Chapter 3

Review of Probability and Random Variables (Z&T 7th ch 6)

Contents

.

3.1 Introduction

Before random processes can be introduced, we need to review the foundational topics of *probability and random variables*. We begin with probability, the concepts of equally likely, relative frequency, and the axioms of probability. Next we consider random variables, and how they extend basic probability beyond events into outcomes that lie on the real line in one or more dimensions. A sequence of random variables can be viewed as a discrete-time random process, so the progression is logical. A continuous-time random process is actually more abstract than the discrete-time case. There are many definitions that must be made as well.

3.2 Probability

3.2.1 Probability Definitions

To begin with consider some possible definitions of probability.

- The equally likely or *classical definition* which is obtained without experimentation, states that the probability of event A is

  $$P(A) = \frac{N_A}{N}$$

  where N_A is the number of outcomes favorable to event A and N is the number of possible outcomes

  - The basic assumption (without a priori knowledge) is that all outcomes are equally likely
  - N and N_A may be difficult to determine

- *Relative frequency*: The probability of event A is

$$P(A) = \lim_{n\to\infty} \frac{n_A}{n}$$

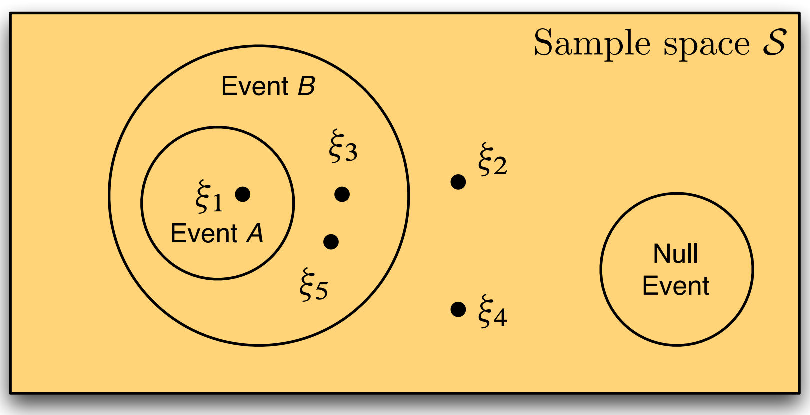
  where n_A is the number of occurrences of A and n is the total number of trials

  - The limit is a hypothesis since it cannot actually be determined experimentally
  - The relative frequency concept is essential in applying probability theory to the physical world

- *Axiomatic Definition*: Here it is assumed that the probabilities $P(A_i)$ of events A_i satisfy certain axioms (3.5 required)

- Consider an experiment which has possible outcomes ξ_i.

- The ξ_i's are elements of the space $\mathcal{S}$ known as the sample space



Discrete sample space $\mathcal{S}$

- A subset of $\mathcal{S}$ is a collection of elements ξ_i, i.e. for

$$A \subset \mathcal{S} \qquad B \subset \mathcal{S}$$

  we might have

$$A = \{\xi_1\} \qquad B = \{\xi_1,\ \xi_3,\ \xi_5\}$$

  Note that here $A \subset B$

- **Events:**
  - Subsets of $\mathcal{S}$ which make up a σ-field are *events* on $\mathcal{S}$
  - The sample space $\mathcal{S}$ is called the *certain event*
  - The empty set, $\emptyset$, is called the *impossible event*

- Associated with each event is a measure called *probability*

- **Axioms of Probability:**
  1. $P(A) \geq 0$ for all A in $\mathcal{S}$
  2. $P(\mathcal{S}) = 1$
  3. If $AB = A \cap B = \emptyset$ then

$$P(A + B) = P(A \cup B) = P(A) + P(B)$$

  Note that (3) extends to

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

  if $A_i A_j = \emptyset \; \forall \; i \neq j$

Useful Relationships

- **Complement:** Since $A \cup \bar{A} = \mathcal{S}$, $A \cap \bar{A} = \emptyset$, and from Axiom 2 and 3 $P(A) + P(\bar{A}) = 1$,

$$P(\bar{A}) = 1 - P(A)$$

- **Joint Probability:**

$$P(AB) = P(A \cap B) = P(A, B)$$

- **Generalization of Axiom 3:**

  For any events A and B we can write that

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

  why?

- **Conditional Probability:**

  Given two events A and B, the probability of A given B is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(AB)}{P(B)}$$

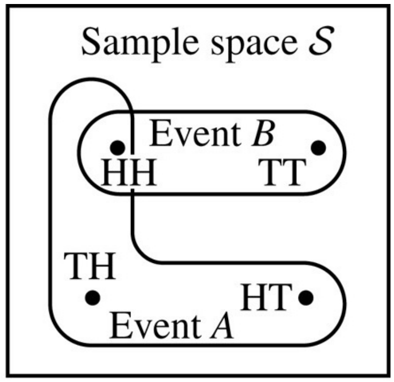
  also

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(AB)}{P(A)}$$

  $\Longrightarrow$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad \text{(Bayes Rule)}$$

Example 3.1: Simultaneous Tossing of Two Fair Coins



Experiment sample space

- Let A denote the event of *at least* one head: HH, HT, and TH are valid outcomes

- Let B denotes a match: HH and TT

- Find $P(A)$ and $P(B)$

- We assume a fair coin or equally likely model, thus it follows that each outcome has probability 1/4, and in particular

$$P(A) = 3 \times \frac{1}{4} = \frac{3}{4}$$
$$P(B) = 2 \times \frac{1}{4} = \frac{1}{2}$$

- Consider $P(A|B)$

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{1/4}{1/2} = \frac{1}{2}$$

- Consider $P(B|A)$

$$P(B|A) = \frac{P(AB)}{P(A)} = \frac{1/4}{3/4} = \frac{1}{3}$$

- $P(A + B) = P(A \cup B)$?

- **Theorem of Total Probability:**

  Given n mutually exclusive (disjoint) events

$$A_i A_j = \emptyset \;\; i \neq j = 1, 2, \ldots, n$$

  such that $\sum_{i=1}^{n} A_i = \mathcal{S}$, then for an arbitrary $B \subset \mathcal{S}$

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$$

  proof

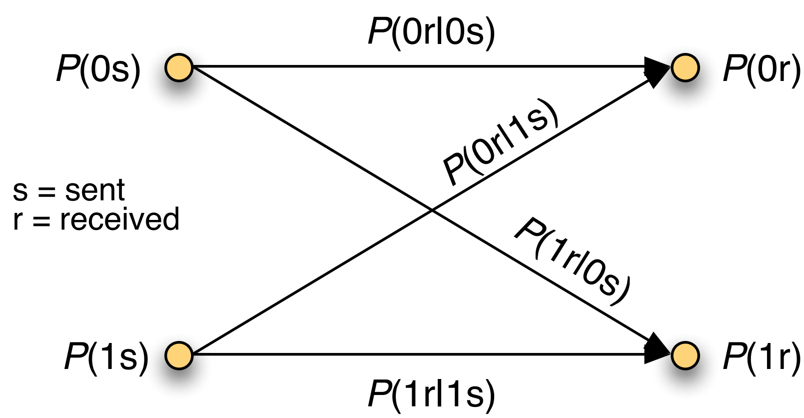
$$B = B\mathcal{S} = B\left(\sum_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} BA_i$$

  so

$$P(B) = P\left(\sum_{i=1}^{n} BA_i\right) = \sum_{i=1}^{n} P(BA_i)$$

  but $P(BA_i) = P(B|A_i)P(A_i)$

Example 3.2: Binary Communication



Binary communications link probability model

- The above is a model for a binary communications link
- Bits of type 0 and 1 are sent from the left side over the channel and received at the right side
- Various events involving transmitted and received bits can be defined, such as $P(1\text{s}) = p$ and $P(0\text{s}) = q$; Note that $q \overset{\text{must}}{=} 1 - p$ why?
- Conditional probabilities relate to the channels ability to successfully convey bits from the sending end to the receiving end
- For a symmetric channel model we have that

$$P(0\text{r}|0\text{s}) = P(1\text{r}|1\text{s})$$

  Note: $P(1\text{r}|0\text{s}) = 1 - P(0\text{r}|0\text{s})$, etc.

- Suppose that $P(0\text{r}|0\text{s}) = P(1\text{r}|1\text{s}) = 0.9$

- Find $P(1\text{s}|1\text{r})$ and $P(1\text{r})$
- Using Bayes' rule

$$P(1\text{s}|1\text{r}) = \frac{P(1\text{r}|1\text{s})P(1\text{s})}{P(1\text{r})}$$

- To continue we need to find $P(1\text{r})$, which can be obtained using total probability where we expand over 1s and 0s which are mutually exclusive and cover the sample space,

$$\begin{aligned} P(1\text{r}) &= P(1\text{r}|1\text{s})P(1\text{s}) + P(1\text{r}|0\text{s})P(0\text{s}) \\ &= 0.9 \cdot p + (1 - 0.9) \cdot (1 - p) \\ &= 0.8p + 0.1 \overset{\text{if } p=0.2}{=} 0.26 \end{aligned}$$

- Now we can continue to find $P(1\text{s}|1\text{r})$

$$P(1\text{s}|1\text{r}) = \frac{0.9p}{0.8p + 0.1} \overset{\text{if } p=0.2}{=} 0.69$$

- **Independence:**

  Events A and B are statistically independent if and only if

  $$P(AB) = P(A)P(B)$$

  It then follows that

  $$P(A|B) = P(A) \quad \text{and} \quad P(B|A) = P(B)$$

- **Bernoulli Trials:**

  Consider repeating an experiment n times, where on each trial we are interested in some event A_i, $i = 1, 2, \ldots, n$, which has probability $P(A_i) = p$ and $P(\bar{A}_i) = q$

- For this *combined experiment* the sample space is a *cartesian product* composed of n-tuples of the form $\{\xi_1, \xi_2, \ldots, \xi_n\}$

- Since we have assumed that the sub-experiments are independent, it follows from the definition of independence that

  $$P(A_1, A_2, \ldots, A_n) = P(A_1)P(A_2)\cdots P(A_n)$$

- Given that A occurs k times in a particular order, and $\bar{A}$ occurs $n-k$ times elsewhere, the probability of this combined experiment event is

  $$P\{A \text{ occurs } k \text{ times in a specific order}\} = p^k q^{n-k}$$

- We are interested in the event A occurs k times in any order, which involves the *combinations* counting formula

  $$\left\{ \begin{array}{l} \text{\# of sequences where} \\ A \text{ occurs } k \text{ times} \end{array} \right\} = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

- Finally,

$$P\{A \text{ occurs } k \text{ times in any order}\} = p_n(k) = \binom{n}{k} p^k q^{n-k}$$

which is known as the *binomial probability*

Example 3.3: Bit Patterns

A random bit sequence of 1's and 0's is observed until two 1's occur. Find the probability that k bits must be observed. Assume that $P(1) = p$ and $P(0) = q = 1 - p$.

- The event of interest is of the form

$$\{\text{event two 1's in } k \text{ bits}\} = \{\underbrace{0, 0, \ldots, 1, 0, \ldots, 0}_{k-1}, 1\}$$

where the '1' in the block of $k-1$ bits can occur in one of $k-1$ positions

- The second '1' must occur in position k

- Due to the assumed independence between bits, we can write

$$p(k|\text{two 1's}) = (k-1)(1-p)^{k-2}p^2$$

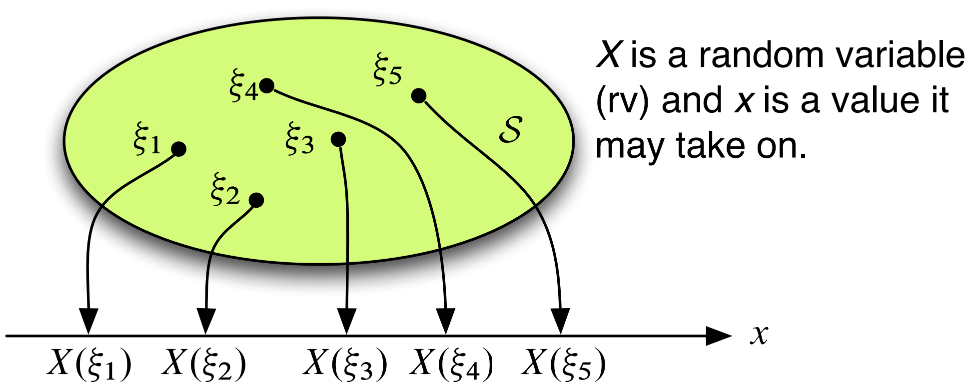
- Observation: The probability that two or more bits must be observed until two ones occurs should be one

- To prove this we can evaluate the infinite sum:

$$\begin{aligned} P(k \geq 2) &= \sum_{k=2}^{\infty} p(k|\text{two 1's}) = \sum_{m=0}^{\infty} (m+1) q^m p^2 \\ &= p^2 \left[\sum_{m=0}^{\infty} q^m + \sum_{m=0}^{\infty} m q^m \right] \\ &= p^2 \left[\frac{1}{1-q} + \frac{q}{(q-1)^2} \right] \\ &= p^2 \left[\frac{1}{p} + \frac{q}{p^2} \right] = p + q = 1 \end{aligned}$$

3.3 Random Variables

A *random variable* is a rule that assigns a numerical value to outcomes of a chance experiment



Elementary outcomes mapped to the real line

It may be continuous, discrete, or mixed. A random variable (rv) will be denoted as $X(\xi_i) = X$ (in some texts $X = \mathbf{x}$, notably Papoulis). The values it takes on will be denoted by x (authors agree here).

