Probability Theory

S.R.S.Varadhan
Courant Institute of Mathematical Sciences
New York University

Fall 1999

Chapter 1

Measure Theory

Introduction

The evolution of probability theory was based more on intuition rather than mathematical axioms during its early development. In the 1930's A. N. Kolmogorov provided an axiomatic basis for probability theory and it is now the universally accepted model. There are certain 'non commutative' versions that have their origins in quantum mechanics that are generalizations of the Kolmogorov Model. We shall however use exclusively Kolmogorov's framework.

The basic intuition in probability theory is the notion of randomness. There are experiments whose results are not predicatable and can be determined only after performing it and then observing the outcome. The simplest familiar examples are, the tossing of a fair coin, or the throwing of a balanced die. In the first experiment the result could be either a head or a tail and the throwing of a die could result in a score of any integer from 1 through 6. These are experiments with only a finite number of alternate outcomes. It is not difficult to imagine experiments that have countably or even uncountably many alternatives as possible outcomes.

Abstractly then, there is a space Ω of all possible outcomes and each individual outcome is represented as a point ω in that space Ω. Subsets of Ω are called events and they correspond to a collection of outcomes. If the outcome ω is in the subset A, then the event A

is said to have occurred. For example in the case of a die the set $A = \{1, 3, 5\} \subset \Omega$ corresponds to the event 'an odd number shows up'. With this terminology it is clear that union of sets corresponds to 'or', intersection to 'and', and complementation to 'negation'.

One would expect that probabilities should be associated with each outcome and there should be a 'Probability Function' $f(\omega)$ which is the probabilty that ω occurs. In the case of coin tossing we may expect $\Omega = \{H, T\}$ and

$$f(T) = f(H) = \frac{1}{2}.$$

Or in the case of a die

$$f(1) = f(2) = \cdots = f(6) = \frac{1}{6}$$

Since 'Probaility' is normalized so that certainty corresponds to a Probability of 1, one expects

$$\sum_{\omega \in \Omega} f(\omega) = 1 \tag{1.1}$$

If Ω is uncountable this is a mess. There is no reasonable way of adding up an uncountable set of numbers each of which is 0. This suggests that it may not be possible to start with probabilities associated with individual outcomes and build a meaningful theory. The next best thing is to start with the notion that probabilities are already defined for events. In such a case, $P(A)$ is defined for a class $\mathcal{B}$ of subsets $A \subset \Omega$. The question that arises naturally is what should $\mathcal{B}$ be and what properties should $P(\cdot)$ defined on $\mathcal{B}$ have? It is natural to demand that the class $\mathcal{B}$ of sets for which probabilities are to be defined satisfy the following properties:

The sets Ω and Φ are in $\mathcal{B}$. For any two sets A and B in $\mathcal{B}$, the sets $A \cup B$ and $A \cap B$ are again in $\mathcal{B}$. If $A \in \mathcal{B}$, then the complement A^c is again in $\mathcal{B}$. Any class of sets satisfying these properties is called a *field.*

Definition 1.1. A 'probability' or more precisely 'a finitely additive probability measure' is a nonnegative set function defined for sets in $\mathcal{B}$ that satisfies the following properties:

$$P(A) \geq 0 \quad \text{for all} \quad A \in \mathcal{B} \tag{1.2}$$

$$P(\Omega) = 1 \quad \text{and} \quad P(\Phi) = 0 \tag{1.3}$$

If $A \in \mathcal{B}$ and $B \in \mathcal{B}$ are *disjoint* then

$$P(A \cup B) = P(A) + P(B) \tag{1.4}$$

In particular

$$P(A^c) = 1 - P(A) \tag{1.5}$$

for all $A \in \mathcal{B}$

A condition which is some what more technical, but important from a mathematical viewpoint is that of countable additivity. The class $\mathcal{B}$, in addition to being a field is assumed to be closed under countable union (or equivalently, intersection); i.e. if $A_n \in \mathcal{B}$ for every n, then $A = \cup_n A_n \in \mathcal{B}$. Such a class is called a *σ-field.* The 'probability' itself is presumed to be defined on a *σ-field* $\mathcal{B}$.

A set function P defined on a σ-field is called a 'countably additive probability measure' if in addition to satsfying equations (1.2), (1.3) and (1.4) it satisfies the following countable additivity property: for any sequence of pairwise *disjoint* sets A_n with $A = \cup_n A_n$

$$P(A) = \sum_n P(A_n) \tag{1.6}$$

Exercise 1.1. The limit of an increasing (or decreasing) sequence A_n of sets is defined as its union $\cup_n A_n$ (or the intersection $\cap_n A_n$). A monotone class is defined as a class that is closed under monotone limits of an increasing or decreasing sequence of sets. Show that a field $\mathcal{B}$ is a σ-field if and only if it is a monotone class.

Exercise 1.2. Show that a finitely additive probability measure $P(\cdot)$ defined on a σ-field $\mathcal{B}$, is countably additive, i.e. satisfies equation (1.6), if and only if either of the following two conditions is satsfied.

If A_n is any nonincreasing sequence of sets in $\mathcal{B}$ and $A = \lim_n A_n = \cap_n A_n$ then

$$P(A) = \lim_n P(A_n).$$

If A_n is any nondecreasing sequence of sets in $\mathcal{B}$ and $A = \lim_n A_n = \cup_n A_n$ then

$$P(A) = \lim_n P(A_n).$$

Exercise 1.3. If $A, B \in \mathcal{B}$, and P is a finitely additive probability measure show that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. How does this generalize to $P(\cup_{j=1}^{n} A_j)$?

