Chapter 1: Probability

Fall 2014
The course provides an introduction to the mathematics of data analysis and a detailed overview of statistical models for inference and prediction.

**Course Structure:**
- Weekly lectures [notes: www.ece.udel.edu/~arce/Courses]
- Homework & computer assignments [30%]
- Midterm & Final examinations [70%]

**Textbooks:**
- Papoulis and Pillai, Probability, random variables, and stochastic processes.
- Hastie, Tibshirani and Friedman, The elements of statistical learning.
- Haykin, Adaptive Filter Theory.
Assumption: Many methods take \( \{x(n)\} \) to be deterministic.

Reality: Real world signals are usually statistical in nature.

Thus,

\[ \ldots x(-1), x(0), x(1), \ldots \]

can be interpreted as a sequence of random variables.

We begin by analyzing each observation \( x(n) \) as a R.V.

Then, to capture dependencies, we consider random vectors

\[ \ldots x(n), x(n+1), \ldots, x(n+N-1), x(n+N), \ldots \]
Random Variables

Definition
For a space $S$, the subsets, or events of $S$, have associated probabilities.

- To every event $\delta$, we assign a number $x(\delta)$, which is called a R.V.
- The distribution function of $x$ is

$$\Pr\{x \leq x_0\} = F_x(x_0) \quad -\infty < x_0 < \infty$$

Properties:
1. $F(+\infty) = 1$, $F(-\infty) = 0$
2. $F(x)$ is continuous from the right

$$F(x^+) = F(x)$$
3. $\Pr\{x_1 < x \leq x_2\} = F(x_2) - F(x_1)$
Example
Fair toss of two coins: H=heads, T=Tails

Define numerical assignments:

<table>
<thead>
<tr>
<th>Events(δ)</th>
<th>Prob.</th>
<th>X(δ)</th>
<th>Y(δ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HH</td>
<td>1/4</td>
<td>1</td>
<td>-100</td>
</tr>
<tr>
<td>HT</td>
<td>1/4</td>
<td>2</td>
<td>-100</td>
</tr>
<tr>
<td>TH</td>
<td>1/4</td>
<td>3</td>
<td>-100</td>
</tr>
<tr>
<td>TT</td>
<td>1/4</td>
<td>4</td>
<td>500</td>
</tr>
</tbody>
</table>

This assignments yield different distribution functions

\[ F_x(2) = \text{Pr}\{HH, HT\} \]
\[ F_y(2) = \text{Pr}\{HH, HT, TH\} \]

How do we attain an intuitive interpretation of the distribution function?
**Distribution Plots**

Note properties hold:

1. \( F(\infty) = 1, \ F(-\infty) = 0 \)
2. \( F(x) \) is continuous from the right
   \[
   F(x^+) = F(x)
   \]
3. \( \Pr\{x_1 < x \leq x_2\} = F(x_2) - F(x_1) \)
Definition

The probability density function is defined as,

\[ f(x) = \frac{dF(x)}{dx} \]

or

\[ F(x) = \int_{-\infty}^{x} f(x) \, dx \]

Thus \( F(\infty) = 1 \Rightarrow \int_{-\infty}^{\infty} f(x) \, dx = 1 \)

Types of distributions:

- Continuous: \( \Pr\{x = x_0\} = 0 \quad \forall x_0 \)

- Discrete: \( F(x_i) - F(x_i^-) = \Pr\{x = x_i\} = P_i \)
  - In which case \( f(x) = \sum_i P_i \delta(x - x_i) \)

- Mixed: discontinuous but not discrete
Distribution examples

Uniform: $x \sim U(a, b) \quad a < b$

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in [a, b] \\ 0 & \text{else} \end{cases}$$
Gaussian: \( x \sim N(\mu, \sigma) \)

\[
f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}
\]
Gaussian Distribution Example

Example

Consider the Normal (Gaussian) distribution PDF and CDF for $\mu = 0, \sigma^2 = 0.2, 1.0, 5.0$ and $\mu = -2, \sigma^2 = 0.5$
Binomial: \( x \sim B(p, q) \quad p + q = 1 \)

Example

Toss a coin \( n \) times. What is the probability of getting \( k \) heads?

For \( p + q = 1 \), where \( q \) is probability of a tail, and \( p \) is the probability of a head:

\[
\Pr\{x = k\} = \binom{n}{k} p^k q^{n-k} \quad \left[ \text{NOTE:} \binom{n}{k} = \frac{n!}{(n-k)!k!} \right]
\]

\[
\Rightarrow f(x) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \delta(x - k)
\]

\[
\Rightarrow F(x) = \sum_{k=0}^{m} \binom{n}{k} p^k q^{n-k} \quad m \leq x < m+1
\]
Example

Toss a coin $n$ times. What is the probability of getting $k$ heads? For $n = 9$, $p = q = \frac{1}{2}$ (fair coin)
Example

Toss a coin $n$ times. What is the probability of getting $k$ heads? For $n = 20, p = 0.5, 0.7$ and $n = 40, p = 0.5$. 
Conditional Distributions

Definition
The conditional distribution of $x$ given event “$M$” has occurred is

$$F_x(x_0 | M) = \frac{\Pr\{ x \leq x_0, M \}}{\Pr\{ M \}}$$

Example
Suppose $M = \{ x \leq a \}$, then

$$F_x(x_0 | M) = \frac{\Pr\{ x \leq x_0, M \}}{\Pr\{ x \leq a \}}$$

If $x_0 \geq a$, what happens?
Special Cases

Special Case: $x_0 \geq a$

\[
\Pr\{x \leq x_0, x \leq a\} = \Pr\{x \leq a\}
\]

\[
\Rightarrow F_x(x_0|M) = \frac{\Pr\{x \leq x_0, M\}}{\Pr\{x \leq a\}} = \frac{\Pr\{x \leq a\}}{\Pr\{x \leq a\}} = 1
\]

Special Case: $x_0 \leq a$

\[
\Rightarrow F_x(x_0|M) = \frac{\Pr\{x \leq x_0, M\}}{\Pr\{x \leq a\}} = \frac{\Pr\{x \leq x_0\}}{\Pr\{x \leq a\}}
\]

\[
= \frac{F_x(x_0)}{F_x(a)}
\]
Conditional Distribution Example

Example

Suppose

What does \( F_x(x|M) \) look like? Note \( M = \{x \leq a\} \).

\[
\Rightarrow F_x(x_0|M) = \begin{cases} 
\frac{F_x(x_0)}{F_x(a)} & x \leq a \\
1 & a \leq x
\end{cases}
\]
Distribution properties hold for conditional cases:

- Limiting cases: $F(\infty|M) = 1$ and $F(-\infty|M) = 0$
- Probability range: $\text{Pr}\{x_0 \leq x \leq x_1| M\} = F(x_1|M) - F(x_0|M)$
- Density–distribution relations:

\[ f(x|M) = \frac{\partial F(x|M)}{\partial x} \]

\[ F(x_0|M) = \int_{-\infty}^{x_0} f(x|M) \, dx \]
Example (Fair Coin Toss)

Toss a fair coin 4 times. Let $x$ be the number of heads. Determine $\Pr\{x = k\}$.