- **Cumulative Distribution Function (cdf):** or probability distribution function

$$F_X(x) = P[\{\xi : X(\xi) \leq x\}] = P[X \leq x]$$

<u>Properties</u>

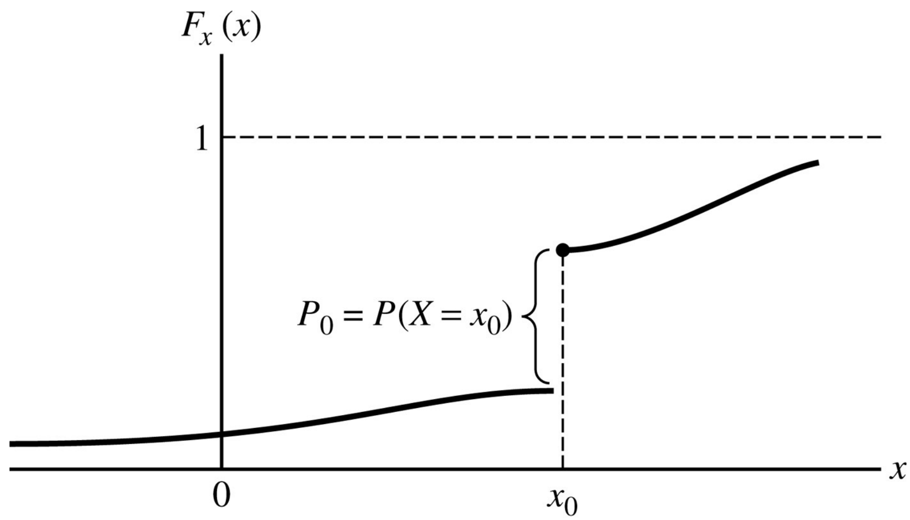
1. $F_X(-\infty) = 0, \quad F_X(\infty) = 1$
2. Right continuous, i.e.,

$$\lim_{x \to x_o^+} F_X(x) = F_X(x_o)$$

3. A nondecreasing function, i.e.,

$$F_X(x_1) \leq F_X(x_2), \text{ if } x_1 < x_2$$

  - It is worth noting that when discrete probability is involved, that is the underlying sample space maps a nonzero probability P_0 to a point x_0, and elsewhere the mapping is continuous; at x_0, $F_X(x)$ will contain a jump



cdf jump due to $P(X = x_0) > 0$

- **Probability Density Function (pdf):**

$$f_X(x) = \frac{dF_X(x)}{dx}$$

Note that

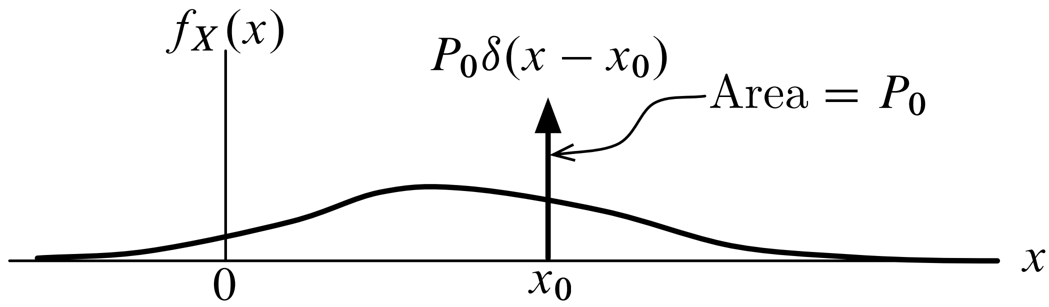
$$F_X(x) = \int_{-\infty}^{x} f_X(u)\,du$$

  – The cdf and pdf both provide a complete description of an rv, but the pdf is often more convenient to work with

Properties

1. $f_X(x) \geq 0$
2. $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$
3. $P(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1) = \int_{x_1}^{x_2} f_X(x)\,dx$

- The cdf jump noted earlier becomes a delta function, $P_0\delta(x - x_0)$, in the pdf



cdf jump results in delta function in pdf

- Property 1 and 2 are special, since they constitute an *existence* theorem; any function that satisfies these properties is a valid pdf and hence a valid rv description

- For $f_X(x)$ continuous at x_1, we also note that

$$\lim_{\Delta x \to 0} \int_{x_1}^{x_1+\Delta x} f_X(u)\,du = 0$$

What does this mean? The probability at a point is zero in spite of the pdf being nonnegative at that point

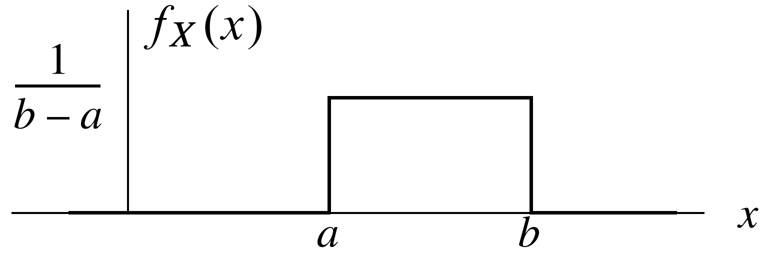
3.3.1 Important pdfs

- <u>Uniform</u>

If rv X is uniform on the interval $[a, b]$, then

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & \text{otherwise} \end{cases}$$

A shorthand notation is simply to say $X \sim \mathcal{U}[a, b]$.



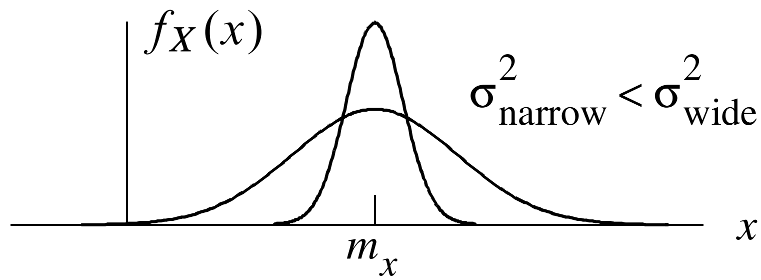
The rv X uniform on $[a, b]$

Gaussian or Normal:

If rv X is normal with parameters: mean m_X and variance σ_X^2, then

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

A shorthand notation is simply to say $X \sim \mathcal{N}[m_X, \sigma_X^2]$.



The rv X normal with parameters m_x and σ_x^2

- The area under the tail of a Gaussian pdf can be written in terms of the $Q(\,)$-function, that is

$$P[X > x] = Q\left(\frac{x - m_X}{\sigma_X}\right)$$

where

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2}\, dt$$

Note that $Q(0) = 1/2$ and $Q(-x) = 1 - Q(x)$.

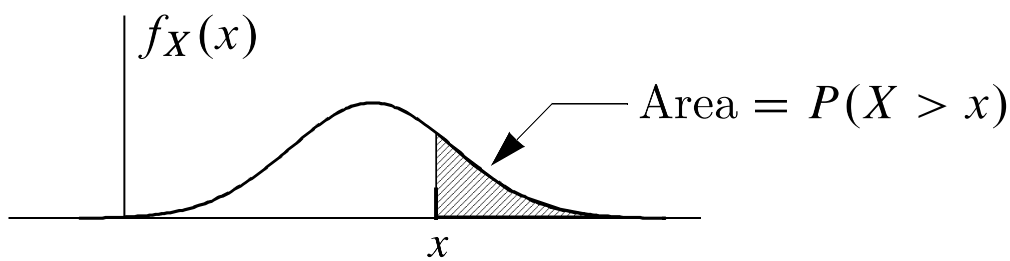
- In MATLAB and in other comm texts you may run into

$$\mathrm{erf}(u) = \frac{2}{\sqrt{\pi}} \int_0^u e^{-y^2}\, dy$$

$$\mathrm{erfc}(u) = 1 - \mathrm{erf}(u) = \frac{2}{\sqrt{\pi}} \int_u^\infty e^{-y^2}\, dy$$

- In particular,

$$Q(u) = \frac{1}{2}\mathrm{erfc}\left(\frac{u}{\sqrt{2}}\right) \text{ or } \mathrm{erfc}(v) = 2Q\left(\sqrt{2}v\right)$$



Tail of normal pdf

Example 3.4: **Approximation for $Q(x)$**

- A simple, reasonably accurate approximation for $Q(x)$ over $0 < x < \infty$ is

$$Q(x) \simeq \left[\frac{1}{(1-a)x + a\sqrt{x^2+b}}\right] \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

where $a = 1/\pi$ and $b = 2\pi$

- A short table of values for $Q(x)$

x	$Q(x)$	Approx.	x	$Q(x)$	Approx.
0.5	3.09E-01	3.05E-01	3.0	1.35E-03	1.35E-03
1.0	1.59E-01	1.57E-01	3.5	2.33E-04	2.32E-04
1.5	6.68E-02	6.63E-02	4.0	3.17E-05	3.16E-05
2.0	2.28E-02	2.26E-02	4.5	3.40E-06	3.40E-06
2.5	6.21E-03	6.19E-03	5.0	2.87E-07	2.87E-07

3.3.2 Joint cdfs and pdfs

It is not uncommon to have more that one random variable present in the chance experiment. As we move toward random processes, we will see that vector random variables are also a possibility. Presently we consider just two random variables.

- cdf

$$F_{X,Y}(x, y) = F[X \leq x, Y \leq y]$$

- pdf

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

Note:

$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u, v)\, du\, dv$$

Properties

1. $f_{X,Y}(x, y) \geq 0$
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y)\, dx dy = 1$

3. $P(x_1 < X \leq x_2, y_1 < Y \leq y_2) = \int_{x_1}^{x_2} \int_{y_1}^{y_2} f_{XY}(x, y)\, dx dy$
4. $F_X(x) = P(X \leq x, Y < \infty) = F_{XY}(x, \infty)$
5. $F_Y(y) = P(X < \infty, Y \leq y) = F_{XY}(\infty, y)$
6. $f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, u)\, du$
7. $f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(u, y)\, du$

Example 3.5: **Existence of a Joint pdf**

Consider the function

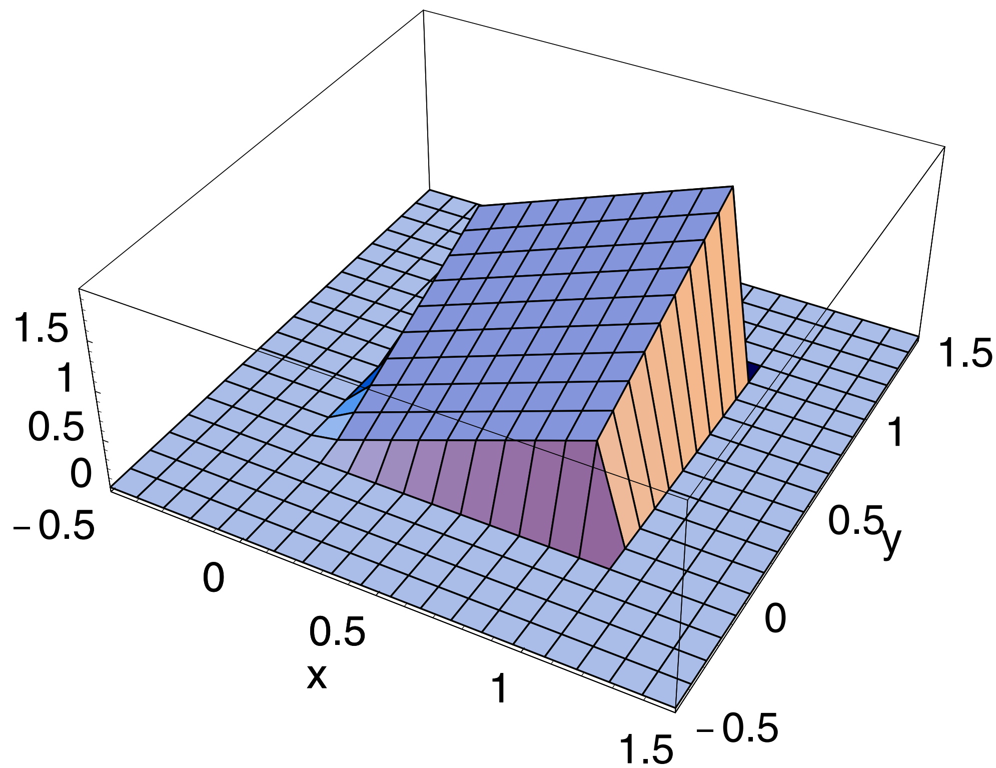
$$f_{XY}(x, y) = \begin{cases} A(x + y), & 0 \leq x \leq 1,\ 0 \leq y \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