Exercise 1.4. If P is a countably additive probability measure, show that for any sequence $A_n \in \mathcal{B}$, $P(\cup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} P(A_n)$.

Although we would like our 'probability' to be a countably additive probability measure, on a σ- field $\mathcal{B}$ of subsets of a space Ω it is not clear that are plenty of such things. As a first small step show

Exercise 1.5. If for $n \geq 1$, $\omega_n \in \Omega$ are distinct points and $p_n \geq 0$ with $\sum_n p_n = 1$ then

$$P(A) = \sum_{n:\omega_n \in A} p_n$$

defines a countably additive probability measure on Ω. (This is still cheating because the measure P lives on a countable set.)

Definition 1.2. A probability measure P on a field $\mathcal{F}$ is said to be countably additive on $\mathcal{F}$ if for any sequence $A_n \in \mathcal{F}$ with $A_n \downarrow \Phi$, we have $P(A_n) \downarrow 0$.

Exercise 1.6. Given any class $\mathcal{F}$ of subsets of Ω there is a unique σ-field $\mathcal{B}$ such that it is the smallest σ-field that contains $\mathcal{F}$.

Definition 1.3. The σ-field in the above exercise is called the σ-field generated by $\mathcal{F}$.

Construction of Measures

The following theorem is important for the construction of countably additive probability measures. A detailed proof can be found in any text on real variables. In an effort to be complete we will sketch the standard proof.

Theorem 1.1. (Caratheodory Extension Theorem). *Any countably additive probabilty measure P on a field $\mathcal{F}$ extends uniquely as a countably additive probability measure to the σ-field $\mathcal{B}$ generated by $\mathcal{F}$.*

Proof: The proof proceeds along the following steps:

Step 1. Define an object P^* called the *outer measure* for all sets A.

$$P^*(A) = \inf_{\cup_j A_j \supset A} \sum_j P(A_j) \tag{1.7}$$

where the infimum is taken over all countable collections $\{A_j\}$ of sets from $\mathcal{F}$ that cover A. Without loss of generality we can assume that $\{A_j\}$ are disjoint. (Replace A_j by $(\cap_{i=1}^{j-1} A_i^c) \cap A_j$).

Step 2. Show that P^* has the following properties:

1. P^* is countably sub-additive,i.e.

$$P^*(\cup_j A_j) \le \sum_j P^*(A_j)$$

2. For $A \in \mathcal{F}$, $P^*(A) \le P(A)$. (Trivial)

3. For $A \in \mathcal{F}$, $P^*(A) \ge P(A)$. (Need to use the countable additivity of P on $\mathcal{F}$)

Step 3. Define a set E to be *measurable* if

$$P^*(A) \ge P^*(A \cap E) + P^*(A \cap E^c)$$

holds for all sets A, and establish the following properties for the class $\mathcal{M}$ of measurable sets. The class of measurable sets $\mathcal{M}$ is a σ-field and P^* is a countably additive measure on it.

Step 4. Finally show that $\mathcal{M} \supset \mathcal{F}$. This implies that $\mathcal{M} \supset \mathcal{B}$ and P^* is an extension of P from $\mathcal{F}$ to $\mathcal{B}$.

Uniqueness is quite simple. Let P_1 and P_2 be two countably additive probability measures on a *σ-field* $\mathcal{B}$ that agree on a field $\mathcal{F} \subset \mathcal{B}$. Let us define $\mathcal{A} = \{A : P_1(A) = P_2(A)\}$. Then $\mathcal{A}$ is a monotone class i.e., if $A_n \in \mathcal{A}$ is increasing (decreasing) then $\cup_n A_n$ $(\cap_n A_n) \in \mathcal{A}$. We will leave the following elementary fact as an

Exercise 1.7. The smallest monotone class generated by a field is the same as the σ-field generated by the field.

It now follows that $\mathcal{A}$ must contain the σ-field generated by $\mathcal{F}$ and that proves uniqueness.

The extension Theorem does not quite solve the problem of constructing countably additive probability measures. It reduces it to constructing them on fields. The following theorem is important in the theory of Lebesgue integrals and is very useful for the construction of countably additive probability measures on the real line. The proof will again be only sketched. The natural σ-field on which to define a probability measure on the line is the Borel σ-field. This is defined as the smallest σ-field containing all intervals and includes in particular all open sets.

Let us consider the class of subsets of the real numbers, $\mathcal{I} = \{I_{a,b} : -\infty \leq a < b \leq \infty\}$ where $I_{a,b} = \{x : a < x \leq b\}$ if $b < \infty$, and $I_{a,\infty} = \{x : a < x < \infty\}$. In other words $\mathcal{I}$ is the collection of intervals that are left-open and right-closed. The class of sets that are finite disjoint unions of members of $\mathcal{I}$ is a field $\mathcal{F}$, if the empty set is added to the class. If we are given a function $F(x)$ on the real line which is nondecreasing, continuous from the right and satisfies

$$\lim_{x \to -\infty} F(x) = 0 \quad \text{and} \quad \lim_{x \to \infty} F(x) = 1$$

we can define a finitely additive probability measure P by first defining

$$P(I_{a,b}) = F(b) - F(a)$$

for intervals and then extending it to $\mathcal{F}$ by defining it as the sum for disjoint unions from $\mathcal{I}$. Let us note that the Borel σ-field $\mathcal{B}$ on the real line is the σ-field generated by $\mathcal{F}$.

