2) \end{bmatrix} \\
&= \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}
\end{aligned}
$$

$$
\begin{aligned}
\sigma_1^2 &= \mathrm{VAR}(X_1) \\
\sigma_2^2 &= \mathrm{VAR}(X_2) \\
\rho &= \frac{\mathrm{COV}(X_1, X_2)}{\sqrt{\mathrm{VAR}(X_1)}\sqrt{\mathrm{VAR}(X_2)}}
\end{aligned}
$$

Alternative Notation

The RVs X_1 and X_2 are said to be jointly Gaussian (or jointly normal) if their joint pdf is given by

$$f_{\vec{X}}(\vec{x}) = \frac{1}{2\pi\sqrt{|\mathbf{C}|}} \exp\left\{-\frac{1}{2}\left(\vec{x}-\vec{m}\right)^T \mathbf{C}^{-1}\left(\vec{x}-\vec{m}\right)\right\}$$

where

$$\vec{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \qquad \vec{m} = E[\vec{X}] = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix}$$

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad \mathbf{C} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

$$|\mathbf{C}| = \det\mathbf{C} \text{ (the determinant of } \mathbf{C}\text{)}$$

$$=(1\text{-}\rho^2)\sigma_1^2\sigma_2^2$$

Justification (FYI)

$$\mathbf{C} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

$$|\mathbf{C}| = \sigma_1{}^2\sigma_2{}^2(1-\rho^2)$$

$$\mathbf{C}^{-1} = \frac{1}{(1-\rho^2)\sigma_1{}^2\sigma_2{}^2}\begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix}$$

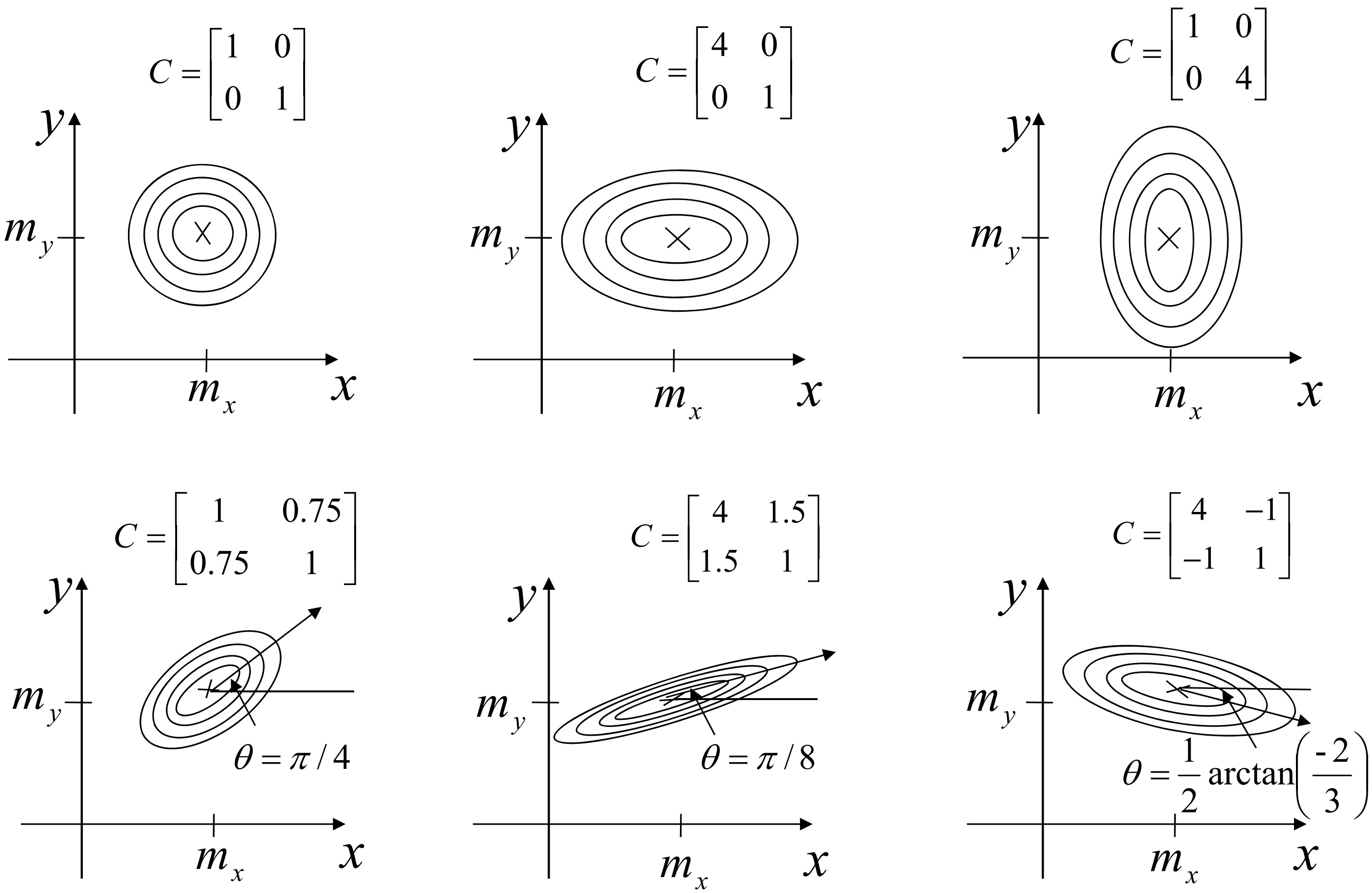
$$\frac{1}{2}(\vec{x}-\vec{m})^T\,\mathbf{C}^{-1}(\vec{x}-\vec{m}) =$$

$$= \frac{1}{2\sigma_1^2\sigma_2^2(1-\rho^2)}\begin{bmatrix} x_1-m_1 & x_2-m_2 \end{bmatrix}\begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix}\begin{bmatrix} x_1-m_1 \\ x_2-m_2 \end{bmatrix}$$

$$= \frac{1}{2\sigma_1^2\sigma_2^2(1-\rho^2)}\begin{bmatrix} x_1-m_1 & x_2-m_2 \end{bmatrix}\begin{bmatrix} \sigma_2^2(x_1-m_1)-\rho\sigma_1\sigma_2(x_2-m_2) \\ -\rho\sigma_1\sigma_2(x_1-m_1)+\sigma_1^2(x_2-m_2) \end{bmatrix}$$

$$= \frac{1}{2(1-\rho^2)}\left[\frac{(x_1-m_1)^2}{\sigma_1{}^2} - 2\rho\frac{x_1-m_1}{\sigma_1}\frac{x_2-m_2}{\sigma_2} + \frac{(x_2-m_2)^2}{\sigma_2{}^2}\right]$$

Contour diagrams of the pdf



Properties of Pairs of Gaussian RVs

Assume that X_1 and X_2 are jointly Gaussian.

- The marginal densities of X_1 and X_2 are Gaussian.
- If X_1 and X_2 are uncorrelated, they are also independent.
- The conditional density X_1 given X_2 is Gaussian.
- Any affine combination of X_1 and X_2, i.e., $Y = aX_1 + bX_2 + c$ is Gaussian.

  Note that this last property implies that we **do not need to integrate** to find the density of Y. Just compute the mean and variance of Y, and plug them into the formula for the Gaussian density

$$E[Y] = aE[X_1] + bE[X_2] + c$$
$$\mathrm{Var}[Y] = a^2\mathrm{Var}[X_1] + 2ab\mathrm{Cov}[X_1, X_2] + b^2\mathrm{Var}[X_2]$$

Example 5.18: Marginal density of X

Suppose that X and Y have joint distribution given by $f_{X,Y}(x,y) = \dfrac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(\dfrac{-(x^2 - 2\rho xy + y^2)}{2(1-\rho^2)}\right)$

Find the marginal density of X.

Solution

$$
\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy \\
&= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \exp\left(\frac{-(x^2 - 2\rho xy + y^2)}{2(1-\rho^2)}\right) dy \\
&= \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(\frac{-x^2}{2(1-\rho^2)}\right) \int_{-\infty}^{\infty} \exp\left(\frac{-(y^2 - 2\rho xy)}{2(1-\rho^2)}\right) dy \\
&= \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(\frac{-x^2 + \rho^2 x^2}{2(1-\rho^2)}\right) \int_{-\infty}^{\infty} \exp\left(\frac{-(y^2 - 2\rho xy + \rho^2 x^2)}{2(1-\rho^2)}\right) dy \\
&= \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right) \times \underbrace{\frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \exp\left(\frac{-(y-\rho x)^2}{2(1-\rho^2)}\right) dy}_{\substack{=\,1 \text{ since Gaussian with mean } \rho x \\ \text{and variance } 1-\rho^2.}} \\
&= \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right)
\end{aligned}
$$

factor out x^2

complete the square

simplify

Example: Uncorrelated Gaussian RVs

Suppose that X_1 and X_2 are jointly Gaussian: $f_{X_1X_2}(x_1,x_2) = \dfrac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{ -\dfrac{\frac{(x_1-m_1)^2}{\sigma_1^2} - 2\rho \frac{x_1-m_1}{\sigma_1}\frac{x_2-m_2}{\sigma_2} + \frac{(x_2-m_2)^2}{\sigma_2^2}}{2(1-\rho^2)} \right\}$

Find their joint density if they are uncorrelated.

Solution

Since they are uncorrelated, $\rho = 0$. Thus,

$$
\begin{aligned}
f_{X_1X_2}(x_1,x_2) &= \frac{1}{2\pi\sigma_1\sigma_2} \exp\left\{ -\frac{(x_1-m_1)^2}{2\sigma_1^2} - \frac{(x_2-m_2)^2}{2\sigma_2^2} \right\} \\
&= \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{ -\frac{(x_1-m_1)^2}{2\sigma_1^2} \right\} \times \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left\{ -\frac{(x_2-m_2)^2}{2\sigma_2^2} \right\}
\end{aligned}
$$

Since marginal densities of X_1 and X_2 are Gaussian, the joint density is the product of the marginals. Thus they must be independent.

Note that the covariance matrix is diagonal.

Example: Sum of Independent Gaussian RVs

Suppose that X_1 and X_2 are independent Gaussian RVs with zero mean and the same variance σ^2:

$$f_{X_1X_2}(x_1,x_2)=\frac{1}{2\pi\sigma^2}\exp\left\{-\frac{x_1^{\,2}+x_2^{\,2}}{2\sigma^2}\right\}$$

Find the density of $Z = X_1 + X_2$.

Solution

$$\begin{aligned}
f_Z(z) &= \int_{-\infty}^{\infty} f_{X_1X_2}(x,z-x)dx \\
&= \int_{-\infty}^{\infty}\frac{1}{2\pi\sigma^2}\exp\left\{-\frac{x^2+(z-x)^2}{2\sigma^2}\right\}dx \\
&= \frac{1}{\sqrt{2\pi}(\sqrt{2}\sigma)}\exp\left\{-\frac{z^2}{2(2\sigma^2)}\right\}\times\overbrace{\frac{1}{\sqrt{2\pi}\left(\frac{\sigma}{\sqrt{2}}\right)}\int_{-\infty}^{\infty}\exp\left\{-\frac{(x-\frac{z}{2})^2}{2(\frac{\sigma}{\sqrt{2}})^2}\right\}dx}^{=1} \\
&= \underbrace{\frac{1}{\sqrt{2\pi}(\sqrt{2}\sigma)}\exp\left\{-\frac{z^2}{2(2\sigma^2)}\right\}}_{\text{Gaussian with variance } 2\sigma^2}
\end{aligned}$$

$$\begin{aligned}
x^2+(z-x)^2 &= x^2+z^2-2xz+x^2 \\
&= 2x^2-2xz+z^2 \\
&= 2x^2-2xz+\tfrac{z^2}{2}+\tfrac{z^2}{2} \\
&= 2\left(x^2-xz+\tfrac{z^2}{4}\right)+\tfrac{z^2}{2} \\
&= 2\left(x-\tfrac{z}{2}\right)^2+\tfrac{z^2}{2}
\end{aligned}$$

Affine Functions of Gaussian Random Variables

- The previous example **proved** that the sum of two independent Gaussian Random variables is a Gaussian random variable by integrating along the line $Z = X_1 + X_2$.
- However, unless you want a rigorous proof, **avoid integrating the density.**
- Instead, use the fact that many operations with Gaussian random variables result in Gaussian variables (see slide on properties of Gaussians).
  - Find the means, variances and covariances of the resulting Gaussians.
  - Plug them into the formula for the density.

- In particular, if X_1 and X_2 are Gaussian, then $Y = aX_1 + bX_2 + c$ is Gaussian for any constants a, b and c, with

$$E[Y] = aE[X_1] + bE[X_2] + c$$
$$\mathrm{Var}[Y] = a^2\mathrm{Var}[X_1] + 2ab\mathrm{Cov}[X_1, X_2] + b^2\mathrm{Var}[X_2]$$

Example: Sum of Independent Gaussian RVs

Suppose that X_1 and X_2 are independent Gaussian RVs with zero mean and the same variance σ^2:

Find the density of $Z = X_1 + X_2$.

