Set Theory

lim inf: If something is in the lim inf of a Series of Sets, That means it is in every set in that sequence after some fixed value N. It is in 'all but finitely many'

$$\liminf_{n\to\infty} (F_n(x)) = \sup_{m\geq 1} \left(\inf_{n\geq m} F_n(x) \right)$$

lim sup: If something is in lim sup of a sequence of sets, it is in infinitely many of them.

$$\limsup_{n\to\infty} (F_n(x)) = \inf_{m\geq 1} \left(\sup_{n\geq m} F_n(x) \right)$$

Example: Imagine a sequence of sets that alternates $[-1-\frac{1}{n}, 0]$, $[0, 1+\frac{1}{n}]$, as n goes from 1 to infinity, so the first few sets are

$$[-2,0], [0,1.5], [-\tfrac{4}{3},0], [0,1.25], [-1.2,0]\ldots$$

The lim inf contains only 0, since any positive number will be contained in at most half of all future sets after any fixed number of sets N, and same for negative numbers.
The lim sup contains $[-1,1]$, since every value in $[-1,1]$ will be repeated infinitely many times.

Set Theory

A property holds <u>almost everywhere</u> (a.e.) if it holds for all elements in a set except a subset of measure 0 (measure to be defined later)

A property holds <u>infinitely often</u> (i.o.) if it holds for infinitely many values of n

A property holds <u>Eventually</u> if it holds always, except for finitely many n. There is a threshold N such that it holds for all $n > N$.

Connection to lim inf, lim sup: i.o. 'corresponds' to lim sup, eventually 'corresponds' to lim inf

If A_n is the sequence alternating $[-1-\frac{1}{n}, 0], [0, 1+\frac{1}{n}]$ Then the Property $x = 0$ and $\forall x, x \in A_n$ holds eventually, and the Property $x \in [0, 1]$ and $\forall x, x \in A_n$ holds infinitely often.

Fields and σ-Fields

A field $\mathcal{F}$ is a collection of subsets of another set Ω that satisfies:

A. $\Omega \in \mathcal{F}$
B. If $A \in \mathcal{F}$, $A^c \in \mathcal{F}$
C. If $A_1, A_2, \dots, A_n \in \mathcal{F}$, $\bigcup_{i=1}^{n} A_i \in \mathcal{F}$

Example: Suppose Ω is the set of results of 2 coin flips, so it has three elements: $\Omega \equiv [TT, HT, HH]$. Let $\mathcal{F}_1$ be

$\mathcal{F}_1 \equiv [[TT, HT, HH], \emptyset]$, Checking A, B, and C, we can see they are all true, so $\mathcal{F}_1$ is a field. Now, suppose $\mathcal{F}_2$ contains all subsets of Ω, $\mathcal{F}_2 \equiv [[TT, HT, HH], \emptyset, [TT, HH], [HH], [TT], [HT], [TT, HT], [HH, HT]]$. Checking A, B, C, they are all still satisfied, so $\mathcal{F}_2$ is a field. Now suppose $\mathcal{F}_3 \equiv [[TT], [HT], [HH], [HH, TT], [HT, HH], [HT, TT]]$ A and C are both violated, B is not, so $\mathcal{F}_3$ is not a field.

Fields and σ-Fields

A <u>σ-Field</u> on Ω is a field that also has D below:

A. $\Omega \in \mathcal{F}$
B. $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$
C. $A_1, A_2, \dots, A_n \in \mathcal{F}$ then $\bigcup_{i=1}^{n} A_i \in \mathcal{F}$

D. $A_1, A_2, \dots \in \mathcal{F}$ then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

Fields are defined as they are so that the notions correspond to the idea of results of a random experiment, where Ω contains all possible <u>outcomes</u> and $\mathcal{F}$ contains all possible <u>events</u>. For example, in the previous note, the example of Heads and Tails, Ω is the set of outcomes of a random experiment, flipping a coin twice, and $\mathcal{F}_2$ is the set of possible events, HH is 2 heads, HT is 1 head 1 tail, TT is two tails. D is included because it extends notions to infinity in a logical way, and a richer mathematical theory is obtained.

Measures

A **measure** is a set function, an assignment of a number $\mu(A)$ to each set A in a set of sets. A measure on $\sigma(\mathcal{F})$ is a non-negative, extended real valued function denoted μ on $\sigma(\mathcal{F})$ such that whenever $A_1, A_2, \ldots$ form a finite or countably infinite collection of disjoint sets in $\sigma(\mathcal{F})$, we have $\mu(\cup_n A_n) = \Sigma_n \mu(A_n)$.

If $\mu(\Omega) = 1$, μ is a **probability measure**

For example, a probability measure on $\Omega = [HH, HT, TT]$, $\mathcal{F}_2$ as defined previously, could assign $\mu(HT) = \frac{1}{2}$, $\mu(HH) = \frac{1}{4}$, $\mu(TT) = \frac{1}{4}$, and it would follow that $\mu(\cup(HH, TT)) = \mu(HH) + \mu(TT) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$, or we could do $\mu(\cup([HH, TT, HT], \emptyset)) = \mu(\Omega) + \mu(\emptyset) = 1$

However, $\mu(\cup([HH, HT], [HT, TT]))$ is not necessarily $\mu([HH, HT]) + \mu(HT, TT)$ because $[HH, HT]$ and $[HT, TT]$ are not disjoint sets.

