# Review of Probability

# Random Variable

- **Definition**

  Numerical characterization of outcome of a random event

- **Examples**

  1) Number on rolled dice
  2) Temperature at specified time of day
  3) Stock Market at close
  4) Height of wheel going over a rocky road

# Random Variable

- **Non-examples**

  1) ‘Heads’ or ‘Tails’ on coin

  2) Red or Black ball from urn

**But we can make these into RV’s**

- **Basic Idea** – don’t know how to completely determine what value will occur
  - Can only specify probabilities of RV values occurring.

# Two Types of Random Variables

**Random Variable**

**Discrete RV**
- Die
- Stocks

**Continuous RV**
- Temperature
- Wheel height

# PDF for Continuous RV

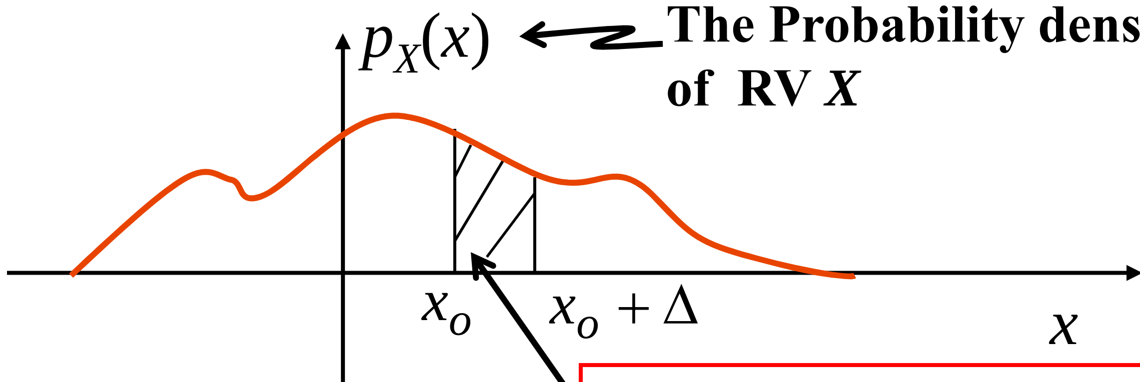
Given Continuous RV $X$…

What is the probability that $X = x_0$ ?

- Oddity : $P(X = x_0) = 0$

  Otherwise the Prob. "Sums" to infinity

- Need to think of <u>Prob. *Density* Function</u> (PDF)




**The Probability density function of RV *X***

$$\boxed{\begin{aligned} P(x_0 < X < x_0 + \Delta) &= \text{area shown} \\ &= \int_{x_o}^{x_o+\Delta} p_X(x)\,dx \end{aligned}}$$

# Most Commonly Used PDF: Gaussian

**A RV *X* with the following PDF is called a Gaussian RV**

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

$m$ & $\sigma$ are parameters of the Gaussian pdf

$\boldsymbol{m}$ **= Mean of RV *X***

$\boldsymbol{\sigma}$ **= Standard Deviation of RV *X* (Note: $\boldsymbol{\sigma > 0}$)**

$\boldsymbol{\sigma^2}$ **= Variance of RV *X***

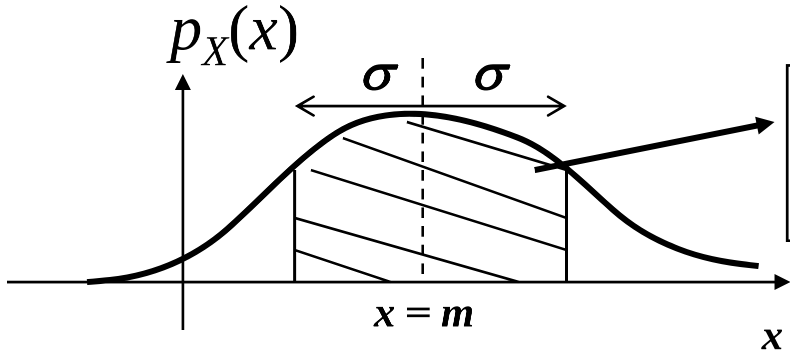
**Notation**: When $X$ has Gaussian PDF we say $X \sim N(m, \sigma^2)$