Solution

Z is Gaussian with mean $E[Z] = E[X_1] + E[X_2] = 0$ and variance $\mathrm{Var}[Z] = \mathrm{Var}[X_1] + \mathrm{Var}[X_2] = 2\sigma^2$ (because independent)

Thus, the density is given by

$$f_Z(z) = \frac{1}{\sqrt{2\pi}\left(\sqrt{Var[Z]}\right)} \exp\left\{-\frac{(z - E[Z])^2}{2Var[Z]}\right\} = \frac{1}{\sqrt{2\pi}(\sqrt{2}\sigma)} \exp\left\{-\frac{z^2}{2(2\sigma^2)}\right\}$$

(compare with previous example)

Example: Affine function of two Gaussian RVs

Suppose that X_1 and X_2 are jointly Gaussian with $E[X_1] = 1$, $\mathrm{Var}[X_1] = 4$, $E[X_2] = -2$, $\mathrm{Var}[X_2] = 9$, and correlation coefficient $\rho = 0.5$. Find the density of

$$Y = -2X_1 + X_2 + 3$$

Solution

$$\begin{aligned} E[Y] &= -2E[X_1] + E[X_2] + 3 \\ &= -2 \times 1 + (-2) + 3 \\ &= -1 \end{aligned}$$

Since $\mathrm{Cov}[X_1, X_2] = \rho\sqrt{\mathrm{Var}[X_1]}\sqrt{\mathrm{Var}[X_2]} = 0.5 \times 2 \times 3 = 3$,

$$\begin{aligned} \mathrm{Var}[Y] &= (-2)^2\mathrm{Var}[X_1] + 2(-2)(1)\mathrm{Cov}[X_1, X_2] + (1)^2\mathrm{Var}[X_2] \\ &= 4 \times 4 - 4 \times 3 + 1 \times 9 = 13 \end{aligned}$$

Thus,

$$f_Y(z) = \frac{1}{\sqrt{2\pi}\left(\sqrt{\mathrm{Var}[Y]}\right)} \exp\left\{-\frac{(y - E[Y])^2}{2\mathrm{Var}[Y]}\right\} = \frac{1}{\sqrt{2\pi}\sqrt{13}} \exp\left\{-\frac{(y+1)^2}{26}\right\}$$

Major Points from this Lecture:

- Gaussian Random Variables

- Mean Vector and Covariance Matrix

- Many operations on Gaussian Random Variables result in new Gaussian RVs
  - This implies that you do not need to do integrations!
  - Just compute means and variances and use the Gaussian formula.

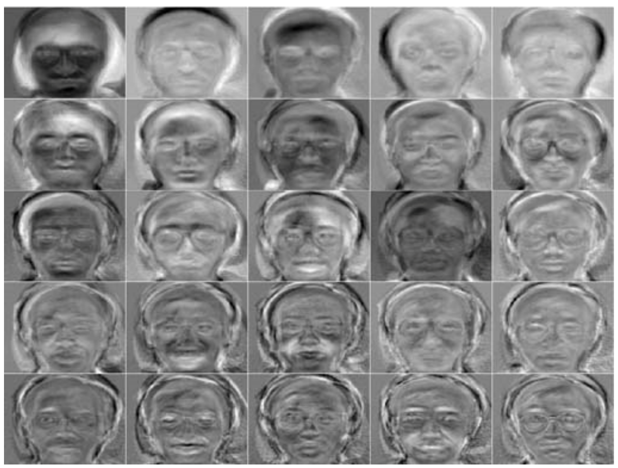
ELEC 2600H: Probability and Random Processes in Engineering

Part III: Multiple Random Variables

- Lecture 10: Pairs of Discrete Random Variable
  - Out-of-Class Reading: 2D Calculus
- Lecture 11: Pairs of Continuous Random Variable
- Lecture 12: Independence, Joint Moments
- Lecture 13: Correlation Coefficients and Their Properties
- Lecture 14: Conditional PDF, Conditional Expectation
- Lecture 15: Sum of Two Random Variables
- Lecture 16: Pairs of Jointly Gaussian Random Variables
- **Lecture 17: More than Two Random Variables**
- Lecture 18: Laws of Large Numbers
- Lecture 19: Central Limit Theorem and Characteristic Function

Elec2600H: Lecture 17

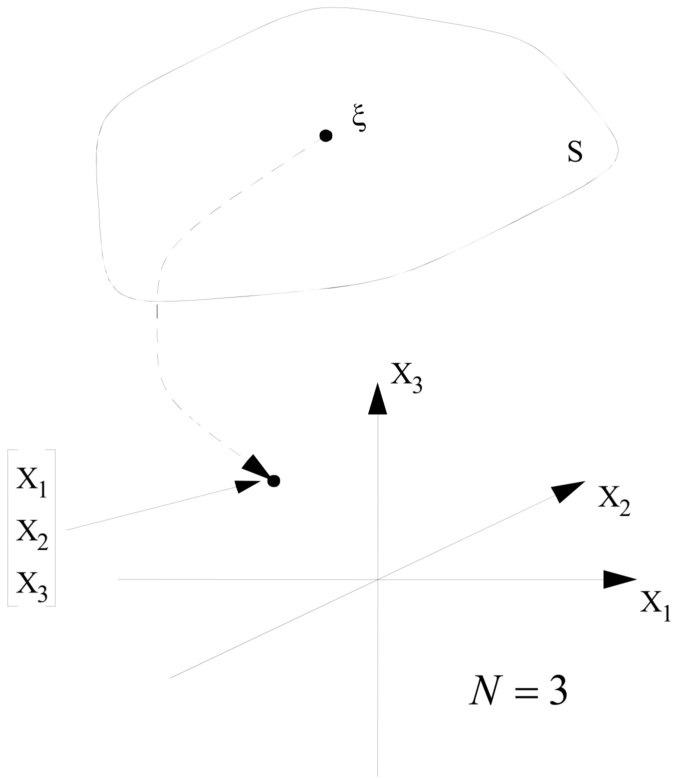
- **Random Vectors**
  - **Joint distribution/density/mass functions**
  - Marginal statistics
  - Conditional densities
  - Independence and Expectation



http://www.cs.princeton.edu/~cdecoro/eigenfaces/

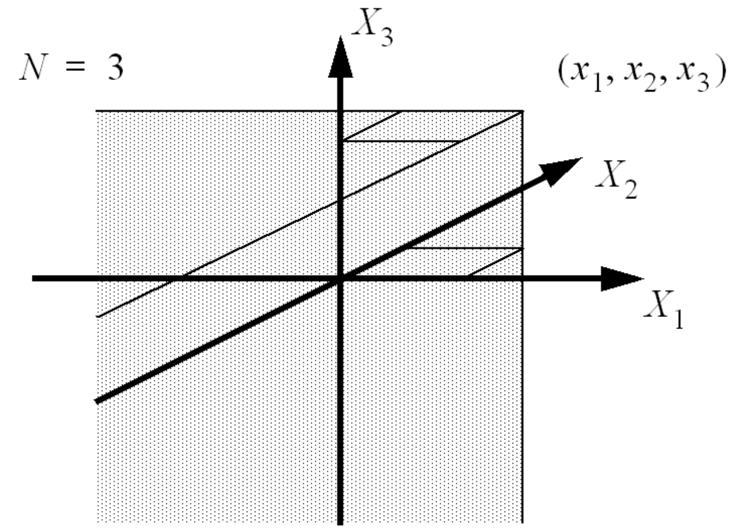
N Random Variables

- An N dimensional random vector is a mapping from a probability space S to R^N
- It is often convenient to think of n RVs as a single n-dimensional random vector: $\vec{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$



Joint Cumulative Distribution Function

- $F_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n) = P[X_1 \le x_1, X_2 \le x_2, \ldots\ldots, X_n \le x_n]$



- To eliminate a variable:

$$F_{X_1,X_2,\ldots,X_{n-1}}(x_1,x_2,\ldots,x_{n-1}) = F_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_{n-1},\infty)$$

Joint Probability Density/Mass Functions

- Joint Probability Density Function $f_{X_1,X_2,...,X_n}(x_1,x_2,...,x_n) = \frac{\partial^n}{\partial x_1 ... \partial x_n} F_{X_1,X_2,...,X_n}(x_1,x_2,...,x_n)$
  - For any set A, $P[A] = \iiint\limits_A \ldots \int f_{X_1,X_2,...,X_n}(x_1,x_2,...,x_n)dx_1 dx_2 \ldots dx_n$

- Joint Probability Mass Function $p_{X_1,X_2,...,X_n}(k_1,k_2,...,k_n) = P[X_1 = k_1, X_2 = k_2,....., X_n = k_n]$
  - For any set A, $P[A] = \sum_{\{k_1,k_2,...,k_n\} \in A} p_{X_1 X_2 \ldots X_n}(k_1,k_2,...,k_n)$

The Multinomial Distribution

- The multinomial distribution is a generalization of the binomial distribution, where instead of two possible outcomes for each of the n trials, we have m possible outcomes, each with probability p_j for $j \in \{1, \ldots, m\}$.

- Since the outcomes are mutually exclusive, $p_1 + p_2 + ... + p_m = 1$

- Let X_j be the number of times outcome j occurs. Note that $X_1 + X_2 + ... + X_m = n$

- The **multinomial distribution** gives the probability of the random vector $\vec{X} = [X_1 \quad X_2 \quad \ldots \quad X_m]^T$

$$p_{X_1, X_2, \ldots, X_m}(k_1, k_2, \ldots, k_m) = \frac{n!}{k_1!k_2!....k_m!} p_1^{k_1} p_2^{k_2} ....p_m^{k_m}$$

if $k_1 + k_2 + \ldots + k_m = n$ and $k_i \geq 0$ for all i, and zero otherwise

Elec2600H: Lecture 17

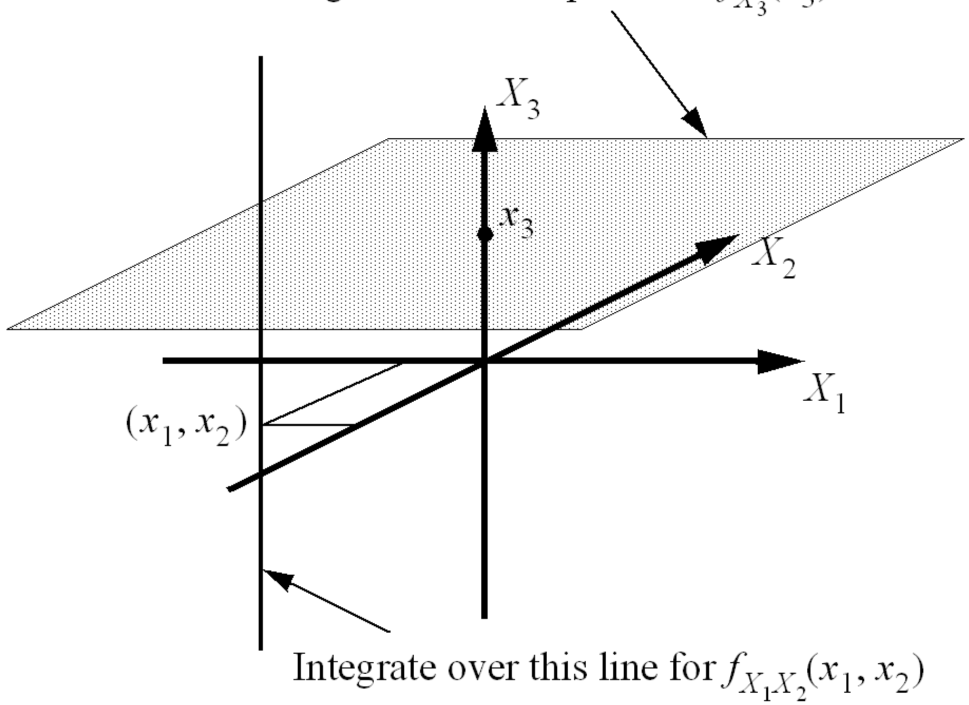
- **Random Vectors**
  - Joint distribution/density/mass functions
  - **Marginal statistics**
  - Conditional densities
  - Independence and Expectation

Marginal Statistics

- Eliminate variables from a pdf (pmf) by integrating (summing)

$$f_{X_1X_2\ldots X_{n-1}}(x_1,x_2,\ldots,x_{n-1}) = \int_{-\infty}^{\infty} f_{X_1X_2\ldots X_n}(x_1,x_2,x_3,\ldots,x_n)\,dx_n$$

$$p_{X_1X_2\ldots X_{n-1}}(k_1,k_2,\ldots,k_{n-1}) = \sum_{k_n=-\infty}^{\infty} p_{X_1X_2\ldots X_n}(k_1,k_2,k_3,\ldots,k_n)$$



- Get marginal by *successive integration*:

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_{X_1X_2\ldots X_n}(x_1,x_2,x_3,\ldots,x_n)\,dx_2\ldots dx_n$$

$$p_{X_1}(k_1) = \sum_{k_2=-\infty}^{\infty}\cdots\sum_{k_n=-\infty}^{\infty} p_{X_1X_2\ldots X_n}(k_1,k_2,k_3,\ldots,k_N)$$

Example: Computer System

A computer system receives messages over three communications lines.
Let X_j be the number of messages received on line j in one hour.

