Homework 1 posted, due February 9
We will often find it convenient to represent collections of random variables in terms of random vectors.

\[ \mathbf{X} = \left( \frac{X_1}{X_2}, \ldots, \frac{X_n}{X} \right) \quad \overrightarrow{X} \in \mathbb{R}^n \]

Joint probability distribution:

\[ P_{\mathbf{X}}(B) = \operatorname{Prob}(\overrightarrow{X} \in B) \]

where \( B \subseteq \mathbb{B} \) and \( \mathbb{B} \) is a \( \sigma \)-algebra of subsets of \( \mathbb{R}^n \).

\( \sigma \)-field (usually Borel or Lebesgue measurable subsets)

CDF is awkward to generalize to multi r.v.'s

If the probability measure is absolutely continuous with respect to Lebesgue measure, then can express the probability measure in terms of joint probability density.

\[ P_{\mathbf{X}}(\mathbf{x}) = \int_B P_{\mathbf{X}}(\mathbf{x}) \, d\mathbf{X} \]

for \( \mathbf{B} \in \mathbb{B} \)

Expectation (averages or means) work the same way as in one dimension:
Expectation (averages or means) work the same way as in one dimension:

\[ E f(\vec{x}) = \langle f(\vec{x}) \rangle = \int_{\mathbb{R}^n} f(\vec{x}) \, p_{\vec{x}}(\vec{x}) \, d\vec{x} \]

The joint pdf contains information describing how the random variables are related to each other.

Independence means that events or random variables don’t have a statistical connection.

Events: \( A, B \epsilon \mathbb{B} \)

\[ \text{Prob}(A \text{ and } B) = \text{Prob}(A) \, \text{Prob}(B) \]

\[ \iff \]

Events \( A \) and \( B \) are independent

Random variables: \( \vec{X} \) and \( \vec{Y} \)

\[ \iff \]

For all \( \vec{x} \epsilon \mathbb{X} \) and \( \vec{y} \epsilon \mathbb{Y} \),

\[ \text{Prob}(\vec{X} \epsilon A \text{ and } \vec{Y} \epsilon B) = \text{Prob}(\vec{X} \epsilon A) \, \text{Prob}(\vec{Y} \epsilon B) \]

In more technical terms,

\[ p_{\vec{x},\vec{y}}(A \times B) = p_{\vec{x}}(A) \, p_{\vec{y}}(B) \]

If \( \vec{X} \) and \( \vec{Y} \) have a joint probability density function, then independence of
If $\mathbf{X}$ and $\mathbf{Y}$ have a joint probability density function, then independence of these random vectors can be described more simply in terms:

$$p_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y}) = p_{\mathbf{X}}(\mathbf{x}) p_{\mathbf{Y}}(\mathbf{y})$$

Important consequence of independence of $\mathbf{X}$ and $\mathbf{Y}$:

For any nice functions $f, g$:

$$\langle f(\mathbf{x}) g(\mathbf{y}) \rangle = \langle f(\mathbf{x}) \rangle \langle g(\mathbf{y}) \rangle$$

Not true if the random variables $\mathbf{X}, \mathbf{Y}$ are dependent.

If we have dependent random variables, how do we describe the dependence in an easy to understand way? The dependence is implicit in the joint PDF, but it is often more helpful to look at some simple measures of dependence.

Covariance of $\mathbf{X}, \mathbf{Y}$ (one dimensional random variables to begin)

$$\text{Cov}(\mathbf{X}, \mathbf{Y}) = \langle (\mathbf{X} - \mu_\mathbf{X})(\mathbf{Y} - \mu_\mathbf{Y}) \rangle$$

where $\mu_\mathbf{X} = \langle \mathbf{X} \rangle$ and $\mu_\mathbf{Y} = \langle \mathbf{Y} \rangle$

The sign of the covariance is meaningful but the size isn't generally meaningful because it is typically dimensional and dependent on the units we use to measure it. Better to look at nondimensionalized (scaled, normalized) with respect to some
sensible reference quantities. (Barenblatt, Dimensional Analysis, Scaling, Similarity and Self-Similarity).

Correlation coefficient:

$$\rho_{X,Y} = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

Cauchy-Schwarz: 

$$-1 \leq \rho_{X,Y} \leq 1$$

$$\rho_{X,Y} = 0 \iff X, Y \text{ uncorrelated}$$

This is not in general the same as saying the random variables are independent, because correlation or covariance just gives one rough measure of the dependence of the random variables. However, when we are discussing jointly Gaussian random variables then for these, being uncorrelated is equivalent to independence.

If \( \rho_{X,Y} \) is:

- small and positive: weak positive correlation
- Nearly 1: strong positive correlation

Small and negative: weak negative correlation
- Nearly -1: strong negative correlation

If discussing more than 2 random variables at a time, then use covariance
matrix and correlation matrix:

\[ \text{Cov} \left( \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \right) = \begin{pmatrix} \text{Cov}(X_1, Y_1) & \text{Cov}(X_1, Y_2) \\ \text{Cov}(X_2, Y_1) & \text{Cov}(X_2, Y_2) \end{pmatrix} \]

Matrix with entries:

\[ \begin{pmatrix} (X_i - \mu_{X_i})(Y_j - \mu_{Y_j}) \end{pmatrix} \]

\[ \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \]

Usually, by covariance matrix one is referring to one vector \( \overrightarrow{X} \) and considering the covariance between its components.

\[ \text{Cov}(\overrightarrow{X}) = \begin{pmatrix} \sigma^2_X & \text{Cov}(X_1, X_2) \\ \text{Cov}(X_1, X_2) & \sigma^2_X \end{pmatrix} \]

Sometimes, the relationship between two random variables needs to be described more precisely than the covariance can.

**Conditional probability:**

Given two events A and B, we write the probability of event A given event B as:

\[ \text{Prob}(A \mid B) = \frac{\text{Prob}(A \text{ and } B)}{\text{Prob}(B)} \]
This intuitive definition of conditional probability works well in many cases but it doesn’t work well when the event B has zero probability. But this is exactly what we need to worry about if we condition on a continuous random variable taking on a certain value.

\[ \Pr(A \mid Y = y) = ? \]

Some \( \Pr(Y = y) = 0 \) for continuous r.v.s.

In general, to handle this one has to formulate the definition of conditional probability involving the values of random variables in “weak” sense:

Define: \( \Pr_{\text{weak}}(A \mid Y = y) \) to be the function on the slice space of \( Y \) (call it \( S_Y \)) such that:

For any \( B \in S_Y \)

\[ \Pr(A \text{ and } Y \in B) = \int_B \Pr(A \mid Y = y) \, d\Pr_Y(y) \]

This is just a continuous version of the law of total probability which for discrete random variables looks like:

\[ \Pr(A \text{ and } Y = y) = \sum \Pr(A \mid Y = y) \, \Pr(Y = y) \]
\[ y \in B \]