Theorem 1.2. *(Lebesgue). P is countably additive on $\mathcal{F}$ if and only if $F(x)$ is a right continuous function of x. Therefore for each right continuous nondecreasing function $F(x)$ with $F(-\infty) = 0$ and $F(\infty) = 1$ there is a unique probability measure P on the Borel subsets of the line, such that $F(x) = P(I_{-\infty,x})$. Conversely every countably additive probability measure P on the Borel subsets of the line comes from some F. The correspondence between P and F is one-to-one.*

Proof: The onlly difficult part is to establish the countable additivity of P on $\mathcal{F}$ from the right continuity of $F(\cdot)$. Let $A_j \in \mathcal{F}$ and $A_j \downarrow \Phi$, the empty set. Let us assume, $P(A_j) \geq \delta > 0$, for all j and then establish a contradiction.

Step 1. Take a large interval $[-\ell, \ell]$ and replace A_j by $B_j = A_j \cap [-\ell, \ell]$. In other words one assumes without loss of generality that $A_j \subset [-\ell, \ell]$ for some $\ell < \infty$.

Step 2. If

$$A_j = \cup_{i=1}^{k_j} I_{a_{j,i}, b_{j,i}}$$

use the right continuity of F to replace A_j by B_j which is a union of left open right closed intervals with the same right end points, but with left end points moved ever so slightly to the right. Achieve this in such a way that

$$P(A_j - B_j) \leq \frac{\delta}{10.2^j}$$

for all j.

Step 3. Define C_j to be the closure of B_j by adding the left end points of the intervals that make up B_j. Let $E_j = \cap_{i=1}^{j} B_i$ and $D_j = \cap_{i=1}^{j} C_i$. Then, (i) D_j are closed bounded sets, (ii) since $A_j \supset D_j$, $D_j \downarrow \Phi$, (iii) because $D_j \supset E_j$ and $P(E_j) \geq \delta - \sum_j P(A_j - B_j) \geq \frac{9}{10}\delta$, each D_j is nonempty and this violates the finite intersection property that every decreasing sequence of bounded nonempty closed sets on the real line has a nonempty intersection, i.e. has atleast one common point.

The rest of the proof is left as an exercise.

The function F is called the distribution function corresponding to the probability measure P.

Examples.

1. Suppose $x_1, x_2, \cdots, x_n, \cdots$ is a sequence of points and we have probabilities p_n at these ponts then for the discrete measure

$$P(A) = \sum_{n : x_n \in A} p_n$$

we have the distribution function

$$F(x) = \sum_{n : x_n \leq x} p_n$$

that only increases by jumps, the jump at x_n being p_n. The points $\{x_n\}$ themselves can be discrete like integers or dense like the rationals.

2. If $f(x)$ is a nonnegative integrable function with $\int_{-\infty}^{\infty} f(y)dy = 1$ then $F(x) = \int_{-\infty}^{x} f(y)dy$ is a distribution function which is continuous. In this case f is the density of the measure P and can be calculated as $f(x) = F'(x)$.

3. There are (messy) examples of F that are continuous, but do not come from any density. More on this later.

Exercise 1.8. Let us try to construct the Lebesgue measure on the rationals $\mathcal{Q} \subset [0, 1]$. We would like to have

$$P[\mathcal{I}_{a,b}] = b - a$$

for all rational $0 \le a \le b \le 1$. Show that it is impossible by showing that $P[\{q\}] = 0$ for the set $\{q\}$ containing the single rational q while $P[\mathcal{Q}] = P[\cup_{q\in\mathcal{Q}}\{q\}] = 1$. Where does the earlier proof break down? Once we have a countably additive probability measure P on a space (Ω, Σ) we will call the triple (Ω, Σ, P) a probabilty space.

Integration

An important notion is that of a random variable or a measurable function.

Definition 1.4. A random variable or measurable function is map from $\Omega \to R$ i.e. a real valued function on Ω such that for every Borel set $B \subset R$, $f^{-1}(B) = \{\omega : f(\omega) \in B\}$ is a measurable subset of Ω i.e $f^{-1}(B) \in \Sigma$.

Exercise 1.9. It is enough to check the requirement for sets $B \subset R$ that are intervals or even just sets of the form $(-\infty, x]$ for $-\infty < x < \infty$.

A function that is measurable and satisfies $|f(\omega)| \le M$ all $\omega \in \Omega$ for for some finite M is called a bounded measurable function. The following statements are the essential steps in developing an integration theory. Details can be found in any book on real variables.

1. If $A \in \Sigma$, the indicator function $\mathbf{1}_A(x)$ defined to be 1 on A and 0 on A^c, is bounded and measurable.

2. Sums, products, limits, compositions and reasonable elementary operations like min and max performed on measurable functions lead to measurable functions.

3. If $\{A_j : 1 \le j \le n\}$ is a disjoint partition of Ω into measurable sets, the function $f(\omega) = \sum_j c_j \mathbf{1}_{A_j}(\omega)$ is a measurable function and is referred to as a 'simple' function.

