REU PROJECT

Robert Prag

August 12, 2006

1 Preface

I hope that this document will be a coherent summary of elementary measure theory and statistics, as well as a rough treatment of stochastic processes.

That said, it was very interesting coming at this topic with a strong grasp on the analytic aspects and a minimal knowledge of statistics. Of course, this will not be much of an issue, as most of the proofs are borrowed directly from *Probability Theory* by S.R.S.Varadhan.

A copy of the text is available on the internet at:
`http://math.nyu.edu/faculty/varadhan/limittheorems.html`

All page numbers are from that book. In the not entirely unlikely event that an error should be suspected in this document, the corresponding section of that book should clarify the matter.

2 Some Measure Theory

The major idea here is to create a device by which we can measure sets in a fairly general context. Measure theory will allow us to do so. This is necessary for the follow probability theory, because we will be building our probability distributions into measure and then integrating random variables to get expectation values. We give a few important

Definitions:

- *σ-field* A class $\mathcal{B}$ such that:

  (i) The sets $\emptyset, \Omega$ are in $\mathcal{B}$,

  (ii) If $A \in \mathcal{B}$ then $A^c \in \mathcal{B}$,

  (iii) If $A_j \in \mathcal{B}, \forall j \in \mathbb{N}$ then $\cup_{j \in \mathbb{N}} A_j \in \mathcal{B}$ and $\cap_{j \in \mathbb{N}} A_j \in \mathcal{B}$.

- *Measure* Let $\mathcal{B}$ be a σ-field over Ω. The function $\mu : \mathcal{B} \to R_+$ is called a measure on $\mathcal{B}$ if:

(i) $\mu(\emptyset) = 0$,

(ii) If $\{A_j\}_{j\in\mathbb{N}} \subset \mathcal{B}, A_i \cap A_j = \emptyset, \forall i \neq j$ then

$$P(\cup_{j\in\mathbb{N}} A_j) = \Sigma_{j\in\mathbb{N}} P(A_j)$$

• *Probability Measure* A measure such that $P(\Omega) = 1$.

• *Convergence in measure* A sequence $(f_n)_{n\in\mathbb{N}}$ is said to converge in measure to the function f if

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

for all $\epsilon > 0$.

• *Almost everywhere convergence* A sequence $(f_n)_{n\in\mathbb{N}}$ is said to converge in measure to the function f if

$$f_n(\omega) \to f(\omega)$$

for all ω except for a set of measure zero.

We have the following relation between the last two types of convergence:

Lemma`(1.3, p.18)`

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

Proof. The fact that $f_n \to f$ outside the set N is equivalent to

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

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.□

On the other hand we have the following

Exercise`(p.18)`

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 $j \to \infty$.

Proof. From Lemma 1.3 (p. 18): $f_n \to f$ outside N is equivalent to

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

We know that $f_n \to f$ as $n \to \infty$ in measure, which means that

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

Note also that

$$\{\omega : |f_n(\omega) - f(\omega)| \geq \epsilon\} \subset \cup_{m\geq n}\{\omega : |f_m(\omega) - f(\omega)| \geq \epsilon\}$$

Define

$$E_j = \{\omega : |f_{n_j}(\omega) - f(\omega)| > 2^{-j}\}$$

where n_j is sufficiently large so that that

$$P(E_j) \leq 2^{-j}$$

If $\omega \notin \cap_{n\geq 0} \cup_{k\geq n} E_k$ then $\exists n_0 \in \mathbb{N}$ such that $\forall k > n_0, \omega \notin E_k$. Therefore $f_{n_j}(\omega) \to f(\omega)$.

Additionally, because $P(\{\omega : |f_{n_j}(\omega) - f(\omega)| > 2^{-j}\}) \leq 2^{-j}$, we know that $P(\cup_{k\geq n} E_k) \leq 2^{-n+1}$, which goes to zero as n goes to zero.

The above imply that $f_{n_k} \to f$ almost everywhere. QED

3 Convergence Theorems

The purpose of these theorems is allow us to go from a sequence of functions to a limit function, and ultimately to be able to be a little bit more free when integrating with respect to various probability measures. These will turn up in later proofs, so their inclusion here is necessary for completeness.

Theorem 1.4`(Bounded Convergence Theorem)`

If the sequence f_n of measurable functions is uniformly bounded and if $f_n \to f$ in measure as $n \to \infty$, then

$$lim_{n\to\infty} \int f_n dP = \int f dP.$$

Proof. Since

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

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

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

and since $\epsilon > 0$ is arbitrary we are done. See also p. 19 □

Theorem 1.5 `(Fatou's Lemma)`

If for each $n \geq 1$, $f_n \geq 0$ is measurable and $f_n \to f$ in measure as $n \to \infty$ then

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

Proof. Suppose g is bounded and satisfies $0 \leq g \leq f$. Then the sequence $h_n = f_n \wedge g = \min(f_n, g)$ is uniformly bounded and

$$h_n \to h = f \wedge g = g.$$

Therefore, by the bounded convergence theorem,

$$\int g dP = \lim_{n\to\infty} \int h_n dP.$$

Since $\int h_n dP \leq \int f_n dP$ for every n it follows that

$$\int g dP \leq \lim_{n\to\infty}\inf \int f_n dP.$$

As g satisfying $0 \leq g \leq f$ is arbitrary we are done. See also p. 20 □

Corollary 1.6 `(Monotone Convergence Theorem)`

If for a sequence f_n of nonnegative functions, we have $f_n \uparrow f$ monotonically then

$$\int f_n dP \to \int f dP \quad \textit{as } n \ \to \infty$$

Proof. Obviously $\int f_n dP \leq \int f dP$ and the other half follows from Fatou's lemma. See also p. 20. □

