Chapter 1-4: Review of Probability Theory

STAT 6313

Outline

Probability Theory

Random variables

Some common random variables

Multiple random variables

Introduction

- Probability theory is the study of uncertainty.
- The mathematical theory of probability is very sophisticated
- We provide a basic treatment of probability that does not address these finer details.

Define probability on sets

1. Sample space Ω: The set of all the outcomes of a random experiment. Each outcome $\omega \in \Omega$ can be thought of as a complete description of the state of the real world at the end of the experiment.
2. Set of events (or event space) $\mathcal{F}$: A set whose elements $A \in \mathcal{F}$ (called events) are subsets of Ω (i.e., $A \subseteq \Omega$ is a collection of possible outcomes of an experiment).
    - $\mathcal{F}$ should satisfy: (1) $\emptyset \in \mathcal{F}$, (2) $A \in \mathcal{F} \Rightarrow \Omega \backslash A \in \mathcal{F}$, and (3) $A_1, A_2, \cdots \in \mathcal{F} \Rightarrow \cup_i A_i \in \mathcal{F}$.
3. Probability measure: A function $P : \mathcal{F} \rightarrow \mathbb{R}$ with following properties:
    - $P(A) \geq 0$, for all $A \in \mathcal{F}$
    - $P(\Omega) = 1$
    - If $A_1, A_2, \cdots$ are disjoint events (i.e., $A_i \cap A_j = \emptyset$ whenever $i \neq j$), then $$P(\cup_i A_i) = \sum_i P(A_i)$$

These three properties are called the Axioms of Probability.

Example

Consider the event of tossing a six-sided die.

- The sample space is $\Omega = \{1,\ 2,\ 3,\ 4,\ 5,\ 6\}$.
- We can define different event spaces on this sample space. For example, the simplest event space is the trivial event space $\mathcal{F} = \{\emptyset, \Omega\}$.
  - unique probability measure satisfying the requirements above is given by $P(\emptyset) = 0, P(\Omega) = 1$.
- Another event space is the set of all subsets of Ω.
  - One valid probability measure is to assign the probability of each set in the event space to be $\frac{i}{6}$ where i is the number of elements of that set;
  - for example, $P(\{1,2,3,4\}) = \frac{4}{6}$ and $P(\{1,2,3\}) = \frac{3}{6}$

Properties:

1. If $A \subseteq B \Rightarrow P(A) \leq P(B)$.
2. $P(A \cap B) \leq \min(P(A),\ P(B))$.
3. (Union Bound) $P(A \cup B) \leq P(A) + P(B)$.
4. $P(\Omega \backslash A) = 1 - P(A)$.
5. If $A_1, \ . \ . \ . \ , A_k$ are a set of disjoint events such that $\bigcup_{i=1}^{k} A_i = \Omega$, then

$$\sum_{i=1}^{k} P(A_k) = 1.$$

Conditional probability

Let B be an event with non-zero probability. The conditional probability of any event A given B is defined as,

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

In other words, $P(A|B)$ is the probability measure of the event A after observing the occurrence of event B.

Independence

Two events are called independent if and only if

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

or equivalently, $P(A|B) = P(A)$ if $P(B) > 0$.
Therefore, independence is equivalent to saying that observing B does not have any effect on the probability of A.

Law of total probability and Bayes' Theorem

1. The *law of total probability*

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

2. *Bayes' Theorem* is useful when given the conditional probability of an event, say $P(E|A_i)$, we want to find the " reverse" conditional probability $P(A_i|E)$:

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

Reversing Conditional Probabilities with Bayes' Theorem

An eyewitness observes a hit-and-run taxi-cab accident in a city in which 95% of the cabs are green and 5% are blue. The witness is 80% sure the cab was blue. Given all this information, how likely is it that the cab actually was blue?
Answer:
Let B be the even that a Blue cab staged the hit and run accident. Let EB be that the Eyewitness reported a Blue cab, and EG be Eyewitness reporting a Green taxicab.
The entire universe of event is contained in $EB \cup EG$. We know $P(EB|B) = 0.8$ and $P(EG|G) = 0.8$. Hence $P(EB|G) = 0.2$.