# Zero-Mean Gaussian PDF

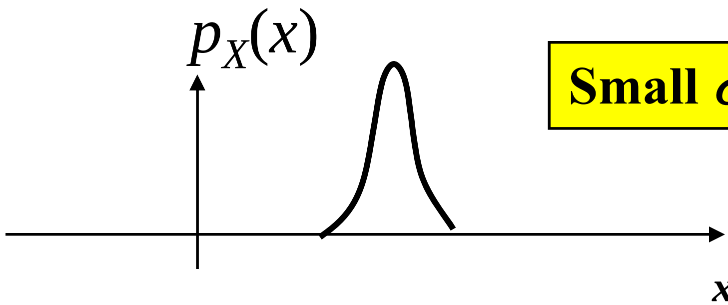
- **Generally: take the noise to be Zero Mean**

$$p_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{x^2 / 2\sigma^2}$$

# Effect of Variance on Gaussian PDF

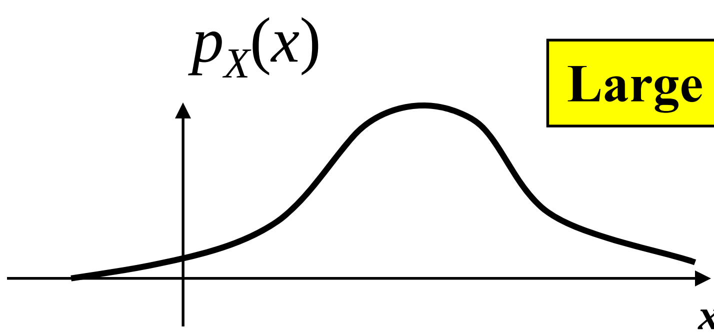



Area within ±1 $\sigma$ of mean = 0.683
= 68.3%




**Small $\sigma$**

Small Variability
(Small Uncertainty)




**Large $\sigma$**

Large Variability
(Large Uncertainty)

# Why Is Gaussian Used?

- **Central Limit theorem (CLT)**

  **The sum of $N$ independent RVs has a pdf that tends to be Gaussian as $N \to \infty$**

- **So What! Here is what : Electronic systems generate internal noise due to random motion of electrons in electronic components. The noise is the result of summing the random effects of lots of electrons.**

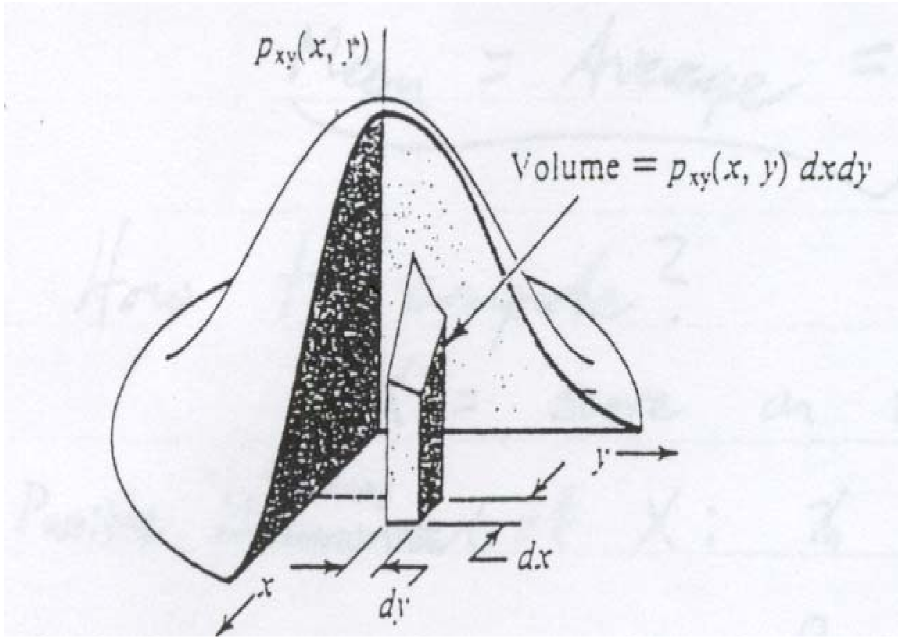
$\Rightarrow$ **CLT applies** $\Rightarrow$ **Guassian Noise**