Measure Spaces & Borel Sets

A <u>measure space</u> is a triple $(\Omega, \mathcal{F}, \mu)$ where Ω is a set, $\mathcal{F}$ is a σ-field, and μ is a measure on $\mathcal{F}$

A <u>probability space</u> is a measure space where μ is a probability measure.

The Borel set is defined as the smallest σ-field containing all intervals $(a, b]$, $a, b \in \mathbb{R}$. From properties of σ-fields, it can be shown that from the above definition, the Borel set must also contain all open intervals, closed intervals, left closed intervals, intervals from a to ∞ or $-b$ to ∞, and all singletons

The Borel set can be thought of as the set of all subsets of $\mathbb{R}$ you'd typically think about, and is a σ-field typically used when Ω is $\mathbb{R}$. The borel set is often denoted $\mathbb{B}(\mathbb{R})$

LBGS Measures and Distribution Functions

A Lebesgue-Stieljes Measure on $\Omega \equiv \mathbb{R}$ and $\mathcal{F} \equiv \mathbb{B}(\mathbb{R})$ is a measure such that $\mu(I) < \infty$ for each bounded interval I.

A Distribution Function on $\mathbb{R}$ is a map $F: \mathbb{R} \to \mathbb{R}$ that is increasing [$a<b$ implies $F(a) \leq F(b)$] and right continuous [$\lim_{x \to x_0^+} F(x) = F(x_0)$]

As it turns out, there is a 1:1 correspondence between LBGS measures and distribution functions given by $\mu((a,b]) = F(b) - F(a)$ for $a>b$

Now we are approaching math we are more familiar with. If F is a distribution function and μ is the LBGS measure, then

$$\mu((a,b]) = F(b) - F(a), \quad \mu((a,b)) = F(b^-) - F(a)$$

$$\mu([a,b]) = F(b) - F(a^-), \quad \mu([a,b)) = F(b^-) - F(a^-)$$

where $F(x^-)$ denotes $\lim_{y \to x^-} F(y)$

LBGS-Meas Dist Func Example

Suppose $F(x) = \begin{cases} x-1 & x \in [-\infty, 1) \\ 2x-1 & x \in [1, 2) \\ x^2 & x \in [2, \infty) \end{cases}$

Let μ be the corresponding LBGS-Meas

What is

A) $\mu((1,2])$?

$= F(2) - F(1) = 4 - 1 = 3$

B) $\mu((-3, 1))$?

$= F(1^-) - F(-3) = \left(\lim_{x \to 1^-} x-1\right) - (-3-1) = 0+4 = 4$

C) $\mu([2,3])$?

$= F(3) - F(2^-) = 9 - \lim_{x \to 2^-} 2x-1 = 9 - 3 = 6$

D) $\mu([1,2))$?

$= F(2^-) - F(1^-) = \left(\lim_{x \to 2^-} 2x-1\right) - \left(\lim_{x \to 1^-} x-1\right)$

$= 3 - 0 = 3$ Question like this has appeared on past Basic exams!

Lebesque Measure

One kind of Lebesgue-Stieljes Measure commonly used to assign a measure to a subset of an n-space is called the <u>Lebesgue Measure</u> (not at all confusing, am I right?). The Lebesgue Measure in dimensions 1, 2, and 3 corresponds to length, area, and volume. For example, the lebesgue measure of $\square_{(0,0),(2,0),(0,2),(2,2)}$ in $\mathbb{R}^2$ would be the area of that box, $= 4$.

The Lebesgue Measure on the measure space $(\mathbb{R}, \mathbb{B}(\mathbb{R}), u)$ is $u((a,b]) = b - a$. All sets that can be given a Lebesgue Measure are called <u>Lebesgue Measurable</u>. All closed and open intervals are Lebesgue Measurable.

Measurability of Functions

Suppose h is a function that maps a set $\Omega_1 \rightarrow \Omega_2$. Suppose further Ω_1 has a σ-field $\mathcal{F}_1$, Ω_2 has a σ-field $\mathcal{F}_2$

We say that the function h is <u>measurable</u> relative to $\mathcal{F}_1, \mathcal{F}_2$ if and only if $h^{-1}(A) \in \mathcal{F}_1$ for each $A \in \mathcal{F}_2$

For example, let's use the heads and tails example again, letting $\mathcal{F}_1$ be $[\emptyset, [HT], [HH], [TT], [HT, HH], [HT, TT], [HH, TT], [HH, HT, TT]]$, and let $\mathcal{F}_2$ be the number of heads where Ω_2 is the set $[0, 1, 2]$. So our h is

$$
\begin{array}{ccc}
 & \Omega_1 & & \Omega_2 \\
 & HH & \longrightarrow & 2 \\
h \equiv & HT & \longrightarrow & 1 \\
 & TT & \longrightarrow & 0
\end{array}
$$

Now, $\mathcal{F}_2$ is $[\emptyset, 0, 1, 2, [0,1], [1,2], [0,2], [1,2,3]]$. h is measurable because for each A in $\mathcal{F}_2$, $h^{-1}(A)$ is in $\mathcal{F}_1$

e.g. if $A = 1$ then $h^{-1}(A) = HT$, if $A = [1,2]$, $h^{-1}(A) = [HT, HH]$, if $A = \emptyset$, $h^{-1}(A) = \emptyset$, for every A in $\mathcal{F}_2$, plugging A into $h^{-1}(A)$ gets something in $\mathcal{F}_1$