$$\begin{aligned} P(B|EB) &= \frac{P(EB|B) \cdot P(B)}{P(EB|B) \cdot P(B) + P(EB|G) \cdot P(G)} \\ &= \frac{0.8 \cdot 0.05}{0.8 \cdot 0.05 + 0.2 \cdot 0.95} = 0.173. \end{aligned}$$

From probability space to random variables

Consider an experiment in which we flip 10 coins, and we want to know the number of coins that come up heads. Here, the elements of the sample space Ω are 10-length sequences of heads and tails. For example, we might have

$$\omega_0 = \{H, H, T, H, T, H, H, T, T, T\} \in \Omega.$$

However, in practice, we usually do not care about the probability of obtaining any particular sequence of heads and tails. Instead we usually care about real-valued functions of outcomes, such as the number of heads that appear among our 10 tosses, or the length of the longest run of tails. These functions, under some technical conditions, are known as random variables.

Definition of random variables

A random variable X is a function

$$X : \Omega \to \mathbb{R}.$$

- Typically, we will denote random variables using upper case letters $X(\omega)$ or more simply X (where the dependence on the random outcome ω is implied).
- We will denote the value that a random variable may take on using lower case letters x.

Example of discrete random variable

In our experiment above, suppose that $X(\omega)$ is the number of heads which occur in the sequence of tosses ω. Given that only 10 coins are tossed, $X(\omega)$ can take only a finite number of values, so it is known as a discrete random variable. Here, the probability of the set associated with a random variable X taking on some specific value k is

$$P(X = k) := P(\{\omega : X(\omega) = k\})$$

Example: $\omega_0 = \{H, H, T, H, T, H, H, T, T, T\}$

- A RV is $X : \Omega \to \mathbb{R}$
    1. # of heads: $X(\omega_0) = 5$
    2. # of tosses until a tail occurs: $\mathrm{X}(\omega_0) = 3$

Example of continuous random variable

Suppose that $X(\omega)$ is a random variable indicating the amount of time it takes for a radioactive particle to decay.
$X(\omega)$ takes on a infinite number of possible values, so it is called a continuous random variable.
We denote the probability that X takes on a value between two real constants a and b (where $a < b$) as

$$P(a \leq X \leq b) := P(\{\omega : a \leq X(\omega) \leq b\})$$

Probability mass functions (PMF)

If X takes on a finite set of possible values (i.e., X is a discrete RV), a *probability mass function* (*PMF*) is a function $p_X : \Omega \to \mathbb{R}$ such that

$$P_X(x) := P(X = x).$$

We use $Val(X)$ for the set of possible values that the random variable X may assume. For example, if $X(\omega)$ is a random variable indicating the number of heads out of ten tosses of coin, then $Val(X) = \{0, 1, 2, ..., 10\}$.

Properties:

1. $0 \leq P_X(x) \leq 1$.
2. $\sum_{x \in Val(X)} P_X(x) = 1$.
3. $\sum_{x \in A} P_X(x) = P(X \in A)$

Cumulative distribution functions

In order to specify the probability measures used when dealing with random variables, it is often convenient to specify alternative functions (CDFs, PDFs, and PMFs) from which the probability measure governing an experiment immediately follows.

A cumulative distribution function (CDF) is a function $F_X : \mathbb{R} \to [0,\ 1]$ which specifies a probability measure as,

$$F_X(x) := P(X \leq x)$$

By using this function one can calculate the probability of any event in $\mathcal{F}$.

A cumulative distribution function (CDF)

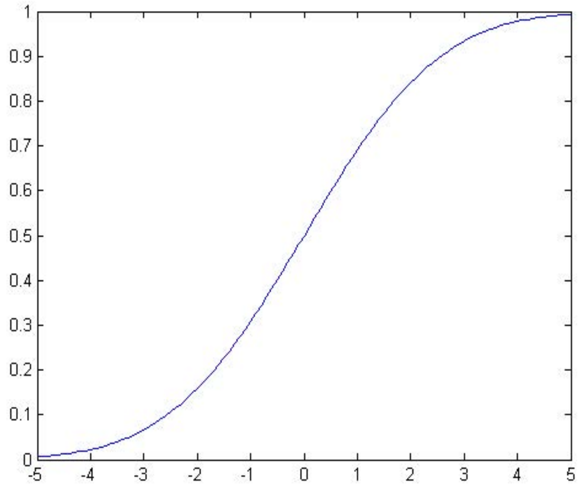


Figure: A cumulative distribution function (CDF).