4. Any bounded measurable function f is a uniform limit of simple functions. To see this, if f is bounded by M, divide $[-M, M]$ into n subintervals I_j of length $\frac{2M}{n}$ with midpoints c_j. Let $A_j = f^{-1}(I_j) = \{\omega : f(\omega) \in I_j\}$ and $f_n = \sum_{j=1}^n c_j \mathbf{1}_{A_j}$. Clearly f_n is simple, $\sup_\omega |f_n(\omega) - f(\omega)| \le \frac{M}{n}$, and we are done.

5. For simple functions $f = \sum c_j \mathbf{1}_{A_j}$ the integral $\int f(\omega) dP$ is defined to be $\sum_j c_j P(A_j)$. It enjoys the following properties:

   (a) If f and g are simple so is any linear combination $af + bg$ for real constants a and b and

   $$\int (af + bg) dP = a \int f dP + b \int g dP.$$

   (b) If f is simple so is $|f|$ and $|\int f dP| \le \int |f| dP \le \sup_\omega |f(\omega)|$.

6. If f_n is a sequence of simple functions converging to f uniformly, $a_n = \int f_n dP$ is a Cauchy sequence of real numbers and therefore has a limit a as $n \to \infty$. The integral $\int f dP$ of f is defined to be this limit a. One can verify that a depends only on f and not on the sequence f_n chosen to approximate f.

7. Now the integral is defined for all bounded measurable functions and enjoys the following properties.

   (a) If f and g are bounded measurable functions and a, b are real constants then for the linear combination $af + bg$ which is a bounded measurable function

   $$\int (af + bg) dP = a \int f dP + b \int g dP.$$

(b) If f is a bounded measurable function so is $|f|$ and $|\int f dP| \leq \int |f| dP \leq \sup_\omega |f(\omega)|$.

(c) In fact a slightly stronger inequality is true. For any bounded measurable f,

$$\int |f| dP \leq P(\{\omega : |f(\omega)| > 0\}) \sup_\omega |f(\omega)|$$

(d) If f is a bounded measurable function and A is a measurable set one defines

$$\int_A f(\omega) dP = \int \mathbf{1}_A(\omega) f(\omega) dP$$

and we can write for any measurable set A,

$$\int f dP = \int_A f dP + \int_{A^c} f dP$$

In addition to uniform convergence there are other weaker notions of convergence.

Definition 1.5. A sequence f_n functions is said to converge to a function f everywhere or poinwise if

$$\lim_{n\to\infty} f_n(\omega) = f(\omega)$$

for every $\omega \in \Omega$.

In dealing with sequences of functions on a space that has a measure defined on it, often it does not matter if the sequence fails to converge on a set of points that is insignificant. For example if we are dealing with the Lebesgue measure on the interval $[0,1]$ and $f_n(x) = x^n$ then $f_n(x) \to 0$ for all x except $x = 1$. A single point, being an interval of length 0 should be insignificant for the Lebesgue measure.

Definition 1.6. A sequence f_n of measurable functions is said to converge to a measurable function f almost everywhere (usually abbreviated as a.e.) if there exists a measurable set N with $P(N) = 0$ such that

$$\lim_{n\to\infty} f_n(\omega) = f(\omega)$$

for every $\omega \in N^c$.

Note that almost everywhere convergence is always relative to a probability measure.

Another notion of convergence is the following:

Definition 1.7. A sequence f_n of measurable functions is said to converge to a measurable function f in measure or in probability if

$$\lim_{n\to\infty} P[\omega : |f_n(\omega) - f(\omega)| \geq \epsilon] = 0$$

for every $\epsilon > 0$.

Let us examine these notions in the context of indicator functions of sets $f_(x) = \mathbf{1}_{A_n}(x)$. As soon as $A \neq B$, $\sup_\omega |\mathbf{1}_A(\omega) - \mathbf{1}_B(\omega)| = 1$, so that uniform convergence never really takes place. On the other hand one can verify that $\mathbf{1}_{A_n}(\omega) \to \mathbf{1}_A(\omega)$ for every ω if and only if the two sets

$$\limsup_n A_n = \cap_n \cup_{m\geq n} A_m$$

and

$$\liminf_n A_n = \cup_n \cap_{m\geq n} A_m$$

both coincide with A. Finally $\mathbf{1}_{A_n}(\omega) \to \mathbf{1}_A(\omega)$ in measure if and only if

$$\lim_{n\to\infty} P(A_n \Delta A) = 0$$

where for any two sets A and B the symmetric difference $A\Delta B$ is defined as $A\Delta B = (A \cap B^c) \cup (A^c \cap B) = A \cup B - A \cap B$. It is the set of points that belong to either set but not to both. For instance $\mathbf{1}_{A_n} \to 0$ in measure if and only if $P(A_n) \to 0$.

Exercise 1.10. There is a difference between almost every where convergence and convergence in measure. The first is really stronger. Consider the interval $[0, 1]$ and devide it successively into $2, 3, 4 \cdots$ parts and enumerate the intervals in succession. That is, $I_1 = [0, \frac{1}{2}]$, $I_2 = [\frac{1}{2}, 1]$, $I_3 = [0, \frac{1}{3}]$, $I_4 = [\frac{1}{3}, \frac{2}{3}]$, $I_5 = [\frac{2}{3}, 1]$, and so on. If $f_n(x) = \mathbf{1}_{I_n}(x)$ it easy to check that f_n tends to 0 in measure but not almost everywhere.