Intuition Behind Measurability

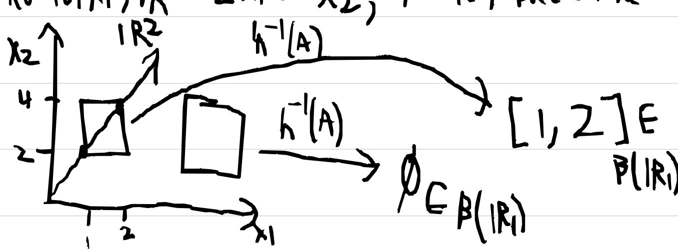
The reason we care about $A \in \mathcal{F}_2$ always having an $h^{-1}(A) \in \mathcal{F}_1$ is because the idea behind probability is to take some "random experiment" and map it to the real line, and we want to measure the probability of something showing up on the real line in terms of that "set" of things showing up on the real line mapping directly to an event in the random experiment. For example, with coin flipping, if I want the probability that I get 2 heads or 1 head, I want that guy in $\mathcal{F}_2$, that is $A \in \mathcal{F}_2$ $A = [1, 2]$ to map to an event in $\mathcal{F}_1$, $h^{-1}(A) \in \mathcal{F}_1$, $h^{-1}(A) = [HT, HH]$ so that I can get the probability of $[1, 2]$ as the probability of $[HT, HH]$ which is 0.75, assuming $\mu([HT]) = 0.5$, $\mu([TT]) = 0.25$, $\mu([HH]) = 0.25$. These notions will be formalized in the note on Random Variables.

Borel Measurability

If $(\Omega, \mathcal{F})$ is a measurable space and $h : \Omega \to \mathbb{R}^n$, then h is said to be <u>Borel Measurable</u> on $(\Omega, \mathcal{F})$ iff h is measurable relative to the σ-fields $\mathcal{F}$ and $\mathbb{B}(\mathbb{R}^n)$. That is, for each $A \in \mathbb{B}(\mathbb{R}^n)$, $h^{-1}(A) \in \mathcal{F}$

If Ω is a Borel subset of $\mathbb{R}^k$ and the term "Borel Measurable" is used, it can be assumed $\mathcal{F} = \mathbb{B}(\mathbb{R}^k)$

For example, suppose $\Omega = \mathbb{R}^1$, $\mathcal{F} = \mathbb{B}(\mathbb{R}_1)$, is mapped into $\mathbb{R}^2$, $\mathbb{B}(\mathbb{R}^2)$ with the function $f(x) = \binom{x}{2x}$. f is borel Measurable because for any shape we can draw in $\mathbb{R}^2$, the points in the shape that lie on $2x_1 = x_2$ are mapped back to $B(\mathbb{R}_1)$ and are always going to be in it. If you draw a shape with no points in $2x_1 = x_2$, you get back the null set which is still in $B(\mathbb{R}_1)$



Lebesgue Integration

We will now go through some tedious steps to outline a more general form of integration than Riemann Integration which can be used to define integrals across certain sets that are not Riemann Integrable. Before we do this, it should be noted that Lebesgue Integration is <u>strictly more general</u> than Riemann integration. If a Riemann integral of something is defined, the Lebesgue integral of that thing is the <u>same</u> as the Riemann integral. Only in cases where Riemann integration is not defined do the two differ.

Let's give an example. Suppose we have a function $f(x) = \begin{cases} 1 & x \in \mathbb{Q} \\ 0 & x \notin \mathbb{Q} \end{cases}$ where $\mathbb{Q}$ is the set of rational numbers. Since the rationals are only countably infinite, intuitively $\int f(x) = 0$. But the Riemann integral is not defined for a situation like this. Now, let us define Lebesgue Integration. LbG integration will allow us later to express $F(x)$ and $E[X]$ for 'discrete' & 'continuous' r.v.'s in a unified framework

Lebesgue Integration

Define a simple function to be a function that is Borel measurable and takes on only finitely many distinct values. Thus it can be written as $h(x)=\sum_{i=1} h(x_i) I_{A_i}$ where A_i are disjoint sets in $\mathcal{F}$.

If $h: \Omega \to \overline{\mathbb{R}}$ is a simple function ($\overline{\mathbb{R}}$ is $\mathbb{R}$ and $-\infty, \infty$), $h(x)=\sum_{i=1}^{n} h(x_i) I_{A_i}$ then the Lebesgue integral is defined as

$$\int_{\Omega} h \, d\mu = \sum_{i=1}^{n} h(x_i)\, \mu(A_i)$$, μ is Lebesgue measure

So long as $-\infty, +\infty$ are not in the sum, the integral exists. Suppose h is not simple, then we define the Lebesgue integral to be

$$\int_{\Omega} h \, d\mu = \sup \int_{\Omega} s \, d\mu : s \text{ is simple, } 0 \leq s \leq h$$

Since h is borel measurable, so are its + and − parts, so we further define

$$\int_{\Omega} h \, d\mu = \int_{\Omega} h^{+} \, d\mu - \int_{\Omega} h^{-} \, d\mu$$

Lebesque Integration

Back to our example: $f(x) = \begin{cases} 1 & x \in \mathbb{Q} \\ 0 & x \notin \mathbb{Q} \end{cases}$

This function only takes on values 0 and 1, so it is a simple function. Ω is $\mathbb{R}$, we are mapping into $\mathbb{R}$, and the measure is the Lebesgue measure. So we do

$$\int_\Omega f\,d\mu = \int_{-\infty}^{\infty} f(x)\,d\mu = \sum_{i=1}^{r} f(x_i)\,\mu(A_i)$$