Recall

$$\Pr\{x = k\} = \binom{n}{k} p^k q^{n-k}$$

In this case

$$\Pr\{x = k\} = \binom{4}{k} \left(\frac{1}{2}\right)^4$$

$$\Pr\{x = 0\} = \Pr\{x = 4\} = \frac{1}{16}$$

$$\Pr\{x = 1\} = \Pr\{x = 3\} = \frac{1}{4}$$

$$\Pr\{x = 2\} = \frac{3}{8}$$
What type of distribution is this? Discrete. Thus,

\[ F(x_i) - F(x_{i-}) = \Pr\{x = x_i\} = P_i \]

\[ F(x) = \int_{-\infty}^{x} f(x) dx = \int_{-\infty}^{x} \sum_i P_i \delta(x - x_i) dx \]
Conditional Case

Example (Conditional Fair Coin Toss)

Toss a fair coin 4 times. Let $x$ be the number of heads. Suppose $M = \{\text{at least one flip produces a head}\}$. Determine $\Pr\{x = k | M\}$.

Recall,

$$\Pr\{x = k | M\} = \frac{\Pr\{x = k, M\}}{\Pr\{M\}}$$

Thus first determine $\Pr\{M\}$

$$\Pr\{M\} = 1 - \Pr\{\text{No heads}\}$$

$$= 1 - \frac{1}{16}$$

$$= \frac{15}{16}$$
Next determine $\Pr\{x = k|M\}$ for the individual cases, $k = 0, 1, 2, 3, 4$

\[
\begin{align*}
\Pr\{x = 0|M\} &= \frac{\Pr\{x = 0, M\}}{\Pr\{M\}} = 0 \\
\Pr\{x = 1|M\} &= \frac{\Pr\{x = 1, M\}}{\Pr\{M\}} \\
&= \frac{\Pr\{x = 1\}}{\Pr\{M\}} = \frac{1/4}{15/16} = \frac{4}{15} \\
\Pr\{x = 2|M\} &= \frac{\Pr\{x = 2\}}{\Pr\{M\}} = \frac{3/8}{15/16} = \frac{6}{15} \\
\Pr\{x = 3|M\} &= \frac{4}{15} \\
\Pr\{x = 4|M\} &= \frac{1}{15}
\end{align*}
\]
Conditional and Unconditional Density Functions

Are they proper density functions?
Total Probability and Bayes’ Theorem

Let $M_1, M_2, \ldots, M_n$ forms a partition of $S$, i.e.

$$\bigcup_i M_i = S \quad \text{and} \quad M_i \cap M_j = \emptyset \quad (i \neq j)$$

Then

$$F(x) = \sum_i F_x(x|M_i)\Pr(M_i)$$

and

$$f(x) = \sum_i f_x(x|M_i)\Pr(M_i)$$

Aside

$$\Pr\{A|B\} = \frac{\Pr\{A, B\}}{\Pr\{B\}} = \frac{\Pr\{B, A\}\Pr\{A\}}{\Pr\{B\}\Pr\{A\}} = \frac{\Pr\{B|A\}\Pr\{A\}}{\Pr\{B\}}$$
From this we get

\begin{align*}
\Pr\{M|x \leq x_0\} &= \frac{\Pr\{x \leq x_0|M\} \Pr\{M\}}{\Pr\{x \leq x_0\}} \\
&= \frac{F(x_0|M) \Pr\{M\}}{F(x_0)}
\end{align*}

and

\begin{align*}
\Pr\{M|x = x_0\} &= \frac{f(x_0|M) \Pr\{M\}}{f(x_0)}
\end{align*}

By integration

\begin{align*}
\int_{-\infty}^{\infty} \Pr\{M|x = x_0\} f(x_0) \, dx_0 &= \int_{-\infty}^{\infty} f(x_0|M) \Pr\{M\} \, dx_0 \\
&= \Pr\{M\} \int_{-\infty}^{\infty} f(x_0|M) \, dx_0 = \Pr\{M\}
\end{align*}

\[ \Rightarrow \Pr\{M\} = \int_{-\infty}^{\infty} \Pr\{M|x = x_0\} f(x_0) \, dx_0 \]
Bayes’ Theorem:

\[ f(x_0|M) = \frac{\Pr\{M|x = x_0\}f(x_0)}{\Pr\{M\}} = \frac{\Pr\{M|x = x_0\}f(x_0)}{\int_{-\infty}^{\infty} \Pr\{M|x = x_0\}f(x_0)\,dx_0} \]
Functions of a R.V.

Problem Statement
Let $x$ and $g(x)$ be RVs such that

$$y = g(x)$$

Question: How do we determine the distribution of $y$?

Note

$$F_y(y_0) = \Pr\{y \leq y_0\} = \Pr\{g(x) \leq y_0\} = \Pr\{x \in R_{y_0}\}$$

where

$$R_{y_0} = \{x : g(x) \leq y_0\}$$

Question: If $y = g(x) = x^2$, what is $R_{y_0}$?
Example

Let \( y = g(x) = x^2 \). Determine \( F_y(y_0) \).

Note that

\[
F_y(y_0) = \Pr(y \leq y_0) = \Pr(-\sqrt{y_0} \leq x \leq \sqrt{y_0}) = F_x(\sqrt{y_0}) - F_x(-\sqrt{y_0})
\]
Example

Let $x \sim N(\mu, \sigma)$ and

$$y = U(x) = \begin{cases} 
1 & \text{if } x > \mu \\
0 & \text{if } x \leq \mu
\end{cases}$$

Determine $f_y(y_0)$ and $F_y(y_0)$. 

![Graphs of $f_y(y)$ and $F_y(y)$](image)
General Function of a Random Variable Case

To determine the density of \( y = g(x) \) in terms of \( f_x(x_0) \), look at \( g(x) \)

\[
f_y(y_0)dy_0 = \Pr(y_0 \leq y \leq y_0 + dy_0) \\
= \Pr(x_1 \leq x \leq x_1 + dx_1) + \Pr(x_2 + dx_2 \leq x \leq x_2) \\
+ \Pr(x_3 \leq x \leq x_3 + dx_3)
\]
\[ f_y(y_0)dy_0 = \Pr(x_1 \leq x \leq x_1 + dx_1) + \Pr(x_2 + dx_2 \leq x \leq x_2) \\
+ \Pr(x_3 \leq x \leq x_3 + dx_3) \\
= f_x(x_1)dx_1 + f_x(x_2)|dx_2| + f_x(x_3)dx_3 \quad (*) \]

Note that

\[ dx_1 = \frac{dx_1}{dy_0} dy_0 = \frac{dy_0}{dy_0/dx_1} = \frac{dy_0}{g'(x_1)} \]

Similarly

\[ dx_2 = \frac{dy_0}{g'(x_2)} \quad \text{and} \quad dx_3 = \frac{dy_0}{g'(x_3)} \]

Thus (*) becomes

\[ f_y(y_0)dy_0 = \frac{f_x(x_1)}{g'(x_1)} dy_0 + \frac{f_x(x_2)}{|g'(x_2)|} dy_0 + \frac{f_x(x_3)}{g'(x_3)} dy_0 \]

or

\[ f_y(y_0) = \frac{f_x(x_1)}{g'(x_1)} + \frac{f_x(x_2)}{|g'(x_2)|} + \frac{f_x(x_3)}{g'(x_3)} \]
Function of a R.V. Distribution General Result

Set $y = g(x)$ and let $x_1, x_2, \ldots$ be the roots, i.e.,

$$y = g(x_1) = g(x_2) = \ldots$$

Then

$$f_y(y) = \frac{f_x(x_1)}{|g'(x_1)|} + \frac{f_x(x_2)}{|g'(x_2)|} + \ldots$$

Example

Suppose $x \sim U(-1, 2)$ and $y = x^2$. Determine $f_y(y)$. 