- Can $f_{XY}(x, y)$ be a pdf corresponding to the random pair (X, Y)?
  1. Observe that $f_{XY}(x, y) \geq 0$ provided $A > 0$
  2. We can force $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y)\, dy dx = 1$ by proper choice of A

  So yes, we have a valid pdf

- To find A we integrate

$$\begin{aligned} \int_0^1 \int_0^1 A(x + y)\, dy dx &= A \int_0^1 \left[xy\big|_{y=0}^{1} + \frac{y^2}{2}\bigg|_{y=0}^{1} \right] dx \\ &= A \left[\frac{x^2}{2}\bigg|_0^1 + \frac{x}{2}\bigg|_0^1 \right] = A \\ &\Rightarrow\ A = 1 \end{aligned}$$



The joint (bivariate) pdf $f_{XY}(x, y)$

– Find the joint cdf $F_{XY}(x, y)$ by considering various ranges for x and y

Case 1:

$$F_{XY}(x, y) = 0, \ x < 0, \ y < 0$$

Case 2:

$$\begin{aligned} F_{XY}(x, y) &= \int_0^y \int_0^x (u + v)\, du dv, \ 0 \leq x, y \leq 1 \\ &= \int_0^y \left[\frac{u^2}{2} + vu\right]\Bigg|_0^x dv = \int_0^y \left(\frac{x^2}{2} + vx\right) dv \\ &= \left[\frac{x^2}{2}v + x\frac{v^2}{2}\right]\Bigg|_0^y = \frac{x^2 y + xy^2}{2} \end{aligned}$$

Case 3:

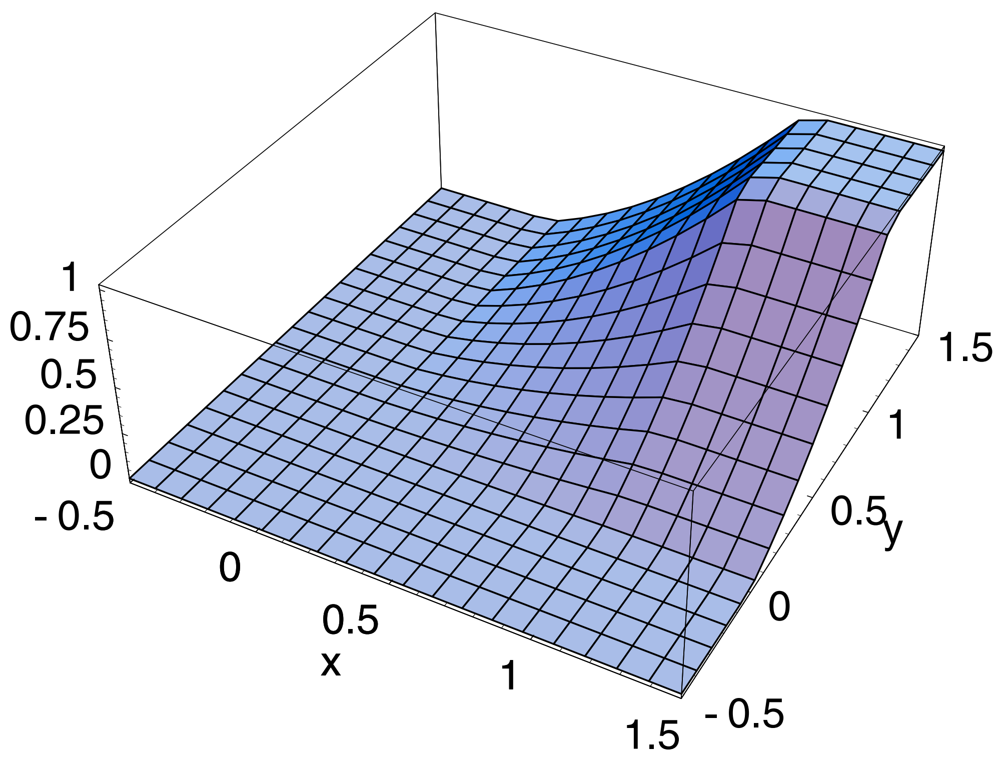
$$F_{XY}(x, y) = \frac{x^2 + x}{2},\ 0 \le x \le 1,\ y \ge 1$$

Case 4:

$$F_{XY}(x, y) = \frac{y^2 + y}{2},\ x \ge 1, 0 \le y \le 1$$

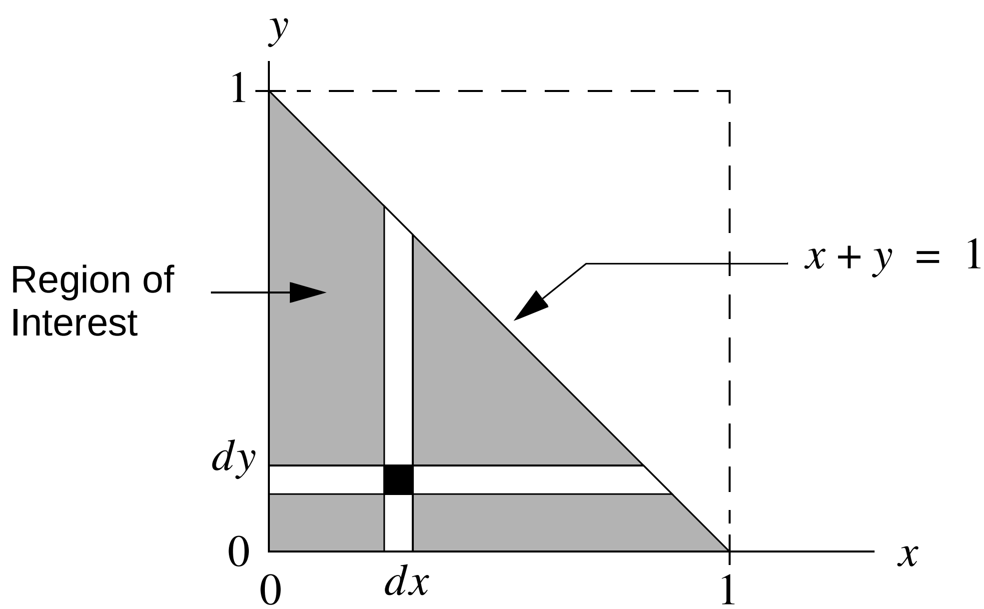
Case 5:

$$F_{XY}(x, y) = 1,\ x, y \ge 1$$



The bivariate cdf $F_{XY}(x, y)$

– Find $P(X+Y \leq 1)$ by first considering the area of interest



The region of integration

– The integral may be set up in one of two ways, $dxdy$ or $dydx$

$$
\begin{aligned}
P(X+Y \leq 1) &= \int_0^1 \int_0^{1-y} (x+y)\,dxdy \\
&= \int_0^1 \left[\frac{x^2}{2} + xy\right]\Bigg|_0^{1-y} dy \\
&= \int_0^1 \left[\frac{(1-y)^2}{2} + (1-y)y\right] dy \\
&= \int_0^1 \left[\frac{1}{2} - \frac{1}{2}y^2\right] dy = \frac{1}{2}[1 - \frac{1}{3}] = \frac{1}{3}
\end{aligned}
$$

- **Conditional pdfs:**

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

and Bayes rule

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

- **Independent RVs:**

$$\begin{aligned} F_{X,Y}(x,y) &= F_X(x)F_Y(y) \\ f_{X,Y}(x,y) &= f_X(x)f_Y(y) \end{aligned}$$

3.3.3 Transformation of Random Variables

In communications and signal processing applications we often need to consider

$$Y = g(X)$$

The pdf Method: Direct determination of the pdf on Y from the pdf on X and $g(X)$ is possible.
Theorem:

$$f_Y(y) = \sum_{i=1}^{n} \frac{f_X(x_i)}{|g'(x_i)|}$$

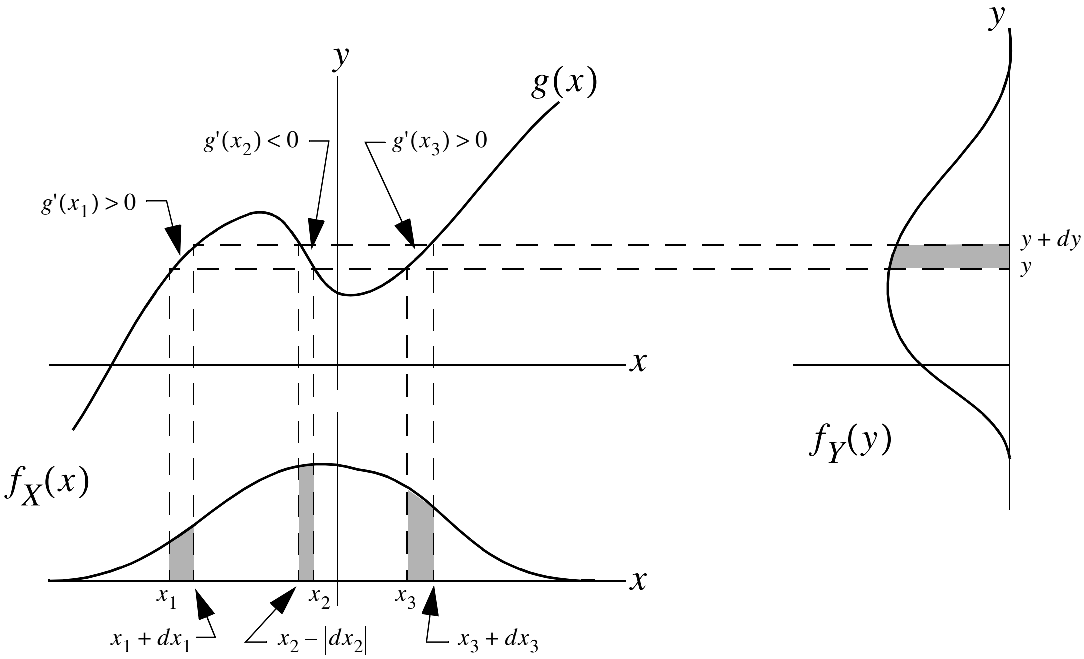
where $x_1, x_2, \ldots, x_n$ are <u>real roots</u> of $y = g(x)$, e.g., $x_i = g^{-1}(y)$ or $y - g(x_i) = 0$ and $g'(x_i) = g'(x)$ at $x = x_i$.

1. For some values of y there are no real roots, that is $g^{-1}(\{y\}) = \emptyset$, thus

$$f_Y(y) = 0 \text{ for these values}$$

2. If $g(x) = y_1$, a constant, for say $x \in (x_0, x_1)$, then $f_Y(y)$ contains an impulse at $y = y_1$, that is

$$\big[F_X(x_1) - F_X(x_0)\big]\delta(y - y_1)$$



Direct method theorem pictorially

- From the above picture we see that

$$P(y < Y \leq y + dy) = f_Y(y)|dy| = \sum_{i=1}^{n} f_X(x_i)|d(x_i)|$$

or

$$f_Y(y) = \sum_{i=1}^{n} \frac{f_X(x_i)}{\left|\frac{dy}{dx_i}\right|} = \sum_{i=1}^{n} \frac{f_X(x_i)}{|g'(x)|_{x=x_i}}$$

- <u>Note</u>: When we insert a root x_i into the right side, we always express it in terms of y so the entire expression for $f_Y(y)$ is in terms of y (the x's are gone)