Properties of CDF:

1. $0 \leq F_X(x) \leq 1$.
2. $\lim_{x \to -\infty} F_X(x) = 0$.
3. $\lim_{x \to \infty} F_X(x) = 1$.
4. $x \leq y \Rightarrow F_X(x) \leq F_X(y)$.

Probability density functions (PDF)

For some continuous random variables, the cumulative distribution function $F_X(x)$ is differentiable everywhere. In these cases, we define the Probability Density Function or PDF as the derivative of the CDF, i.e.,

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

Note here, that the PDF for a continuous random variable may not always exist (i.e., if $F_X(x)$ is not differentiable everywhere).

Properties of PDF

According to the properties of differentiation, for very small $\triangle x$,

$$P(x \leq X \leq x + \triangle x) \approx f_X(x) \triangle x$$

The value of PDF at any given point x is not the probability of that event, i.e., $f_X(x) \neq P(X = x)$. For example, $f_X(x)$ can take on values larger than one (but the integral of $f_X(x)$ over any subset of $\mathbb{R}$ will be at most one).

Properties of PDF:

1. $f_X(x) \geq 0$.
2. $\int_{-\infty}^{\infty} f_X(x) = 1$.
3. $\int_{x \in A} f_X(x) dx = P(X \in A)$.

Discrete vs. Continuous RV

Discrete RV: $Val(X)$ countable

$P(\mathrm{X} = k) := P(\{\omega | \mathrm{X}(\omega) = k\})$

Probability mass functions (PMF)

$P_X : Val(X) \rightarrow [0, 1]$

$P_X(x) := P(X = x)$

$\sum\limits_{x \in Val(X)} p_X(x) = 1$

Continuous RV: $Val(X)$ uncountable

$P(\mathrm{a} \leq \mathrm{X} \leq b) := P(\{\omega | \mathrm{a} \leq \mathrm{X}(\omega) \leq b\})$

Probability density functions (PDF)

$f_{X:}\mathrm{R} \rightarrow \mathrm{R}$

$f_X(x) := \frac{d}{dx} F_X(x)$

$\int_{-\infty}^{\infty} f_X(x) dx = 1$ where $f_X(x) dx = P(x \leq X \leq x + dx)$

Expectation

Suppose that X is a discrete random variable with PMF $p_X(x)$ and $g : \mathbb{R} \to \mathbb{R}$ is an arbitrary function.
In this case, $g(X)$ can be considered a random variable, and we define the expectation or expected value of $g(X)$ as

$$E[g(X)] := \sum_{x \in Val(X)} g(x) p_X(x).$$

If X is a continuous random variable with PDF $f_X(x)$, then the expected value of $g(X)$ is defined as,

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

Properties of expectations

$E[g(X)]$ can be thought of as a "weighted average" of the values that $g(x)$ can taken on for different values of x, where the weights are given by $P_X(x)$ or $f_X(x)$.
As a special case of the above, note that the expectation, $E[X]$ of a random variable itself is found by letting $g(x) = x$; this is also known as the mean of the random variable X.
Properties:

1. $E[a] = a$ for any constant $a \in \mathbb{R}$.
2. $E[af(X)] = aE[f(X)]$ for any constant $a \in \mathbb{R}$.
3. (Linearity of Expectation)
   $E[f(X) + g(X)] = E[f(X)] + E[g(X)]$.
4. For a discrete random variable $X, E[1\{X = k\}] = P(X = k)$.

Variance

The variance of a random variable X is a measure of how concentrated the distribution of a random variable X is around its mean.

$$Var[X] := E[(X - E(X))^2]$$

We can derive an alternate expression for the variance:

$$\begin{aligned} E[(X - E[X])^2] &= E[X^2 - 2E[X]X + E[X]^2] \\ &= E[X^2] - 2E[X]E[X] + E[X]^2 \\ &= E[X^2] - E[X]^2 \end{aligned}$$

Properties of Variance:

1. $Var[a] = 0$ for any constant $a \in \mathbb{R}$.
2. $Var[af(X)] = a^2 Var[f(X)]$ for any constant $a \in \mathbb{R}$.

Example

Calculate the mean and the variance of the uniform random variable X with PDF $f_X(x) = 1, \forall x \in [0,\ 1]$, and 0 elsewhere.