![Graph of $f_x(x)$ and $y = x^2$]

Gonzalo R. Arce (ECE, Univ. of Delaware)
Note that

\[ g(x) = x^2 \Rightarrow g'(x) = 2x \]

Consider special cases separately:

**Case 1: \( 0 \leq y \leq 1 \)**

\[ y = x^2 \Rightarrow x = \pm \sqrt{y} \]

\[
\begin{align*}
f_y(y) &= \frac{f_x(x_1)}{|g'(x_1)|} + \frac{f_x(x_2)}{|g'(x_2)|} \\
&= \frac{1/3}{2\sqrt{y}} + \frac{1/3}{-2\sqrt{y}} = \frac{1/3}{\sqrt{y}}
\end{align*}
\]

**Case 2: \( 1 \leq y \leq 4 \)**

\[ y = x^2 \Rightarrow x = \sqrt{y} \]

\[
\begin{align*}
f_y(y) &= \frac{f_x(x_1)}{|g'(x_1)|} = \frac{1/3}{2\sqrt{y}} = \frac{1/6}{\sqrt{y}}
\end{align*}
\]
Result: For $x \sim U(-1, 2)$ and $y = x^2$

$$f_y(y) = \begin{cases} 
\frac{1/3}{\sqrt{y}} & 0 \leq y \leq 1 \\
\frac{1/6}{\sqrt{y}} & 1 < y \leq 4 
\end{cases}$$
Example

Let $x \sim N(\mu, \sigma)$ and $y = e^x$. Determine $f_y(y)$.

Note $g(x) \geq 0$ and $g'(x) = e^x$

Also, there is a single root (inverse solution):

$$x = \ln(y)$$

Therefore,

$$f_y(y) = \frac{f_x(x)}{|g'(x)|} = \frac{f_x(x)}{e^x}$$

Expressing this in terms of $y$ through substitution yields:

$$f_y(y) = \frac{f_x(\ln(y))}{e^{\ln(y)}} = \frac{f_x(\ln(y))}{y}$$
Note that $x$ is Gaussian:

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

\[
\Rightarrow f_y(y) = \frac{1}{\sqrt{2\pi y\sigma}} e^{-\frac{(\ln(y)-\mu)^2}{2\sigma^2}}, \text{ for } y > 0
\]
Distribution of $F_x(x)$

For any RV with continuous distribution $F_x(x)$, the RV $y = F_x(x)$ is uniform on $[0, 1]$.

Proof: Note $0 < y < 1$. Since

\[
g(x) = F_x(x) \\
g'(x) = f_x(x)
\]

Thus

\[
f_y(y) = \frac{f_x(x)}{g'(x)} = \frac{f_x(x)}{f_x(x)} = 1
\]

![Graph of $f_y(y)$ and $F_y(y)$]
Thus the function

\[ g(x) = F_x(x) \]

performs the mapping:

The converse also holds:

Combining operations yields

Synthesis:
Mean, Median and variance

Definitions

Mean
\[ E\{x\} = \int_{-\infty}^{\infty} xf(x) \, dx \]

Conditional Mean
\[ E\{x|M\} = \int_{-\infty}^{\infty} xf(x|M) \, dx \]

Example

Suppose \( M = \{x \geq a\} \). Then

\[ E\{x|M\} = \int_{-\infty}^{\infty} xf(x|M) \, dx \]
\[ = \frac{\int_{a}^{\infty} xf(x) \, dx}{\int_{a}^{\infty} f(x) \, dx} \]
For a function of a RV, $y = g(x)$,

$$E\{y\} = \int_{-\infty}^{\infty} y f_y(y) \, dy = \int_{-\infty}^{\infty} g(x) f_x(x) \, dx$$

### Example

Suppose $g(x)$ is a step function: Determine $E\{g(x)\}$.

$$E\{g(x)\} = \int_{-\infty}^{\infty} g(x) f_x(x) \, dx = \int_{-\infty}^{x_0} f_x(x) \, dx = F_x(x_0)$$
Median

Definitions

Median = \[ m = \int_{-\infty}^{m} f(x) dx = \int_{m}^{\infty} f(x) dx = \frac{1}{2} \]

Median \[ Pr\{x \leq m\} = Pr\{x \geq m\} \]

Example

Let \[ x \sim \lambda \exp^{-\lambda x} U(x) \]. Then \[ m = \frac{\ln(2)}{\lambda} \]
Definition (Variance)

\[
\text{Variance} \quad \sigma^2 = \int_{-\infty}^{\infty} (x - \eta)^2 f(x) \, dx
\]

where \( \eta = E\{x\} \). Thus,

\[
\sigma^2 = E\{(x - \eta)^2\} = E\{x^2\} - E^2\{x\}
\]

Example

For \( x \sim N(\eta, \sigma^2) \), determine the variance.

\[
f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\eta)^2}{2\sigma^2}}
\]

Note: \( f(x) \) is symmetric about \( x = \eta \) \( \Rightarrow E\{x\} = \eta \)

Also

\[
\int_{-\infty}^{\infty} f(x) \, dx = 1 \Rightarrow \int_{-\infty}^{\infty} e^{-\frac{(x-\eta)^2}{2\sigma^2}} \, dx = \sqrt{2\pi\sigma}
\]
\[ \int_{-\infty}^{\infty} e^{-\frac{(x-\eta)^2}{2\sigma^2}} \, dx = \sqrt{2\pi}\sigma \]

Differentiating w.r.t. \( \sigma \):

\[ \Rightarrow \int_{-\infty}^{\infty} \frac{(x-\eta)^2}{\sigma^3} e^{-\frac{(x-\eta)^2}{2\sigma^2}} \, dx = \sqrt{2\pi} \]

Rearranging yields

\[ \int_{-\infty}^{\infty} (x-\eta)^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\eta)^2}{2\sigma^2}} \, dx = \sigma^2 \]

or

\[ E\{ (x-\eta)^2 \} = \sigma^2 \]
Definition (Moments)

- Moments
  \[ m_n = E\{x^n\} = \int_{-\infty}^{\infty} x^n f(x)dx \]

- Central Moments
  \[ \mu_n = E\{(x-\eta)^n\} = \int_{-\infty}^{\infty} (x-\eta)^n f(x)dx \]

From the binomial theorem

\[ \mu_n = E\{(x-\eta)^n\} = E\left\{ \sum_{k=0}^{n} \binom{n}{k} x^k (-\eta)^{n-k} \right\} \]

\[ = \sum_{k=0}^{n} \binom{n}{k} m_k (-\eta)^{n-k} \]

\[ \Rightarrow \mu_0 = 1, \quad \mu_1 = 0, \quad \mu_2 = \sigma^2, \quad \mu_3 = m_3 - 3\eta m_2 + 2\eta^3 \]
Example

Let $x \sim N(0, \sigma^2)$. Prove

$$E\{x^n\} = \begin{cases} 
0 & n = 2k + 1 \\
1 \cdot 3 \cdots (n-1)\sigma^n & n = 2k
\end{cases}$$

For $n$ odd

$$E\{x^n\} = \int_{-\infty}^{\infty} x^n f(x) = 0$$

since $x^n$ is an odd function and $f(x)$ is an even function.