Example 3.6: $g(X) = aX + b$

- For this transformation there is only one real root,

$$x_1 = \frac{y-b}{a}$$

- Using the theorem we observe that

$$g'(x_1) = \frac{d(ax_1 + b)}{dx_1} = a\big|_{x_1=(y-b)/a} = a$$

  so it follows that

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

- As a more specific example suppose further that X is Gaussian (normal) of the form $\mathcal{N}(\mu_x, \sigma_x^2)$

$$\begin{aligned} f_Y(y) &= \frac{1}{|a|} \cdot \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left[-\frac{((y-b)/a - \mu_x)^2}{2\sigma_x^2}\right] \\ &= \frac{1}{\sqrt{2\pi\ a^2\sigma_x^2}} \exp\left[-\frac{(y-(a\mu_x+b))^2}{2a^2\sigma_x^2}\right] \\ &= \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left[-\frac{(y-\mu_y)^2}{2\sigma_y^2}\right],\ y \text{ is also Gaussian!} \end{aligned}$$

  where $\mu_y = a\mu_x + b$ and $\sigma_y^2 = a^2\sigma_x^2$

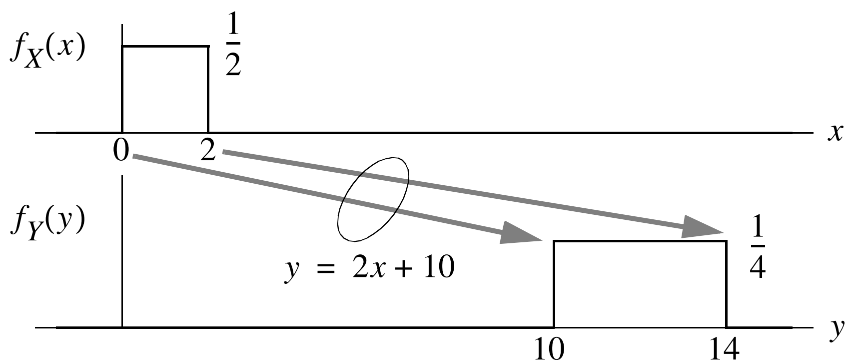
- As a second specific example suppose that $X \sim \mathcal{U}(0, 2)$ and $a = 2$ and $b = 10$

- We need to evaluate $f_X((y-10)/2))/|2|$, but since X is uniform we just need to find where f_X is active, that is for what interval

$$0 \le (y-10)/2 \le 2 \;\Rightarrow\; 10 \le y \le 14$$

- On this interval f_X has value $1/(2-0) = 1/2$, so in summary

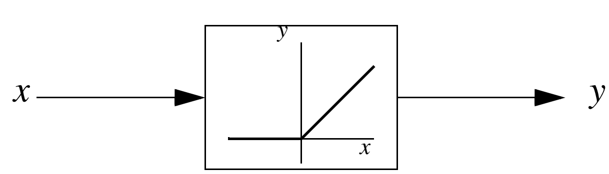
$$f_Y(y) = \begin{cases} 1/(2\cdot 2) = 1/4, & 10 \le y \le 14 \\ 0, & \text{otherwise} \end{cases}$$



Linear transform of a uniform rv

Example 3.7: **Half-wave Rectifier**

The transformation is $y = xu(x)$:



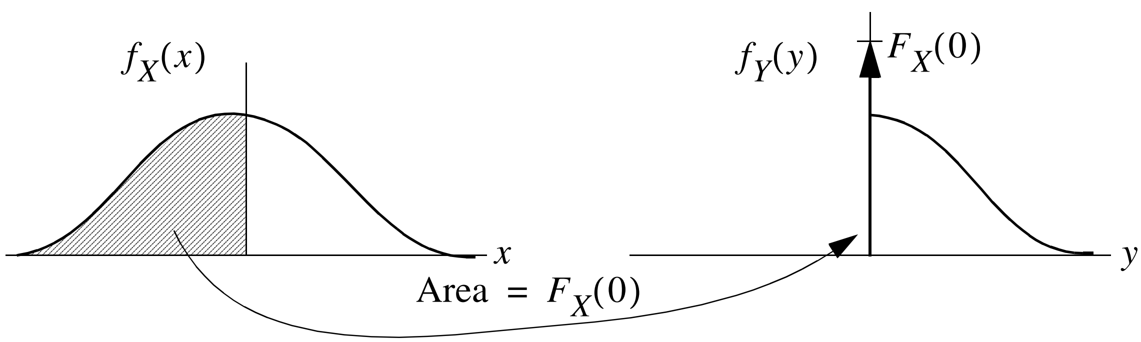
Half-wave Rectifier

- The range of y is $R_y = [0, \infty)$
- On this range there exists a single real root $x_1 = y$ for $y \geq 0$
- Using the theorem

$$g'(x_1) = 1,\ y \geq 0$$

- We also have the interval $x \in (-\infty, 0)$ where $g(x) = 0$, we must also include a delta function in the pdf
- In summary,

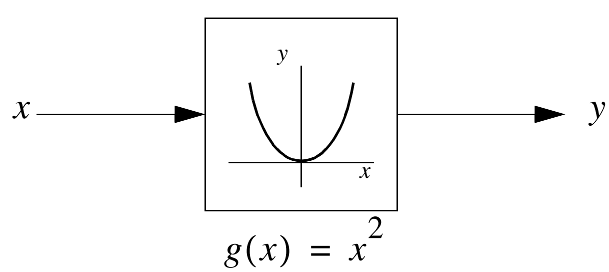
$$\begin{aligned} f_Y(y) &= f_X(y)u(y) + \big[F_X(0) - F_X(-\infty)\big]\delta(y) \\ &= f_X(y)u(y) + F_X(0)\delta(y) \end{aligned}$$



The pdf transformation

Example 3.8: **Square-Law Device with Gain**

The transformation is $y = g(x) = ax^2$:



- The range of y is $R_y = [0, \infty)$
- On this range there exists two real roots:

$$x_1 = \sqrt{\frac{y}{a}}, \quad x_2 = -\sqrt{\frac{y}{a}}$$

- Using the theorem

$$g'(x_1) = 2ax_1 = 2a\sqrt{\frac{y}{a}} = 2\sqrt{ay}$$
$$g'(x_2) = 2ax_2 = -2a\sqrt{\frac{y}{a}} = -2\sqrt{ay}$$

- Combining the above analysis we have

$$f_Y(y) = \frac{1}{2\sqrt{ay}}\left[f_X\left(\sqrt{\frac{y}{a}}\right) + f_X\left(-\sqrt{\frac{y}{a}}\right)\right]u(y)$$

A Special Transformation

Suppose that $y = g(x) = F_x(x)$ then $\mathbf{y}$ is uniform on [0, 1].

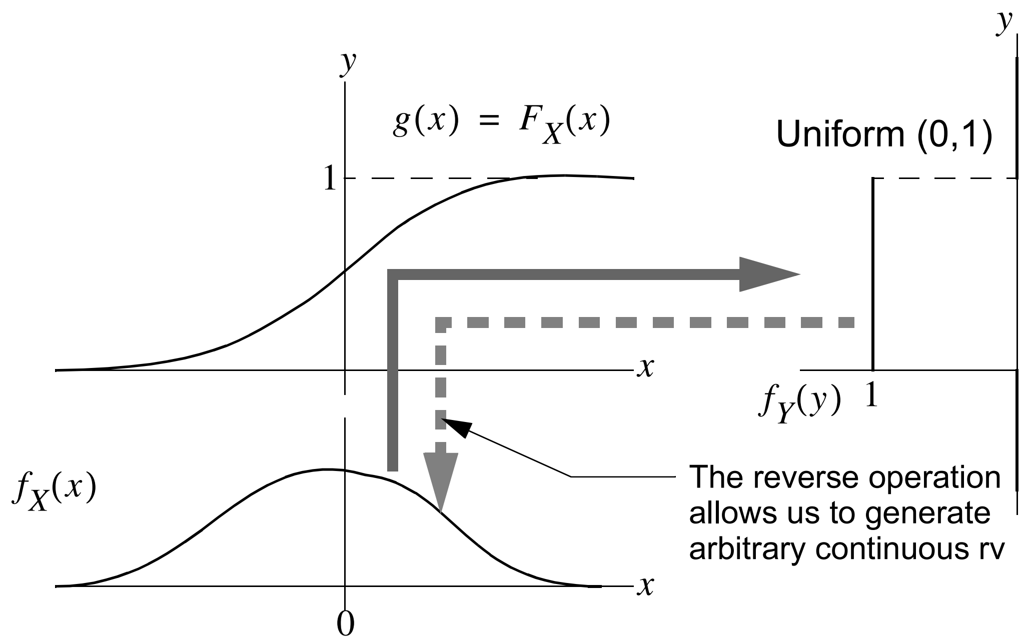
<u>proof</u>:

- Since $F_x(x)$ is monotonic $y = F_x(x)$ has a single root x_1 for $y \in [0, 1]$
- No other real roots exist
- Note also that

$$g'(x_1) = F'_{\mathbf{x}}(x_1) = f_x(x_1)$$

thus

$$f_y(y) = \begin{cases} \frac{f_X(x_1)}{|g'(x_1)|} = \frac{f_X(x_1)}{f_X(x_1)} = 1, & 0 \le y \le 1 \\ 0, & \text{otherwise} \end{cases}$$



Pictoral view of the transformation

3.3.4 One Function of Two RV

Suppose we have

$$Z = X + Y \text{ or } Z = XY$$

To find the pdf on

$$Z = g(X, Y)$$

we first find the cdf.

- Using the joint cdf we can write that

$$F_Z(z) = P(Z \le z) = P(g(X, Y) \le z)$$

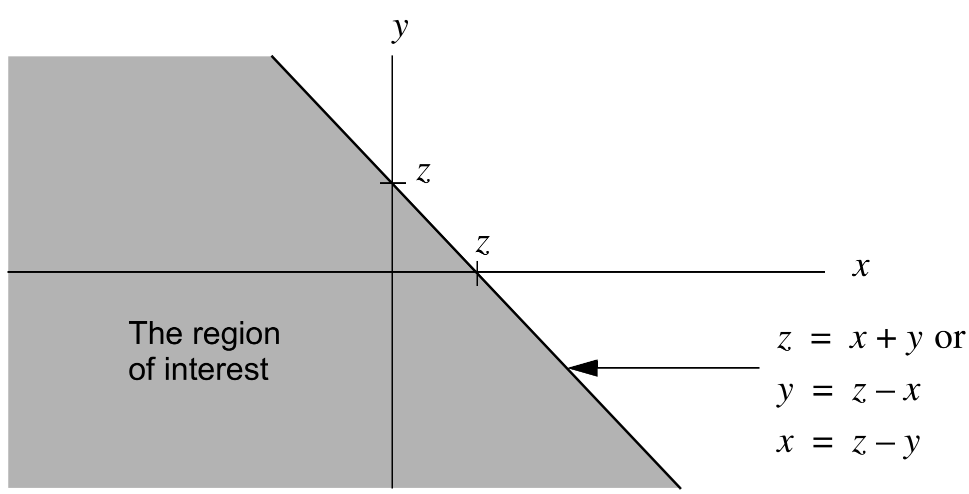
which usually requires that we integrate the joint cdf over some region of the $x{-}y$ plane

- Once we obtain the cdf on Z we then differentiate to obtain the pdf on Z

- The use of this *cdf* method is been understood from an example

---

Example 3.9: $Z = g(X, Y) = X + Y$



The region corresponding to $X + Y \le z$

- We need to integrate over the shaded region to obtain the cdf on rv Z

$$F_Z(z) = P(X + Y \le z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-y} f_{XY}(u, v)\, du dv$$

- Now we need to use *Leibnitz's Rule* to differentiate the integral(s)

Leibnitz's Rule for Differentiating Integrals:

$$\frac{d}{d\alpha}\int_{\phi_1(\alpha)}^{\phi_2(\alpha)} F(x,\alpha)\,dx = \int_{\phi_1(\alpha)}^{\phi_2(\alpha)} \frac{\partial F}{\partial\alpha}\,dx + F(\phi_2,\alpha)\frac{d\phi_2}{d\alpha} - F(\phi_1,\alpha)\frac{d\phi_1}{d\alpha}$$

- We need to apply Leibnitz's rule first for the outer integral and then for the inner integral

$$\begin{aligned} f_Z(z) &= \frac{d}{dz}\int_{-\infty}^{\infty}\left[\int_{-\infty}^{z-y} f_{XY}(x,y)\,dx\right]dy \\ &= \int_{-\infty}^{\infty}\frac{d}{dz}\left[\int_{-\infty}^{z-y} f_{XY}(x,y)dx\right]dy \\ &= \int_{-\infty}^{\infty} f_{XY}(z-y,y)\,dy \stackrel{\text{or}}{=} \int_{-\infty}^{\infty} f_{XY}(x,z-x)\,dx \end{aligned}$$