Exercise 1.11. But the following statement is true. If $f_n \to f$ as $n \to \infty$ in measure, then there is a subsequence f_{n_j} such that $f_{n_j} \to f$ almost everywhere as $j \to \infty$.

Lemma 1.3If $f_n \to f$ almost everywhere then $f_n \to f$ in measure.

Proof: $f_n \to f$ outside N is equivalent to

$$\cap_n \cup_{m \geq n} [\omega : |f_m(\omega) - f(\omega)| \geq \epsilon] \subset N$$

for every $\epsilon > 0$. In particular by countable additivity

$$P[\omega : |f_n(\omega) - f(\omega)| \geq \epsilon] \leq P[\cup_{m \geq n}[\omega : |f_m(\omega) - f(\omega)| \geq \epsilon] \to 0$$

as $n \to \infty$ and we are done.

Theorem 1.4. (Bounded Convergence Theorem). *If f_n is uniformly bounded and $f_n \to f$ in measure then $\int f_n dP \to \int f dP$.*

Proof. Since

$$| \int f_n dP - \int f dP \mid = | \int (f_n - f) dP \mid \leq \int |f_n - f| dP$$

we need only prove that if $f_n \to 0$ in measure and $|f_n| \leq M$ then $\int |f_n| dP \to 0$. To see this

$$\int |f_n| dP = \int_{|f_n| \leq \epsilon} |f_n| dP + \int_{|f_n| > \epsilon} |f_n| dP \leq \epsilon + MP[\omega : |f_n(\omega)| > \epsilon]$$

and taking limits

$$\limsup_{n \to \infty} \int |f_n| dP \leq \epsilon$$

and since $\epsilon > 0$ is arbitrary we are done.

The bounded convergence theorem is the essence of countable additivity. Let us look at the example of $f_n(x) = x^n$ on $0 \leq x \leq 1$ with Lebesgue measure. Clearly $f_n(x) \to 0$ a.e. and therefore in meausre. While the convergence is not uniform, $0 \leq x^n \leq 1$ for all n and x and so the bounded convergence theorem applies. In fact

$$\int_0^1 x^n dx = \frac{1}{n+1} \to 0.$$

However if we replace x^n by nx^n, $f_n(x)$ still goes to 0 a.e., but the sequence is no longer uniformly bounded and the integral does not go to 0. We now proceed to define integrals of nonnegative measurable functions.

Definition 1.8. If f is a nonnegative measurable function we define

$$\int f dP = \{\sup \int g dP : 0 \leq g \leq f,\ g \ \text{ bounded } \}$$

An important result is

Theorem 1.5(Fatou's Lemma). *If $f_n \geq 0$ are measurable and $f_n \to f$ in measure then*

$$\int f dP \leq \liminf_{n\to\infty} \int f_n dP.$$

Proof: Suppose $0 \leq g \leq f$ is bounded. Then the sequence $h_n = f_n \wedge g$ is uniformly bounded and $h_n \to h = f \wedge g = g$. Therefore $\int g dP = \lim_{n\to\infty} \int h_n dP$ and $\int h_n dP \leq \int f_n dP$ for every n. Hence $\int g dP \leq \liminf_{n\to\infty} \int f_n dP$. Since this is true for every bounded measurable $g \leq f$ we are done.

Corollary 1.6. (Monotone Convergence Theorem). *If $f_n \uparrow f$ monotonically then*

$$\int f_n dP \to \int f dP \ \text{ as } \ n \to \infty.$$

Proof: Obviously $\int f_n dP \leq \int f dP$ and the rest is Fatou.

Now we try to define integrals of arbitrary measurable functions. A nonnegative measurable function is said to be **integrable** if $\int f dP < \infty$. A function f is said to be integrable if $|f|$ is integrable and we define $\int f dP = \int f^+ dP - \int f^- dP$ where $f^+ = f \vee 0$ and $f^- = -f \wedge 0$ are the positive and negative parts of f. The integral has the following properties.

1. It is linear. If f and g are integrable so is $af + bg$ for any two real constants and $\int (af + bg) dP = a \int f dP + b \int g dP$.

2. . $|\int f dP| \leq \int |f| dP$ for every integrable f.

3. If $f = 0$ except on a set N of measure 0, then f is integrable and $\int f dP = 0$. In particular if $f = g$ almost everywhere then $\int f dP = \int g dP$.

Another important theorem is

Theorem 1.7. (The Dominated Convergence Theorem). *If a sequence of measurable functions $f_n \to f$ in measure and $|f_n| \leq g$ for all n and ω for some integrable function g, then $\int f_n dP \to \int f dP$ as $n \to \infty$.*

Proof: $g + f_n$ and $g - f_n$ are nonnegative and converge in measure to $g + f$ and $g - f$ respectively. By Fatou's lemma

$$\liminf_{n\to\infty} \int (g + f_n) dP \geq \int (g + f) dP$$

Since $\int g dP$ is finite we can subtract it from both sides and get

$$\liminf_{n\to\infty} \int f_n dP \geq \int f dP.$$

Working the same way with $g - f_n$ yields

$$\limsup_{n\to\infty} \int f_n dP \leq \int f dP$$

and we are done.