Where $A_1 = \mathbb{Q}$ and $A_2 = \mathbb{Q}^c$

$$= \left(f(x_1)\mu(\mathbb{Q}) + f(x_2)\mu(\mathbb{Q}^c)\right) = \left(1\,\mu(\mathbb{Q}) + 0\,\mu(\mathbb{Q}^c)\right)$$

$= \mu(\mathbb{Q})$. Since the Lebesgue Measure is defined as $\mu((a,b]) = b-a$, and since the set of rationals in countably infinite, and since a measure has $\mu(\bigcup_n A_n) = \sum_n \mu(A_n)$, the measure of $\mathbb{Q}$ is the infinite sum of the measure of singletons, which is the sum of infinitely many zeros, so $\mu(\mathbb{Q}) = 0 + 0 + \ldots + 0 = 0$

Important Theorems

Included here are some important theorems from measure theory. Proofs are not included, the reason being that the purpose of these notes is to include as much core information as possible with as little "extra" as necessary, which is why throughout I've tried to keep examples as simple as possible and used them just to demonstrate the concept.

Monotone Convergence Theorem: Let $h_1, h_2, \ldots$ form an increasing sequence of non-neg Borel measurable functions, and let $h(\omega) = \lim_{n\to\infty} h_n(\omega)$, $\omega \in \Omega$. Then $\int_\Omega h_n \, d\mu \to \int_\Omega h \, d\mu$.

Basically this just says if a sequence of h's ≥ 0, then $h_n \uparrow h$ implies that $\int_\Omega h_n \, d\mu \uparrow \int_\Omega h \, d\mu$

The theorems are sourced from Probability and Measure Theory by Ash

Important Theorems

Additivity Theorem: Let f and g be Borel measurable and $f+g$ be not well defined. If $\int_\Omega f\,d\mu$ and $\int_\Omega g\,d\mu$ exist and are well defined, and $\int_\Omega f\,d\mu + \int_\Omega g\,d\mu$ is well defined, then $\int_\Omega f+g\,d\mu = \int_\Omega f\,d\mu + \int_\Omega g\,d\mu$. Basically you can do in Lebesgue integration what you can do in Riemann integration with the sum of the integrals being integral of the sum

Monotone Convergence Theorem the sequel: Let $g_1, \ldots, g, h$ be Borel measurable.

a. If $g_n \geq h$ for all n, $\int_\Omega h\,d\mu > -\infty$, $g_n \downarrow g$, then $\int_\Omega g_n\,d\mu \downarrow \int_\Omega g\,d\mu$

b. If $g_n \leq h$ for all n, $\int_\Omega h\,d\mu < \infty$, $g_n \uparrow g$, then $\int_\Omega g_n\,d\mu \uparrow \int_\Omega g\,d\mu$

Important Theorems

Fatou's Lemma: Let $f_1, f_2, \ldots f$ be Borel measurable.

a. If $f_n \geq f$ for all n, where $\int_\Omega f \, d\mu > -\infty$, then $\liminf_{n\to\infty} \int_\Omega f_n \, d\mu \geq \int_\Omega \liminf_{n\to\infty} f_n \, d\mu$

b. If $f_n \leq f$ for all n, where $\int_\Omega f \, d\mu < \infty$, then $\limsup_{n\to\infty} \int_\Omega f_n \, d\mu \leq \int_\Omega \limsup_{n\to\infty} f_n \, d\mu$

Basically, the lim inf of the integrals is greater or equal to the integral of the lim infs, and the lim sup of the integrals is less than or equal to the integral of the lim sups.

Dominated Convergence Theorem: If $f_1, f_2, \ldots, f, g$ are Borel measurable, $|f_n| \leq g \; \forall n$, g is μ-integrable, and $f_n \to f$ a.e., then f is μ-integrable and $\int_\Omega f_n \, d\mu \to \int_\Omega f \, d\mu$

Bounded Convergence Theorem: Dominated Convergence theorem where g is constant

Random Variables

We now learn what is meant by the joke
"A random variable is neither random nor a variable"

A <u>random variable</u> is a measurable set function $X: \Omega_1 \to \Omega_2$ from a probability space $(\Omega_1, \mathcal{F}_1, \mu)$ to a measurable space $(\Omega_2, \mathcal{F}_2)$, often the real numbers $(\mathbb{R}, \mathbb{B}(\mathbb{R}))$

To be precise, for a probability space $(\Omega, \mathcal{F}, \mu)$ and a real valued function $X: \Omega \to \mathbb{R}$, X is said to be a random variable if for every Borel set $B \in \mathbb{B}(\mathbb{R})$ we have $X^{-1}(B) = \{w : X(w) \in B\} \in \mathcal{F}$

Every Borel set in $\mathbb{R}$ has to map back to some event in $\mathcal{F}$.

Probability

Let $(\Omega, \mathcal{F}, P)$ be a probability space and X be a random variable on that space. The probability of a Borel set $A \in \mathbb{B}(\mathbb{R})$ written $P(X \in A)$ or $P(X^{-1}(A))$, is defined as $\mu(A) = P(X \in A)$, where $X^{-1}(A) \in \mathcal{F}$. That is, $X^{-1}(A)$ is an event in the probability space mapped by the function X to the Borel set A. Note that P here is a probability measure.