The joint pmf of X_1, X_2, X_3 is: $p_{X_1,X_2,X_3}(k_1,k_2,k_3) = \begin{cases} (1-a_1)(1-a_2)(1-a_3)a_1^{k_1}a_2^{k_2}a_3^{k_3} & \text{if } k_1 \geq 0, k_2 \geq 0, k_3 \geq 0, \\ 0 & \text{otherwise} \end{cases}$

Find $p_{X_1,X_2}(k_1,k_2)$ and $p_{X_1}(k_1)$

$a_1 \geq 0, a_2 \geq 0, a_3 \geq 0$ constant

Solution

$$\begin{aligned} p_{X_1,X_2}(k_1,k_2) &= (1-a_1)(1-a_2)(1-a_3)a_1^{k_1}a_2^{k_2} \times \sum_{k_3=0}^{\infty} a_3^{k_3} \\ &= (1-a_1)(1-a_2)(1-a_3)a_1^{k_1}a_2^{k_2} \times \frac{1}{(1-a_3)} \\ &= (1-a_1)(1-a_2)a_1^{k_1}a_2^{k_2} \quad \text{for} \quad k_1, k_2 \geq 0 \end{aligned}$$

$$\begin{aligned} p_{X_1}(k_1) &= (1-a_1)(1-a_2)a_1^{k_1} \times \sum_{k_2=0}^{\infty} a_2^{k_2} \\ &= (1-a_1)(1-a_2)a_1^{k_1} \times \frac{1}{1-a_2} = (1-a_1)a_1^{k_1} \quad \text{for} \quad k_1 \geq 0 \end{aligned}$$

Elec2600H: Lecture 17

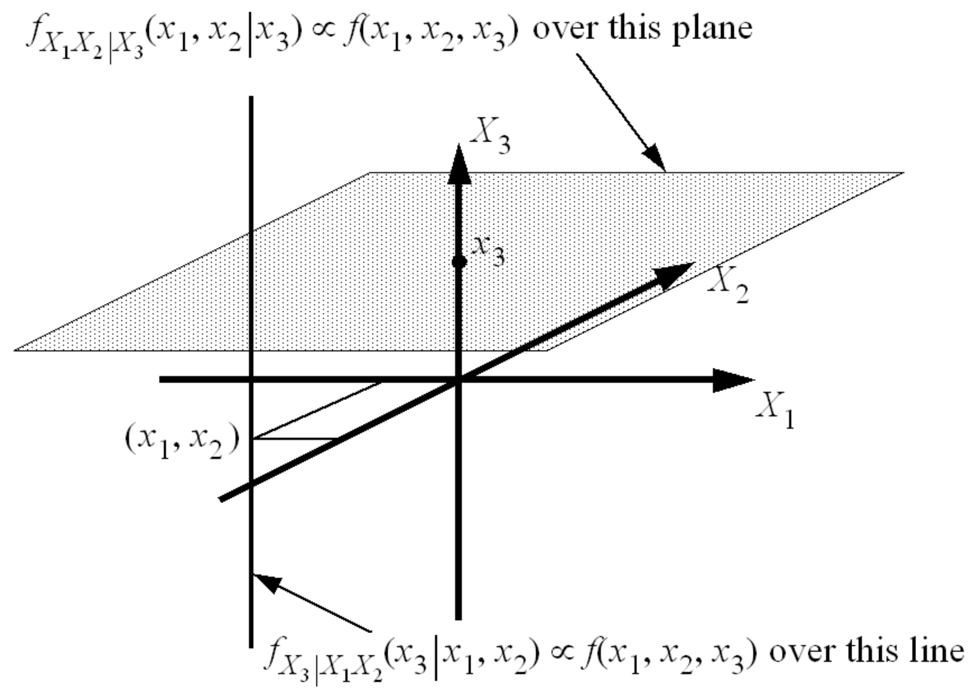
- **Random Vectors**
  - Joint distribution/density/mass functions
  - Marginal statistics
  - **Conditional densities**
  - Independence and expectation

Conditional pdf/pmf

Conditional pdf

$$f_{X_1X_2|X_3}(x_1,x_2 \mid x_3)=\frac{f_{X_1X_2X_3}(x_1,x_2,x_3)}{f_{X_3}(x_3)}$$

$$f_{X_1|X_2X_3}(x_1 \mid x_2,x_3)=\frac{f_{X_1X_2X_3}(x_1,x_2,x_3)}{f_{X_2X_3}(x_2,x_3)}$$



Conditional pmf

$$p_{X_1X_2|X_3}(k_1,k_2 \mid k_3)=\frac{p_{X_1X_2X_3}(k_1,k_2,k_3)}{p_{X_3}(k_3)}$$

$$p_{X_1|X_2X_3}(k_1 \mid k_2,k_3)=\frac{p_{X_1X_2X_3}(k_1,k_2,k_3)}{p_{X_2X_3}(k_2,k_3)}$$

Product rule

- From the definition of the conditional probability, $p_{X_1|X_2X_3}(k_1 \mid k_2, k_3) = \dfrac{p_{X_1X_2X_3}(k_1, k_2, k_3)}{p_{X_2X_3}(k_2, k_3)}$

  we get the following product rule: $p_{X_1X_2X_3}(k_1, k_2, k_3) = p_{X_1|X_2X_3}(k_1 \mid k_2, k_3)\, p_{X_2X_3}(k_2, k_3)$

- Combining with the product rule for two RVs, $p_{X_2X_3}(k_2, k_3) = p_{X_2|X_3}(k_2 \mid k_3)\, p_{X_3}(k_3)$

  we obtain the following formula, which breaks the generation of three random variables into a sequence of three steps:

$$p_{X_1X_2X_3}(k_1, k_2, k_3) = p_{X_1|X_2X_3}(k_1 \mid k_2, k_3)\, p_{X_2|X_3}(k_2 \mid k_3)\, p_{X_3}(k_3)$$

Elec2600H: Lecture 17

- **Random Vectors**
  - Joint distribution/density/mass functions
  - Marginal statistics
  - Conditional densities
  - **Independence and expectation**

Independence

- The following statements are **equivalent!**

1. $X_1, X_2, \ldots, X_n$ are independent, i.e., for all $A_1, A_2, \ldots, A_n$

$$P[X_1 \in A_1, X_2 \in A_2, \ldots. X_n \in A_n] = P[X_1 \in A_1]P[X_2 \in A_2]\ldots.P[X_n \in A_n]$$

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

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

Random Vectors—First Moment

- The **expectation** of a vector is taken element by element:

$$\vec{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \qquad E[\vec{X}] = \begin{bmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_n] \end{bmatrix} \quad \text{the mean vector}$$

- The expectation of a matrix is also taken element by element

$$\mathbf{X} = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix} \qquad E[\mathbf{X}] = \begin{bmatrix} E[X_{11}] & E[X_{12}] \\ E[X_{21}] & E[X_{22}] \end{bmatrix}$$

Linearity of the expectation

- The expectation operator on multiple random variables is linear. Therefore, we can switch the order of expectation and any linear operator, including
  - Multiplication by a constant: $E[aX_i] = aE[X_i]$
  - Addition of a constant: $E[X_i + c] = E[X_i] + c$
  - Addition:
    - $E[X_i + X_j] = E[X_i] + E[X_j]$
    - $E[X_i + X_j + X_k] = E[X_i] + E[X_j] + E[X_k]$
  - Any combination of the above, e.g. if a_i are constants, then
    - $E[a_0 + a_1X_1 + a_2X_2 + a_3X_3] = a_0 + a_1E[X_1] + a_2E[X_2] + a_3E[X_3]$
    - $E[a_0 + \sum_{i=1}^{n} a_iX_i] = a_0 + \sum_{i=1}^{n} a_iE[X_i]$

- However, we cannot switch the order of expectation and any nonlinear function, i.e. if $g(\cdot)$ is nonlinear, then $E[g(X_1, X_2, \dots, X_n)] \neq g(E[X_1], E[X_2], \dots, E[X_n])$

Correlation Matrix

- The **correlation matrix**, $\boldsymbol{R}$, contains all pairwise correlations between the elements of a random vector

$$R = E[\vec{X}\vec{X}^T] = E\left[\begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}\begin{bmatrix} X_1 & \dots & X_n \end{bmatrix}\right]$$

$$= \begin{bmatrix} E[X_1X_1] & E[X_1X_2] & \dots & E[X_1X_n] \\ E[X_2X_1] & E[X_2X_2] & & \\ \vdots & & \ddots & \\ E[X_nX_1] & & & E[X_nX_n] \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} & \dots & R_{1n} \\ R_{21} & R_{22} & & \\ \vdots & & \ddots & \\ R_{n1} & & & R_{nn} \end{bmatrix} \qquad R_{ij} = E[X_iX_j]$$

Covariance Matrix

- The **covariance matrix**, ***C*** or ***K***, contains all pairwise covariances between the elements of a random vector

$$\mathbf{C} = E[(\vec{X} - E[\vec{X}])(\vec{X} - E[\vec{X}])^T]$$

$$= E\left[\begin{bmatrix} X_1 - E[X_1] \\ \vdots \\ X_n - E[X_n] \end{bmatrix} \begin{bmatrix} X_1 - E[X_1] & \dots & X_n - E[X_n] \end{bmatrix}\right]$$

$$= \begin{bmatrix} \mathrm{VAR}(X_1) & \mathrm{COV}(X_1, X_2) & \dots & \mathrm{COV}(X_1, X_n) \\ \mathrm{COV}(X_2, X_1) & \mathrm{VAR}(X_2) & & \\ \vdots & & \ddots & \\ \mathrm{COV}(X_n, X_1) & & & \mathrm{VAR}(X_n) \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} & \dots & C_{1n} \\ C_{21} & C_{22} & & \\ \vdots & & \ddots & \\ C_{n1} & & & C_{nn} \end{bmatrix}$$

$$C_{ij} = \begin{cases} \mathrm{VAR}(X_i) & i = j \\ \mathrm{COV}(X_i, X_j) & i \neq j \end{cases}$$

Relationship between Correlation and Covariance

The correlation and covariance matrices are related by $\mathbf{C} = \mathbf{R} - E[\vec{X}] \cdot E[\vec{X}]^T$

Proof:

$$
\begin{aligned}
\mathbf{C} &= E\left[(\vec{X} - E[\vec{X}]) \cdot (\vec{X} - E[\vec{X}])^T \right] \\
&= E\left[\vec{X} \cdot \vec{X}^T - \vec{X} \cdot E[\vec{X}]^T - E[\vec{X}] \cdot \vec{X}^T + E[\vec{X}] \cdot E[\vec{X}]^T \right] \\
&= E[\vec{X} \cdot \vec{X}^T] - E[\vec{X}] \cdot E[\vec{X}]^T - E[\vec{X}] \cdot E[\vec{X}]^T + E[\vec{X}] \cdot E[\vec{X}]^T \\
&= \mathbf{R} - E[\vec{X}] \cdot E[\vec{X}]^T
\end{aligned}
$$

Note that if the mean vector is zero, then $\mathbf{C} = \mathbf{R}$.

Properties of the Covariance Matrix

- The covariance matrix is symmetric.

  Proof:

$$C_{ij} = E[(X_i - E[X_i])(X_j - E[X_j])]$$
$$= E[(X_j - E[X_j])(X_i - E[X_i])] = C_{ji}$$

$$\begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ C_{21} & C_{22} & & \\ \vdots & & \ddots & \\ C_{n1} & & & C_{nn} \end{bmatrix}$$

e.g. $C_{12} = C_{21}, C_{1n} = C_{n1}$

- The covariance matrix is diagonal if and only if the elements of the random vector are uncorrelated.

  Proof:

  The off diagonal elements, C_{ij}, for $i \neq j$ are zero if and only if $C_{ij} = \mathrm{COV}(X_i, X_j) = 0$

$$\begin{bmatrix} C_{11} & 0 & \cdots & 0 \\ 0 & C_{22} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & C_{nn} \end{bmatrix}$$

the diagonal elements are the variances

Jointly Gaussian (Normal) Random Variables

- The RVs $X_1, X_2, \dots, X_n$ are said to be jointly Gaussian (or jointly normal) if their joint pdf is given by

$$f_{\vec{X}}(\vec{x}) = \frac{1}{(2\pi)^{n/2} |\mathbf{C}|^{1/2}} \exp\left\{ -\frac{1}{2} (\vec{x}-\vec{m})^T \mathbf{C}^{-1} (\vec{x}-\vec{m}) \right\}$$

where $\vec{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$ and $\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$

$\vec{m} = E[\vec{X}]$

$\mathbf{C}$ is the covariance matrix of $\vec{X}$

$|\mathbf{C}| = \det \mathbf{C}$ (the determinant of $\mathbf{C}$)

- Note that jointly Gaussian RVs are completely specified by their means and covariance.