# <u>Joint PDF of RVs *X* and *Y*</u>

$p_{XY}(x, y)$

Describes probabilities of joint events concerning *X* and *Y*. For example, the probability that *X* lies in interval [*a*,*b*] <u>and</u> *Y* lies in interval [*a*,*b*] is given by:

$$\boxed{\Pr\{(a < X < b) \text{ and } (c < Y < d)\} = \int_a^b \int_c^d p_{XY}(x, y)\,dx\,dy}$$




**This graph shows a <u>Joint PDF</u>**

Graph from B. P. Lathi's book: *Modern Digital & Analog Communication Systems*

# Conditional PDF of Two RVs

When you have two RVs… often ask: What is the PDF of $Y$ if $X$ is constrained to take on a specific value.

In other words: What is the PDF of $Y$ conditioned on the fact $X$ is constrained to take on a specific value.

**Ex.**: Husband's salary $X$ conditioned on wife's salary = \$100K?

First find all wives who make EXACTLY \$100K… how are their husband's salaries distributed.

Depends on the joint PDF because there are two RVs… but it should only depend on the slice of the joint PDF at $Y$=\$100K.

Now… we have to adjust this to account for the fact that the joint PDF (even its slice) reflects how likely it is that $X$=\$100K will occur (e.g., if $X=10^5$ is unlikely then $p_{XY}(10^5,y)$ will be small); so… if we divide by $p_X(10^5)$ we adjust for this.

## Conditional PDF (cont.)

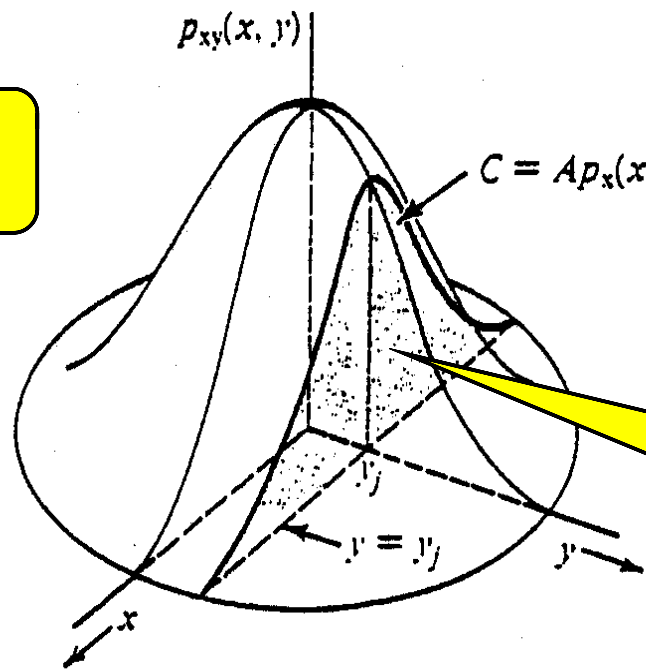
Thus, the conditional PDFs are defined as ("slice and normalize"):

$$p_{Y|X}(y\,|\,x)=\begin{cases}\dfrac{p_{XY}(x,y)}{p_X(x)}, & p_X(x)\neq 0\\ 0, & \text{otherwise}\end{cases}$$

$$p_{X|Y}(x\,|\,y)=\begin{cases}\dfrac{p_{XY}(x,y)}{p_Y(y)}, & p_Y(y)\neq 0\\ 0, & \text{otherwise}\end{cases}$$

*x* is held fixed

*y* is held fixed




*"slice and normalize"*
*y* is held fixed

**This graph shows a <u>Conditional PDF</u>**

Graph from B. P. Lathi's book: *Modern Digital & Analog Communication Systems*

# <u>Independent RV's</u>

Independence should be thought of as saying that:

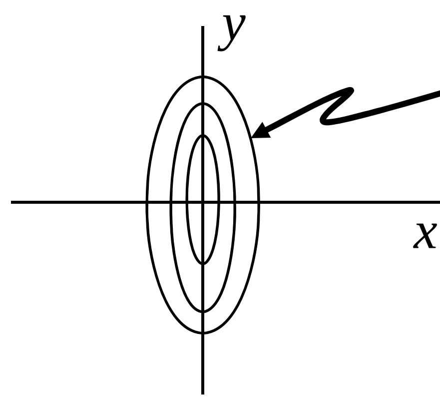
Neither RV impacts the other statistically – thus, the values that one <u>*will* likely take</u> should be irrelevant to the value that the other <u>*has* taken</u>.