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^1 x dx = \frac{1}{2}.$$

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x) dx = \int_0^1 x^2 dx = \frac{1}{3}.$$

Thus, $Var[X] = E[X^2] - E[X]^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$.

Example

Suppose that $g(x) = 1\{x \in A\}$ for some subset $A \subseteq \Omega$. What is $E[g(X)]$?
Discrete case:

$$E[g(X)] = \sum_{x \in Val(X)} 1\{x \in A\} P_X(x) dx = \sum_{x \in A} P_X(x) dx = P(x \in A).$$

Continuous case:

$$E[g(X)] = \int_{-\infty}^{\infty} 1\{x \in A\} f_X(x) dx = \int_{x \in A} f_X(x) dx = P(x \in A).$$

Discrete random variables

$X \sim \mathit{Bernoulli}\ (p)$ (where $0 \leq p \leq 1$): one if a coin with heads probability p comes up heads, zero otherwise.

$$P(x) = \begin{cases} p & \text{if } p = 1 \\ 1 - p & \text{if } p = 0 \end{cases}$$

$X \sim \mathit{Binomial}\ (n, p)$ (where $0 \leq p \leq 1$): the number of heads in n independent flips of a coin with heads probability p.

$$P(x) = \begin{pmatrix} n \\ x \end{pmatrix} p^x (1-p)^{n-x}$$

$X \sim \mathit{Geometric}(p)$ (where $p > 0$): the number of flips of a coin with heads probability p until the first heads.

$$P(x) = p(1-p)^{x-1}$$

$X \sim \mathit{Poisson}(\lambda)$ (where $\lambda > 0$): a probability distribution over the nonnegative integers used for modeling the frequency of rare events.

$$P(x) = e^{-\lambda} \frac{\lambda^x}{x!}$$

Continuous random variables

$X \sim$ *Uniform*$(a,\ b)$ (where $a < b$): equal probability density to every value between a and b on the real line.

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

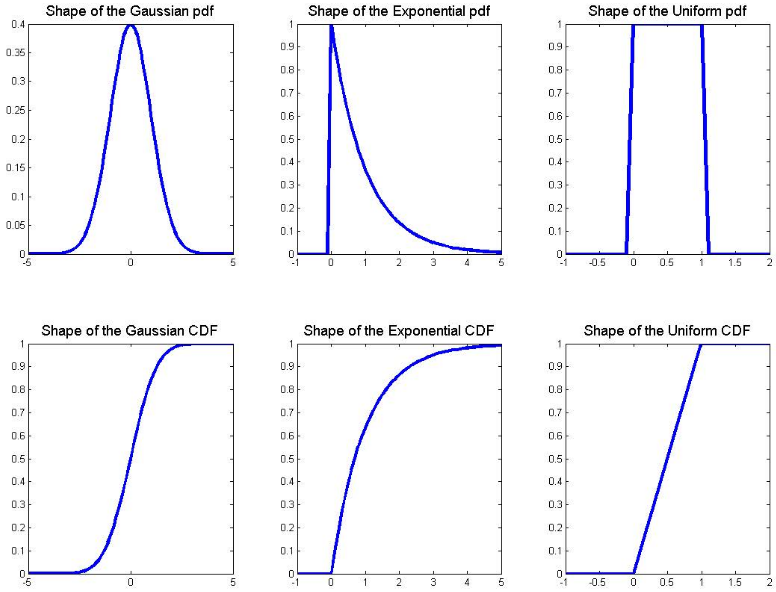
$X \sim$ *Exponential* (λ) (where $\lambda > 0$): decaying probability density over the nonnegative reals.