Example

Suppose X_1, X_2 and X_3 are jointly Gaussian, with the first and second moments shown. Find the joint density.

$$
\begin{array}{lll}
E[X_1]=5 & E[X_1{}^2]=31 & E[X_1X_2]=16 \\
E[X_2]=3 & E[X_2{}^2]=13 & E[X_1X_3]=13 \\
E[X_3]=2 & E[X_3{}^2]=11 & E[X_2X_3]=8
\end{array}
$$

Solution:

The density is $f_{\vec{X}}(\vec{x})=\dfrac{1}{(2\pi)^{n/2}|C|^{1/2}}\exp\left\{-\dfrac{1}{2}(\vec{x}-\vec{m})^T C^{-1}(\vec{x}-\vec{m})\right\}$

where

$$\vec{m}=E[\vec{X}]=\begin{bmatrix}5\\3\\2\end{bmatrix}$$

$$C=R-E[\vec{X}]\cdot E[\vec{X}]^T=\begin{bmatrix}31&16&13\\16&13&8\\13&8&11\end{bmatrix}-\begin{bmatrix}5\\3\\2\end{bmatrix}\begin{bmatrix}5&3&2\end{bmatrix}$$

$$=\begin{bmatrix}31&16&13\\16&13&8\\13&8&11\end{bmatrix}-\begin{bmatrix}25&15&10\\15&9&6\\10&6&4\end{bmatrix}=\begin{bmatrix}6&1&3\\1&4&2\\3&2&7\end{bmatrix}$$

Properties of Gaussian Random Variables

Assume that $X_1, X_2, \ldots, X_n$ are jointly Gaussian.

1. Any subset of $X_1, X_2, \ldots, X_n$ is also jointly Gaussian.

2. In particular, the marginal density of any X_i is Gaussian.

3. If $X_1, X_2, \ldots, X_n$ are uncorrelated, they are also independent.

4. The conditional density of any subset of $X_1, X_2, \ldots, X_n$ conditioned on any other subset is Gaussian.

5. Any affine transformation of a Gaussian vector is also Gaussian,

   i.e. $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$ is Gaussian, where $\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$

Example 6.8

Suppose that the first n samples of a noise waveform, $X_1, X_2, \ldots X_n$ are independent and all have marginal density given by a zero mean, unit variance Gaussian. Find their joint density.

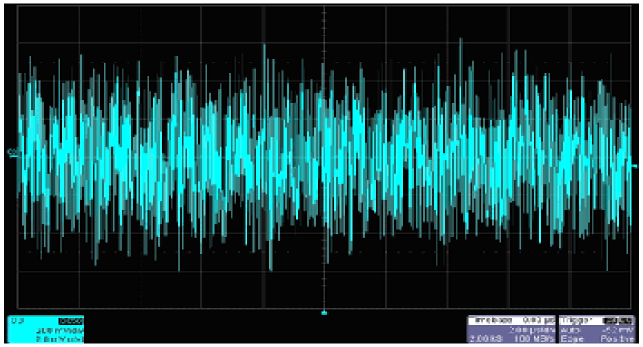
Solution

The marginal density of each X_i is $f_{X_i}(x_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_i^2}$ $\quad -\infty < x_i < \infty$

Thus, the joint density is $f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_i^2}$

$$= \frac{1}{\left(\sqrt{2\pi}\right)^n} e^{-\frac{1}{2}\left(x_1^2 + x_2^2 + \ldots + x_n^2\right)}$$

$$= \frac{1}{\left(\sqrt{2\pi}\right)^n} e^{-\frac{1}{2}\sum_{i=1}^{n} x_i^2}$$



Example

Suppose X_1, X_2 and X_3 are jointly Gaussian with the mean and covariance matrix shown. Find the joint density of X_1 and X_3.

$$\vec{m} = E[\vec{X}] = \begin{bmatrix} 5 \\ 3 \\ 2 \end{bmatrix} \qquad C = \begin{bmatrix} 6 & 1 & 3 \\ 1 & 4 & 2 \\ 3 & 2 & 7 \end{bmatrix}$$

Solution:

We know that X_1 and X_3 are jointly Gaussian, thus the density is given by:

$$f_{\vec{X}}(\vec{x}) = \frac{1}{(2\pi)^{n/2}|C|^{1/2}} \exp\left\{-\frac{1}{2}(\vec{x}-\vec{m})^T C^{-1}(\vec{x}-\vec{m})\right\}$$

The mean vector and covariance matrix are subsets of the mean/covariance above:

$$\vec{m} = E\begin{bmatrix} X_1 \\ X_3 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \end{bmatrix} \qquad C = \begin{bmatrix} 6 & 3 \\ 3 & 7 \end{bmatrix}$$

Conditional density of two Gaussians

Suppose X and Y are jointly Gaussian. Find the conditional density of X given Y.

Solution:

Since Y is Gaussian,

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)}$$

$$= \frac{\dfrac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\exp\left\{-\dfrac{\dfrac{(x-m_X)^2}{\sigma_X{}^2} - 2\rho\dfrac{x-m_X}{\sigma_X}\dfrac{y-m_Y}{\sigma_Y} + \dfrac{(y-m_Y)^2}{\sigma_Y{}^2}}{2(1-\rho^2)}\right\}}{\dfrac{1}{\sqrt{2\pi}\sigma_Y}\exp\left\{-\dfrac{1}{2}\dfrac{(y-m_Y)^2}{\sigma_Y{}^2}\right\}}$$

$$= \frac{1}{\sqrt{2\pi}\sigma_X\sqrt{1-\rho^2}}\exp\left\{-\frac{\dfrac{(x-m_X)^2}{\sigma_X{}^2} - 2\rho\dfrac{x-m_X}{\sigma_X}\dfrac{y-m_Y}{\sigma_Y} + \dfrac{(y-m_Y)^2}{\sigma_Y{}^2}}{2(1-\rho^2)} + \frac{1}{2}\frac{(y-m_Y)^2}{\sigma_Y{}^2}\right\}$$

Rearrange term inside exponential

$$-\frac{\dfrac{(x-m_X)^2}{{\sigma_X}^2}-2\rho\dfrac{x-m_X}{\sigma_X}\dfrac{y-m_Y}{\sigma_Y}+\dfrac{(y-m_Y)^2}{{\sigma_Y}^2}}{2(1-\rho^2)}+\frac{1}{2}\frac{(y-m_Y)^2}{{\sigma_Y}^2}$$

$$=-\frac{\dfrac{(x-m_X)^2}{{\sigma_X}^2}-2\rho\dfrac{x-m_X}{\sigma_X}\dfrac{y-m_Y}{\sigma_Y}+\dfrac{(y-m_Y)^2}{{\sigma_Y}^2}-(1-\rho^2)\dfrac{(y-m_Y)^2}{{\sigma_Y}^2}}{2(1-\rho^2)}$$

$$=-\frac{\dfrac{(x-m_X)^2}{{\sigma_X}^2}-2\rho\dfrac{x-m_X}{\sigma_X}\dfrac{y-m_Y}{\sigma_Y}+\rho^2\dfrac{(y-m_Y)^2}{{\sigma_Y}^2}}{2(1-\rho^2)}$$

$$=-\frac{\left(\dfrac{x-m_X}{\sigma_X}-\rho\dfrac{y-m_Y}{\sigma_Y}\right)^2}{2(1-\rho^2)}$$

$$=-\frac{1}{2}\frac{\left(x-m_X-\rho\dfrac{\sigma_X}{\sigma_Y}(y-m_Y)\right)^2}{\left(\sigma_X\sqrt{1-\rho^2}\right)^2}$$

Substitute the new expression into the exponential

$$f_{X|Y}(x|y) = \frac{1}{\sqrt{2\pi}\sigma_X\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2}\frac{\left(x - m_X - \rho\frac{\sigma_X}{\sigma_Y}(y - m_Y)\right)^2}{\left(\sigma_X\sqrt{1-\rho^2}\right)^2}\right\}$$

$$= \frac{1}{\sqrt{2\pi}\sqrt{\mathrm{Var}[X|y]}} \exp\left\{-\frac{1}{2}\frac{(x - \mathrm{E}[X|y])^2}{\mathrm{Var}[X|y]}\right\}$$

where

mean of X "unit conversion"

$$\mathrm{E}[X|y] = m_X + \rho\frac{\sigma_X}{\sigma_Y}(y - m_Y)$$ (the conditional mean of X given $Y = y$)

dependency between X and Y difference between observed and expected values of Y

$$\mathrm{Var}[X|y] = {\sigma_X}^2(1 - \rho^2)$$ (the conditional variance of X given $Y = y$)

variance of X reduction due to knowledge of Y

Example

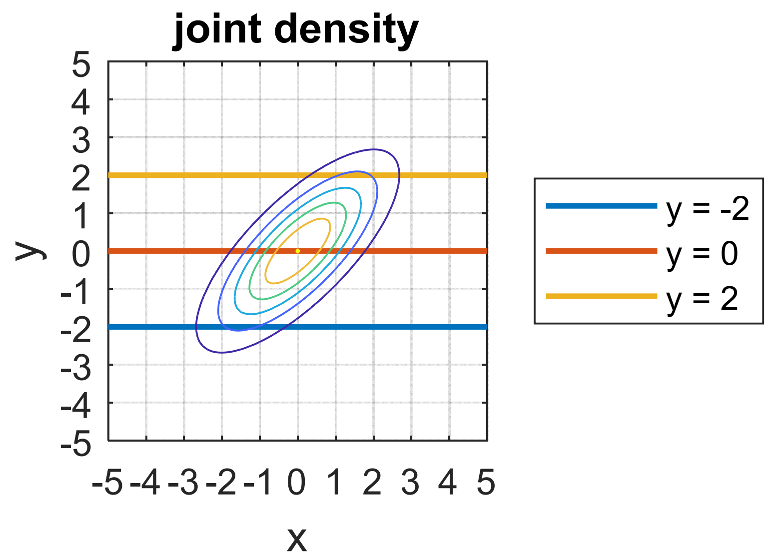
$m_X = 0$

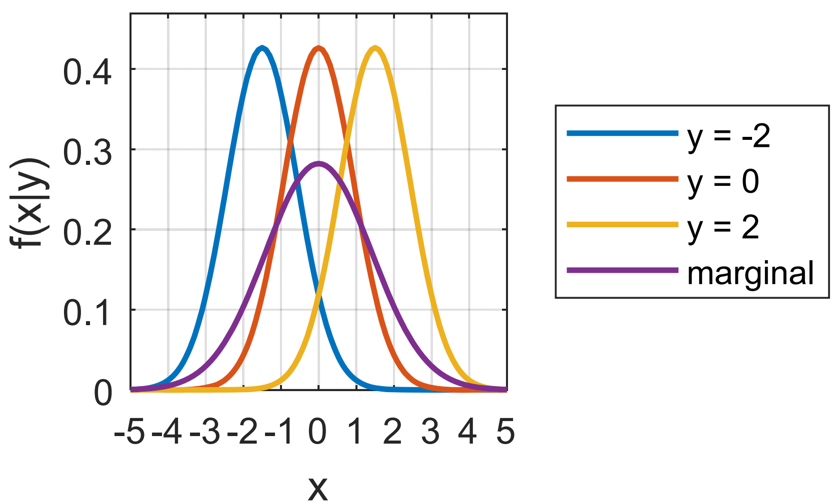
$m_Y = 0$

${\sigma_X}^2 = 2$

${\sigma_Y}^2 = 2$

$\rho_{XY} = 0.75$





Major Points from this Lecture:

- **Introduction to random vectors (beyond TWO variables!!)**
- **Distribution, density, pmf**
- **Marginals**
- **Conditional distributions**
- **Independence and expectation**
- **Covariance and correlation**
- **Gaussian Random Variables**

ELEC 2600H: Probability and Random Processes in Engineering

Part III: Multiple Random Variables

- Lecture 10: Pairs of Discrete Random Variable
  - Out-of-Class Reading: 2D Calculus
- Lecture 11: Pairs of Continuous Random Variable
- Lecture 12: Independence, Joint Moments
- Lecture 13: Correlation Coefficients and Their Properties
- Lecture 14: Conditional PDF, Conditional Expectation
- Lecture 15: Sum of Two Random Variables
- Lecture 16: Pairs of Jointly Gaussian Random Variables
- Lecture 17: More than Two Random Variables
- **Lecture 18: Laws of Large Numbers**
- Lecture 19: Central Limit Theorem and Characteristic Function

Elec2600H: Lecture 18

- **Sums of Random Variables**
- Mean and Variance of Sample Means
- Useful Inequalities
- Laws of Large Numbers

Sums of Random Variables

- For any set of random variables, $X_1, X_2, \ldots., X_n$

$$E\left[\sum_j X_j\right] = \sum_j E[X_j]$$

$$\mathrm{VAR}\left(\sum_{j=1}^{n} X_j\right) = \sum_{j=1}^{n} \mathrm{VAR}(X_j) + \sum_{j=1}^{n}\sum_{k\neq j} \mathrm{COV}(X_j, X_k)$$

- In particular, for two random variables, X and Y,

$$\mathrm{VAR}(X+Y) = \mathrm{VAR}(X) + \mathrm{VAR}(Y) + 2\mathrm{COV}(X,Y)$$

- If the random variables are uncorrelated,

$$\mathrm{VAR}\left(\sum_{j=1}^{n} X_j\right) = \sum_{j=1}^{n} \mathrm{VAR}(X_j)$$

Justification for Variance of Sum

$$\mathrm{VAR}\left(\sum_{j=1}^{n} X_j\right) = E\left[\left(\sum_{j=1}^{n} X_j - E\left[\sum_{j=1}^{n} X_j\right]\right)^2\right] = E\left[\left(\sum_{j=1}^{n} X_j - \sum_{j=1}^{n} E[X_j]\right)^2\right]$$

$$= E\left[\left(\sum_{j=1}^{n} (X_j - E[X_j])\right)^2\right] = E\left[\left(\sum_{j=1}^{n} (X_j - E[X_j])\right)\left(\sum_{k=1}^{n} (X_k - E[X_k])\right)\right]$$

$$= E\left[\sum_{j=1}^{n}\sum_{k=1}^{n} (X_j - E[X_j])(X_k - E[X_k])\right]$$

$$= \sum_{j=1}^{n}\sum_{k=1}^{n} E[(X_j - E[X_j])(X_k - E[X_k])]$$

$$= \sum_{j=1}^{n}\sum_{k=1}^{n} \mathrm{COV}(X_j, X_k) \quad \text{where} \quad \mathrm{COV}(X_j, X_j) = \mathrm{VAR}(X_j)$$

$$= \sum_{j=1}^{n} \mathrm{VAR}(X_j) + \sum_{j=1}^{n}\sum_{k \neq j} \mathrm{COV}(X_j, X_k)$$

Example 7.2: i.i.d. RV's

Suppose $X_1, X_2, \ldots, X_n$ are independent and identically distributed (i.i.d.), i.e., they all have the same distribution with mean and variance given by $\mu = E[X_i]$ and $\sigma^2 = \mathrm{VAR}(X_i)$

Let $S = \sum_{i=1}^{n} X_i$

Find $E[S]$ and $\mathrm{VAR}(S)$.