In other words: conditioning doesn't change the PDF!!!

$$\boxed{\begin{aligned} p_{Y|X=x}(y\,|\,x) &= \frac{p_{XY}(x,y)}{p_X(x)} = p_Y(y) \\[2ex] p_{X|Y=y}(x\,|\,y) &= \frac{p_{XY}(x,y)}{p_Y(y)} = p_X(x) \end{aligned}}$$

# Independent and Dependent Gaussian PDFs

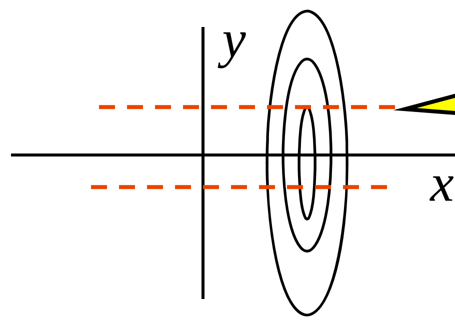
**Independent**
**(zero mean)**




Contours of $p_{XY}(x,y)$.

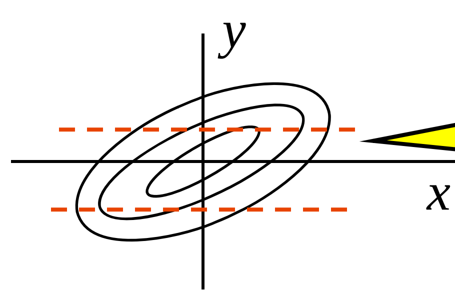
**If $X$ & $Y$ are independent, then the contour ellipses are aligned with either the $x$ or $y$ axis**

**Independent**
**(non-zero mean)**




Different slices give same normalized curves

**Dependent**




Different slices give different normalized curves

# An "Independent RV" Result

RV's $X$ & $Y$ are independent if:

$$\boxed{p_{XY}(x,y) = p_X(x)\,p_Y(y)}$$

Here's why:

$$\boxed{p_{Y|X=x}(y\,|\,x) = \frac{p_{XY}(x,y)}{p_X(x)} = \frac{\cancel{p_X(x)}\,p_Y(y)}{\cancel{p_X(x)}} = p_Y(y)}$$

# Characterizing RVs

- PDF tells everything about an RV
  - but sometimes they are "more than we need/know"
- So… we make due with a few Characteristics
  - Mean of an RV (Describes the centroid of PDF)
  - Variance of an RV (Describes the spread of PDF)
  - Correlation of RVs (Describes "tilt" of joint PDF)

$$\underbrace{\textbf{Mean} = \textbf{Average} = \textbf{Expected Value}}_{\text{Symbolically: } E\{X\}}$$

# Motivating Idea of Mean of RV

Motivation First w/ "Data Analysis View"

Consider RV $X$ = Score on a test     Data: $x_1, x_2, \ldots x_N$

Possible values of RV $X$ : $V_0 \; V_1 \; V_2 \ldots \; V_{100}$

$0 \quad 1 \quad 2 \; \ldots \; 100$

$$\text{Test Average} = \bar{x} = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{N_0 V_0 + N_1 V_1 + \ldots N_n V_{100}}{N} = \sum_{i=0}^{100} V_i \frac{N_i}{N}$$

$\approx P(X = V_i)$

$N_i$ = # of scores of value $V_i$

$N = \sum_{i=1}^{n} N_i$ (Total # of scores)

This is called Data Analysis View

But it motivates the Data Modeling View

Statistics

Probability

# Theoretical View of Mean

Data Analysis View leads to Probability Theory:

Data Modeling

- For Discrete random Variables :

$$E\{X\} = \sum_{n=1}^{n} x_i P_X(x_i)$$

Probability Function

- This Motivates form for Continuous RV:

$$E\{X\} = \int_{-\infty}^{\infty} x \, p_X(x) dx$$

Probability Density Function

$$\boxed{\text{Notation: } E\{X\} = \overline{X}}$$

Shorthand Notation

# Aside: Probability vs. Statistics

## Probability Theory

» Given a PDF Model

» Describe how the data will likely behave

$$E\{X\} = \int_{-\infty}^{\infty} x \; p_X(x)dx$$

Dummy Variable

PDF

## Statistics

» Given a set of Data

» Determine how the data did behave

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

Data

$\approx$

"Law of Large Numbers"

**There is no DATA here!!!**
The PDF models how the data will likely behave

**There is no PDF here!!!**
The Statistic measures how the data did behave

# Variance of RV

**There are similar Data vs. Theory Views here…**

**But let's go right to the theory!!**

**Variance:** Characterizes how much you expect the RV to Deviate Around the Mean

Variance: $$\sigma^2 = E\{(X - m_x)^2\}$$

$$= \int (x - m_x)^2 p_X(x)dx$$

Note : If zero mean…

$$\sigma^2 = E\{X^2\}$$

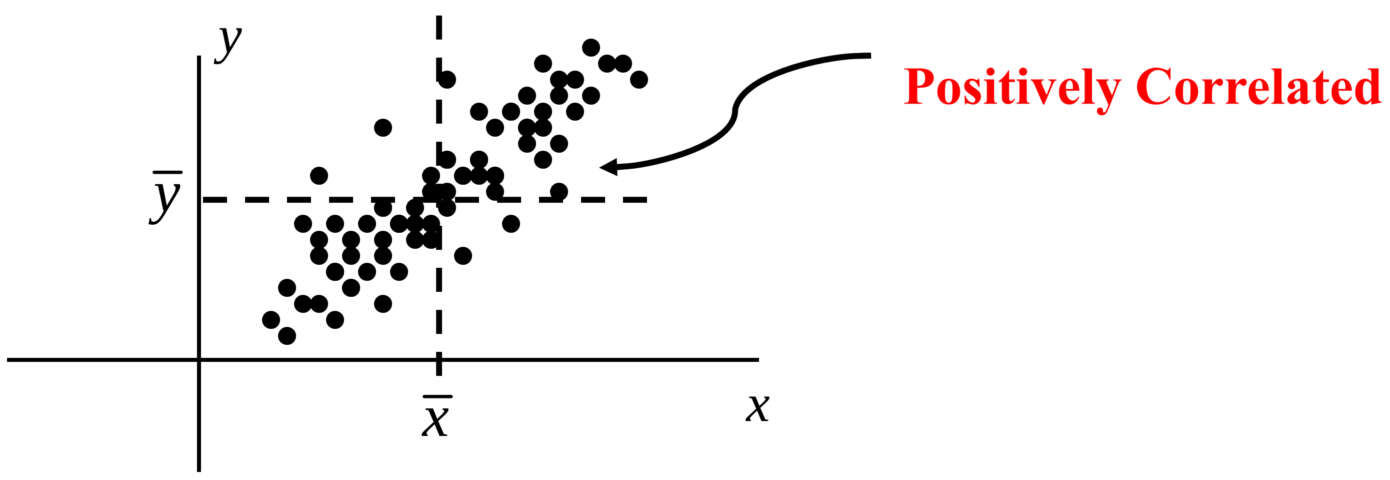
$$= \int x^2 p_X(x)dx$$

# Motivating Idea of Correlation

Motivate First w/ Data Analysis View

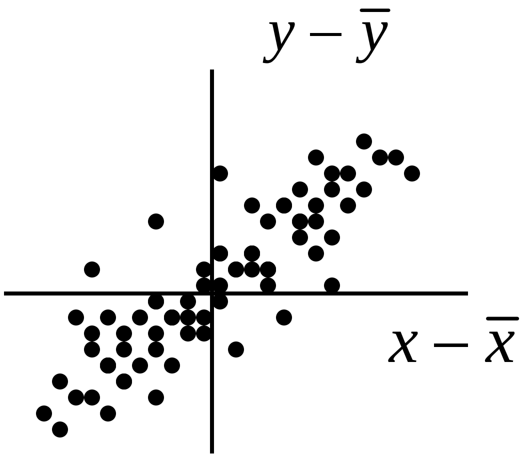
Consider a random experiment that observes the outcomes of two RVs:

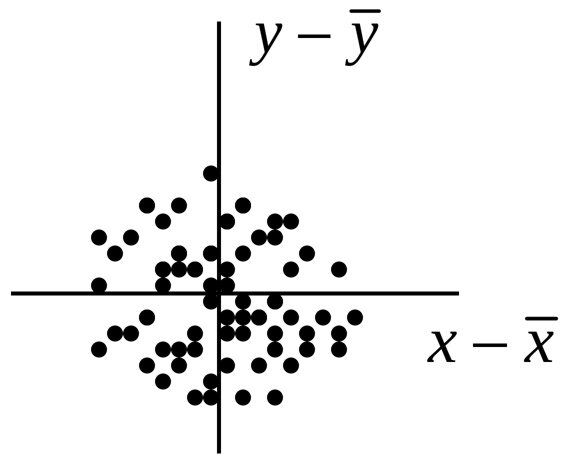
Example: 2 RVs $X$ and $Y$ representing height and weight, respectively

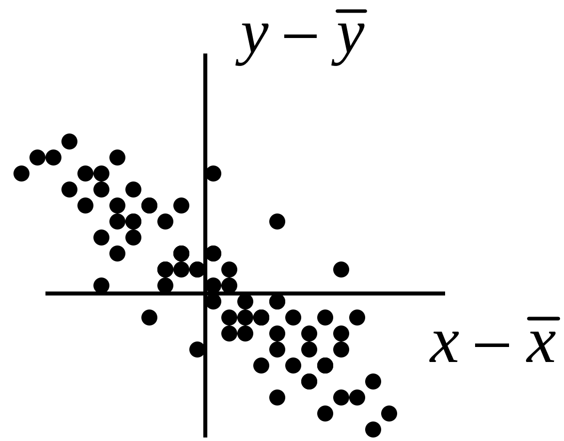


# Illustrating 3 Main Types of Correlation

Data Analysis View: $C_{xy} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})$




**Positive Correlation**
**"Best Friends"**

**Zero Correlation**
**i.e. uncorrelated**
**"Complete Strangers"**

**Negative Correlation**
**"Worst Enemies"**

**GPA**
**&**
**Starting Salary**

**Height**
**&**
**$ in Pocket**

**Student Loans**
**&**
**Parents' Salary**

# Prob. Theory View of Correlation

To capture this, define Covariance :

$$\boxed{\sigma_{XY} = E\{(X - \overline{X})(Y - \overline{Y})\}}$$

$$\boxed{\sigma_{XY} = \int\int (x - \overline{X})(y - \overline{Y})\, p_{XY}(x, y)\, dx\, dy}$$

If the RVs are both Zero-mean : $\boxed{\sigma_{XY} = \mathrm{E}\{XY\}}$

If $X = Y$: $\boxed{\sigma_{XY} = \sigma_X^2 = \sigma_Y^2}$

If $X$ & $Y$ are independent, then: $\boxed{\sigma_{XY} = 0}$

If $\sigma_{XY} = E\{(X - \overline{X})(Y - \overline{Y})\} = 0$

Then… Say that *X* and *Y* are **<u>"uncorrelated"</u>**

If $\sigma_{XY} = E\{(X - \overline{X})(Y - \overline{Y})\} = 0$

Then $\underbrace{E\{XY\}} = \overline{X}\,\overline{Y}$

**Called <u>"Correlation of *X* & *Y*"</u>**

**So… RVs *X* and *Y* are said to be uncorrelated**

**if $\sigma_{XY} = 0$**

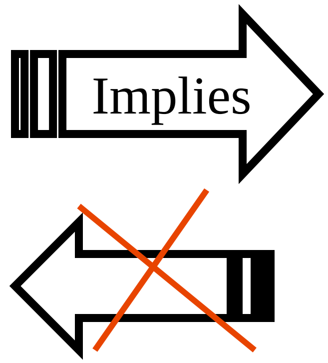
**or equivalently… if $E\{XY\} = E\{X\}E\{Y\}$**