To prove the second part, use the fact that

$$\int_{-\infty}^{\infty} e^{-\alpha x^2} \, dx = \sqrt{\frac{\pi}{\alpha}}$$
Differentiate

\[ \int_{-\infty}^{\infty} e^{-\alpha x^2} \, dx = \sqrt{\frac{\pi}{\alpha}} \]

with respect to \( \alpha \), \( k \) times

\[ \Rightarrow \int_{-\infty}^{\infty} x^{2k} e^{-\alpha x^2} \, dx = \frac{1 \cdot 3 \cdots (2k - 1)}{2^k} \sqrt{\frac{\pi}{\alpha^{2k+1}}} \]

Let \( \alpha = \frac{1}{2\sigma^2} \), then

\[ \int_{-\infty}^{\infty} x^{2k} e^{-\frac{x^2}{2\sigma^2}} \, dx = 1 \cdot 3 \cdots (2k - 1)\sigma^{2k+1} \sqrt{2\pi} \]

Setting \( n = 2k \) and rearranging

\[ \int_{-\infty}^{\infty} x^n \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x^2}{2\sigma^2}} \, dx = 1 \cdot 3 \cdots (n - 1)\sigma^n \quad [QED] \]

**Note:** Variance is a measure of a RVs concentration around its mean
Tchebycheff Inequality

For any $\varepsilon > 0$,

$$
\Pr(|x - \eta| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}
$$

To prove this, note

$$
\Pr(|x - \eta| \geq \varepsilon) = \int_{-\infty}^{\eta - \varepsilon} f(x)dx + \int_{\eta + \varepsilon}^{\infty} f(x)dx
$$

$$
= \int_{|x - \eta| \geq \varepsilon} f(x)dx
$$

Also note that

$$
\sigma^2 = \int_{-\infty}^{\infty} (x - \eta)^2 f(x)dx
$$

$$
\geq \int_{|x - \eta| \geq \varepsilon} (x - \eta)^2 f(x)dx
$$
\[ \sigma^2 \geq \int_{|x-\eta| \geq \varepsilon} (x - \eta)^2 f(x) \, dx \]

Using the fact that \( |x - \eta| \geq \varepsilon \) in the above gives

\[ \sigma^2 \geq \varepsilon^2 \int_{|x-\eta| \geq \varepsilon} f(x) \, dx \]

\[ = \varepsilon^2 \Pr\{ |x - \eta| \geq \varepsilon \} \]

Rearranging gives the desired result

\[ \Rightarrow \Pr\{ |x - \eta| \geq \varepsilon \} \leq \left( \frac{\sigma}{\varepsilon} \right)^2 \]

QED
Jensen’s Inequality

For a real convex function \( \Psi \) and \( x \) a random variable,

\[
E[g(x)] \geq g(E[x]).
\]

with equality when \( g \) not strictly convex, \( g(x) = x \).
Inequality reversed for concave \( g \).

Since \( g \) is convex, it lies above \( L(x) \) so

\[
E(g(x)) \geq E(L(x)) = E(a + bx) = a + bE(x) = L(E(x)) = g(E(x))
\]
Definition (Characteristic Function)

The **characteristic function** of a random variable $x$ with pdf $f_x(x)$ is defined by

$$
\phi_x(\omega) = E\left(e^{i\omega x}\right) = \int_{-\infty}^{\infty} e^{i\omega x} f_x(x) \, dx
$$

If $f_x(x)$ is symmetric about 0 ($f_x(x) = f_x(-x)$), then $\phi_x(x)$ is real.

The magnitude of the characteristic function is bound by

$$
|\phi_x(\omega)| \leq \phi_x(0) = 1
$$

Theorem (Characteristic Function for the sum of independent RVs)

**Let** $x_1, x_2, \ldots, x_N$ be independent (but not necessarily identically distributed) **RVs** and set $s_N = \sum_{i=1}^{n} a_i x_i$ **where** $a_i$ **are constants.** Then

$$
\phi_{s_N}(\omega) = \prod_{i=1}^{N} \phi_{x_i}(a_i \omega)
$$
The theorem can be proved by a simple extension of the following: Let $x$ and $y$ be independent. Then

$$
\phi_{x+y}(\omega) = E \left( e^{j\omega(x+y)} \right)
$$

$$
= E \left( e^{j\omega x} e^{j\omega y} \right) = E \left( e^{j\omega x} \right) E \left( e^{j\omega y} \right)
$$

$$
= \phi_x(\omega) \phi_y(\omega)
$$

**Example**

Determine the characteristic function of the sample mean operating on iid samples.

Note $\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i \Rightarrow a_i = \frac{1}{N}$

$$
\Rightarrow \phi_{\bar{x}}(\omega) = \prod_{i=1}^{N} \phi_{x_i}(a_i \omega) = \left( \phi_{x_i} \left( \frac{\omega}{N} \right) \right)^N
$$
The Moment Generating function is realized by making the substitution $j\omega \rightarrow s$ in the above

**Definition (Moment Generating Function)**

The moment generating function of a random variable $x$ with pdf $f_x(x)$ is defined by

$$\Phi_x(s) = E(e^{sx}) = \int_{-\infty}^{\infty} e^{sx} f_x(x) \, dx$$

Note $\Phi_x(j\omega) = \phi_x(\omega)$

**Theorem (Moment Generation)**

Provided that $\Phi_x(s)$ exists in an open interval around $s = 0$, the following hold

$$m_n = E(x^n) = \Phi_x^{(n)}(0) = \frac{d^n \Phi_x}{ds^n}(0)$$

Simply noting that $\Phi_x^{(n)}(s) = E(x^n e^{sx})$ proves the result
Example

Let $x$ be exponentially distributed,

$$f(x) = \lambda e^{-\lambda x} U(x)$$

Determine $\eta = m_1, m_2$, and $\sigma^2$

Note

$$\Phi_x(s) = \lambda \int_0^\infty e^{sx} e^{-\lambda x} dx = \lambda \int_0^\infty e^{-x(\lambda - s)} dx$$

$$= \frac{\lambda}{\lambda - s}$$

Thus

$$\Phi_x^{(1)}(0) = \frac{1}{\lambda} \quad \text{and} \quad \Phi_x^{(2)}(0) = \frac{2}{\lambda^2}$$

and

$$E\{x\} = \frac{1}{\lambda}, \quad E\{x^2\} = \frac{2}{\lambda^2} \quad \Rightarrow \quad \sigma^2 = \frac{1}{\lambda^2}$$
Bivariate Statistics