In the coin example, lets say $(\Omega, \mathcal{F}, \mu)$ is $\Omega \equiv [TT, HT, HH]$ and $\mathcal{F}$ is all possible subsets of Ω. We can define a random variable X into $(\mathbb{R}, \mathbb{B}(\mathbb{R}))$ as before.

Ω		$\mathbb{R}$
HH	$\longrightarrow$	0
HT	$\longrightarrow$	1
TT	$\longrightarrow$	2

$P(HH) = 0.25$
$P(HT) = 0.5$, $P(TT) = 0.25$

What are the probabilities of the following Borel sets:

$\mathbb{B}((3,10])$? $P(X^{-1}((3,10])) = P(\emptyset) = 0$

$\mathbb{B}((1.5,2.5])$? $P(X^{-1}((1.5,2.5])) = P(HH) = 0.25$

$\mathbb{B}((-3,3])$? $P(X^{-1}((-3,3])) = P([TT, HT, HH])$
$= 0.25 + 0.5 + 0.25 = 1$

Distribution Functions

Note that 'distribution function' as defined here is commonly referred to as the CDF. It is the defining function of a random variable.

Let $(\Omega, \mathcal{F}, P)$ be a probability space and X be a random variable on that probability space. The <u>distribution function</u> $F(X)$ is defined $F(x) = P(X \leq x)$ $= P(X \in (-\infty, x)) = P(X^{-1}((-\infty, x)))$ $= \int_{-\infty}^{x} dF$, the CdF defined in terms of Lebesgue Integration.

A function F is a distribution function of some random variable if and only if:

a. F is non-decreasing

b. $\lim_{x \to \infty} F(x) = 1$ and $\lim_{x \to -\infty} F(X) = 0$

c. F is right continuous, that is $\lim_{y \downarrow x} F(y) = F(x)$

If F is a distr. function, it must also be true that:

d. If $F(x^-) = \lim_{y \uparrow x} F(y)$ then $F(x^-) = P(X < x)$

e. $P(X = x) = F(x) - F(x^-)$

Density Functions & Expected Value

When the distribution function $F(x)$ has the form $F(x) = \int_{-\infty}^{x} f(y)\,dy$ we say that X has <u>density function</u> $f(x)$

If X is a random variable on $(\Omega, \mathcal{F}, P)$, Then the <u>expected value</u> of X is defined as <u>$\int_\Omega X\,dP$</u> $= \int_\Omega X^+\,dP - \int_\Omega X^-\,dP$. Let $x^+ = \max(X, 0)$, the positive part of X, and x^- be $\max(-X, 0)$ be the negative part of X. We say $E[X]$ exists if the subtraction $E[X^+] - E[X^-]$ makes sense, that is it isn't $\infty - \infty$ or $-\infty - (-\infty)$

Alternatively, and more commonly, we use the following definition of Expected Value:

$$E[X] = \int_{-\infty}^{\infty} x\,dF_X(x) = \int_{-\infty}^{\infty} x f(x)\,dx$$

if X has a density function, and if not we use

$$E[X] = \int_\Omega X\,dP = \sum_{i=1}^{n} x_n P(X = x_n)$$

Coin example: $E[X] = \sum_{i=1}^{n} x_n P(X = x_n) = 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = \frac{1}{2} + \frac{1}{2} = 1$

Discrete Probability Spaces

If the sample space Ω is a finite or countably infinite set, as in the coin example, and $\Omega = \{w_1, w_2, \ldots\}$, then let $p_1, p_2, \ldots$ be the probabilities of $w_1, w_2, \ldots$, respectively.

Then $P(A) = \sum_{w_i \in A} p_i$ and $P(\Omega) = 1$

So that P is a probability measure

In this case we call $(\Omega, \mathcal{F}, P)$ a <u>discrete probability space</u>

If $X_1, \ldots, X_n$ is a random vector in a discrete probability space $(\Omega, \mathcal{F}, P)$, then the <u>probability mass function</u> or PMF is a function that gives the probability of any particular $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$ where $X^{-1}(x_1, x_2, \ldots, x_n) = u_{x_1, x_2, \ldots, x_n} \in \Omega$

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = P(u_{x_1, x_2, \ldots, x_n})$$

Conditional Probability

If $(\Omega, \mathcal{F}, P)$ is a probability space
space and $A_1, \dots, A_n, B_1, \dots, B_n \in \mathcal{F}$ are events,
we define the conditional probability of events
$A_1, \dots, A_n$ given $B_1, \dots, B_n$ as

$$P(A_1, \dots, A_n \mid B_1, \dots, B_n) = \frac{P(A_1 \cap A_2 \cap \dots \cap B_1 \cap \dots \cap B_n)}{P(B_1 \cap \dots \cap B_n)}$$

If $A_1, \dots, A_n \in \mathcal{F}$ are events, we say $A_1, \dots, A_n$
are <u>independent</u> if

$$P(A_1, \dots, A_n) = P(A_1) P(A_2) \dots P(A_n)$$

Alternatively, we can say $A_1, \dots, A_n$ are
independent if $\forall i \in n$

$$P(A_i \mid A_{\neq i}) = P(A_i)$$

Where $A_{\neq i}$ indicates $A_1, \dots, A_{i-1}, A_{i+1}, \dots A_n$, all
events A we are considering except A_i

Marginal Probability

The marginal probability of X when X is defined in a joint discrete probability space with Y is

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

And if X and Y have a joint density function $f_{X,Y}(x,y)$, the marginal probability of X is defined