Exercise 1.12. Take the unit interval with the Lebesgue measure and define $f_n(x) = n^a \mathbf{1}_{[0,\frac{1}{n}]}$. Clearly for $x \neq 0$ $f_n(x) \to 0$ while $\int f_n(x)dx = n^{\alpha-1} \to 0$ if and only if $\alpha < 1$. What is $g(x) = \sup_n f_n(x)$ and when is g integrable?

If $F(\omega) = f(\omega) + ig(\omega)$ is a complex valued measurable function with real and imaginary parts $f(\omega)$ and $g(\omega)$ that are integrable we define

$$\int F(\omega) dP = \int f(\omega) dP + i \int g(\omega) dP$$

Exercise 1.13. Show that for any complex function $F(\omega) = f(\omega) + ig(\omega)$ with measurable f and g, $|F(\omega)|$ is integrable, if and only if $|f|$ and $|g|$ are integrable and we then have

$$\left| \int F(\omega)\, dP \right| \leq \int |F(\omega)|\, dP$$

A measurable space $(\Omega, \mathcal{B})$ is a set Ω together with a σ-field $\mathcal{B}$ of subsets of Ω.

Transformations

Definition 1.9. Given two measurable spaces $(\Omega_1, \mathcal{B}_1)$ and $(\Omega_2, \mathcal{B}_2)$, a mapping or a transformation from $F : \Omega_1 \to \Omega_2$, i.e. a function $\omega_2 = F(\omega_1)$ that assigns for each point $\omega_1 \in \Omega_1$ a point $\omega_2 = F(\omega_1) \in \Omega_2$, is said to be *measurable* if for every measurable set $A \in \mathcal{B}_2$, the inverse image

$$F^{-1}(A) = \{\omega_1 : F(\omega_1) \in A\} \in \mathcal{B}_1.$$

Exercise 1.14. Show that, in the above definition, it is enough to verify the property for $A \in \mathcal{A}$ where $\mathcal{A}$ is any class of sets that generates the σ-field $\mathcal{B}_2$.

If F is a measurable map from $(\Omega_1, \mathcal{B}_1)$ into $(\Omega_2, \mathcal{B}_2)$ and P is a probability measure on $(\Omega_1, \mathcal{B}_1)$, the *induced* probability measure Q on $(\Omega_2, \mathcal{B}_2)$ is defined by

$$Q(A) = P(F^{-1}(A)) \quad \text{for} \quad A \in \mathcal{B}_2 \tag{1.8}$$

Exercise 1.15. Verify that Q indeed does define a probability measure on $(\Omega_2, \mathcal{B}_2)$.

Q is called the induced measure and is denoted by $Q = PF^{-1}$.

Theorem 1.8. *If $f : \Omega_2 \to \mathbf{R}$ is a real valued measurable function on Ω_2, then $g(\omega_1) = f(F(\omega_1))$ is a measurable real valued function on $(\Omega_1, \mathcal{B}_1)$. Moreover g is integrable with respect to P if and only if f is integrable with respect to Q, and*

$$\int_{\Omega_2} f(\omega_2)\, dQ = \int_{\Omega_1} g(\omega_1)\, dP \tag{1.9}$$

Proof: If $f(\omega_2) = \mathbf{1}_A(\omega_2)$ is the indicator function of a set $A \in \mathcal{B}_2$, the claim in equation 1.9 is in fact the definition of measurability and the induced measure. We see that by linearity the claim extends easily from indicator functions to simple functions. By uniform limits, the claim can now be extended to bounded measurable functions. Monotone limits then extend it to nonnegative functions. By considering the positive and negative parts separately we are done.

A measurable function is just the generalization of the concept of a random variable introduced in section 2. We can either think of a random variable as special case of a measurable transformation where the target space is the real line. Or we can think of a measurable transformation as a random variable with values in an arbitrary target space. The induced measure $Q = PF^{-1}$ is called the *distribution* of the random variable F under P. In particular, if F takes real values, Q is a probability distribution on R.

Exercise 1.16. When F is real valued show that

$$\int F(\omega)dP = \int x \, dQ$$

When $F = (f_1, f_2, \cdots f_n)$ takes values in R^n, the induced distribution Q on R^n is called the *joint distribution* of the n random variables $f_1, f_2 \cdots, f_n$.

Exercise 1.17. If T_1 is a measurable map from $(\Omega_1, \mathcal{B}_1)$ into $(\Omega_1, \mathcal{B}_1)$ and T_2 is a measurable map from $(\Omega_2, \mathcal{B}_2)$ into $(\Omega_3, \mathcal{B}_3)$, then show that $T = T_2 \circ T_1$ is a measurable map from $(\Omega_1, \mathcal{B}_1)$ into $(\Omega_3, \mathcal{B}_3)$. If P is a probability measure on $(\Omega_1, \mathcal{B}_1)$, then on $(\Omega_3, \mathcal{B}_3)$, the two measures PT^{-1} and $(PT_1^{-1})T_2^{-1}$ are identical.

Product Spaces

Given two sets Ω_1 and Ω_2 the Cartesian product $\Omega = \Omega_1 \times \Omega_2$ is the set of pairs (ω_1, ω_2) with $\omega_1 \in \Omega_1$ and $\omega_2 \in \Omega_2$. If Ω_1 and Ω_2 come with σ-fields $\mathcal{B}_1$ and $\mathcal{B}_2$ respectively, we can define a natural σ-field $\mathcal{B}$ on Ω as the σ-field generated by sets (measurable rectangles) of the form $A_1 \times A_2$ with $A_1 \in \mathcal{B}_1$ and $A_2 \in \mathcal{B}_2$. This σ-field will be called the product σ-field.