- **Special Case**: If X and Y are independent

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

then the pdf on $Z = X + Y$ is the convolution integral

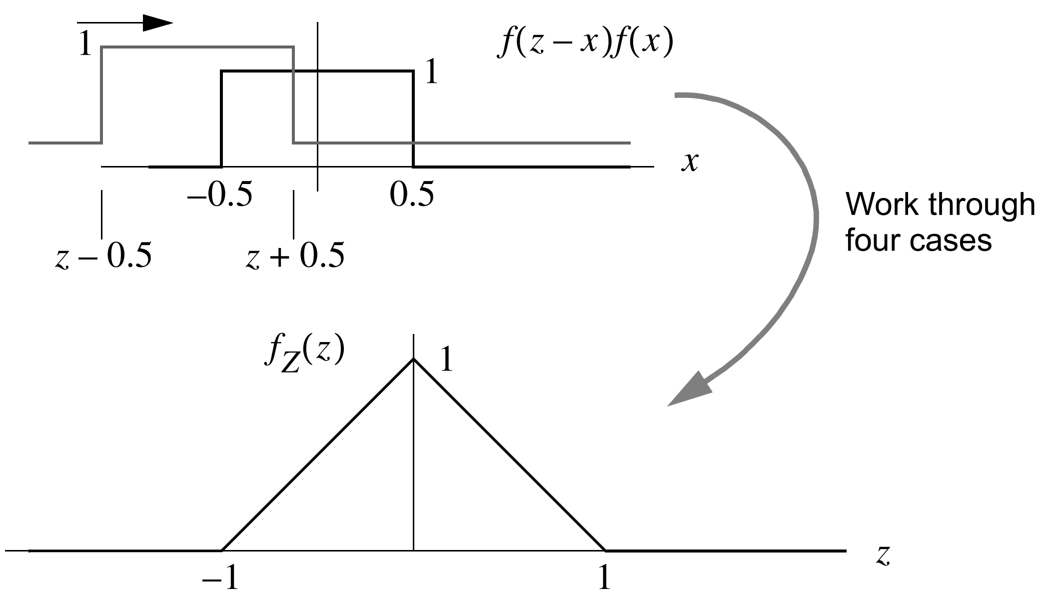
$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z-y)\,f_Y(y)\,dy = f_Y(z) * f_X(z)$$

or

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

- The important conclusion here is that for independent rv the density of the sum equals the *convolution* of the respective densities
- In particular suppose that X and Y are each $\mathcal{U}(-0.5, 0.5)$
- We need to consider four regions in the convolution integral, but can reduce the final answer to

$$f_Z(z) = \begin{cases} 1 - |z|, & |z| \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

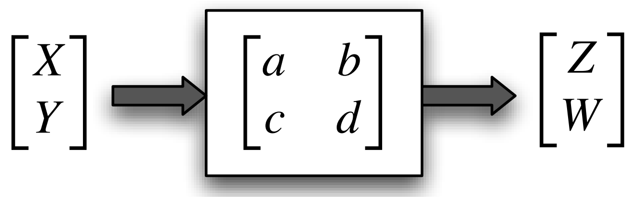


The convolution of two $\mathcal{U}(0, 1)$ pdf's

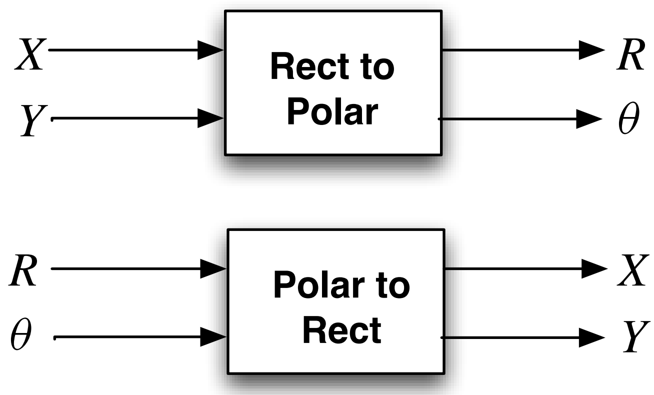
3.3.5 Two Functions of Two RV

We will now briefly consider the transformation of a pair of rv

$$Z = g_1(X, Y), \quad W = g_2(X, Y)$$



General Linear Transform



Special 2D Coordinate Transforms

$(X, Y) \rightarrow (Z, W)$ Functions of Interest

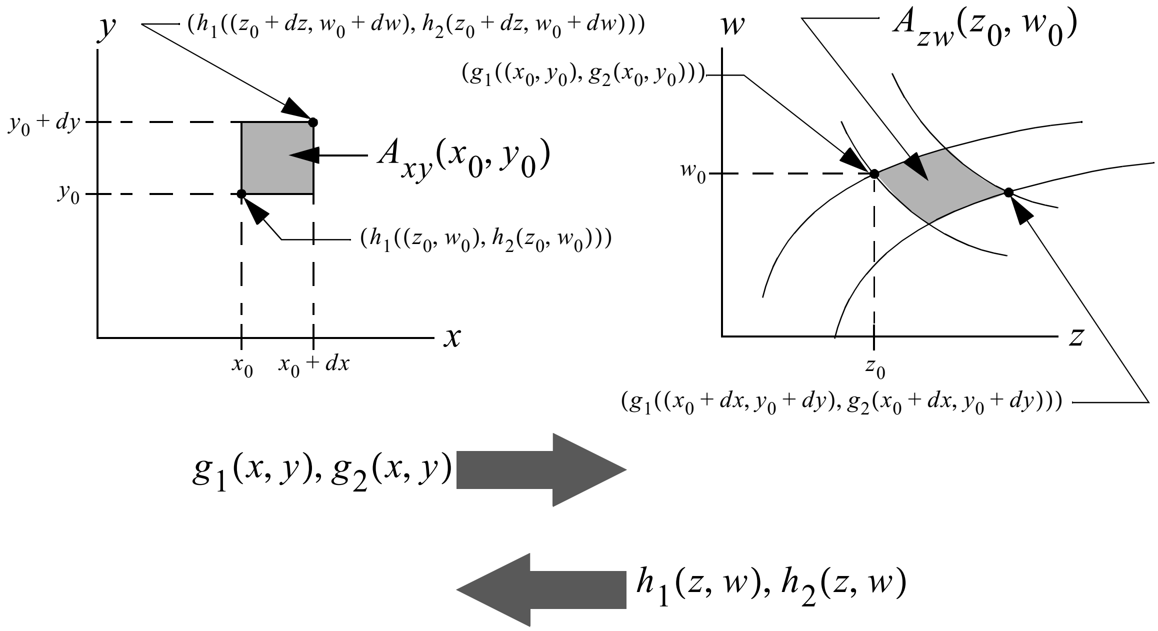
The basic problem is given the above transformation and $f_{XY}(x, y)$, what is the joint pdf on (Z, W)?

- To keep things simple we will assume that a single inverse mapping exits to carry points from the (z, w) plane back to the (x, y) plane, i.e.,

$$x = h_1(z, w), \quad y = h_2(z, w)$$

- To develop a transformation formula equate probabilities between small regions in the (x, y) and (z, w) planes

$$P(x_0 < X \le x_0 + dx, y_0 < Y \le y_0 + dy) = f_{XY}(x_0, y_0)A_{xy}(x_0, y_0) = f_{ZW}(z_0, w_0)A_{wz}(z_0, w_0)$$



Mapping a small rectangle from (x, y) to (z, w) and back

- The area $A_{xy}(x_0, y_0) = dxdy$
- The area $A_{zw}(z_0, w_0)$ is more complicated, but ultimately it can be shown that taking both areas into account

$$f_{ZW}(z, w) = \left.\frac{f_{XY}(x, y)}{|J(x, y)|}\right|_{x=h_1(z,w), y=h_2(z,w)}$$

where $J(x, y)$ is the *Jacobian* of the transformation which is the determinant of the matrix of partial derivatives

$$J(x, y) = \det\begin{bmatrix} \frac{\partial g_1(x,y)}{\partial x} & \frac{\partial g_1(x,y)}{\partial y} \\ \frac{\partial g_2(x,y)}{\partial x} & \frac{\partial g_2(x,y)}{\partial y} \end{bmatrix}$$

Example 3.10: A Linear Transform of Two RV $\to$ Two RV

Consider the linear transform

$$Z = aX + bY, \quad W = cX + dY$$

or in matrix form

$$\begin{bmatrix} Z \\ W \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} = \mathbf{A} \begin{bmatrix} X \\ Y \end{bmatrix}$$

- This transformation has a single inverse which is

$$\begin{aligned} \begin{bmatrix} X \\ Y \end{bmatrix} &= \mathbf{A}^{-1} \begin{bmatrix} Z \\ W \end{bmatrix} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \begin{bmatrix} Z \\ W \end{bmatrix} \\ &= \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} Z \\ W \end{bmatrix} \end{aligned}$$

- The Jacobian in this case is simply the determinant of the transformation matrix

$$J(x, y) = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

- Thus we can write that

$$f_{ZW}(z, w) = \frac{1}{|ad - bc|} f_{XY}(Az + Bw, Cz + Dw)$$

- Let us now see if a jointly Gaussian pdf transforms to another jointly Gaussian pdf

- As a special case suppose that

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

- Following the transformation we have

$$f_{ZW}(z,w) = \frac{1}{2\pi\sigma_x\sigma_y|ad-bc|}$$
$$\exp\left[-\left(\frac{(Az+Bw)^2}{2\sigma_x^2} + \frac{(Cz+Dw)^2}{2\sigma_y^2}\right)\right]$$

which has a quadratic in z and w in the exponential term, which ultimately can be manipulated to look like

$$f_{ZW}(z,w) = \frac{1}{2\pi\sigma_w\sigma_z\sqrt{1-\rho_{zw}^2}}$$
$$\exp\left[-\frac{1}{2(1-\rho_{zw}^2)}\left(\frac{z^2}{2\sigma_z^2} + 2\rho_{zw}\frac{zw}{\sigma_z\sigma_w} + \frac{w^2}{2\sigma_w^2}\right)\right]$$

Example 3.11: **Coordinate Conversion**

Consider the pair for functions that converts from rectangular coordinates to polar coordinates

$$R = \sqrt{X^2+Y^2}, \quad \Theta = \tan^{-1}(Y/X)$$

- This transformation has the single inverse function pair

$$x = r\cos\theta, \quad y = r\sin\theta,\ 0 \le \theta < 2\pi$$

- The Jacobian, $J(x,y)$ can be calculated as in the previous example, or we can use the inverse function pair, since with appropriate variable substitutions,

$$J(x,y) = 1/J(r,\theta) = \begin{vmatrix} \frac{\partial r\cos\theta}{\partial r} & \frac{\partial r\cos\theta}{\partial\theta} \\ \frac{\partial r\sin\theta}{\partial r} & \frac{\partial r\sin\theta}{\partial\theta} \end{vmatrix}^{-1}$$

$$= \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix}^{-1} = \frac{1}{r}$$

- The joint pdf on (R,Θ) is in general terms

$$f_{R\Theta}(r,\theta) = rf_{XY}(r\cos\theta, r\sin\theta),\ r \geq 0,\ 0 \leq \theta < 2\pi$$

- If (X,Y) are jointly Gaussian with zeros mean, equal variance, and uncorrelated (independent),

$$f_{R\Theta}(r,\theta) = \frac{r}{2\pi\sigma^2}e^{-r^2/(2\sigma^2)},\ r \geq 0,\ 0 \leq \theta < 2\pi$$

- By integrating to obtain the marginal pdf's we find that R and Θ are also independent

- As in a previous transformation example R is Rayleigh distributed and Θ is uniform on $[0, 2\pi]$

$$f_R(r) = \frac{r}{\sigma^2}e^{-r^2/(2\sigma^2)}u(r), \quad f_\Theta(\theta) = \frac{1}{2\pi},\ 0 \leq \theta < 2\pi$$

3.3.6 Statistical Averages

- Mean or first moment

$$m_X = \overline{X} = \mu_X = E[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx$$

Fundamental Theorem of Expectation

$$\boxed{E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx}$$

mean square $E[X^2] = \overline{X^2}$ or second moment

variance

$$\sigma_X^2 = E[(X - m_X)^2] = E[X^2] - m_X^2 = \overline{X^2} - \overline{X}^2$$

Note: $\sigma_X = \textit{standard deviation}$

- For more than one rv the fundamental theorem extends to

$$E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{XY}(x,y)\, dxdy$$

- Conditional expectation, which uses conditional pdf's, gives another option for computing $E[g(X,Y)]$

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

- Covariance and Correlation Coefficient

$$\mu_{XY} = E\big[(X - m_X)(Y - m_Y)\big] = E[XY] - E[X]E[Y]$$

- The normalized covariance is known as the correlation coefficient

$$\rho_{XY} = \frac{\mu_{XY}}{\sigma_X \sigma_Y}$$

Note: $-1 \leq \rho_{XY} \leq 1$

- Definition: Random variables X and Y are uncorrelated if

$$E[XY] = E[X]E[Y]$$

Note if either X or Y has zero mean, then uncorrelated rvs are also *orthogonal* i.e. $E[XY] = 0$

Example 3.12: **Uniform rv**

- We are given

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{otherwise} \end{cases}$$

- Find the mean of X

$$E[X] = \int_a^b x \, \frac{dx}{b-a} = \frac{1}{2}(a+b)$$

- Find the the mean-square of X

$$E[X^2] = \int_a^b x^2 \, \frac{dx}{b-a} = \frac{1}{3}(b^2 + ab + a^2)$$

- Find σ_x^2

$$\sigma_x^2 = E[X^2] - \big(E[X]\big)^2 = \frac{1}{3}(b^2 + ab + a^2) - \frac{1}{4}(a+b)^2$$
$$= \frac{1}{12}(a-b)^2$$