Given two RVs, $x$ and $y$, the bivariate (joint) distribution is given by

$$F(x_0, y_0) = \Pr\{x \leq x_0, y \leq y_0\}$$

Properties:

- $F(-\infty, y) = F(x, -\infty) = 0$
- $F(\infty, \infty) = 1$
- $F_x(x) = F(x, \infty)$, $F_y(y) = F(\infty, y)$
Special Cases

Case 1: \( M = \{x_1 \leq x \leq x_2, y \leq y_0\} \)

\[ \Rightarrow \Pr\{M\} = F(x_2, y_0) - F(x_1, y_0) \]

Case 2: \( M = \{x \leq x_0, y_1 \leq y \leq y_2\} \)

\[ \Rightarrow \Pr\{M\} = F(x_0, y_2) - F(x_0, y_1) \]
Case 3: \( M = \{ x_1 \leq x \leq x_2, y_1 \leq y \leq y_2 \} \) Then

\[
\Pr\{M\} = F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1)
\]

Added back because this region was subtracted twice.
Definition (Joint Statistics)

\[ f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y} \]

and

\[ F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(\alpha, \beta) \, d\alpha \, d\beta \]

In general, for some region \( M \), the joint statistics are

\[ \Pr\{(x, y) \in M\} = \int_M \int f(x, y) \, dx \, dy \]

Marginal Statistics: \( F_x(x) = F(x, \infty) \) and \( F_y(y) = F(\infty, y) \)

\[ \Rightarrow f_x(x) = \int_{-\infty}^{\infty} f(x, y) \, dy \]

\[ \Rightarrow f_y(y) = \int_{-\infty}^{\infty} f(x, y) \, dx \]
Independence

**Definition (Independence)**

Two RVs $x$ and $y$ are statistically independent if for arbitrary events (regions) $x \in A$ and $y \in B$,

$$
\Pr\{x \in A, y \in B\} = \Pr\{x \in A\}\Pr\{y \in B\}
$$

Letting $A = \{x \leq x_0\}$ and $B = \{y \leq y_0\}$, we see $x$ and $y$ are independent iff

$$
F_{x,y}(x,y) = F_x(x)F_y(y)
$$

and by differentiation

$$
f_{x,y}(x,y) = f_x(x)f_y(y)
$$
If \( x \) and \( y \) are independent \( RVs \), then

\[
z = q(x) \quad \text{and} \quad w = h(y)
\]

are also independent.

**Function of two \( RVs \)**

Given two \( RVs \), let \( z = g(x, y) \). Define \( D_z \) to be the \( xy \) plane region

\[
\{ z \leq z_0 \} = \{ g(x, y) \leq z_0 \} = \{ (x, y) \in D_z \}
\]

Then

\[
F_z(z_0) = Pr\{ z \leq z_0 \} = Pr\{ (x, y) \in D_z \} = \int \int_{D_z} f(x, y) \, dx \, dy
\]
Example

Let \( z = x + y \). Then, \( z \leq z_0 \) gives the region \( x + y \leq z_0 \) which is delineated by the line \( x + y = z_0 \)

\[
\begin{align*}
F_z(z_0) &= \int \int_{D_z} f(x, y) \, dx \, dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{z_0 - y} f(x, y) \, dx \, dy
\end{align*}
\]
We can obtain $f_z(z)$ by differentiation

$$
\frac{\partial F_z(z)}{\partial z} = \int_{-\infty}^{\infty} \frac{\partial}{\partial z} \int_{-\infty}^{z-y} f(x, y) \, dx \, dy
$$

$$
f_z(z) = \int_{-\infty}^{\infty} f(z - y, y) \, dy \quad (\ast)
$$

Note that if $x$ and $y$ are independent,

$$
f(x, y) = f_x(x) f_y(y) \quad (\ast\ast)
$$

Thus utilizing $(\ast\ast)$ in $(\ast)$

$$
f_z(z) = \int_{-\infty}^{\infty} f_x(z - y) f_y(y) \, dy \quad \text{Convolution}
$$

$$
= f_x(z) * f_y(z)
$$
Example

Let $z = x + y$ where $x$ and $y$ are independent with

$$f_x(x) = \alpha e^{-\alpha x} U(x)$$
$$f_y(y) = \alpha e^{-\alpha y} U(y)$$

Then

$$f_z(z) = \int_{-\infty}^{\infty} f_x(z - y) f_y(y) dy$$

$$= \alpha^2 \int_{0}^{z} e^{-\alpha(z-y)} e^{-\alpha y} dy$$

$$= \alpha^2 e^{-\alpha z} \int_{0}^{z} dy$$

$$= \alpha^2 z e^{-\alpha z} U(z)$$
Example

Let \( z = \max(x, y) \). Determine \( F_z(z_0) \) and \( f_z(z_0) \).

Note

\[
F_z(z_0) = \Pr\{z \leq z_0\} = \Pr\{\max(x, y) \leq z_0\} = F_{xy}(z_0, z_0)
\]
If $x$ and $y$ are independent,

$$F_z(z_0) = F_x(z_0)F_y(z_0)$$

and

$$f_z(z_0) = \frac{\partial F_z(z_0)}{\partial z_0}$$

$$= \frac{\partial F_x(z_0)}{\partial z_0}F_y(z_0) + \frac{\partial F_y(z_0)}{\partial z_0}F_x(z_0)$$

$$= f_x(z_0)F_y(z_0) + f_y(z_0)F_x(z_0)$$
Joint Moments

For RVs $x$ and $y$ and function $z = g(x, y)$

$$E\{z\} = \int_{-\infty}^{\infty} zf_z(z)dz$$
$$E\{g(x, y)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)f(x, y)dxdy$$

Definition (Covariance)

For RVs $x$ and $y$,

$$C_{xy} = \text{Cov}(x, y)$$
$$= E[(x - \eta_x)(y - \eta_y)]$$
$$= E[xy] - \eta_x E[y] - \eta_y E[x] + \eta_x \eta_y$$
$$= E[xy] - \eta_x \eta_y$$
Definition (Correlation Coefficient)

The correlation coefficient is given by

\[ r = \frac{C_{xy}}{\sigma_x \sigma_y} \]

Note that

\[ 0 \leq E \left\{ [a(x - \eta_x) + (y - \eta_y)]^2 \right\} \]
\[ = E\{(x - \eta_x)^2\} a^2 + 2E\{(x - \eta_x)(y - \eta_y)\} a + E\{(y - \eta_y)^2\} \]
\[ = \sigma_x^2 a^2 + 2C_{xy} a + \sigma_y^2 \]

This is a positive quadratic function of \( a \)