# Independence vs. Uncorrelated

$X$ & $Y$ are Independent

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

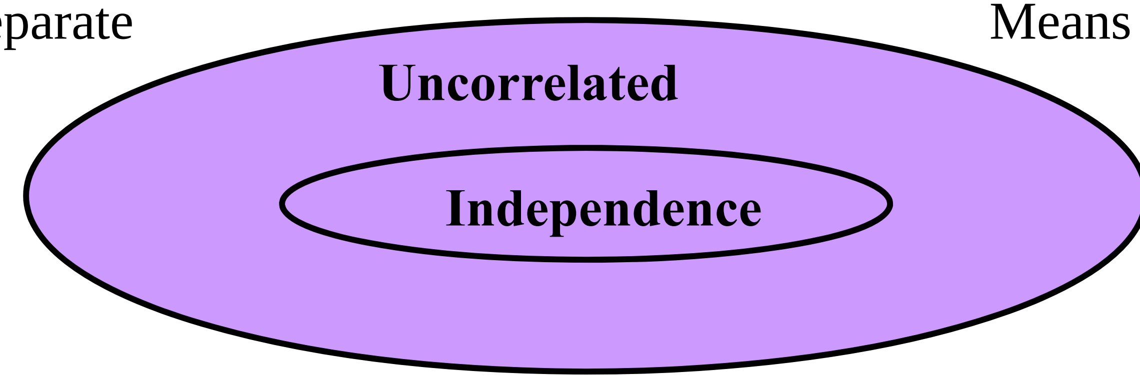
PDFs Separate




$X$ & $Y$ are Uncorrelated

$$E\{XY\} = E\{X\}E\{Y\}$$

Means Separate




**INDEPENDENCE IS A STRONGER CONDITION !!!!**

# Confusing Covariance and Correlation Terminology

Covariance : $\boxed{\sigma_{XY} = E\{(X - \overline{X})(Y - \overline{Y})\}}$

Correlation : $\boxed{E\{XY\}}$

Same if zero mean

Correlation Coefficient : $\boxed{\rho_{XY} = \dfrac{\sigma_{XY}}{\sigma_X \sigma_Y}}$

$$\boxed{-1 \le \rho_{XY} \le 1}$$

# Covariance and Correlation For Random Vectors…

$$\mathbf{x} = [\, X_1 \; X_1 \; \cdots \; X_N \,]^T$$

Correlation Matrix :

$$\boxed{\mathbf{R_x} = E\{\mathbf{x}\mathbf{x}^T\} = \begin{bmatrix} E\{X_1X_1\} & E\{X_1X_2\} & \cdots & E\{X_1X_N\} \\ E\{X_2X_1\} & E\{X_2X_2\} & \cdots & E\{X_2X_N\} \\ \vdots & \vdots & \ddots & \vdots \\ E\{X_NX_1\} & E\{X_NX_2\} & \cdots & E\{X_NX_N\} \end{bmatrix}}$$

Covariance Matrix :

$$\boxed{\mathbf{C_x} = E\{(\mathbf{x} - \overline{\mathbf{x}})(\mathbf{x} - \overline{\mathbf{x}})^T\}}$$

# A Few Properties of Expected Value

$$E\{X+Y\} = E\{X\} + E\{Y\}$$

$$E\{aX\} = aE\{X\}$$

$$E\{f(X)\} = \int f(x) p_X(x) dx$$

$$\mathrm{var}\{X+Y\} = \begin{cases} \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY} \\ \\ \sigma_X^2 + \sigma_Y^2, \quad \text{if } X \,\&\, Y \text{ are uncorrelated} \end{cases}$$

$$\mathrm{var}\{aX\} = a^2 \sigma_X^2$$

$$\mathrm{var}\{a + X\} = \sigma_X^2$$

$$\begin{aligned}
\mathrm{var}\{X+Y\} &= E\left\{\left(X + Y - \overline{X} - \overline{Y}\right)^2\right\} \\
&= E\left\{\left(X_z + Y_z\right)^2\right\} \quad \text{where } X_z = X - \overline{X} \\
&= E\left\{\left(X_z\right)^2 + \left(Y_z\right)^2 + 2X_z Y_z\right\} \\
&= E\left\{\left(X_z\right)^2\right\} + E\left\{\left(Y_z\right)^2\right\} + 2E\{X_z Y_z\} \\
&= \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY}
\end{aligned}$$