The Characteristic Function

- A special statistical average, known as the characteristic function, is obtained by letting $g(X) = e^{jvX}$

$$\boxed{M_X(jv) = E[e^{jvX}] = \int_{-\infty}^{\infty} f_X(x)e^{jvx}\,dx}$$

  - Note that $M_X(jv)$ would be the Fourier transform of $f_X(x)$ if $j\omega$ were replaced with $-jv$

- The connection with the Fourier transform allows us to write the inverse relationship

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} M_X(jv)e^{-jvx}\,dv$$

Example 3.13: **Revisit** $Z = X + Y$

- For X and Y independent we can use the characteristic function find the pdf of $Z = X + Y$

- We start with the characteristic function of Z

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

using the fact that $f_{XY}(x, y) = f_X(x) f_Y(y)$

- The integrals in the last line are separable

$$
\begin{aligned}
M_Z(jv) &= \int_{-\infty}^{\infty} e^{jvx} f_X(x)\, dx \int_{-\infty}^{\infty} e^{jvy} f_Y(y)\, dy \\
&= E\left[e^{jvX}\right] E\left[e^{jvY}\right] \\
&= M_X(jv) M_Y(jv)
\end{aligned}
$$

- From the convolution theorem for Fourier transforms it follows that

$$
f_Z(z) = f_X(x) * f_Y(y) = \int_{-\infty}^{\infty} f_X(z-u) f_Y(u)\, du
$$

- This is the result obtained in the earlier example

3.3.7 Multivariate Gaussian Random Variables

Two rv's are jointly normal (Gaussian) if they have joint pdf of the form

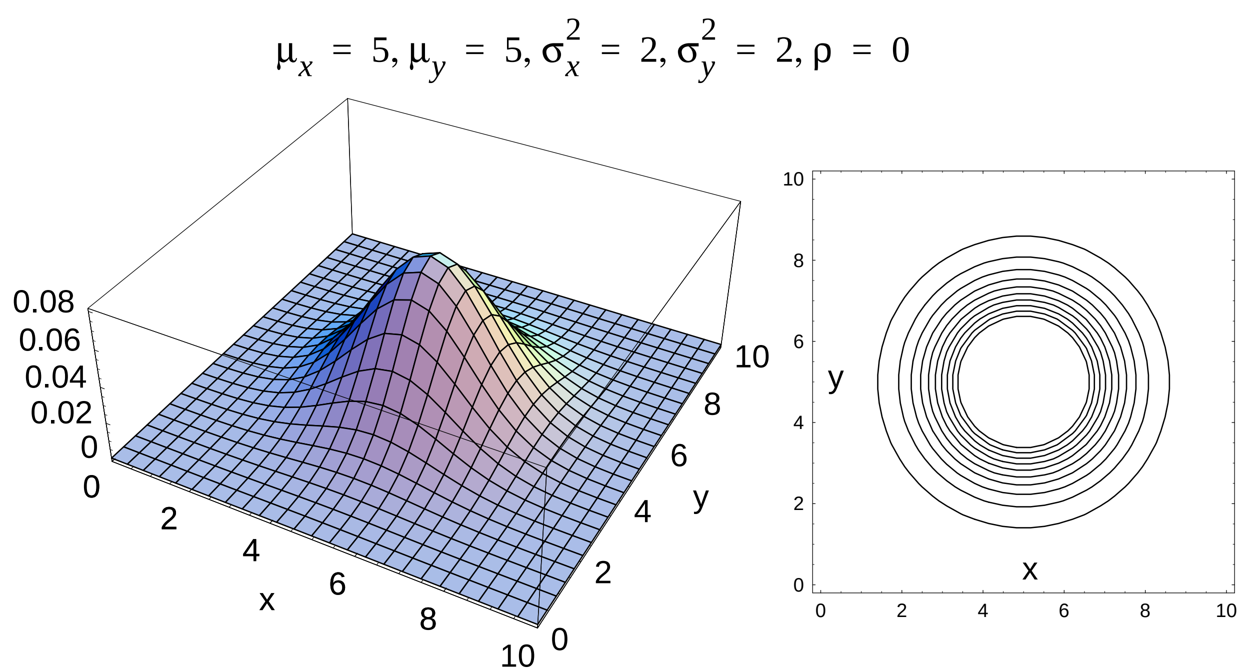
$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}} \exp\left\{-\frac{1}{2(1-\rho_{xy}^2)} \left[\frac{(x-\mu_x)^2}{\sigma_x^2} - \frac{2\rho_{xy}(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right]\right\}$$

where ρ is the correlation coefficient between X and Y

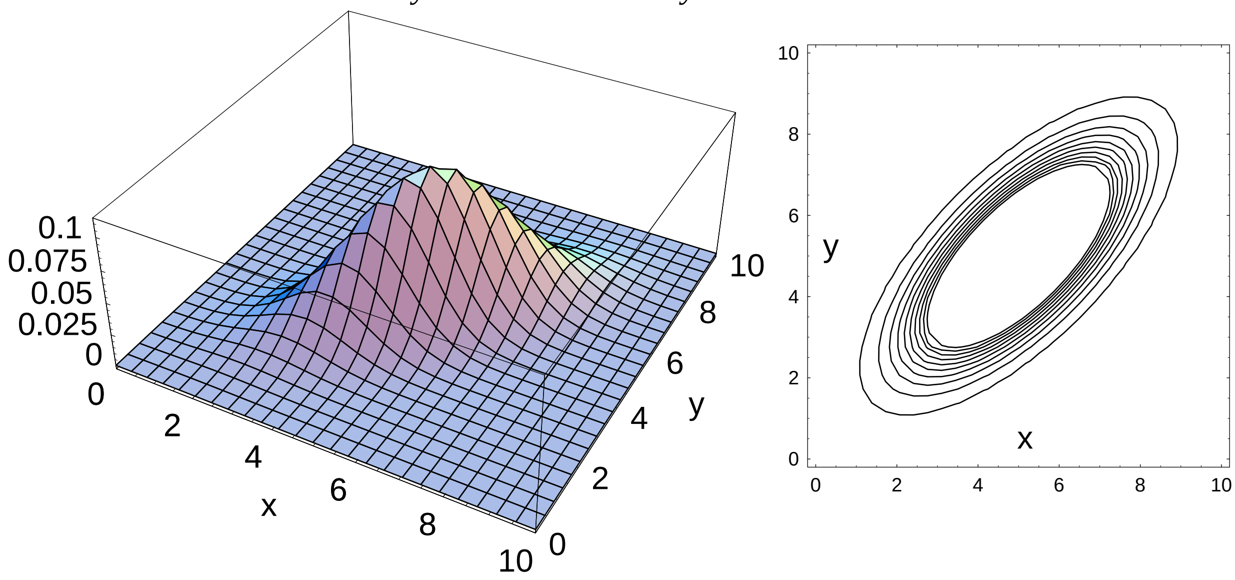
- In shorthand notation jointly Gaussian rv are denoted as

$$(X,Y) \sim \mathcal{N}(\mu_x, \mu_y; \sigma_x^2, \sigma_y^2, \rho_{xy})$$

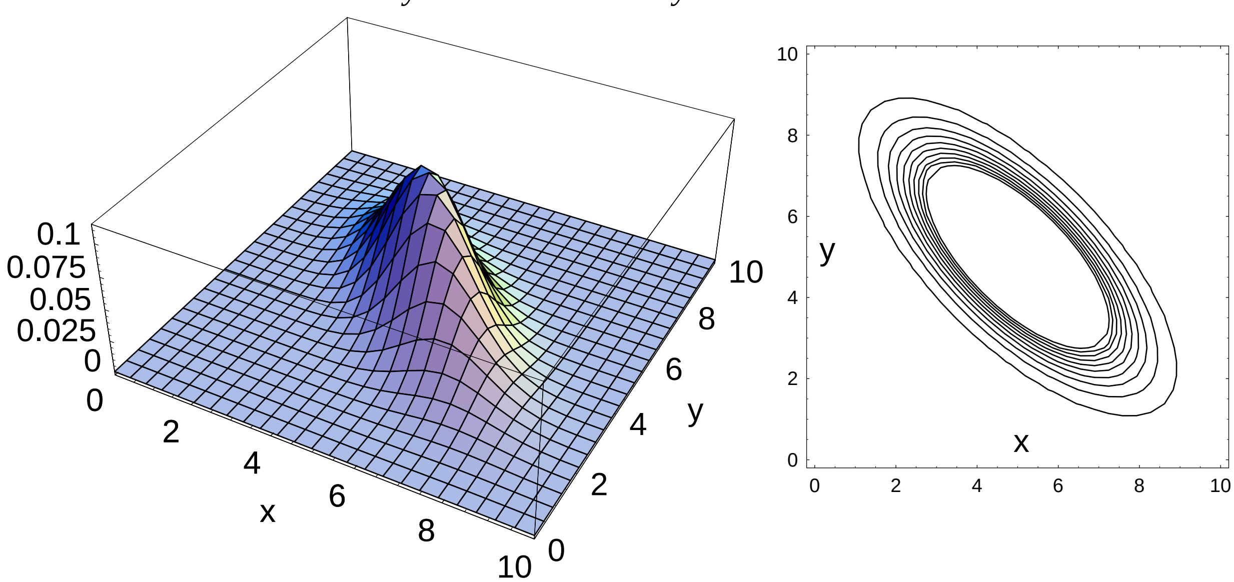
Example 3.14: **Bivariate Gaussian pdf Plots**



$\mu_x = 5, \mu_y = 5, \sigma_x^2 = 2, \sigma_y^2 = 2, \rho = 0.7$



$\mu_x = 5, \mu_y = 5, \sigma_x^2 = 2, \sigma_y^2 = 2, \rho = -0.7$



Properties of Jointly Normal RV

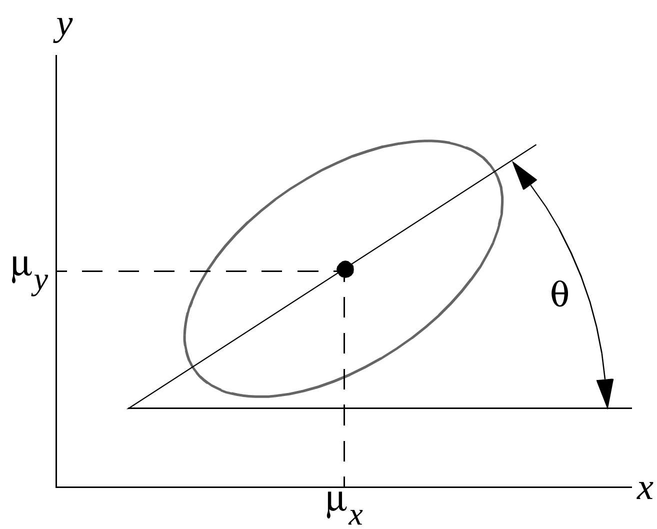
- As the above example shows the *level curves* or equal-pdf contours are ellipses centered at μ_x, μ_y

- The contours reduce to circles when $\rho = 0$ and the variances are equal, $\sigma_x^2 = \sigma_y^2$

- For the case of elliptical contours, it can be shown that the major axis of the ellipse has angle

$$\theta = \frac{1}{2}\tan^{-1}\left(\frac{2\rho\sigma_x\sigma_y}{\sigma_x^2 - \sigma_y^2}\right)$$

with respect to the x axis



Orientation angle of bivariate normal elliptical contours

- When the X and Y variances are equal and $\rho > 0$, the angle is 45°

- The marginal pdf's are also normal and do not depend upon ρ

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

- If $\rho = 0$, that is X and Y are uncorrelated, the cross term in the exponent vanishes and joint pdf becomes the product of the marginal pdf's, thus X and Y are also independent

- **Special Result**: For X and Y jointly Gaussian, they are independent <u>if and only if</u> they are also uncorrelated

  – <u>Note</u>: In general being uncorrelated does not imply independence

- A linear combination of jointly normal rv is another jointly normal random pair, e.g.,

$$Z = aX + bY + c, \quad W = \alpha X + \beta Y + \gamma$$

or in matrix form

$$\begin{bmatrix} Z \\ W \end{bmatrix} = \begin{bmatrix} a & b \\ \alpha & \beta \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} + \begin{bmatrix} c \\ \gamma \end{bmatrix}$$