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

This notion, and the notion of conditional probability, extend the same way for more than two random variables. e.g. if we have X, Y, Z: $f_X(x) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y,Z}(x,y,z)\,dy\,dz$

$$f_{X|Y,Z} = \frac{f_{X,Y,Z}(x,y,z)}{f_{Y,Z}(y,z)}, \quad P(X|Y,Z) = \frac{P(X=x, Y=y, Z=z)}{P(Y=y, Z=z)}$$

etcetera

Independence of Random Variables

Let $X_1, \ldots, X_n$ be random variables (vector) on $(\Omega, \mathcal{F}, P)$. Then $X_1, \ldots, X_n$ are said to be <u>independent</u> iff for all sets $B_1, \ldots, B_n \in \mathcal{B}(\mathbb{R}^n)$, we have

$$P(X_1 \in B_1, X_2 \in B_2, \ldots, X_n \in B_n) = P(X_1 \in B_1) P(X_2 \in B_2) \ldots P(X_n \in B_n)$$

If $X_1, \ldots, X_n$ are independent and each individual random variable X_i has the same distribution function $F(x_i)$, then we say $X_1, \ldots, X_n$ are <u>independently identically distributed</u>, often abbreviated iid

Conditional Expectation

If X is a random variable on $(\Omega, \mathcal{F}, P)$ and $B \in \mathcal{F}$, then the conditional expectation of X given B, $E[X|B]$ is

$$E[X|B] = \frac{E[X 1_B]}{P(B)}$$

If X is a random variable defined on a joint probability space with $Y_1, \dots, Y_n$ on $(\Omega, \mathcal{F}, P)$, then the conditional expectation of $X | Y_1, Y_2, \dots, Y_n$, written as $E[X|G]$ where G is the σ-algebra generated by $Y_1, \dots, Y_n$, is itself a random variable such that

$$\int_A E[X|G]\, dP = \int_A X\, dP \quad \forall A \in G$$

The σ-algebra G can be thought of as the 'information' provided by $Y_1, \dots, Y_n$. Often, $E[X|G]$ can be found by simply integrating the joint pdf times x across x and dividing by the marginal pdf of $\vec{Y}$

Conditional Expecration

If $X, Y_1, \ldots, Y_n$ are each equipped with density functions and defined in $(\Omega, \mathcal{F}, P)$ then we can write the conditional expectation of X given $Y_1, \ldots, Y_n$ as

$$E[X \mid Y_1, \ldots, Y_n] = \int_{-\infty}^{\infty} x \frac{f_{X, Y_1, \ldots, Y_n}(x, \ldots, y_n)}{f_{Y_1, \ldots, Y_n}(y_1, \ldots, y_n)} \, dx$$

Likewise, if $X, Y_1, \ldots, Y_n$ are in a discrete probability space, then we can write the conditional expectation of X given $Y_1, \ldots, Y_n$ as

$$E[X \mid Y_1, \ldots, Y_n] = \sum_{x \in \text{support of } X} x \frac{P(X = x, \ldots, Y_n = y_n)}{P(Y_1 = y_1, \ldots, Y_n = y_n)}$$

Random Vectors

A <u>random vector</u> is a set of random variables associated with the same random experiment. It is a Borel measurable map from Ω to $\mathbb{R}^n$ where the random vector is defined on a 'joint' probability space $(\Omega, \mathcal{F}, P)$. The random vector can be considered as an n-tuple of random variables $(X_1, X_2, \dots, X_n)$.

Example: Suppose X is a coin flip, where heads is 1 and tails is 0, and if we flip heads we flip a three sided dice and let Y be the result of the dice roll, and if we flip tails we set Y to the result of a 4-sided dice then our random vector (function from $(\Omega, \mathcal{F}, P)$ to $\mathbb{R}^2$) is defined to the right

Ω	$\binom{X}{Y}$	$\mathbb{R}^2$
T, •	$\longrightarrow$	0, 1
T, ••	$\longrightarrow$	0, 2
T, •••	$\longrightarrow$	0, 3
T, ⁘	$\longrightarrow$	0, 4
H, •	$\longrightarrow$	1, 1
H, ••	$\longrightarrow$	1, 2
H, •••	$\longrightarrow$	1, 3
		$\uparrow$ X, $\uparrow$ Y

Conditional Probability

If X is one of a vector of random variables $X, Y_1, \dots, Y_n$ all not having density functions (in a discrete probability space), then we can write the conditional probability of $X \mid Y_1, \dots, Y_n$ as

$$P(X \mid Y_1, \dots, Y_n) = \frac{P(X = x, Y_1 = y_1, \dots, Y_n = y_n)}{P(Y_1 = y_1, \dots, Y_n = y_n)}$$

If X only has a density functions, we define $f_{X \mid Y_1, \dots, Y_n}$ as the function that satisfies

$$f_X(x) = \sum_{\vec{y} \in \text{support of } \vec{Y}} f_{X \mid \vec{Y}}(x \mid \vec{y}) \, P(\vec{Y} = \vec{y})$$

where $\vec{Y}$ is shorthand for $Y_1, \dots, Y_n$ and the support of $\vec{Y}$ is the set on which $P(\vec{Y} = \vec{y}) \neq 0$. If X has no density, but $\vec{Y}$ does, then we can write

$$P(X \mid Y_1, \dots, Y_n) = \frac{f_{X, Y_1, \dots, Y_n}(x, \dots, y_n)}{f_{Y_1, \dots, Y_n}(x, \dots, y_n)}$$