# Joint PDF for Gaussian

Let $\mathbf{x} = [X_1\ X_2\ \ldots\ X_N]^T$ be a vector of random variables. These random variables are said to be jointly Gaussian if they have the following PDF

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{N}{2}}\sqrt{\det(\mathbf{C}_x)}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_x)^T \mathbf{C}_x^{-1}(\mathbf{x}-\boldsymbol{\mu}_x)\right\}$$

where $\boldsymbol{\mu}_x$ is the mean vector and $\mathbf{C}_x$ is the covariance matrix:

$$\boldsymbol{\mu}_x = E\{\mathbf{x}\} \qquad \mathbf{C}_x = E\{(\mathbf{x}-\boldsymbol{\mu}_x)(\mathbf{x}-\boldsymbol{\mu}_x)^T\}$$

For the case of two jointly Gaussian RVs $X_1$ and $X_2$ with

$E\{X_i\} = \mu_i \qquad \mathrm{var}\{X_i\} = \sigma_i^2 \qquad E\{(X_1-\mu_1)\,(X_2-\mu_2)\} = \sigma_{12} \qquad \rho = \sigma_{12}/\,(\sigma_1\,\sigma_2)$

Then…

$$p(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]\right\}$$

It is easy to verify that $X_1$ and $X_2$ are uncorrelated (and independent!) if $\rho = 0$

# Linear Transform of Jointly Gaussian RVs

Let $\mathbf{x} = [X_1\ X_2\ \ldots\ X_N]^T$ be a vector of jointly Gaussian random variables with mean vector $\boldsymbol{\mu}_x$ and covariance matrix $\mathbf{C}_x$…

Then the linear transform $\mathbf{y} = \mathbf{Ax} + \mathbf{b}$ is also jointly Gaussian with

$$\boldsymbol{\mu}_y = E\{\mathbf{y}\} = \mathbf{A}\boldsymbol{\mu}_x + \mathbf{b}$$

$$\mathbf{C}_y = E\{(\mathbf{y} - \boldsymbol{\mu}_y)(\mathbf{y} - \boldsymbol{\mu}_y)^T\} = \mathbf{AC}_x\mathbf{A}^T$$

A special case of this is the **sum of jointly Gaussian RVs**… which can be handled using $\mathbf{A} = [1\ \ 1\ \ 1\ \ \ldots\ \ 1]$

# Moments of Gaussian RVs

Let $X$ be zero mean Gaussian with variance $\sigma^2$

Then the moments $E\{X^k\}$ are as follows:

$$E\{X^k\} = \begin{cases} 1 \cdot 3 \cdots (k-1)\sigma^k, & k \text{ even} \\ 0, & k \text{ odd} \end{cases}$$

Let $X_1$ $X_2$ $X_3$ $X_4$ be any four jointly Gaussian random variables with zero mean

Then…

$$E\{X_1X_2X_3X_4\} = E\{X_1X_2\}\mathrm{E}\{X_3X_4\} + E\{X_1X_3\}\mathrm{E}\{X_2X_4\} + E\{X_1X_4\}\mathrm{E}\{X_2X_3\}$$

Note that this can be applied to find $E\{X^2Y^2\}$ if $X$ and $Y$ are jointly Gaussian

# Chi-Squared Distribution

Let $X_1$ $X_2$ … $X_N$ be a set of zero-mean independent jointly Gaussian random variables each with unit variance.

Then the RV $Y = X_1^2 + X_2^2 + \ldots + X_N^2$ is called a chi-squared ($\chi^2$) RV of $N$ degrees of freedom and has PDF given by

$$p(y) = \begin{cases} \dfrac{1}{2^{N/2}\Gamma(N/2)} y^{(N/2)-1} e^{-y/2}, & y \geq 0 \\ 0, & y < 0 \end{cases}$$

For this RV we have that:

$$E\{Y\} = N \quad \text{and} \quad \mathrm{var}\{Y\} = 2N$$