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

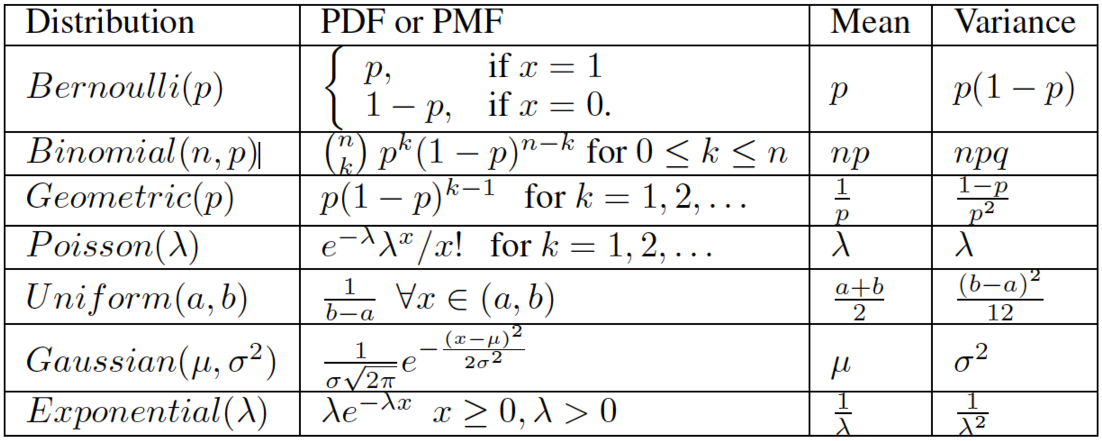
$X \sim$ *Normal* (μ, σ^2) : also known as the Gaussian distribution

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

PDF and CDF of a couple of random variables



Summary of some of the properties of these distributions.



Distribution	PDF or PMF	Mean	Variance
$Bernoulli(p)$	$\begin{cases} p, & \text{if } x = 1 \\ 1-p, & \text{if } x = 0. \end{cases}$	p	$p(1-p)$
$Binomial(n,p)$|	$\binom{n}{k} p^k (1-p)^{n-k}$ for $0 \leq k \leq n$	np	npq
$Geometric(p)$	$p(1-p)^{k-1}$ for $k = 1, 2, \ldots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
$Poisson(\lambda)$	$e^{-\lambda}\lambda^x/x!$ for $k = 1, 2, \ldots$	λ	λ
$Uniform(a,b)$	$\frac{1}{b-a}$ $\forall x \in (a,b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
$Gaussian(\mu,\sigma^2)$	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2
$Exponential(\lambda)$	$\lambda e^{-\lambda x}$ $x \geq 0, \lambda > 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

Two random variables

Thus far, we have considered single random variables. In many situations, however, there may be more than one quantity that we are interested in knowing during a random experiment. For instance, in an experiment where we flip a coin ten times, we may care about both

1. $X(\omega) =$ the number of heads that come up
2. $Y(\omega) =$ the length of the longest run of consecutive heads.

In this section, we consider the setting of two random variables.

Joint distributions

Suppose that we have two random variables X and Y. One way to work with these two random variables is to consider each of them separately. If we do that we will only need $F_X(x)$ and $F_Y(y)$. But if we want to know about the values that X and Y assume simultaneously during outcomes of a random experiment, we require a more complicated structure known as the joint cumulative distribution function of X and Y, defined by

$$F_{XY}(x, y) = P(X \leq x, Y \leq y)$$

It can be shown that by knowing the joint cumulative distribution function, the probability of any event involving X and Y can be calculated.

Marginal distributions

The marginal cumulative distribution functions $F_X(x)$ and $F_Y(y)$ can be derived from the joint CDF $F_{XY}(x, y)$:

$$F_X(x) = \lim_{y \to \infty} F_{XY}(x, y), \; F_Y(y) = \lim_{x \to \infty} F_{XY}(x, y).$$

Properties:

1. $0 \leq F_{XY}(x, y) \leq 1$.
2. $\lim_{x,y \to \infty} F_{XY}(x, y) = 1$.
3. $\lim_{x,y \to -\infty} F_{XY}(x, y) = 0$.
4. $F_X(x) = \lim_{y \to \infty} F_{XY}(x, y)$.

Joint and marginal probability mass functions

If X and Y are discrete random variables, then the joint probability mass function $P_{XY} : \mathbb{R} \times \mathbb{R} \to [0,\ 1]$ is defined by

$$P_{XY}(x,\ y) = P(X = x,\ \ Y = y)\ .$$

Here, $0 \leq P_{XY}(x,\ y) \leq 1$ for all x, y, and $\sum\limits_{x \in Val(X)} \sum\limits_{y \in Val(Y)} P_{XY}(x,\ y) = 1$.