Solution:

$$E[S] = E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} \mu = n\mu$$

Both the mean and variance increase **linearly** with n

Since independence implies uncorrelated,

$$\mathrm{VAR}(S) = \mathrm{VAR}\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} \mathrm{VAR}(X_k) = n\sigma^2$$

Elec2600H: Lecture 18

- Sums of Random Variables
- **Mean and Variance of Sample Means**
- Useful Inequalities
- Laws of Large Numbers

The Sample Mean

- Let X be a random variable whose mean, $\mathrm{E}[X]=m$, is **unknown**.
- Let $X_1, X_2, \ldots, X_n$ be n independent repeated measurements of X.
- Define the sample mean as $M_n = \frac{1}{n}\sum_{j=1}^{n} X_j$
- Question: **How well does M_n estimate the mean m?**

  Note that M_n is a random variable.

The Sample Mean is **Unbiased**

- Suppose Y (a random variable) is an estimate of a constant c.
- Typically, Y is some function of a set of measurements.
- Y is said to be ***unbiased*** if $\mathrm{E}[Y] = c$.
- This is a desirable property, since it says the long term average of the estimate is the value we are trying to estimate.

- The sample mean is an unbiased estimator of $m = \mathrm{E}[X_i]$.

  Proof:

$$E[M_n] = E\left[\frac{1}{n}\sum_{j=1}^{n} X_j\right] = \frac{1}{n}\sum_{j=1}^{n} E[X_j] = \frac{nm}{n} = m$$

The Variance of the Sample Mean Decreases with n

- Intuitively, as we increase the number of measurements, n, that we are using to estimate the mean, our estimate should get "better."
- One way in which the estimate gets "better" is that the variance decreases.

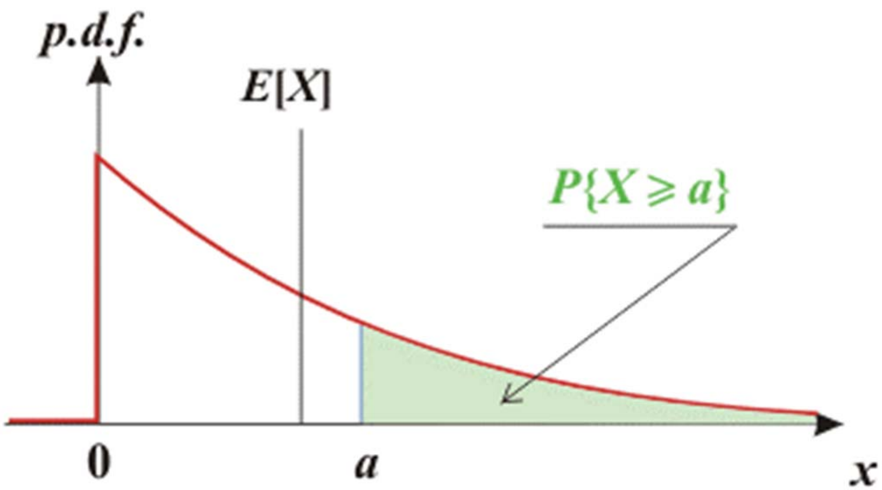
$$\mathrm{VAR}(M_n) = \mathrm{VAR}\left(\sum_{j=1}^{n} \frac{X_j}{n}\right) = \sum_{j=1}^{n} \mathrm{VAR}\left(\frac{X_j}{n}\right) \quad \Longleftarrow \quad \text{because the } X_j \text{ are independent}$$

$$= \sum_{j=1}^{n} \frac{\mathrm{VAR}(X_j)}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

- Thus, the values of the estimate, although random, are clustered closer and closer around m.

Elec2600H: Lecture 18

- Sums of Random Variables
- Mean and Variance of Sample Means
- **Useful Inequalities**
- Laws of Large Numbers



Markov Inequality

Andrey Markov
(1856—1922)

Markov Inequality

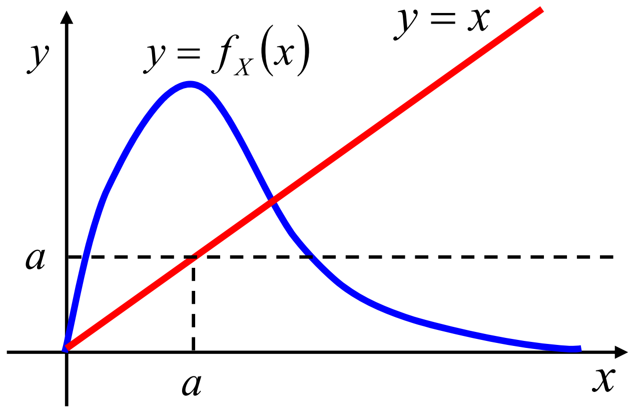
If X is a non-negative random variable, then $P[X \geq a] \leq \dfrac{E[X]}{a}$

Proof:

$$E[X] = \int_0^a x f_X(x)dx + \int_a^\infty x f_X(x)dx$$

$$\geq \int_a^\infty x f_X(x)dx \qquad \text{(throw away first term)}$$

$$\geq \int_a^\infty a f_X(x)dx \qquad \text{(since } x \geq a \text{)}$$



Thus, $\dfrac{E[X]}{a} \geq \int_a^\infty f_X(x)dx = P[X \geq a]$

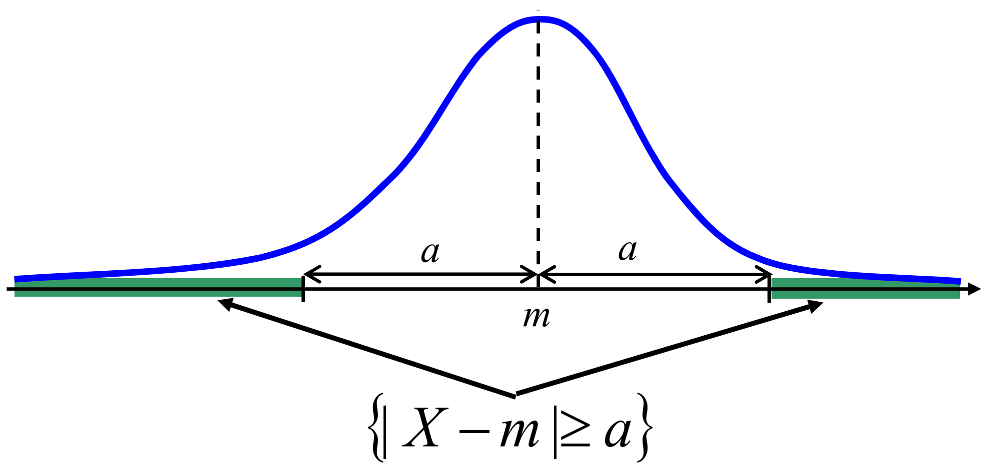
Based on the mean ONLY!

Chebyshev Inequality

Pafnuty Chebyshev
(1821—1894)

Chebyshev Inequality

Let X be a random variable with mean m and variance σ^2. For any $a > 0$, $P[|X - m| \geq a] \leq \frac{\sigma^2}{a^2}$



In other words, for any a large enough (compared the standard deviation), the probability that X is farther than a from the mean is negligible.

Proof of Chebyshev Inequality

Let $D = X - m$ be the deviation of X from the mean.

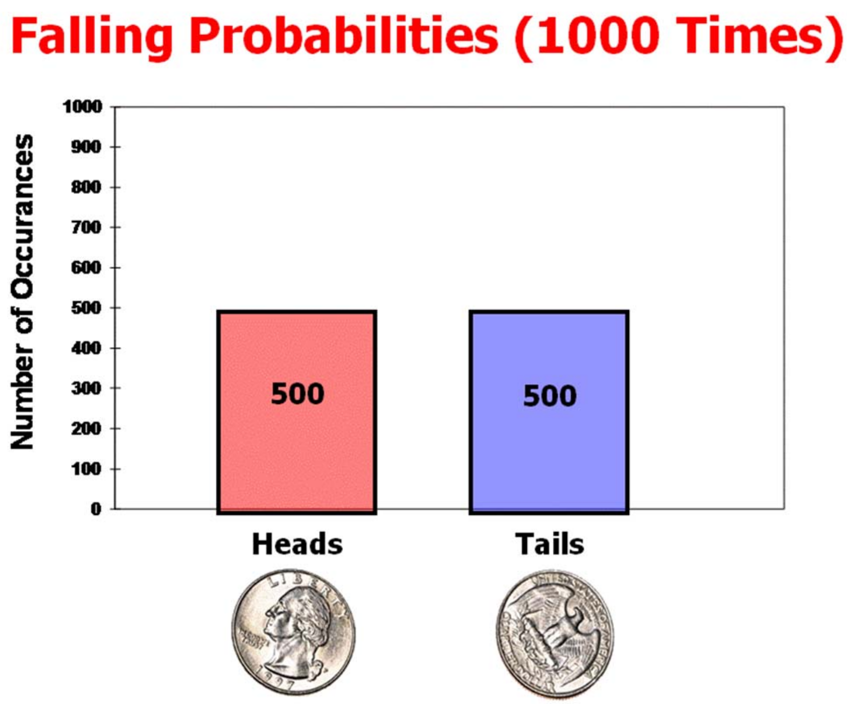
Since D^2 is non-negative, we can apply the Markov Inequality:

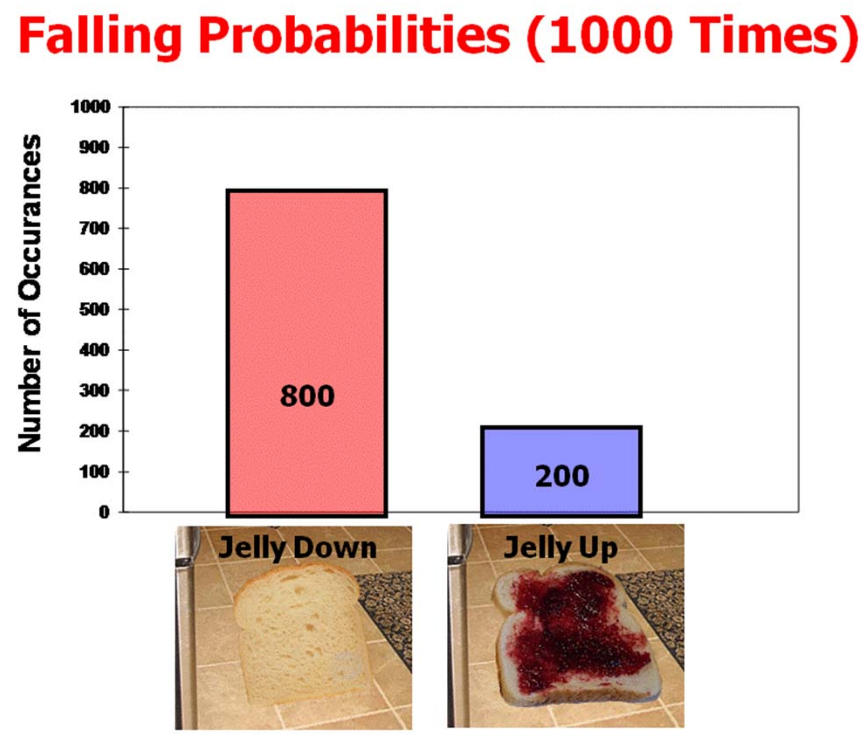
$$P[(X-m)^2 \geq a^2] = \underbrace{P[D^2 \geq a^2] \leq \frac{E[D^2]}{a^2}}_{\text{Markov Inequality}} = \frac{E[(X-m)^2]}{a^2} = \frac{\mathrm{VAR}[X]}{a^2}$$