Conditional & Marginal Summarized

For all: $X, \vec{Y}$ defined in some $(\Omega, \mathcal{F}, P)$

$X, \vec{Y}$ in a discrete probability space:

Marginal: $P(X=x) = \sum_{\substack{y \in \text{support} \\ \text{of } Y}} P(X=x, Y=\vec{y})$

Conditional: $P(X=x \mid \vec{Y}=\vec{y}) = \dfrac{P(X=x, \vec{Y}=\vec{y})}{P(\vec{Y}=\vec{y})}$

$X, \vec{Y}$ all having density functions:

Marginal: $f_X(x) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X,\vec{Y}}(x, \vec{y}) \, dy_n \, dy_{n-1} \ldots dy_1$

Conditional: $f_{X|\vec{Y}}(x, \vec{y}) = \dfrac{f_{X,\vec{Y}}(x, \vec{y})}{f_{\vec{Y}}(\vec{y})}$

Radon Nikodym Theorem

Let μ be a σ-finite measure and λ a signed σ-finite measure on the σ-field $\mathcal{F}$ of subsets of Ω. Assume that λ is absolutely continuous with respect to μ. Then there is a Borel measurable function $g: \Omega \to \overline{\mathbb{R}}$ such that

$$\lambda(A) = \int_A g \, d\mu \quad \forall A \in \mathcal{F}$$

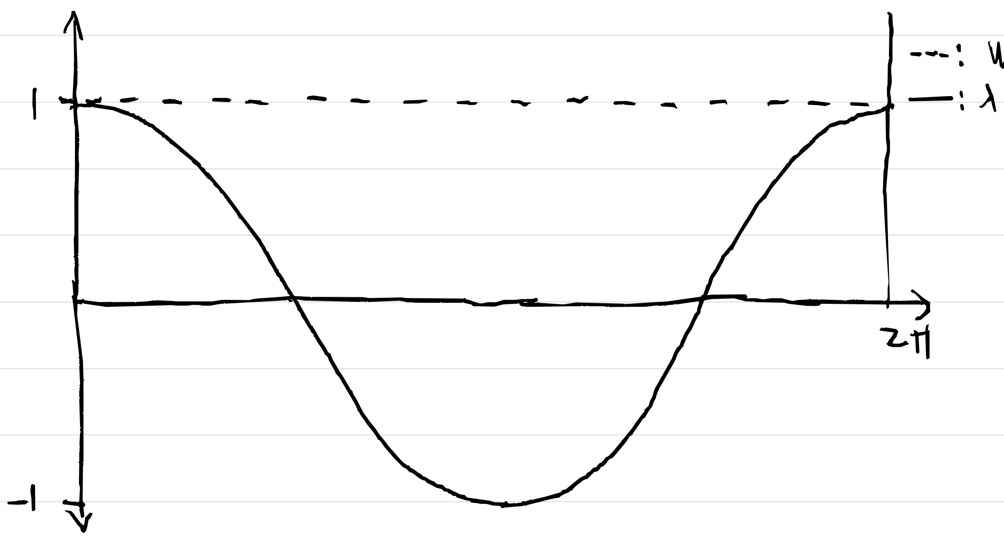
If h is another such function, $g = h$ a.e.

g is often called the radon-nikodym derivative and notated $g = \frac{d\lambda}{d\mu}$

Essentially, $\frac{d\lambda}{d\mu}$ expresses how λ "allocates measure" with respect to how μ "allocates measure". If P_1 and P_2 are probability measures, then $\frac{dP_1}{dP_2}$ the radon-nikodym derivative is the likelihood ratio. λ can be thought of as a kind of "reweighting" of μ

Radon-Nikodym Example

Suppose μ is the lebesgue measure on $\Omega = [0, 2\pi]$ and $\mathcal{F} = \mathbb{B}([0, 2\pi])$, which just measures the length of an interval. Now suppose λ is defined as $\lambda((a,b]) = \cos(b) - \cos(a)$.



We can express $\lambda((a,b])$ as

$\lambda((a,b]) = \int_a^b (-\sin(x))\,d\mu$ since

$\int_a^b (-\sin(x))\,d\mu = \cos(x)\Big|_a^b = \cos(b) - \cos(a)$

thus $\frac{d\lambda}{d\mu} = -\sin(x)$.

Marginal Distributions

For a random vector $X_1, \dots, X_n, Y_1, \dots, Y_n$ the <u>marginal distribution</u> of $X_1, \dots, X_n$ can be thought of as the joint distribution of just $X_1, \dots, X_n$

When $X_1, \dots, X_n, Y_1, \dots, Y_n$ are a part of a discrete probability space, and the joint distribution of $X_1, \dots, X_n, Y_1, \dots, Y_n$ is known, the marginal probability is

$$P(X_1 = x_1, \dots, X_n = x_n) = \sum_{y \in \text{support of } Y} P(X_1 = x_1, \dots, Y_n = y_n)$$

When $X_1, \dots, X_n, Y_1, \dots, Y_n$ all have known joint density function $f_{X_1, \dots, Y_n}(x_1, \dots, y_n)$ then the marginal density function of $X_1, \dots, X_n$ is

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1, \dots, Y_n}(x_1, \dots, y_n)\, dy_1\, dy_2 \dots dy_n$$