- It can be shown (see section on two functions of two random variables) that the above linear transformation results in (Z, W) being jointly normal with parameters

$$\begin{aligned}
\mu_z &= a\mu_x + b\mu_y + c \\
\mu_w &= \alpha\mu_x + \beta\mu_y + \gamma \\
\sigma_z^2 &= a^2\sigma_x^2 + 2abC_{xy} + b^2\sigma_y^2 \\
\sigma_w^2 &= \alpha^2\sigma_x^2 + 2\alpha\beta C_{xy} + \beta^2\sigma_y^2 \\
C_{zw} &= E\Big[(aX + bY + C - a\mu_x - b\mu_y - C) \\
&\qquad \cdot (\alpha X + \beta Y + \gamma - \alpha\mu_x - \beta\mu_y - \gamma)\Big] \\
&= E\Big[\{a(X - \mu_x) + b(Y - \mu_y)\} \\
&\qquad \cdot \{\alpha(X - \mu_x) + \beta(Y - \mu_y)\}\Big] \\
&= a\alpha\sigma_x^2 + b\beta\sigma_y^2 + a\beta C_{xy} + \alpha b C_{xy}
\end{aligned}$$

- This property is an extension of the one dimensional case, that is for $Y = aX + b$ and X Gaussian, we discussed earlier that Y is also Gaussian with $\mu_y = a\mu_x + b$ and $\sigma_y^2 = a^2\sigma_x^2$

Example 3.15: **Linear Transform**

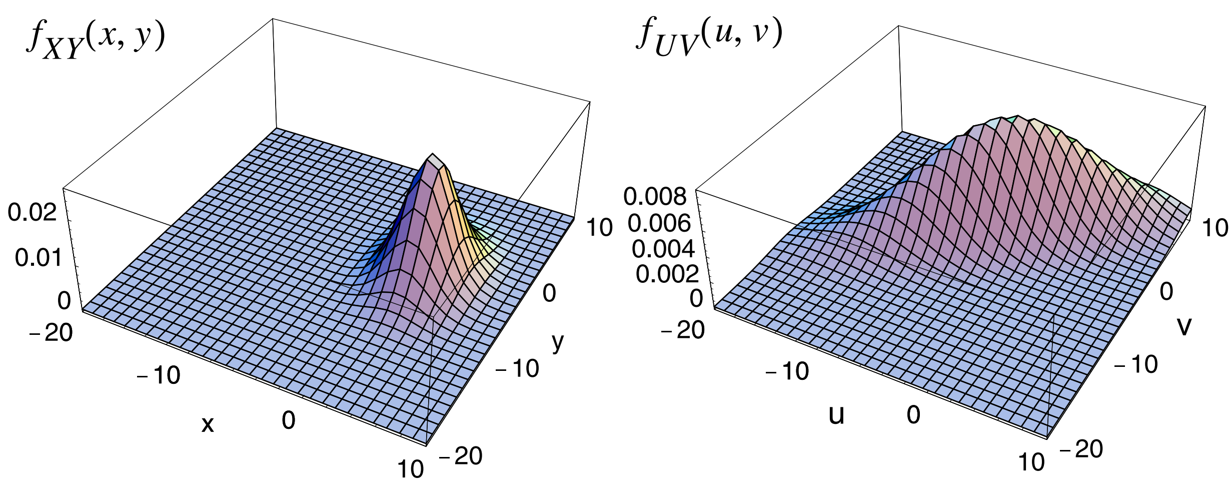
Consider (X, Y) uncorrelated jointly Gaussian rv with

$$X \sim \mathcal{N}(2, 4), \quad Y \sim \mathcal{N}(-3, 9)$$

$$U = X + 2Y, \quad V = 2X + Y$$

- Fully characterize the random pair (U, V)

$$\begin{aligned}
\mu_u &= \mu_x + 2\mu_y = 2 + 2(-3) = -4 \\
\mu_v &= 2\mu_x + \mu_y = 2(2) + (-3) = 1 \\
\sigma_u^2 &= \sigma_x^2 + 4\sigma_y^2 = 4 + 4(9) = 40 \\
\sigma_v^2 &= 4\sigma_y^2 + \sigma_y^2 = 4(4) + 9 = 25 \\
C_{uv} &= 2\sigma_x^2 + 2\sigma_y^2 + (1 + 4)C_{xy} = 2(4) + 2(9) + 0 = 26 \\
\rho_{uv} &= C_{uv}/\sqrt{\sigma_u^2\sigma_v^2} = 26/\sqrt{40 \times 25} = 0.8222
\end{aligned}$$



Bivariate Gaussian before and after linear transform

Example 3.16: Conditional Gaussian RV

- If we begin with a jointly Gaussian random pair (X, Y), e.g.,

$$(X, Y) \sim \mathcal{N}(\mu_x, \mu_y; \sigma_x^2, \sigma_y^2, \rho)$$

it can be shown that the conditional pdf's are also Gaussian

- To show this we plug into the definition and then complete the square on the exponent to recognize that a Gaussian pdf has been formed

- The result is

$$f_{X|Y}(x|y) = \mathcal{N}\Big(\mu_x + \rho\frac{\sigma_x}{\sigma_y}(y - \mu_y), \sigma_x^2(1 - \rho^2)\Big)$$
$$f_{Y|X}(y|x) = \mathcal{N}\Big(\mu_y + \rho\frac{\sigma_y}{\sigma_x}(x - \mu_x), \sigma_y^2(1 - \rho^2)\Big)$$

Example 3.17: Scatter Plots and Correlation Analysis

Suppose we are given a data set $(x_n, y_n),\ n = 1, 2, \ldots, N$ corresponding to samples of the rv pair (X, Y).

- By plotting the point pairs as isolated points on an $x - y$ graph we create what is known as a *scatter plot*

- From the scatter plot we can observe correlation properties between the rv samples x_n and y_n

- A simple MATLAB simulation can be used to generate correlated Gaussian random pairs and then plot the results as scatter plots

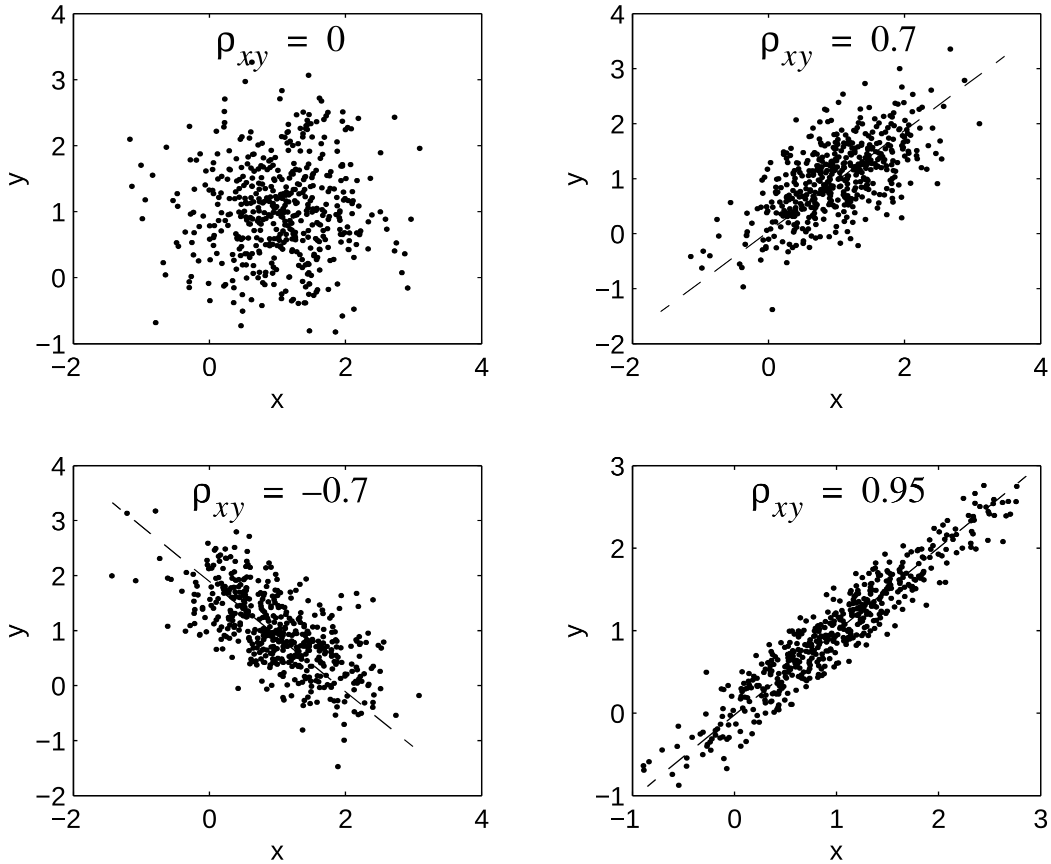
- Here we use the custom function `corrgauss()` to generate simulation values

```
>> help corrgauss

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function    y = corrgauss(mean,var,p,N)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

Generate Correlated Gaussian RV's
mean = mean vector [m1 m2]
var = variance vector [var1 var2]
    p = correlation coefficient
    N = number of random pairs to generate
========================================
>> y1 = corrgauss([1 1],[.5 .5],0,500);
>> y2 = corrgauss([1 1],[.5 .5],.7,500);
>> y3 = corrgauss([1 1],[.5 .5],-.7,500);
>> y4 = corrgauss([1 1],[.5 .5],.95,500);
>> subplot(221)
>> plot(y(1,:),y(2,:),'.')
>> xlabel('x')
>> ylabel('y')
>> subplot(222)
>> plot(y2(1,:),y2(2,:),'.')
>> xlabel('x')
>> ylabel('y')
>> subplot(223)
>> plot(y3(1,:),y3(2,:),'.')
>> xlabel('x')
>> ylabel('y')
>> subplot(224)
>> plot(y4(1,:),y4(2,:),'.')
>> xlabel('x')
>> ylabel('y')
```

- In Python the Scipy function `scipy.stats.multivariate_normal.rvs` generates correlated random pairs

- That code that follows generates 10000 random pairs given the five defining parameters $[m_x, m_y, \sigma_x^2, \sigma_y^2, \rho_{xy}]$

```
Cxy = [[varx,rhoxy*sqrt(varx*vary)],[rhoxy*sqrt(varx*vary),vary]]
xy = stats.multivariate_normal.rvs([mx,my],Cxy,10000)
```



Scatter plots of Gaussian random pairs

- Experimentally obtain some of the *descriptive statistics* associated with data set `y2`

- When MATLAB processes multiple columns of data it returns either a vector or matrix of results

- In the case of `mean()` two columns of input data in returned as a mean vector, e.g.,

$$\texttt{mean(xy\_col\_data)} = [\mu_x\ \mu_y]$$

- In the case of `cov()` a matrix of variances and covariances is returned, e.g.,

$$\texttt{cov(xy\_col\_data)} = \begin{bmatrix} \sigma_x^2 & C_{xy} \\ C_{yx} & \sigma_y^2 \end{bmatrix}$$

- Using the elements returned from `cov()` the correlation coefficient can be calculated

- Alternatively `corrcoef()` calculated a matrix correlation coefficients directly, that is

$$\texttt{corrcoef(xy\_col\_data)} = \begin{bmatrix} \rho_{xx} & \rho_{xy} \\ \rho_{yx} & \rho_{yy} \end{bmatrix}$$

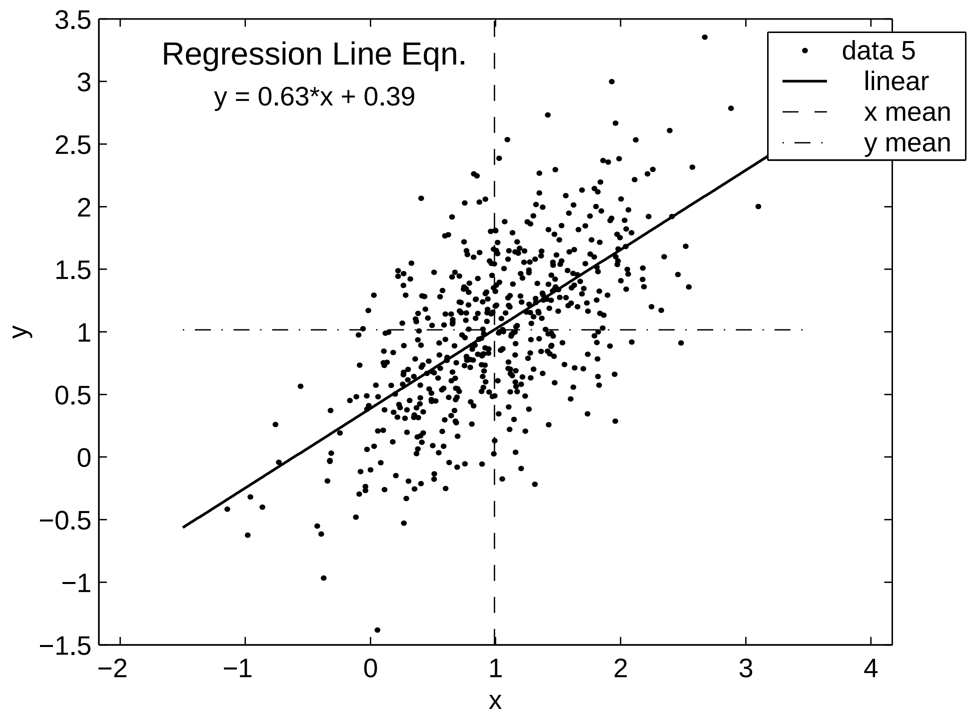
  Note that in theory we should have $\rho_{xx} = \rho_{yy} = 1$

- MATLAB processing of the data in `y2` which is in row format, so is transposed to column format (`y2 -> y2'`)

```
>> plot(y2(1,:),y2(2,:),'.')
>> mean(y2')

ans =
0.9912507617   1.0167419235 % Expect: 1.0 1.0

>> cov(y2')

ans =
0.4461329930   0.2831388209 % Expect: 0.5     0.7*.5
0.2831388209   0.4516831261 %         0.7*.5  0.5

>> corrcoef(y2')

ans =
1.0000000000   0.6307399144 % Expect: 1.0   0.7
0.6307399144   1.0000000000 %         0.7   1.0
```

- With a sample size of only 500 the results are close to theory

- Using the plot features found in MATLAB (since version 6, release 12), certain statistics can be plotted directly on the scatter plot
  - Using Data Statistics under the Tools menu, we can for example overlay plots of the mean
  - Using Basic Fitting under the Tools menu we can fit a linear model to the data which gives us the *linear regression line* whose slope is related to the correlation coefficient



Plot window statistics in MATLAB

3.3.8 Multiple Continuous RV

For the case of more than two continuous rv the extension is natural, but the details become challenging except for certain special cases. More than two discrete rv may also be considered under this framework if higher dimensionality delta functions are utilized. It also becomes convenient to introduce vector and matrix notation, e.g., for random variables $X_1, X_2, \ldots, X_p$ we can write

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}$$

We now what is known as a *random vector*.