⇒ Roots are imaginary and discriminant is non-positive

\[ \sqrt{4C_{xy}^2 - 4\sigma_x^2 \sigma_y^2} \rightarrow \text{imaginary} \]
\[ \Rightarrow 4C_{xy}^2 - 4\sigma_x^2 \sigma_y^2 \leq 0 \]
\[ \Rightarrow C_{xy}^2 \leq \sigma_x^2 \sigma_y^2 \]
Thus,

\[ |C_{xy}| \leq \sigma_x \sigma_y \quad \text{and} \quad |r| = \frac{|C_{xy}|}{\sigma_x \sigma_y} \leq 1 \]

Definition (Uncorrelated)

Two RVs are uncorrelated if their covariance is zero

\[ C_{xy} = 0 \]

\[ \Rightarrow r = \frac{C_{xy}}{\sigma_x \sigma_y} = 0 \]

\[ = \frac{E\{xy\} - E\{x\}E\{y\}}{\sigma_x \sigma_y} = 0 \]

\[ \Rightarrow E\{xy\} = E\{x\}E\{y\} \]

Thus

\[ C_{xy} = 0 \iff E\{xy\} = E\{x\}E\{y\} \]
Result

If $x$ and $y$ are independent, then

$$E\{xy\} = E\{x\}E\{y\}$$

and $x$ and $y$ are uncorrelated

Note: Converse is not true (in general)
- Converse only holds for Gaussian RVs
- Independence is a stronger condition than uncorrelated

Definition (Orthogonality)

Two RVs are orthogonal if

$$E\{xy\} = 0$$

Note: If $x$ and $y$ are correlated, they are not orthogonal
Example

Consider the correlation between two RVs, $x$ and $y$, with samples shown in a scatter plot.
Sequences and Vectors of Random Variables

Definition (Vector Distribution)

Let \( \{x\} \) be a sequence of RVs. Take \( N \) samples to form the random vector

\[
x = [x_1, x_2, \ldots, x_N]^T
\]

Then the vector distribution function is

\[
F_x(x^0) = \Pr\{x_1 \leq x_1^0, x_2 \leq x_2^0, \ldots, x_N \leq x_N^0\}
\]

\[
\triangleq \Pr\{x \leq x^0\}
\]

Special Case: For complex data

\[
x = x_r + jx_i
\]
The distribution in the complex case is defined as

\[ F_{\mathbf{x}}(\mathbf{x}^0) = \Pr\{ \mathbf{x}_r \leq \mathbf{x}_r^0, \mathbf{x}_i \leq \mathbf{x}_i^0 \} \]
\[ \triangleq \Pr\{ \mathbf{x} \leq \mathbf{x}^0 \} \]

The density function is given by

\[ f_{\mathbf{x}}(\mathbf{x}) = \frac{\partial^N F_{\mathbf{x}}(\mathbf{x})}{\partial x_1 \partial x_2 \ldots \partial x_N} \]

\[ F_{\mathbf{x}}(\mathbf{x}^0) = \int_{-\infty}^{\mathbf{x}_1^0} \int_{-\infty}^{\mathbf{x}_2^0} \ldots \int_{-\infty}^{\mathbf{x}_N^0} f_{\mathbf{x}}(\mathbf{x}) \, dx_1 \, dx_2 \ldots \, dx_N \]
Properties:

\[
F_{\mathbf{x}}([\infty, \infty, \cdots, \infty]^T) = 1 \\
\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x} = 1 \\
F_{\mathbf{x}}([x_1, x_2, \cdots, -\infty, \cdots, x_N]^T) = 0
\]

Also

\[
F([\infty, x_2, x_3, \cdots, x_N]^T) = F([x_2, x_3, \cdots, x_N]^T) \\
\int_{-\infty}^{\infty} f([x_1, x_2, x_3, \cdots, x_N]^T) dx_1 = f([x_2, x_3, \cdots, x_N]^T)
\]

- Setting \( x_i = \infty \) in the cdf eliminates this sample
- Integrating over \((-\infty, \infty)\) along \( x_i \) in the pdf eliminates this sample
Joint Distribution

Definitions (Joint Distribution and Density)

Given two random vectors \( \mathbf{x} \) and \( \mathbf{y} \), the joint distribution and density are

\[
F_{xy}(\mathbf{x}^0, \mathbf{y}^0) = \Pr\{ \mathbf{x} \leq \mathbf{x}^0, \mathbf{y} \leq \mathbf{y}^0 \}
\]

\[
f_{xy}(\mathbf{x}, \mathbf{y}) = \frac{\partial^N \partial^M F_{xy}(\mathbf{x}, \mathbf{y})}{\partial x_1 \partial x_2 \cdots \partial x_N \partial y_1 \partial y_2 \cdots \partial y_M}
\]

Definition (Vector Independence)

The vectors are independent iff

\[
F_{xy}(\mathbf{x}, \mathbf{y}) = F_x(\mathbf{x})F_y(\mathbf{y})
\]

or equivalently

\[
f_{xy}(\mathbf{x}, \mathbf{y}) = f_x(\mathbf{x})f_y(\mathbf{y})
\]
Expectations & Moments

**Objective:** Obtain partial description of process generating $\mathbf{x}$

**Solution:** Use moments

The first moment, or mean, is

$$m_x = E\{\mathbf{x}\} = \left[ m_1, m_2, \ldots, m_N \right]^T$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \mathbf{x} f_x(\mathbf{x}) d\mathbf{x}$$

$$\Rightarrow m_k = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} x_k f_x(\mathbf{x}) dx_1 dx_2 \ldots dx_N$$

$$= \int_{-\infty}^{\infty} x_k f_{x_k}(x_k) dx_k$$

$\uparrow$ marginal distribution of $x_k$
**Definition (Correlation Matrix)**

A complete set of second moments is given by the correlation matrix

\[
R_x = E\{xx^H\} = E\{xx^*\}^T
\]

\[
= \begin{bmatrix}
E\{|x_1|^2\} & E\{x_1x_2^*\} & \cdots & E\{x_1x_N^*\} \\
E\{x_2x_1^*\} & E\{|x_2|^2\} & \cdots & E\{x_2x_N^*\} \\
\vdots & \vdots & \ddots & \vdots \\
E\{x_Nx_1^*\} & E\{x_Nx_2^*\} & \cdots & E\{|x_N|^2\}
\end{bmatrix}
\]

**Result**

The correlation matrix is Hermitian symmetric

\[
(R_x)^H = (E\{xx^H\})^H
\]

\[
= E\{(xx^H)^H\}
\]

\[
= E\{xx^H\} = R_x
\]
Definition (Covariance Matrix)

The set of second central moments is given by the covariance

\[
C_x = E\{(x - m_x)(x - m_x)^H\} = E\{xx^H\} - m_x E\{x^H\} - E\{x\}m_x^H + m_x m_x^H = R_x - m_x m_x^H
\]

Result

The covariance is Hermitian symmetric

\[
C_x = C_x^H
\]
Result

The correlation and covariance matrices are positive semi-definite

\[ a^H R_x a \geq 0 \quad a^H C_x a \geq 0 \quad (\forall a) \]