Conditional Distributions

For a random vector $X_1, \ldots, X_n, Y_1, \ldots, Y_n$, the <u>conditional distribution</u> of $X_1, \ldots, X_n$ given $Y_1, \ldots, Y_n$, notated $X_1, \ldots, X_n \mid Y_1, \ldots, Y_n$ can be thought of as the distribution of $X_1, \ldots, X_n$ when it is known that $Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n$

When $X_1, \ldots, X_n, Y_1, \ldots, Y_n$, are in a discrete probability space the conditional PMF of $X_1, \ldots, X_n \mid Y_1, \ldots, Y_n$ is

$$P(X_1 = x_1, \ldots, X_n = x_n \mid Y_1 = y_1, \ldots, Y_n = y_n) = \frac{P(X_1 = x_1, \ldots, Y_n = y_n)}{P(Y_1 = y_1, \ldots, Y_n = y_n)}$$

When $X_1, \ldots, X_n, Y_1, \ldots, Y_n$ have joint density function $f_{X_1, \ldots, X_n, Y_1, \ldots, Y_n}(x_1, \ldots, y_n)$ then the joint conditional density function of $X_1, \ldots, X_n \mid Y_1, \ldots, Y_n$ is given by

$$f_{X_1, \ldots, X_n \mid Y_1, \ldots, Y_n}(x_1, \ldots, y_n) = \frac{f_{X_1, \ldots, X_n, Y_1, \ldots, Y_n}(x_1, \ldots, x_n, y_1, \ldots, y_n)}{f_{Y_1, \ldots, Y_n}(y_1, \ldots, y_n)}$$

Borel-Cantelli Lemma

If $A_1, A_2, \ldots$ is a sequence of events in a probability space $(\Omega, \mathcal{F}, P)$, and $\sum_{i=1}^{\infty} P(A_i) < \infty$, then $P(\limsup_{n\to\infty} A_n) = 0$, that is, the set of elements of Ω that occur infinitely often has measure zero. This is often stated "The probability that A_n occurs infinitely often is zero" because it implies $\sum_{i=1}^{\infty} I_{A_i} < \infty$ with probability 1 when I_{A_i} is the random variable that is 1 when A_i occurs and 0 when A_i does not occur.

Additionally, if $\sum_{i=1}^{\infty} P(A_i) = \infty$ and the events $(A_i)_{i=1}^{\infty}$ are pairwise independent, then $P(\limsup_{n\to\infty} A_n) = 1$, "the probability that A_n occurs infinitely often is 1", $\sum_{i=1}^{\infty} I_{A_i} = \infty$. This is known as the second Borel-Cantelli Lemma

Proof of Borel Cantelli Lemma

$$P(\limsup_{n\to\infty} A_n) = P(\lim_{n\to\infty} \bigcup_{i=n}^{\infty} A_i) = \lim_{n\to\infty} P(\bigcup_{i=n}^{\infty} A_i)$$

The first equality is true by definition, The second is true because the sequence $\{\bigcup_{i=n}^{\infty} A_i\}_{n=1}^{\infty}$ is non-increasing, $\bigcup_{i=1}^{\infty} A_i \supseteq \bigcup_{i=2}^{\infty} A_i \ldots \supseteq \lim_{n\to\infty} \bigcup_{i=n}^{\infty} A_i$. By countable additivity, for disjoint $E_1, E_2, \ldots$ $P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$. Let $E_1 \equiv \bigcup_{i=1}^{\infty} A_i - \bigcup_{i=2}^{\infty} A_i$, $E_2 \equiv \bigcup_{i=2}^{\infty} A_i - \bigcup_{i=3}^{\infty} A_i, \ldots,$ so $E_1, E_2, \ldots$ are disjoint, but $\forall n$, $P(\bigcup_{i=n}^{\infty} E_i) = P(\bigcup_{i=n}^{\infty} A_i)$. Thus $P(\bigcup_{i=n}^{\infty} A_i) = \sum_{i=n}^{\infty} P(E_i)$, and $P(\lim_{n\to\infty} \bigcup_{i=n}^{\infty} A_i) = \lim_{n\to\infty} \sum_{i=n}^{\infty} P(E_i) = \lim_{n\to\infty} P(\bigcup_{i=1}^{\infty} A_i)$

Now suppose $\sum_{i=1}^{\infty} P(A_i) < \infty$. Let $C \equiv \sum_{i=1}^{\infty} P(A_i) < \infty$.

$\forall n$ $\sum_{i=1}^{\infty} P(A_i) = \sum_{i=1}^{n-1} P(A_i) + \sum_{i=n}^{\infty} P(A_i)$, thus $\lim_{n\to\infty} (\sum_{i=1}^{n-1} P(A_i) + \sum_{i=n}^{\infty} P(A_i)) = C + \lim_{n\to\infty} \sum_{i=n}^{\infty} P(A_i)$. But $\lim_{n\to\infty} (\sum_{i=1}^{\infty} P(A_i)) = C$. So $C = C + \lim_{n\to\infty} \sum_{i=1}^{\infty} P(A_i) \Rightarrow \lim_{n\to\infty} \sum_{i=n}^{\infty} P(A_i) = 0$. Since $P(\bigcup_{i=n}^{\infty} A_i) \leq \sum_{i=n}^{\infty} P(A_i)$, $\lim_{n\to\infty} \sum_{i=1}^{\infty} P(A_i) = 0 \Rightarrow \lim_{n\to\infty} P(\bigcup_{i=n}^{\infty} A_i) = 0$

Thus $P(\limsup_{n\to\infty} A_n) = 0$ qed