Since $\{(X-m)^2 \geq a^2\}$ and $\{|X-m| \geq a\}$ are equivalent events, $P[|X-m| \geq a] \leq \dfrac{VAR[X]}{a^2}$

Elec2600H: Lecture 18

- Sums of Random Variables
- Mean and Variance of Sample Means
- **Laws of Large Numbers**

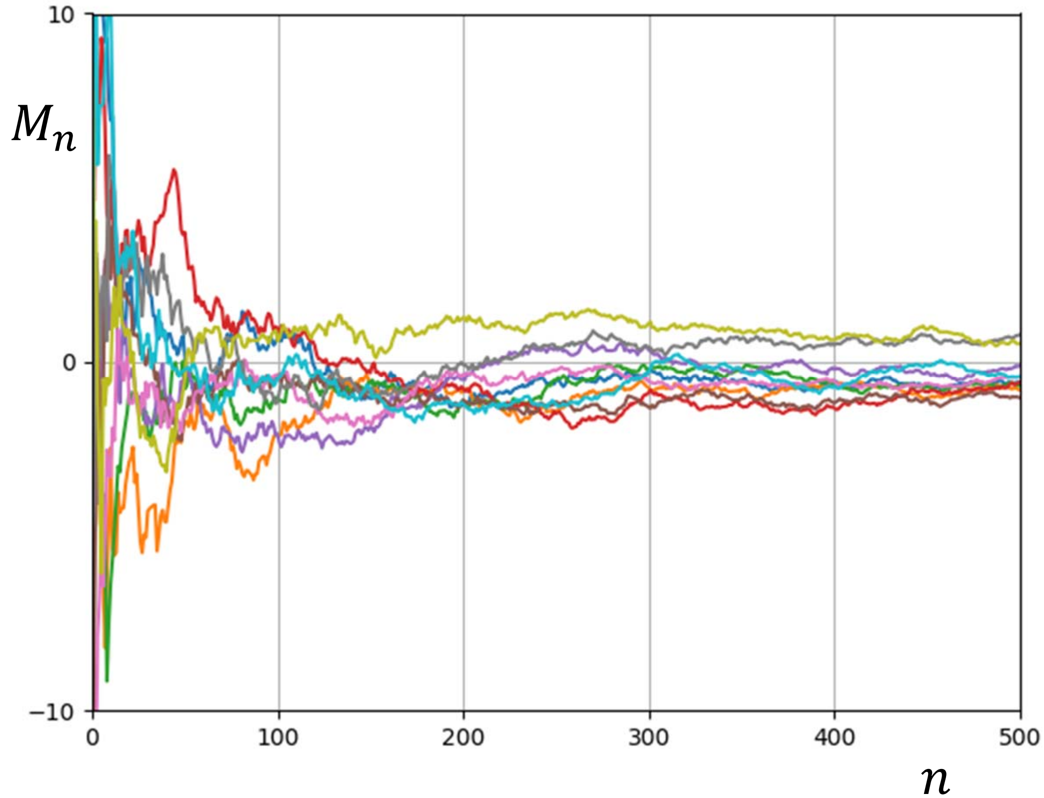




Laws of Large Numbers

- **Weak Law of Large Numbers**
  - Let $X_1, X_2, X_3, \ldots$ be a sequence of i.i.d. random variables with finite mean μ.
  - For any $\varepsilon > 0$,

$$\lim_{n\to\infty} P[|M_n - \mu| < \varepsilon] = 1$$



- **Strong Law of Large Numbers**
  - Let $X_1, X_2, X_3, \ldots$ be a sequence of i.i.d. random variables with finite mean μ, and finite variance.
  - Then,

$$P\left[\lim_{n\to\infty} M_n = \mu\right] = 1$$

  - Proof is outside scope of this class.

Proof of the Weak Law of Large Numbers

Let $X_1, X_2, X_3, \ldots$ be a sequence of i.i.d. random variables with finite mean μ. For any e > 0,

$$\lim_{n\to\infty} P[|M_n - \mu| < \varepsilon] = 1$$

Proof:

Assume the X_i have finite variance σ. By the Chebyshev Inequality, $P[|M_n - E[M_n]| \geq \varepsilon] \leq \frac{Var\ (M_n)}{\varepsilon^2}$

We have already seen that $E[M_n] = \mu$ and Var$(M_n) = \frac{\sigma^2}{n}$

Thus, $P[|M_n - \mu| < \varepsilon] > 1 - \frac{Var(M_n)}{\varepsilon^2} = 1 - \frac{\sigma^2}{n\varepsilon^2}$

This implies that $\lim_{n\to\infty} P[|M_n - \mu| < \varepsilon] = 1$

In fact, the theorem holds even for infinite variance.

Proof of the Strong Law of Large Numbers

- Let $X_1, X_2, X_3, \ldots$ be a sequence of i.i.d. random variables with finite mean μ and finite variance, then

$$P\left[\lim_{n\to\infty} M_n = \mu\right] = 1$$

- **Proof:** Beyond scope of this course.

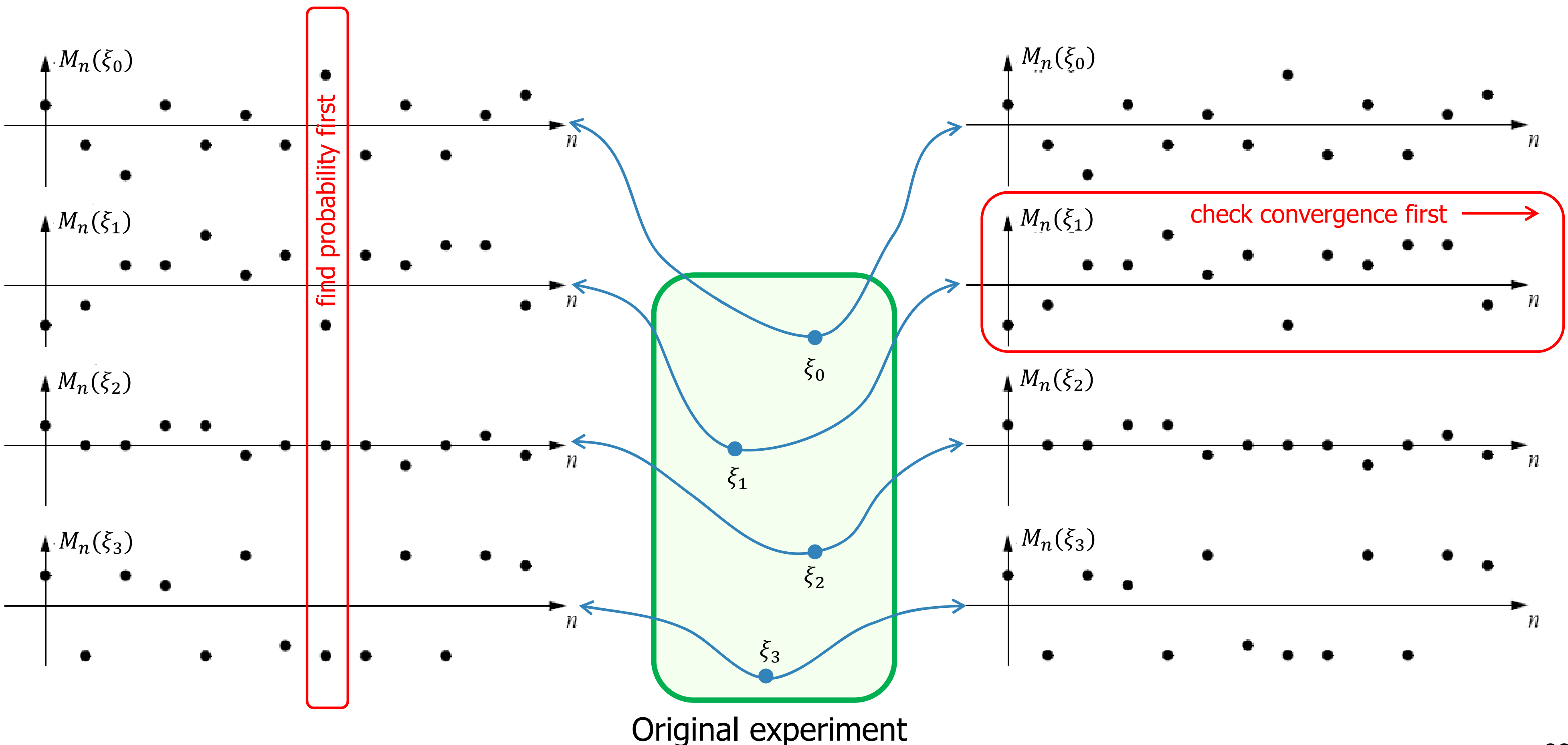
- The Weak Law says that by choosing enough measurements n, the sample mean will be close (within ε) of the true mean with high probability.

- However, it does not ensure that making additional measurements will make the sample mean closer to the true mean for any particular sequence of measurements.

- In contrast, the strong law says that with probability one, every sequence of sample mean calculations eventually approaches and stays close to the true mean.

Weak Law $\lim_{n\to\infty} P[|M_n - \mu| < \varepsilon] = 1$

Strong Law $P\left[\lim_{n\to\infty} M_n = \mu\right] = 1$



Example 7.8 - Relative Frequency

- Let I be the indicator function for an event A, i.e., if ξ is the outcome of the experiment,

$$I(\xi) = \begin{cases} 1 & \text{if } \xi \in A \\ 0 & \text{otherwise} \end{cases}$$

- Let I_j for $j = 1,2,\dots,n$ be the indicator in trial j of n trials. The sum, $N_n = \sum_{j=1}^{N} I_j$ is a count of how many times A occurs in the n trials. Thus,

$$M_n = \frac{1}{n}\sum_{j=1}^{n} I_j = \frac{N_n}{n} = f_A(n)$$

- In other words, the sample mean of I is the relative frequency of A.

- Similarly, the expected value of I is the probability of A:

$$E[I] = 1 \times P[\xi \in A] + 0 \times P[\xi \notin A] = P[A]$$

Convergence of the Relative Frequency

- The relative frequency is a special case of sample average.

- Applying the weak law of large numbers, $\lim_{n\to\infty} P[| f_A(n) - P[A] | < \varepsilon] = 1$

  If we perform many experiments, it is very likely (the probability is close to one) that the relative frequency of A will be close to (within ε of) the probability of A.

- Applying the strong law of large numbers, $P[\lim_{n\to\infty} f_A(n) = P[A]] = 1$

  We are very confident (with probability 1) that as we perform more and more experiments, the relative frequency of A will converge to the probability of A.

Statistical regularity and convergence

- In Chapter 1, we noted that we observe **statistical regularity** in many physical phenomena, i.e., the averages obtained in many repetitions of an experiments consistently yield approximately the same value.

- Based on this observation, we used the properties of the relative frequency to define a set of axioms for probability theory.

- We have **now come full circle**, showing that the theory predicts the observed statistical regularity. For example, the relative frequency of an event should converge to the probability of that event.

- This consistency between the theoretical predictions and observed behavior is one of the reasons that probability theory is so useful.

Major Points from this Lecture:

- **Mean and variance of sums of independent variables**
- **Mean and variance of sample means**
- **Markov and Chebyshev inequalities**
- **Weak and strong laws of large numbers**

ELEC 2600H: Probability and Random Processes in Engineering

Part III: Multiple Random Variables

- Lecture 10: Pairs of Discrete Random Variable
  - Out-of-Class Reading: 2D Calculus
- Lecture 11: Pairs of Continuous Random Variable
- Lecture 12: Independence, Joint Moments
- Lecture 13: Correlation Coefficients and Their Properties
- Lecture 14: Conditional PDF, Conditional Expectation
- Lecture 15: Sum of Two Random Variables
- Lecture 16: Pairs of Jointly Gaussian Random Variables
- Lecture 17: More than Two Random Variables
- Lecture 18: Laws of Large Numbers
- **Lecture 19: Central Limit Theorem and Characteristic Function**

Elec2600H: Lecture 19

- Central Limit Theorem
- The characteristic function
- Proof of the Central Limit Theorem

https://www.youtube.com/watch?v=Vo9Esp1yaC8

Central Limit Theorem

- Suppose X_i for $i \in \{1,2,\ldots,n\}$ are independent and identically distributed with mean μ and variance σ^2.