The pmf of X is

$$P_X(x) = \sum_y P_{XY}(x,\ y)\ .$$

and similarly for $P_Y(y)$. In this case, we refer to $P_X(x)$ as the marginal probability mass function of X.

In statistics, the process of forming the marginal distribution with respect to one variable by summing out the other variable is often known as "marginalization."

Joint probability density function

Let X and Y be two continuous random variables with joint distribution function F_{XY}. In the case that $F_{XY}(x, y)$ is everywhere differentiable in both x and y, then we can define the joint probability density function,

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

Like in the single-dimensional case, $f_{XY}(x, y) \neq P(X = x,\ \ Y = y)$, but rather

$$\int \int_{x \in A} f_{XY}(x,\ y) dx dy = P((X, Y) \in A).$$

Note that the values of the probability density function $f_{XY}(x, y)$ are always nonnegative, but they may be greater than 1.

Nonetheless, it must be the case that $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x,\ y) = 1$.

Marginal probability density function

Define

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

as the marginal probability density function (or marginal density) of X, and similarly for $f_Y(y)$.

Conditional distributions

Conditional distributions seek to answer the question, what is the probability distribution over Y, when we know that X must take on a certain value x? In the discrete case, the conditional probability mass function of X given Y is simply

$$P_{Y|X}(y|x) = \frac{P_{XY}(x, y)}{P_X(x)},$$

assuming that $P_X(x) \neq 0$.

Continuous case

In the continuous case, the situation is technically a little more complicated because the probability that a continuous random variable X takes on a specific value x is equal to zero.
We start by finding the conditional CDF,

$$F_{Y|X}(y|x) = \lim_{\triangle x \to 0} P(Y \leq y | x \leq X \leq x + \triangle x).$$

It can be easily seen that if $F(x, y)$ is differentiable in both x, y then,

$$F_{Y|X}(y|x) = \int_{-\infty}^{y} \frac{f_{X,Y}(x, \alpha)}{f_X(x)} d\alpha.$$

Therefore, we define the conditional PDF of Y given $X = x$ in the following way,

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

provided $f_X(x) \neq 0$.

Bayes' rule

A useful formula that often arises when trying to derive expression for the conditional probability of one variable given another, is Bayes's rule.
In the case of discrete random variables X and Y,

$$P_{Y|X}(y|x) = \frac{P_{XY}(x, y)}{P_X(x)} = \frac{P_{X|Y}(x|y) P_Y(y)}{\sum_{y' \in Val(Y)} P_{X|Y}(x|y') P_Y(y')}.$$

If the random variables X and Y are continuous,

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

Independence

Two random variables X and Y are independent if $F_{XY}(x, y) = F_X(x)F_Y(y)$ for all values of x and y. Equivalently,

- For discrete random variables,
  - $p_{XY}(x, y) = p_X(x)p_Y(y)$ for all $x \in Val(X)$, $y \in Val(Y)$.
  - $p_{Y|X}(y|x) = p_Y(y)$ whenever $p_X(x) \neq 0$ for all $y \in Val(Y)$.
- For continuous random variables,
  - $f_{XY}(x, y) = f_X(x)f_Y(y)$ for all $x, y \in \mathbb{R}$.
  - $f_{Y|X}(y|x) = f_Y(y)$ whenever $f_X(x) \neq 0$ for all $y \in \mathbb{R}$.

Independence of functions of (X, Y)

Informally, two random variables X and Y are independent if "knowing" the value of one variable will never have any effect on the conditional probability distribution of the other variable, that is, you know all the information about the pair (X, Y) by just knowing $f(x)$ and $f(y)$.

Lemma

If X and Y are independent then for any subsets $A, B \subseteq \mathbb{R}$, we have,

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B).$$

By using the above lemma one can prove that if X is independent of Y then any function of X is independent of any function of Y.

Expectation

Suppose that we have two discrete random variables X, Y and $g : \mathbb{R}^2 \to \mathbb{R}$ is a function of these two random variables:

$$E[g(X,Y)] := \sum_{x \in Val(X)} \sum_{y \in Val(Y)} g(x,y) P_{XY}(x,y).$$

For continuous random variables (X, Y), the analogous expression is

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

Covariance

The covariance of two random variables X and Y is defined as

$$Cov[X,Y] := E[(X - E[X])(Y - E[Y])]$$

Using an argument similar to that for variance, we can rewrite this as,

$$\begin{aligned} Cov[X,Y] &= E[(X - E[X])(Y - E[Y])] \\ &= E[XY - XE[Y] - YE[X] + E[X]E[Y]] \\ &= E[XY] - E[X]E[Y] - E[Y]E[X] + E[X]E[Y]] \\ &= E[XY] - E[X]E[Y]. \end{aligned}$$

When $Cov[X,Y] = 0$, we say that X and Y are uncorrelated.