To prove this, note

\[
\begin{align*}
a^H R_x a &= a^H E \{ x x^H \} a \\
&= E \{ a^H x x^H a \} \\
&= E \{ (a^H x)(a^H x)^H \} \\
&= E \{ |a^H x|^2 \} \geq 0
\end{align*}
\]

For most cases, \( R \) and \( C \) are positive definite

\[ a^H R_x a > 0 \quad a^H C_x a > 0 \]

\[ \Rightarrow \text{ no linear dependencies in } R_x \text{ or } C_x \]
Definitions (Cross-Correlation and Cross-Covariance)

For random vectors $\mathbf{x}$ and $\mathbf{y}$,

- **Cross-correlation** $\triangleq R_{xy} = E\{\mathbf{x}\mathbf{y}^H\}$
- **Cross-covariance** $\triangleq C_{xy} = E\{ (\mathbf{x} - \mathbf{m}_x)(\mathbf{y} - \mathbf{m}_y)^H \} = R_{xy} - \mathbf{m}_x\mathbf{m}_y^H$

Definition (Uncorrelated Vectors)

Two vectors $\mathbf{x}$ and $\mathbf{y}$ are uncorrelated if

$$C_{xy} = R_{xy} - \mathbf{m}_x\mathbf{m}_y^H = 0$$

or equivalently

$$R_{xy} = E\{\mathbf{x}\mathbf{y}^H\} = \mathbf{m}_x\mathbf{m}_y^H$$
Note that as in the scalar case

\[
\text{independence } \Rightarrow \text{ uncorrelated}
\]

\[
\text{uncorrelated } \not\Rightarrow \text{ independence}
\]

Also, \( \mathbf{x} \) and \( \mathbf{y} \) are orthogonal if

\[
R_{xy} = E\{\mathbf{x}\mathbf{y}^H\} = 0
\]

**Example**

Let \( \mathbf{x} \) and \( \mathbf{y} \) be the same dimension. If

\[
\mathbf{z} = \mathbf{x} + \mathbf{y}
\]

find \( R_z \) and \( C_z \)
By definition

\[ R_z = E\{(x + y)(x + y)^H\} \]
\[ = E\{xx^H\} + E\{xy^H\} + E\{yx^H\} + E\{yy^H\} \]
\[ = R_x + R_{xy} + R_{yx} + R_y \]

Similarly

\[ C_z = C_x + C_{xy} + C_{yx} + C_y \]

**Note:** If \( x \) and \( y \) are uncorrelated,

\[ R_z = R_x + m_x m_y^H + m_y m_x^H + R_y \]

and

\[ C_z = C_x + C_y \]
**Definition (Multivariate Gaussian Density)**

For a $N$ dimensional random vector $\mathbf{x}$ with covariance $\mathbf{C}_x$, the multivariate Gaussian pdf is

$$f_{\mathbf{x}}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{N}{2}} |\mathbf{C}_x|^{\frac{1}{2}}} e^{-\frac{1}{2} (\mathbf{x} - \mathbf{m}_x)^H \mathbf{C}_x^{-1} (\mathbf{x} - \mathbf{m}_x)}$$

Note the similarity to the univariate case

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

**Example**

Let $N = 2$ (bivariate case) and $\mathbf{x}$ be real. Then

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \text{and} \quad \mathbf{m}_x = E\{\mathbf{x}\} = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix}$$
\[ \mathbf{C}_x = E \left\{ (\mathbf{x} - \mathbf{m}_x)(\mathbf{x} - \mathbf{m}_x)^T \right\} \]
\[ = E \{ \mathbf{xx}^T \} - \mathbf{m}_x \mathbf{m}_x^T \]
\[ = E \left\{ \begin{bmatrix} x_1^2 & x_1 x_2 \\ x_2 x_1 & x_2^2 \end{bmatrix} \right\} - \begin{bmatrix} m_1^2 & m_1 m_2 \\ m_2 m_1 & m_2^2 \end{bmatrix} \]
\[ = \begin{bmatrix} E \{ x_1^2 \} - m_1^2 & E \{ x_1 x_2 \} - m_1 m_2 \\ E \{ x_2 x_1 \} - m_2 m_1 & E \{ x_2^2 \} - m_2^2 \end{bmatrix} \]

Recall that
\[ \sigma_x^2 = E \{ x^2 \} - E^2 \{ x \} \]
and
\[ r = \frac{E \{ x_1 x_2 \} - m_1 m_2}{\sigma_{x_1} \sigma_{x_2}} \]
Rearranging: \[ C_x = \begin{bmatrix} \sigma^2_{x1} & r\sigma_{x1}\sigma_{x2} \\ r\sigma_{x1}\sigma_{x2} & \sigma^2_{x2} \end{bmatrix} \]

Also,

\[
C_x^{-1} = \frac{1}{\sigma^2_{x1} \sigma^2_{x2} - r^2 \sigma^2_{x1} \sigma^2_{x2}} \begin{bmatrix} \sigma^2_{x2} & -r\sigma_{x1}\sigma_{x2} \\ -r\sigma_{x1}\sigma_{x2} & \sigma^2_{x1} \end{bmatrix}
\]

\[
= \frac{1}{\sigma^2_{x1} \sigma^2_{x2} (1 - r^2)} \begin{bmatrix} \sigma^2_{x2} & -r\sigma_{x1}\sigma_{x2} \\ -r\sigma_{x1}\sigma_{x2} & \sigma^2_{x1} \end{bmatrix}
\]

Substituting into the Gaussian pdf and simplifying

\[
f_x(x) = \frac{1}{2\pi |C_x|^{1/2}} e^{-\frac{1}{2}(x - m_x)^T C_x^{-1} (x - m_x)}
\]

\[
= \frac{1}{2\pi \sigma_{x1} \sigma_{x2} (1 - r^2)^{1/2}} e^{-\frac{1}{2(1 - r^2)} \left[ \frac{(x_1 - m_1)^2}{\sigma^2_{x1}} - 2r \frac{(x_1 - m_1)(x_2 - m_2)}{\sigma_{x1} \sigma_{x2}} + \frac{(x_2 - m_2)^2}{\sigma^2_{x2}} \right]}
\]
Note: If uncorrelated, $r = 0$

$$\Rightarrow f_{\mathbf{x}}(\mathbf{x}) = \frac{1}{2\pi\sigma_{x_1}\sigma_{x_2}} e^{-\frac{1}{2} \left[ \frac{(x_1 - m_1)^2}{\sigma_{x_1}^2} + \frac{(x_2 - m_2)^2}{\sigma_{x_2}^2} \right]}$$

$$= f_{x_1}(x_1) f_{x_2}(x_2)$$

Gaussian special case result:

uncorrelated $\Rightarrow$ independent

Example

Examine the contours defined by

$$(\mathbf{x} - \mathbf{m}_x)^T \mathbf{C}_x^{-1} (\mathbf{x} - \mathbf{m}_x) = \text{constant}$$

Why? For all values on the contour

$$f_{\mathbf{x}}(\mathbf{x}) = \text{constant}$$
\( r = 0 \quad \sigma_{x_1} = \sigma_{x_2} \)