Exercise 1.18. Show that sets that are finite disjoint unions of measurable rectangles constitute a field.

Given two probability measures P_1 and P_2 on $(\Omega_1, \mathcal{B}_1)$ and $(\Omega_2, \mathcal{B}_2)$ respectively we try to define on the product space $(\Omega, \mathcal{B})$ a probability measure P by defining for a measurable rectangle $A = A_1 \times A_2$

$$P(A_1 \times A_2) = P_1(A_1) \times P_2(A_2)$$

and extending it to the field $\mathcal{F}$ of finite disjoint unions of measurable rectangles as the obvious sum.

Exercise 1.19. If $E \in \mathcal{F}$ has two representations as disjoint finite unions of measurable rectangles

$$E = \cup_i \,(A_1^i \times A_2^i) = \cup_j \,(B_1^j \times B_2^j)$$

then

$$\sum_i P_1(A_1^i) \times P_2(A_2^i) = \sum_j P_1(B_1^j) \times P_2(B_2^j)$$

so that $P(E)$ is well defined. P is a finitely additive probability measure on $\mathcal{F}$.

Lemma 1.9. *The measure P is countably additive on the field $\mathcal{F}$.*

Proof: For any set $E \in \mathcal{F}$ let us define the section E_{ω_2} as

$$E_{\omega_2} = \{\omega_1 : (\omega_1, \omega_2) \in E\} \tag{1.10}$$

Then $P_1(E_{\omega_2})$ is a measurable function of ω_2 (is in fact a simple function) and

$$P(E) = \int_{\Omega_2} P_1(E_{\omega_2})\, dP_2 \tag{1.11}$$

Now let $E_n \in \mathcal{F} \downarrow \Phi$, the empty set. Then it is easy to verify that E_{n,ω_2} defined by

$$E_{n,\omega_2} = \{\omega_1 : (\omega_1, \omega_2) \in E_n\}$$

satisfies $E_{n,\omega_2} \downarrow \Phi$ for each $\omega_2 \in \Omega_2$. From the countable additivity of P_1 we conclude that $P_1(E_{n,\omega_2}) \to 0$ for each $\omega_2 \in \Omega_2$ and since, $0 \le P_1(E_{n,\omega_2}) \le 1$ for $n \ge 1$, it follows from equation 1.11 and the bounded convergence theorem that

$$P(E_n) = \int_{\Omega_2} P_1(E_{n,\omega_2})\, dP_2 \to 0$$

establishing the countable additivity of P on $\mathcal{F}$. We conclude that P extends uniquely as a countably additive measure to the σ-field $\mathcal{B}$ generated by $\mathcal{F}$. We will call this the *Product Measure* P.

Corollary 1.10. *For any $A \in \mathcal{B}$ if we denote by A_{ω_1} and A_{ω_2} the respective sections*

$$A_{\omega_1} = \{\omega_2 : (\omega_1, \omega_2) \in A\}$$

and

$$A_{\omega_2} = \{\omega_1 : (\omega_1, \omega_2) \in A\}$$

then the functions $P_1(A_{\omega_2})$ and $P_2(A_{\omega_1})$ are measurable and

$$P(A) = \int P_1(A_{\omega_2}) dP_2 = \int P_2(A_{\omega_1})\, dP_1.$$

In particular for a measurable set A, $P(A) = 0$ if and only if for almost all ω_1 with respect to P_1, the sections A_{ω_1} have measure 0 or equivalently for almost all ω_2 with respect to P_2, the sections A_{ω_2} have measure 0.

Proof: The assertion is clearly valid if A is rectangle of the form $A_1 \times A_2$ with $A_1 \in \mathcal{B}_1$ and $A_2 \in \mathcal{B}_2$. If $A \in \mathcal{F}$, then it is a finite disjoint union of such rectangles and the assertion is extended to such a set by simple addition. Clearly, the class of sets for which the assertion is valid is a monotone class and since it contains the field $\mathcal{F}$ it also contains the σ-field $\mathcal{B}$ generated by the field $\mathcal{F}$.

Warning. It is possible that a set A may not be *measurable* with respect to the product σ-field, but nevertheless the sections A_{ω_1} and A_{ω_2} are all measurable, $P_2(A_{\omega_1})$ and $P_1(A_{\omega_2})$ are measurable functions, but

$$\int P_1(A_{\omega_2}) dP_2 \neq \int P_2(A_{\omega_1})\, dP_1.$$

In fact there is a rather nasty example where $P_1(A_{\omega_2})$ ia identically 1 whereas $P_2(A_{\omega_1})$ is identically 0. The next result concerns the equality of the double integral, (i.e. the integral with respect to the product measure) and the repeated integrals in any order.