Properties

1. (Linearity of expectation) $E[f(X,Y)+g(X,Y)] = E[f(X,\ Y)] + E[g(X,Y)$
2. $Var[X+Y] = Var[X] + Var[Y] + 2Cov[X,Y]$.
3. If X and Y are independent, then $Cov[X,Y] = 0$.
4. If X and Y are independent, then $E[f(X)g(Y)] = E[f(X)]E[g(Y)]$.

Multiple continuous random variables

Consider n continuous random variables $X_1(\omega), X_2(\omega), \cdots, X_n(\omega)$. (The generalization to discrete random variables works similarly) the Joint CDF of $X_1, X_2, \ . \ . \ . \ . \ , X_n$ is

$$F_{X_1,X_2,\cdots,X_n}(x_1,x_2,\cdots,x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \cdots, X_n \leq x_n).$$

The joint PDF is:

$$f_{X_1,X_2,\cdots,x_n}(x_1,x_2,\cdots,x_n) = \frac{\partial^n F_{X_1,X_2,\cdots,x_n}(x_1,x_2,\cdots,x_n)}{\partial x_1 \cdots \partial x_n}.$$

To calculate the probability of an event $A \subseteq \mathbb{R}^n$,

$$P((X_1,\ldots,X_n) \in A) = \underset{(x_1,\ldots,x_n)\in A}{\int \cdots \int} f_{X_1,\cdots,X_n}(x_1,x_2,\ldots,x_n)dx_1 dx_2 \ldots dx_n$$

Marginal and conditional probability density functions

The marginal probability density function of X_1, and the conditional probability density function of X_1 given X_2, . . . , X_n are:

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1, X_2, \cdots, x_n}(x_1, x_2, \cdots, x_n) dx_2 \dots dx_n$$

$$f_{X_1|X_2, \cdots, x_n}(x_1|x_2, \cdots, x_n) = \frac{f_{X_1, X_2, \cdots, x_n}(x_1, x_2, \cdots, x_n)}{f_{X_2,.,X_n}(x_2, \cdots, x_n)}$$

Chain rule:

$$\begin{aligned} f(x_1, x_2, \cdots, x_n) &= f(x_n|x_1, x_2, \dots, x_{n-1}) f(x_1, x_2, \dots, x_{n-1}) \\ = f(x_n|x_1, x_2, \dots, x_{n-1}) &f(x_{n-1}|x_1, x_2, \dots, x_{n-2}) f(x_1, x_2, \dots, x_{n-2}) \\ &= \dots = f(x_1) \prod_{i=2}^{n} f(x_i|x_1, \dots, x_{i-1}) \end{aligned}$$

Independence

Multiple events $A_1, \cdots, A_k$ are *mutually independent* if and only if for any subset $S \subseteq \{1, \cdots, k\}$,

$$P(\bigcap_{i \in S} A_i) = \prod_{i \in S} P(A_i).$$

Random variables X_1, X_2, . . . , X_n are *mutually independent* if and only if

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

Multivariate Gaussian distribution

Multivariate Gaussian or multivariate normal distribution:

$$f(x_1, x_2, \cdots, x_n; \mu, \Sigma) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

with mean $\mu \in \mathbb{R}^n$ and covariance matrix $\Sigma \in \mathrm{S}^n_{++}$ (where S^n_{++} refers to the space of symmetric positive definite $n \times n$ matrices).
We write this as $X \sim \mathcal{N}(\mu, \Sigma)$.
Note that in the case $n = 1$, this reduces the regular definition of a normal distribution with mean parameter μ_1 and variance Σ_{11}.

Central limit theorem

Let $X_1, ..., X_n$ be IID random variables from a distribution with mean μ and variance σ^2. Define

$$\bar{X} = \frac{1}{n}\sum_i x_i.$$

The Central limit theorem states that as n goes to infinity,

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n}).$$