\( r = 0 \quad \sigma_{x_1} > \sigma_{x_2} \)

\( r > 0 \quad \sigma_{x_1} > \sigma_{x_2} \)

\( r > 0 \quad \sigma_{x_1} < \sigma_{x_2} \)
\( r < 0 \) and \( \sigma_{x_1} > \sigma_{x_2} \)

- Integrating over \( x_2 \) yields \( f_{x_1}(x_1) \)
- Integrating over \( x_1 \) yields \( f_{x_2}(x_2) \)
Additional Gaussian (surface) examples:

\[ r = 0 \quad \sigma_{x_1} = \sigma_{x_2} \]

\[ r = 0 \quad \sigma_{x_1} < \sigma_{x_2} \]

\[ r > 0 \quad \sigma_{x_1} < \sigma_{x_2} \]

\[ r < 0 \quad \sigma_{x_1} < \sigma_{x_2} \]
Transformations of a vector

Let the $N$ functions $g_1(\cdot), g_2(\cdot), \ldots, g_N(\cdot)$ map $x$ to $z$, where

\[
egin{align*}
  z_1 &= g_1(x_1, x_2, \ldots, x_N) \\
  z_2 &= g_2(x) \\
  &\vdots \\
  z_N &= g_N(x)
\end{align*}
\]

Forward mapping

Let $g_1(\cdot), g_2(\cdot), \ldots, g_N(\cdot)$ be independent and yield a one-to-one transformation such that $\exists$ a set of functions

\[
\begin{align*}
  x_1 &= h_1(z), x_2 = h_2(z), \ldots, x_N = h_N(z)
\end{align*}
\]

where $z = [z_1, z_2, \ldots, z_N]^T$.

Reverse mapping

**Question:** How do we determine the distribution $f_z(z)$?
Let $N = 2$ and consider the probability of being in the region defined by

$$[z_1, z_1 + dz_1] \quad \text{and} \quad [z_2, z_2 + dz_2]$$

Identify an equivalent area in the $x_1, x_2$ domain and equate the probabilities

$$\Pr\{(z_1, z_2) \in A_z\} = \Pr\{(x_1, x_2) \in A_x\}$$

$$f_{z_1 z_2}(z_1, z_2) \text{Area}(A_z) = f_{x_1 x_2}(x_1, x_2) \text{Area}(A_x)$$
\[
\frac{\text{Area}(A_x)}{\text{Area}(A_z)} = \text{abs} \left( J \left( \begin{array}{cc} x_1 & x_2 \\ z_1 & z_2 \end{array} \right) \right) \frac{1}{\text{abs} \left( J \left( \begin{array}{cc} z_1 & z_2 \\ x_1 & x_2 \end{array} \right) \right)}
\]

The Jacobian is defined as

\[
J \left( \begin{array}{cc} x_1 & x_2 \\ z_1 & z_2 \end{array} \right) = \begin{vmatrix} \frac{\partial x_1}{\partial z_1} & \frac{\partial x_1}{\partial z_2} \\ \frac{\partial x_2}{\partial z_1} & \frac{\partial x_2}{\partial z_2} \end{vmatrix}
\]

and

\[
J \left( \begin{array}{cc} z_1 & z_2 \\ x_1 & x_2 \end{array} \right) = \begin{vmatrix} \frac{\partial z_1}{\partial x_1} & \frac{\partial z_1}{\partial x_2} \\ \frac{\partial z_2}{\partial x_1} & \frac{\partial z_2}{\partial x_2} \end{vmatrix}
\]

Note that

\[
\frac{\partial x_1}{\partial z_1} = \frac{\partial h_1(z)}{\partial z_1} \quad \text{and} \quad \frac{\partial z_1}{\partial x_1} = \frac{\partial g_1(x)}{\partial x_1}
\]
Thus

\[ f_{z_1z_2}(z_1, z_2) \text{Area}(A_z) = f_{x_1x_2}(x_1, x_2) \text{Area}(A_x) \]

\[ \Rightarrow f_{z_1z_2}(z_1, z_2) = \frac{f_{x_1x_2}(x_1, x_2)}{\text{Area}(A_z)/\text{Area}(A_x)} = \frac{f_{x_1x_2}(x_1, x_2)}{\text{abs}(J(\begin{bmatrix} z_1 & z_2 \\ x_1 & x_2 \end{bmatrix}))} \]

**General Case Result (Functions of Vectors)**

\[ f_z(z) = \frac{f_x(x)}{\text{abs}(J(\begin{bmatrix} z \\ x \end{bmatrix}))} \]

where

\[ J(\begin{bmatrix} z \\ x \end{bmatrix}) = \begin{vmatrix} \frac{\partial z_1}{\partial x_1} & \cdots & \frac{\partial z_1}{\partial x_N} \\ \vdots & \ddots & \vdots \\ \frac{\partial z_N}{\partial x_1} & \cdots & \frac{\partial z_N}{\partial x_N} \end{vmatrix} \]
Example (Linear Transformation)

Let \( z \) and \( x \) be linearly related

\[
\begin{align*}
z_1 &= a_{11} x_1 + a_{12} x_2 \\
z_2 &= a_{21} x_1 + a_{22} x_2
\end{align*}
\]

or \( z = Ax \) and \( x = A^{-1} z \)

Then

\[
J \begin{pmatrix} z_1 \\ z_2 \\ x_1 \\ x_2 \end{pmatrix} = \begin{vmatrix}
\frac{\partial z_1}{\partial x_1} & \frac{\partial z_1}{\partial x_2} \\
\frac{\partial z_2}{\partial x_1} & \frac{\partial z_2}{\partial x_2}
\end{vmatrix} = \begin{vmatrix}
a_{11} & a_{12} \\
a_{21} & a_{22}
\end{vmatrix} = |A|
\]
Let \( A^{-1} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \). Then

\[
\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}
\]

and

\[
f_{z_1 z_2}(z_1, z_2) = \frac{f_{x_1 x_2}(x_1, x_2)}{|A|}
= \frac{f_{x_1 x_2}(b_{11} z_1 + b_{12} z_2, b_{21} z_1 + b_{22} z_2)}{|A|}
\]

**General Case Result (Linear Transformations)**

For case where \( z = Ax \)

\[
f_z(z) = \frac{1}{|A|} f_x(A^{-1}z)
\]
Vector Statistics for Linear Transformations

For such linear transformations $\mathbf{z} = \mathbf{Ax}$

$$E\{\mathbf{z}\} = E\{\mathbf{Ax}\} = \mathbf{Am}_x$$

Similarly

$$E\{\mathbf{zz}^H\} = E\{\mathbf{Ax}(\mathbf{Ax})^H\} = E\{\mathbf{Axx}^HA^H\}$$

$$\Rightarrow \mathbf{R}_z = \mathbf{AR}_x\mathbf{A}^H$$

By similar arguments it is easy to show

$$\mathbf{C}_z = \mathbf{AC}_x\mathbf{A}^H$$

Note: Results in simple linear transformations of statistics