Theorem 1.11. (Fubini's Theorem). *Let $f(\omega) = f(\omega_1, \omega_2)$ be a measurable function of ω on $(\Omega, \mathcal{B})$. Then f can be considered as a function of ω_2 for each fixed ω_1 or the other way around. The functions $g_{\omega_1}(\cdot)$ and $h_{\omega_2}(\cdot)$ defined respectively on Ω_2 and Ω_1 by*

$$g_{\omega_1}(\omega_2) = h_{\omega_2}(\omega_1) = f(\omega_1, \omega_2)$$

are measurable for each ω_1 and ω_2. If f is integrable then the functions $g_{\omega_1}(\omega_2)$ and $h_{\omega_2}(\omega_1)$ are integrable for almost all ω_1 and ω_2 respectively. Their integrals

$$G(\omega_1) = \int_{\Omega_2} g_{\omega_1}(\omega_2)\, dP_2$$

and

$$H(\omega_2) = \int_{\Omega_1} h_{\omega_2}(\omega_1)\, dP_1$$

are measurable, finite almost everywhere and integrable with repect to P_1 and P_2 respectively. Finally

$$\int f(\omega_1, \omega_2)\, dP = \int G(\omega_1) dP_1 = \int H(\omega_2) dP_2$$

Conversely for a nonnegative measurable function f if either G are H, which are always measurable, has a finite integral so does the other and f is integrable with its integral being equal to either of the repeated integrals, namely integrals of G and H.

Proof: The proof follows the standard pattern. It is a restatement of the earlier corollary if f is the indicator function of a measurable set A. By linearity it is true for simple functions and by passing to uniform limits, it is true for bounded measurable functions f. By monotone limits it is true for nonnegative functions and finally by taking the positive and negative parts seperately it is true for any arbitrary integrable function f.

Warning. The following could happen. f is a measurable function that takes both positive and negative values that is not integrable. Both the repeated integrals exist and are unequal. The example is not hard.

Exercise 1.20. Construct a measurable function $f(x, y)$, on the product space $[0, 1] \times [0, 1]$ of two copies of the unit interval with Lebesgue measure, which is not integrable but the repeated integrals make sense and or unequal, i.e.

$$\int_0^1 dx \int_0^1 f(x, y)\, dy \neq \int_0^1 dy \int_0^1 f(x, y)\, dx$$

Distributions and Expectations

Let us recall that a triplet $(\Omega, \mathcal{B}, P)$ is a Probability Space if Ω is a set, $\mathcal{B}$ is a σ-field of subsets of Ω and P is a (countably additive) probability measure on $\mathcal{B}$. A random variable X is a real valued measurable function on $(\Omega, \mathcal{B})$. Given such a function X it induces a probability distribution α on the Borel subsets of the line $\alpha = PX^{-1}$. The distribution function $F(x)$ corresponding to α is obviously

$$F(x) = \alpha((-\infty, x]) = P[\,\omega : X(\omega) \leq x\,]$$

The measure α is called the distibution of α and $F(x)$ is called the distribution function of X. If g is a measurable function of the real variable x, then $Y(\omega) = g(X(\omega))$ is again a random variable and its distribution $\beta = PY^{-1}$ can be obtained as $\beta = \alpha\, g^{-1}$ from α. The *Expectation* of a random variable is defined if it is integrable and

$$E[X] = E^P[X] = \int X(\omega)\, dP$$

By the change of variables formula (Exercise 3.3) it can be obtained directly from α as

$$E[X] = \int x\, d\alpha$$

Here we are taking advantage of the fact that on the real line x is a very special real valued function. The value of the integral in this context is referred to as the expectation or mean of α. Of course it exists if and only if

$$\int |x|\, d\alpha < \infty$$

and

$$\left|\int x\, d\alpha\right| \leq \int |x|\, d\alpha$$

Similarly

$$E(g(X)) = \int g(X(\omega))\, dP = \int g(x)\, d\alpha$$

and anything concerning X can be calculated from α. The statement X is a random variable with distribution α has to be interpreted in the sense that somewhere in the background there is a Probability Space and a random variable X on it, which has α for its distribution. Usually only α matters and the underlying $(\Omega, \mathcal{B}, P)$ never emerges from the background and in a pinch we can always say Ω is the real line, $\mathcal{B}$ are the Borel sets , P is nothing but α and the random variable $X(x) = x$.

Some other related quantities are

$$Var(X) = \sigma^2(X) = E[X^2] - [E[X]]^2. \tag{1.12}$$

$Var(X)$ is called the variance of X.

Exercise 1.21. Show that if it is defined $Var(X)$ is always nonnegative. Moreover $Var(X) = 0$ if and only if for some value a, which is necessarily equal to $E[X]$, $P[X = a] = 1$.

Some what more generally we can consider a measurable mapping $X = (X_1, \cdots, X_n)$ of a probability space $(\Omega, \mathcal{B}, P)$ into R^n as a vector of n random variables $X_1(\emptyset), X_2(\emptyset), \cdots, X_n(\emptyset)$. These are caled random vectors or vector valued random variables and the induced distribution $\alpha = PX^{-1}$ on R^n is called the distribution of X or the *joint distribution* of $(X_1, \cdots, X_n)$. If we denote by π_i the coordinate maps $(x_1, \cdots, x_n) \to x_i$ from $R^n \to R$, then

$$\alpha_i = \alpha \pi_i^{-1} = PX_i^{-1}$$

are called the marginals of α.

The *Covariance* between two random variables X and Y is defined as

$$Cov(X, Y) = E[(X - E(X))(Y - E(Y))] = E[XY] - E[X]E[Y] \tag{1.13}$$

Exercise 1.22. If $X_1, \cdots, X_n$ are n random variables the matrix

$$C_{i,j} = Cov(X_i, X_j)$$

is called the *covariance matrix*. Show that it is a symmetric positive semi-definite matrix. Is every positive semi-definite the covariance matrix of some random vector?