- Define $S_n = \sum_{i=1}^{n} X_i$

- We have already seen that the mean and variance of S_n increase linearly, i.e. $E[S_n] = n\mu$ and $\mathrm{Var}(S_n) = n\sigma^2$. However, if we define

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

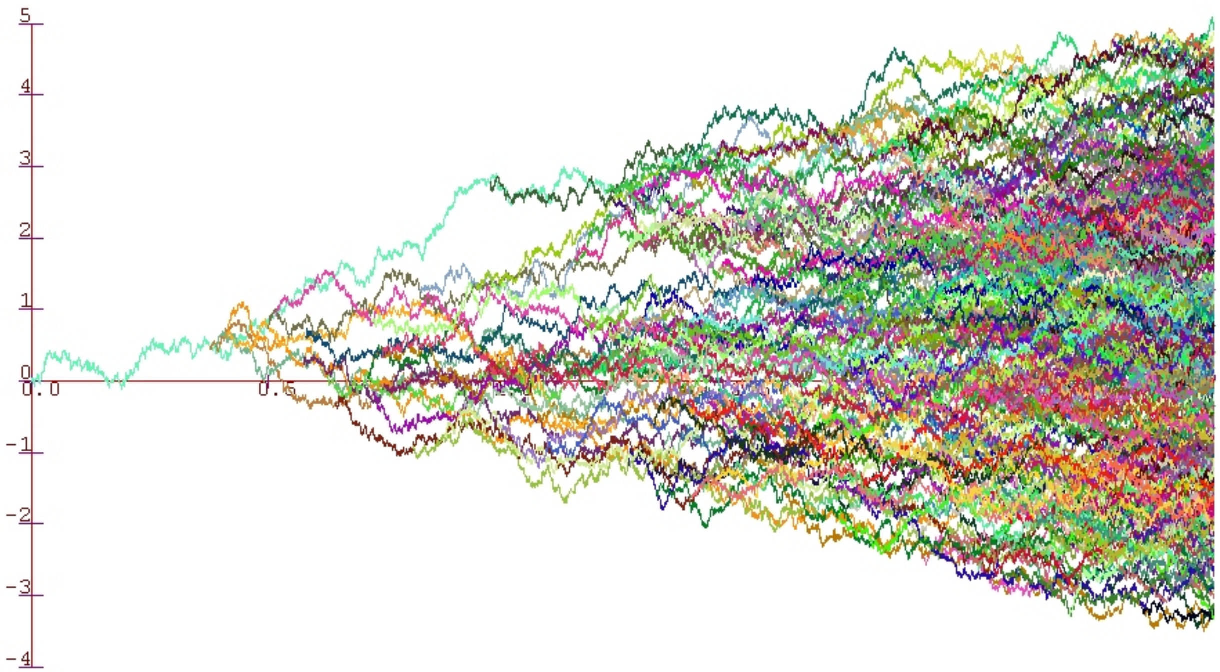
  Then $E[Z_n]=0$ and $VAR(Z_n)=1$.

- **Central Limit Theorem:** $\lim_{n\to\infty} P[Z_n < z] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{\frac{-x^2}{2}} dx$

  In other words, the *distribution* of Z_n approaches the distribution of a Gaussian with zero mean and unit variance. Note that we made **no assumption** on the distribution of the X_i!

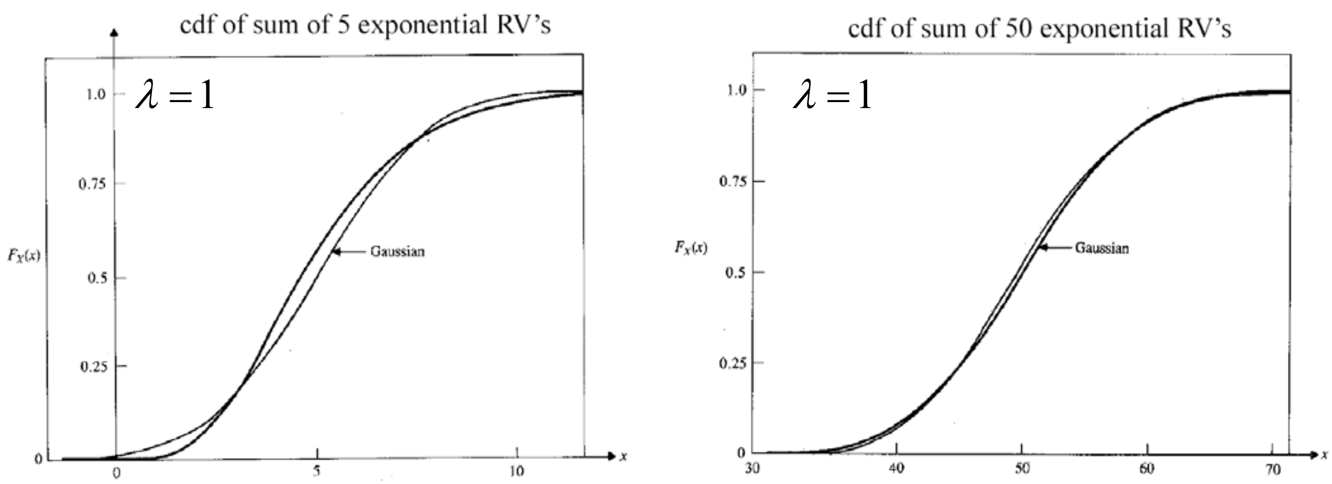
Discussion

- Equivalently, the distribution of S_n approaches the distribution of a Gaussian with mean $n\mathrm{m}$ and variance $n\mathrm{s}^2$.
- The CLT is one reason we use Gaussian random variables so frequently.
- In nature, many macroscopic phenomena result from the addition of numerous independent but non-Gaussian processes.



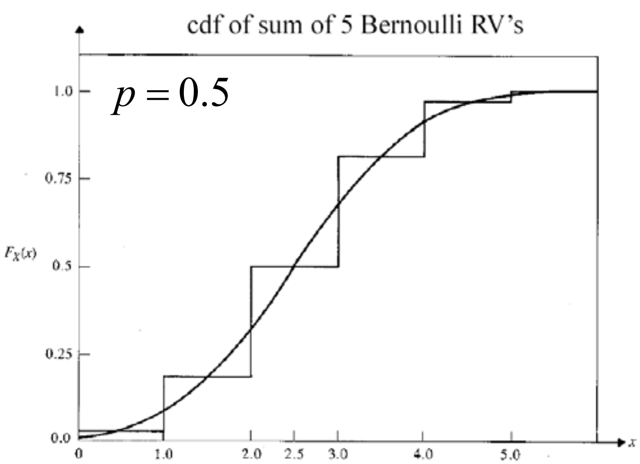
Example: Sum of Exponential Random Variables

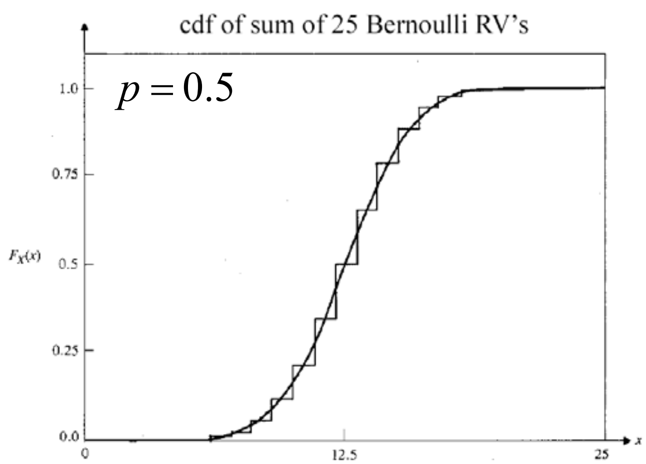
- The sum of m exponential random variables is an m-Erlang random variable, which has mean m/λ and variance m/λ^2
- The distribution of the m-Erlang approaches a Gaussian distribution with the same mean and variance.



Example: Sum of Bernoulli Random Variables

- The sum of n Bernoulli random variables with parameter p is a binomial random variable with parameters n and p.
- The bionomial random variable has mean np.
- The variance is $np(1\text{-}p)$.





Special Case I:

If the X_i are continuous, then the density of S_n approaches the density of a Gaussian with mean $n\mu$ and variance $n\sigma^2$.

Example

Suppose that the processing times for forms at an insurance agency are independent with mean 8 and standard deviation 2. Estimate the probability that it takes more than 840 minutes to process the first 100 forms.

Solution:

Let T_i be the time to process the i-th form. The time to process 100 forms is $S = \sum_{i=1}^{100} T_i$

By the Central Limit Theorem, the density of S approaches a Gaussian with mean $100 \cdot 8 = 800$ and variance $100 \cdot 4 = 400$ (standard deviation 20). Thus,

$$P[S > 840] \approx Q(\frac{840 - 800}{20}) = Q(2) = 2.28 \times 10^{-2}$$

Review of the Q function (Survival Function)

- The Q function is the integral of the tail of a Gaussian pdf with m = 0 and σ = 1.
- If X is Gaussian with $m = 0$ and σ = 1, then $P[\{X > x\}] = Q(x)$

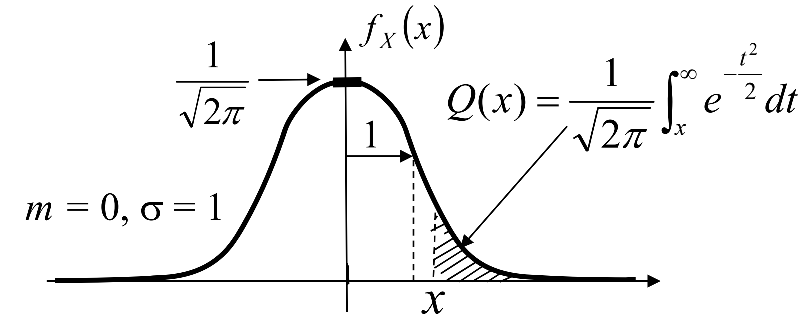
- If Y is a Gaussian with mean m and standard deviation σ,

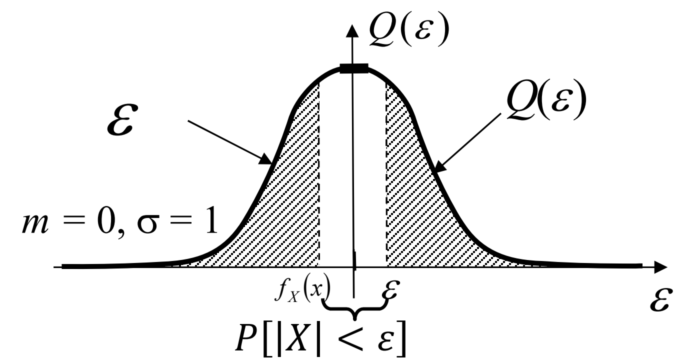
$$P[\{Y > y\}] = Q\left(\frac{y-m}{\sigma}\right)$$

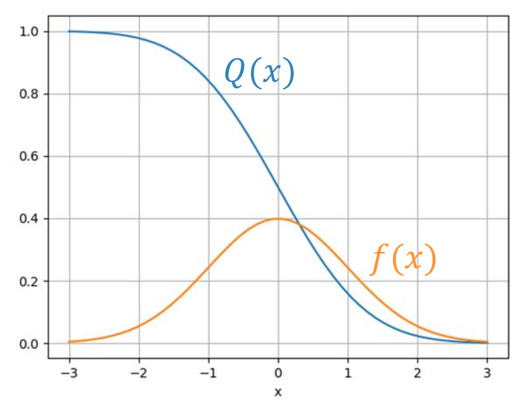
- By symmetry of the Gaussian: $P\left[\{|X| < \varepsilon\}\right] = 1 - 2Q(\varepsilon)$

- Similarly, $P\left[\{|Y - m| < \varepsilon\}\right] = 1 - 2Q\left(\frac{\varepsilon}{\sigma}\right)$

- Using Python:
  - $P[X > x] = \text{scipy.stats.norm.sf}(x)$
  - $P[Y > y] = \text{scipy.stats.norm.sf}(y, \text{loc} = m, \text{scale} = \sigma)$







Example 7.13

The times between events in a random experiment are i.i.d. exponential random variables with mean m seconds. Find the probability that the 1000^{th} event occurs between times $950m$ and $1050m$.

Solution

Let X_i be the time between the $(i\text{-}1)^{\text{th}}$ and the i^{th} events. $E[X_i] = m$ $\quad \mathrm{Var}[X_i] = m^2$

The time of the nth event, S_n, is $S_n = \sum_{i=1}^{n} X_i$

By the CLT, S_n is approximately Gaussian with mean and variance $E[S_n] = nm$ $\quad \mathrm{Var}[S_n] = nm^2$

Thus,

$$P[950m \le S_{1000} \le 1050m] = P[\frac{950m - 1000m}{m\sqrt{1000}} \le \frac{S_n - 1000m}{m\sqrt{1000}} \le \frac{1050m - 1000m}{m\sqrt{1000}}] = 1 - 2Q(1.58) = 0.8866$$

As n becomes large, S_n is very likely to be close (within 5%) of its mean nm.