Existence Theorem

Consider p random variables defined over the region R of p-dimensional space. The existence theorem states that the joint pdf on rvs X_1, X_2, ..., X_p must satisfy

1. $f_{\mathbf{X}}(\mathbf{x}) = f_{x_1 x_2 \cdots X_p}(x_1, x_2, \ldots, x_p) \geq 0$
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{x_1 x_2 \cdots X_p}(x_1, x_2, \ldots, x_p)\, dx_1\, dx_2 \cdots dx_p = 1$

Properties

- The joint cdf can be found by integrating all the variables up from $-\infty$ up to some threshold value $\mathbf{x}$

$$F_{\mathbf{X}}(\mathbf{x}) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_p} f_{\mathbf{X}}(\mathbf{u})\, d\mathbf{u}$$

- To find the probability that $\mathbf{X}$ lies within some region B we write

$$P(B) = \underbrace{\int \cdots \int}_{\mathbf{X} \in B} f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x}$$

- The marginal pdf on any $X_i, i = 1, 2, \ldots, p$ can be found as in the 2-dimensional case by integrating out the unwanted rv

$$f_{X_i}(x_i) = \underbrace{\int \cdots \int}_{\text{omit } x_i} f_{x_1 \cdots x_p}(x_1, \ldots, x_p)\ \underbrace{dx_1 \cdots dx_p}_{\text{omit } x_i}$$

- The mean and variance of each X_i can be obtained using the standard moment formulas which require the use of the corresponding marginal pdf's

- The joint pdf over a subset of rv, say $X_1, X_2, \ldots, X_k,\ k < p$ can be obtained by integrating the joint pdf over just the unwanted rvs $X_{k+1}, X_{k+2}, \ldots, X_p$,

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{x_1 \ldots x_k}(x_1, \ldots, x_k)\, dx_{k+1} \cdots dx_p$$

- Conditional pdfs can be defined such as

$$f_{X_1 X_2 X_3 | x_4 x_5}(x_1, x_2, x_3 | x_4, x_5) = \frac{f_{x_1 \cdots x_5}(x_1, \ldots, x_5)}{f_{x_4 x_5}(x_4, x_5)}$$

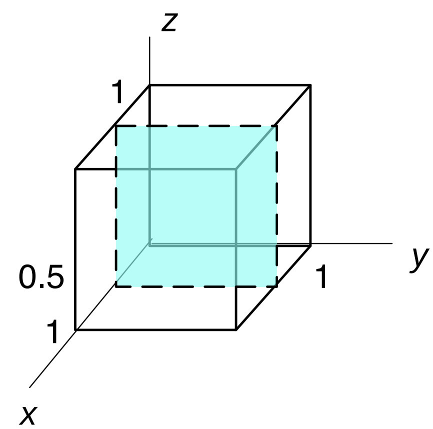
- If $X_1, \ldots, X_p$ are mutually independent, it then follows (if and only if) that

$$f_{X_1 \cdots X_p}(x_1, \ldots, x_p) = \prod_{i=1}^{p} f_{x_i}(x_i) = f_{x_1}(x_1) \cdots f_{x_p}(x_p)$$

Example 3.18: **Multiple Continuous RV**

Consider the joint pdf

$$f_{XYZ}(x,y,z) = \begin{cases} 8xyz, & 0 < x, y, z < 1 \\ 0, & \text{otherwise} \end{cases}$$



Support region of $f_{XYZ}(x,y,z)$

- Find $P(X < 0.5)$

$$\begin{aligned} P(X < 0.5) &= \int_0^{0.5}\int_0^1\int_0^1 8xyz\,dy\,dz\,dx \\ &= 8 \times \frac{x^2}{2}\Big|_0^{0.5} \times \frac{y^2}{2}\Big|_0^1 \times \frac{z^2}{2}\Big|_0^1 \\ &= 8 \times \frac{1}{8} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{4} \end{aligned}$$

- Find $f_{X|yz}(x|y = 0.5, z = 0.8)$

$$f_{X|yz}(x|y,z) = \frac{f_{XYZ}(x,y,z)}{f_{YZ}(y,z)} = \frac{8xyz}{f_{YZ}(y,z)}$$

- The joint pdf on Y, Z is

$$f_{YZ}(y,z) = \int_0^1 f_{XYZ}(x,y,z)\,dx = 8\left(\frac{x^2}{2}\bigg|_0^1\right) yz = 4yz$$

- Now,

$$f_{X|yz}(x|y,z) = \begin{cases} 2x, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

independent of specific values for y and z

- Find $P(X < 0.5|y = 0.5, z = 0.8)$

$$\begin{aligned} P(X < 0.5|y = 0.5, z = 0.8) &= P(X < 0.5|y,z) \\ &= \int_0^{0.5} 2x\,dx = 2\frac{x^2}{2}\bigg|_0^{0.5} \\ &= 0.25 \end{aligned}$$

3.3.9 Linear Combinations of RVs

There are times when we may wish to work with a collection of p random variables, $X_1, X_2, \ldots, X_p$. Let

$$Y = a_1X_1 + a_2X_2 + \cdots + a_pX_p = \sum_{i=1}^{p} a_iX_i$$

We may in particular be interested in just the mean and the variance.

- Find the mean of Y

$$E[Y] = \mu_y = E\left[\sum_{i=1}^{n} a_iX_i\right] = \sum_{i=1}^{p} a_iE[X_i] = \sum_{i=1}^{p} a_i\mu_{x_i}$$

- Find the variance of Y

$$\begin{aligned}
\sigma_y^2 &= E\left[\left(\sum_{i=1}^{p} a_i X_i - \sum_{i=1}^{p} a_i \mu_{x_i}\right)^2\right] \\
&= \sum_{i=1}^{p}\sum_{j=1}^{p} a_i a_j E\big[(X_i - \mu_{x_i})(X_j - \mu_{x_j})\big] \\
&= \sum_{i=1}^{p}\sum_{j=1}^{p} a_i a_j C_{x_i x_j} \\
&= \sum_{i=1}^{p} a_i^2 \sigma_{x_i}^2 + 2\sum_{i=1}^{p}\sum_{j>i}^{p} a_i a_j C_{x_i x_j}
\end{aligned}$$

where $C_{x_i x_j}$ is the covariance between X_i and X_j

  - For the special case of the rv's X_i mutually uncorrelated $C_{x_i x_j} = 0$ unless $i = j$, thus

$$\sigma_y^2 = \sum_{i=1}^{p} a_i^2 \sigma_{x_i}^2$$

  - Which says that the variance of a sum of uncorrelated rv's is the weighted sum of the variances
  - **Special Case 1**: $a_i = 1, i = 1, 2, \ldots, p$, then the variance of the sum is just the sum of the variances, i.e.,

$$\sigma_y^2 = \sigma_{x_1}^2 + \sigma_{x_2}^2 + \cdots + \sigma_{x_n}^2$$

  - **Special Case 2**: Consider the *average*

$$\bar{X} = (X_1 + X_2 + \cdots + X_p)/p = \frac{1}{p}\sum_{i=1}^{p} X_i$$

with $E[X_i] = \mu, i = 1, 2, \ldots, p$

– The mean and variance of the average is

$$E(\bar{X}) = \frac{1}{p}[p \cdot \mu] = \mu$$

If the X_i are also mutually uncorrelated, with equal variance $V(X_i) = \sigma^2, i = 1, 2, \ldots, p$, then

$$V(\bar{X}) = \frac{1}{p^2}[p \cdot \sigma^2] = \frac{\sigma^2}{p}$$

Sums of Normal RV

- Sums of normal rv frequency arise, in particular suppose that $X_1, X_2, \ldots, X_p$ are *uncorrelated* normal rv each having mean $E[X_i] = \mu_i$ and variance $V(X_i) = \sigma_i^2$ (independence allows the same results, but is a stronger assumption)

- Form the sum

$$Y = \sum_{i=1}^{p} a_i X_i$$

- A linear combination p normal rv is again a normal rv, so Y is normal with mean and variance given by

$$\mu_y = E[Y] = \sum_{i=1}^{p} a_i \mu_i$$

$$\sigma_y^2 = V(Y) = \sum_{i=1}^{p} a_i^2 \sigma_i^2$$

3.3.10 Table of pdf Properties

Probability-density or mass function	Mean	Variance
Uniform:		
$f_X(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & \text{otherwise} \end{cases}$	$\frac{1}{2}(a+b)$	$\frac{1}{12}(b-a)^2$
Gaussian:		
$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-(x-m)^2/2\sigma^2\right]$	m	σ^2
Rayleigh:		
$f_R(r) = \frac{r}{\sigma^2} \exp\left(-r^2/2\sigma^2\right), \ r \ge 0$	$\sqrt{\frac{\pi}{2}}\sigma$	$\frac{1}{2}(4-\pi)\sigma^2$
Laplacian:		
$f_X(x) = \frac{\alpha}{2} \exp(-\alpha\lvert x\rvert), \ \alpha > 0$	0	$2/\alpha^2$
One-sided exponential:		
$f_X(x) = \alpha \exp(-\alpha x) u(x)$	$1/\alpha$	$1/\alpha^2$
Hyperbolic:		
$f_X(x) = \frac{(m-1) h^{m-1}}{2(\lvert x\rvert + h)^m}, \ m > 3, \ h > 0$	0	$\frac{2h^2}{(m-3)(m-2)}$
Nakagami-m:		
$f_X(x) = \frac{2m^m}{\Gamma(m)} x^{2m-1} \exp\left(-mx^2\right), \ x \ge 0$	$\frac{1 \times 3 \times \cdots \times (2m-1)}{2^m \Gamma(m)}$	$\frac{\Gamma(m+1)}{\Gamma(m)\sqrt{m}}$
Central Chi-square (n = degrees of freedom)[1]:		
$f_X(x) = \frac{x^{n/2-1}}{\sigma^n 2^{n/2} \Gamma(n/2)} \exp\left(-x/2\sigma^2\right)$	$n\sigma^2$	$2n\sigma^4$
Lognormal[2]:		
$f_X(x) = \frac{1}{x\sqrt{2\pi\sigma_y^2}} \exp\left[-\left(\ln x - m_y\right)^2/2\sigma_y^2\right]$	$\exp\left(m_y + 2\sigma_y^2\right)$	$\exp\left(2m_y + \sigma_y^2\right)$ $\times\left[\exp\left(\sigma_y^2\right) - 1\right]$
Binomial:		
$P_n(k) = \binom{n}{k} p^k q^{n-k}, \ k = 0, 1, 2, \cdots, n, \ p + q = 1$	np	npq
Poisson:		
$P(k) = \frac{\lambda^k}{k!} \exp(-\lambda), \ k = 0, 1, 2, \cdots$	λ	λ
Geometric:		
$P(k) = pq^{k-1}, \ k = 1, 2, \cdots$	$1/p$	q/p^2

[1] $\Gamma(m)$ is the gamma function and equals $(m-1)!$ for m an integer.

[2] The lognormal random variable results from the transformation $Y = \ln X$ where Y is a Gaussian random variable with mean m_y and variance σ_y^2.