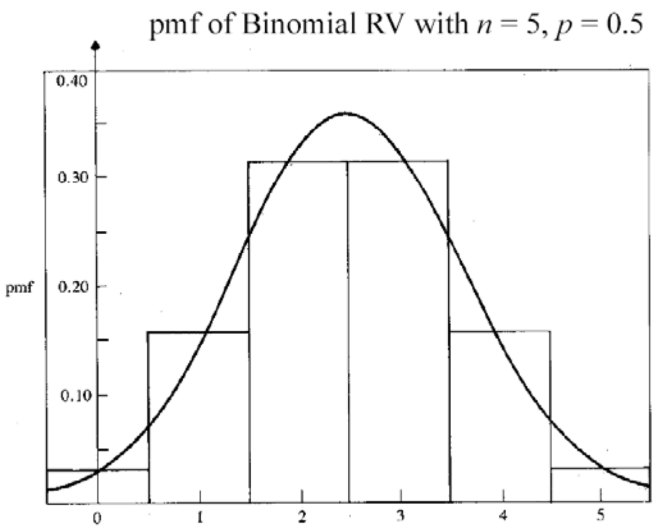
Special Case II:

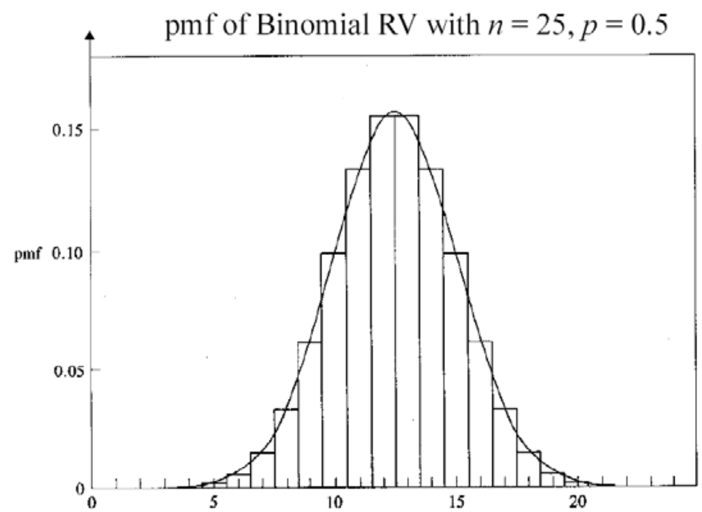
If the X_i are discrete and assume only integer values, then $p_k = p[S_n = k] \approx \frac{1}{\sqrt{2\pi n}\sigma} e^{-\frac{(k-n\mu)^2}{2n\sigma^2}}$

Example:

Since a binomial RV is a sum of n independent Bernoulli random variables, $p_k \approx \frac{1}{\sqrt{2\pi}\sqrt{np(1-p)}} e^{-\frac{(k-np)^2}{2np(1-p)}}$

This result is known as the De Moivre-Laplace Theorem





Examples 7.10 and 7.14

Suppose X_i are independent Bernoulli random variables with parameter p.

Thus, each X_j assumes values 0 and 1 with probability p and 1-p and

$$E[X_j] = 0 \cdot (1-p) + 1 \cdot p = p \qquad \mathrm{VAR}(X_j) = E[X_j^{\,2}] - E[X_j]^2$$

$$E[X_j^{\,2}] = 0 \cdot (1-p) + 1 \cdot p = p \qquad = p - p^2 = p(1-p)$$

Suppose we wish to estimate the value of p using the sample mean, $M_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

How large should n be so that we estimate p within 0.01 with at least 95% probability?

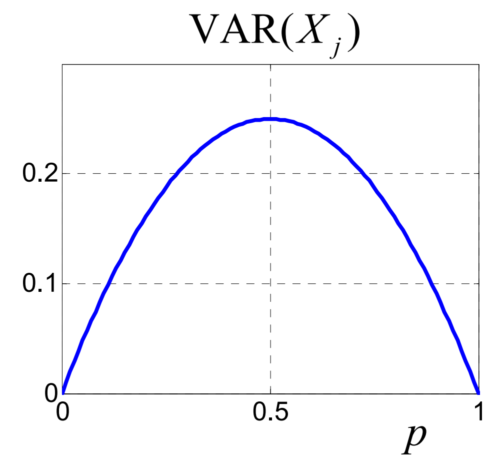
We approach this in two ways

1. Using the Chebyshev inequality (Example 7.10).
2. Using the Central Limit Theorem (Example 7.14).

Using the Chebyshev Inequality

Although the p is unknown, $\mathrm{VAR}(X_j) = p(1-p) \le \frac{1}{4}$

From the proof of the Weak Law, $P[|M_n - p| < \varepsilon] \ge 1 - \frac{\mathrm{VAR}[X_j]}{n\varepsilon^2} \ge 1 - \frac{1}{4n\varepsilon^2}$



Thus, $1 - \frac{1}{4n(0.01)^2} = 0.95 \Rightarrow \frac{1}{4n(0.01)^2} = 0.05 \Rightarrow n = 50{,}000$

This **guarantees** that we estimate p to within 0.01 with at least 95% probability.

However, it requires a very large number of samples.

Using the Central Limit Theorem

- Recall that the sample mean has mean $m = \mathrm{E}[X_j] = p$ and variance $\sigma^2 = \dfrac{\mathrm{VAR}(X_j)}{n} = \dfrac{p(1-p)}{n} \le \dfrac{1}{4n}$

- For large enough n, distribution of the sample mean M_n approaches a Gaussian distribution with mean p and variance σ^2. Thus, (recall that $Q(x)$ <u>decreases</u> with x)

$$P[|M_n - p| < \varepsilon] \approx 1 - 2Q\left(\frac{\varepsilon}{\sigma}\right) \ge 1 - 2Q\left(\frac{\varepsilon}{1/\sqrt{4n}}\right) = 1 - 2Q\left(2\varepsilon\sqrt{n}\right)$$

- Thus, $1 - 2Q(2\varepsilon\sqrt{n}) = 0.95 \Rightarrow Q(2\varepsilon\sqrt{n}) = 0.025 \Rightarrow 2\varepsilon\sqrt{n} = 1.95 \Rightarrow n = \left(\dfrac{1.95}{2\times 0.01}\right)^2 = 9506$

scipy.stats.norm.isf(0.025)

- This is **not** a guarantee but results in a much more manageable number of samples.

Elec2600H: Lecture 19

- Central Limit Theorem
- The characteristic function
- Proof of the Central Limit Theorem

Characteristic Function

- The *characteristic function* of a continuous random variable X with density $f_X(x)$ is defined as

$$\Phi_X(\omega) = E[e^{j\omega X}] = \int_{-\infty}^{\infty} f_X(x) e^{j\omega x} dx \quad \text{where } j = \sqrt{-1}$$

- The pdf can be computed by

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_X(\omega) e^{-j\omega x} d\omega$$

- The above two equations imply that the characteristic function and the probability density function are equivalent (knowing one is the same as knowing the other).
- For a discrete integer valued random variable X,

$$\Phi_X(\omega) = \sum_k p_X(k) e^{j\omega k} \qquad p_X(k) = \frac{1}{2\pi} \int_0^{2\pi} \Phi_X(\omega) e^{-j\omega k} d\omega$$

- The characteristic function is primarily used to prove theorems, especially those involving sums of random variables.

Examples

- Find the characteristic function of the exponential random variable

Solution

The pdf is $f_X(x) = \lambda e^{-\lambda x} u(x)$

Thus,

$$\Phi_X(\omega) = \int_0^{\infty} \lambda e^{-\lambda x} e^{j\omega x} dx$$

$$= \left. \frac{\lambda e^{-(\lambda - j\omega)x}}{-(\lambda - j\omega)} \right|_0^{\infty}$$

$$= 0 - \frac{\lambda}{-(\lambda - j\omega)}$$

$$= \frac{\lambda}{\lambda - j\omega}$$

- Find the characteristic function of the geometric random variable.

Solution

The pmf is $p_X(k) = pq^k \quad \text{for } k \in \{0,1,2,\ldots\}$

Thus, $\Phi_X(\omega) = \sum_{k=0}^{\infty} pq^k e^{j\omega k}$

$$= p \sum_{k=0}^{\infty} (qe^{j\omega})^k$$

$$= \frac{p}{1 - qe^{j\omega}}$$

The Moment Theorem

$$E[X^n] = (-j)^n \frac{d^n}{d\omega^n} \Phi_X(\omega)\Big|_{\omega=0}$$

Proof:

Substituting the power series expansion into the characteristic function definition

$$e^{j\omega x} = 1 + j\omega x + \frac{(j\omega x)^2}{2!} + \frac{(j\omega x)^3}{3!} + \ldots$$

$$\Phi_X(\omega) = \int_{-\infty}^{\infty} f_X(x) e^{j\omega x} dx = \int_{-\infty}^{\infty} f_X(x) \left\{ 1 + j\omega x + \frac{(j\omega x)^2}{2!} + \frac{(j\omega x)^3}{3!} + \ldots\ldots \right\} dx$$

$$= 1 + j\omega E[X] + \frac{(j\omega)^2}{2!} E[X^2] + \frac{(j\omega)^3}{3!} E[X^3] + \ldots\ldots..$$

Knowing all moments is equivalent to knowing the characteristic function!

Differentiating and setting $\omega = 0$,

$$\frac{d}{d\omega} \Phi_X(\omega)\Big|_{\omega=0} = jE[X] + \frac{2j\omega \cdot j}{2!} E[X^2] + \frac{3(j\omega)^2 \cdot j}{3!} E[X^3] + \ldots\ldots = jE[X]$$

Similarly, differentiating n times:

$$\frac{d^n}{d\omega^n} \Phi_X(\omega)\Big|_{\omega=0} = j^n E[X^n]$$

Char. Function of a Sum of Independent RVs

- Suppose X_1 and X_2 are independent.
- The characteristic function of the sum $Z = X_1 + X_2$ is

$$\begin{aligned}\Phi_Z(\omega) &= E[e^{j\omega Z}] = E[e^{j\omega(X_1+X_2)}] \\ &= E[e^{j\omega X_1} e^{j\omega X_2}] \\ &= E[e^{j\omega X_1}]E[e^{j\omega X_2}] \\ &= \Phi_{X_1}(\omega)\Phi_{X_2}(\omega)\end{aligned}$$

By independence

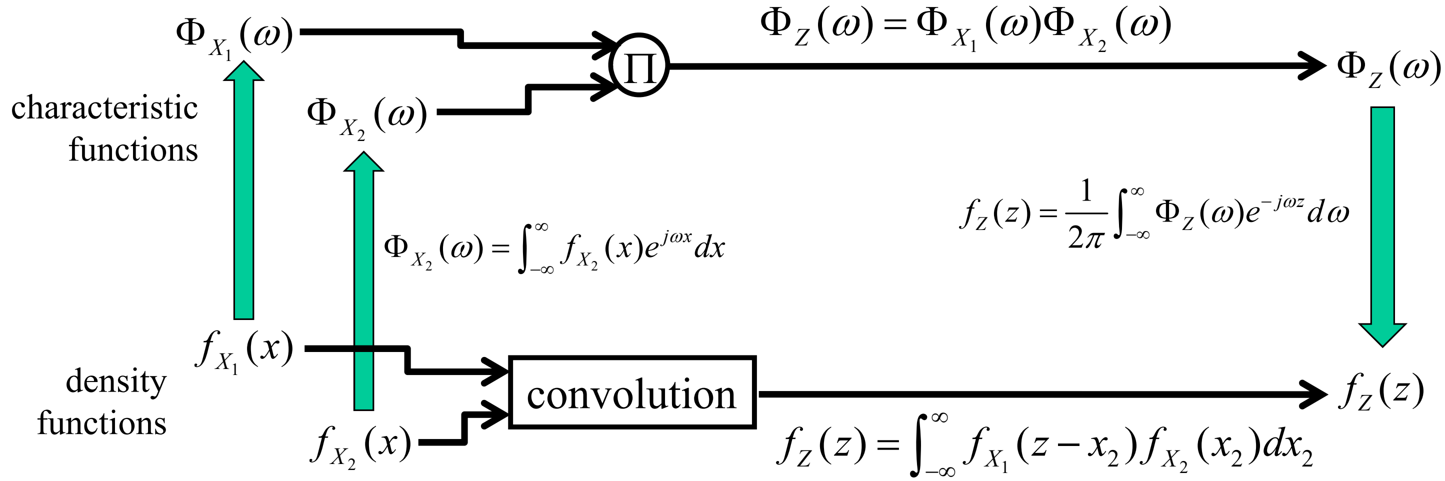
- More generally, if $X_1, X_2, \ldots, X_n$ are independent, then

$$Z = \sum_i X_i \Longrightarrow \Phi_Z(\omega) = \prod_i \Phi_{X_i}(\omega)$$

- Characteristic functions are used when dealing with sums of RVs, because products are easier than integrals.

Transformations

- The characteristic function can be thought of as a transformation of the probability density function into a new domain where the effect of adding two random variables can be computed more easily.



- For those who have taken ELEC2100: The characteristic function is essentially the Fourier Transform of the pdf. The previous result is equivalent to the statement that the Fourier Transform of the convolution of two functions is the product of their Fourier Transforms

Elec2600H: Lecture 19

- Central Limit Theorem
- The characteristic function
- Proof of the Central Limit Theorem

Proof of Central Limit Theorem

- We sketch the proof of the central limit theorem.

- Steps
  - We find the characteristic function of a Gaussian RV with zero mean.
  - We find the characteristic function of

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

  - We show that the characteristic function of Z_n approaches the characteristic function of a Gaussian as n approaches infinity.

Major Points from this Lecture:

- Central Limit Theorem
- Review of the Q function (survival function)
- The characteristic function
- Proof of the Central Limit Theorem