Corollary 1.11 *For any $A \in \mathcal{B}$ if we denote by $A_{\omega_1} and A_{\omega_2}$ 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 $A_1 \in B_1$ and $A_2 \in B_2$. If $A \in F$, then it is a finite disjoint union of such rectangles and the assertion is extended to such a set by simple addition. Clearly, by the monotone convergence theorem, the class of sets for which the assertion is valid is a monotone class and since it contains the field F it also contains the field $\mathcal{B}$ generated by the field $\mathcal{F}$. See also p. 26.□

Theorem 1.12`(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 $g_{\omega_1}()$ and $h_{\omega_2}()$ 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_1. 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 respect to P_1 and P_2 respectively. Finally

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

Conversely for a nonnegative measurable function f if either G or 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 separately it is true for any arbitrary integrable function f. See also p. 27 □

4 Distribution Functions

Here we introduce distribution functions. Also we are showing a very important property: the points of continutiy tell the whole story for a distribution function (which is good, considering that the discontinuities are at worst countably many).

Definition A *distribution function* f is a right-continuous non-decreasing function $f : R \to [0, 1]$ such that $\lim_{x\to-\infty} f(x) = 0$ and $\lim_{x\to\infty} f(x) = 1$. (See also p.13)

Exercise*Prove that if two distribution functions agree on the set of points at which they are both continuous, they agree everywhere.*

Proof. Call the distribution functions F and G. Because these functions are non-decreasing, they have denumerably many discontinuities. As an immediate consequence of this, if F is discontinuous at c and d, then $\exists x \in (c, d)$ such that F is continuous at x. As a result of this, F and G clearly must have exactly the same points of discontinuity, otherwise they would have to differ at a point of continuity. Any distribution function must have a right limit at all points. If $F(x) = G(x)$ for all $x \in \{x|F \text{ cts at } x\}$ then $\lim_{x\to c^+} F(x) = \lim_{x\to c^+} G(x)$. However, as previously stated, $\lim_{x\to c^+} F(x)$ exists for all c. Therefore $F(x) = G(x)$ for all $x \in \mathbb{R} \Rightarrow F = G$.□

5 Weak Law of Large Numbers

A rough paraphrasing of the law of large numbers would be that reality will eventually start to act like our statistical models predict, as long as our models were right to begin with.

Definition If α is a probability distribution on the line, its *characteristic function* is defined by

$$\phi(t) = \int \exp[itx]d\alpha$$

Theorem 3.3`(Weak Law of Large Numbers)`

If $X_1, X_2, \ldots, X_n$ are independent and identically distributed with a finite first moment and $E(X_i) = m$, then $\frac{X_1+X_2+\ldots+X_n}{n}$ converges to m in probability as $n \to \infty$. *(See also p. 56)*

Proof. We can use characteristic functions. If we denote the characteristic function of X_i by $\phi(t)$, then the characteristic function of $\frac{1}{n}\Sigma_{1\leq i\leq n}X_i$ is given by $\psi_n(t) = [\phi(\frac{t}{n})]^n$. The existence of the first moment assures us that $\phi(t)$ is differentiable at $t = 0$ with a derivative equal to im where $m = E(X_i)$.

Therefore by Taylor expansion

$$\phi(\frac{t}{n}) = 1 + \frac{imt}{n} + o(\frac{1}{n}).$$

Whenever $na_n \to z$ it follows that $(1 + a_n)^n \to e^z$. Therefore,

$$\lim_{n\to\infty} \psi_n(t) = exp[imt]$$

which is the characteristic function of the distribution degenerate at m. Hence the distribution of $\frac{S_n}{n}$ tends to the degenerate distribution at the point m. The weak law of large numbers is thereby established. (See also p. 57) □

6 Central Limit Theorem

We present The Central Limit Theorem, from which the importance of the bell-curve becomes visible. This is especially useful in trying to pick models to which reality is likely to conform.

Theorem 3.17`(The central limit theorem)`

Let $\{X_1, X_2, \ldots, X_n, \ldots\}$ *be a sequence of independent, identically distributed random variables of mean zero and finite variance* $\sigma^2 > 0$. *The distribution of* $\frac{S_n}{\sqrt{n}}$ *converges as* $n \to \infty$ *to the normal distribution with density*

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp[-\frac{x^2}{2\sigma^2}]$$

Proof. If we denote by $\phi(t)$ the characteristic function of any X_i then the characteristic function of $\frac{S_n}{\sqrt{n}}$ is given by

$$\psi_n(t) = [\phi(\frac{t}{\sqrt{n}})]^n$$

We can use the expansion

$$\phi(t) = 1 - \frac{\sigma^2 t^2}{2} + o(t^2)$$

to conclude that

$$\phi(\frac{t}{\sqrt{n}}) = 1 - \frac{\sigma^2 t^2}{2n} + o(\frac{1}{n})$$

and it then follows that

$$\lim_{n\to\infty} \psi_n(t) = \psi(t) = \exp[-\frac{\sigma^2 t^2}{2}]$$

Since $\psi(t)$ is the characteristic function of the normal distribution with density $p(x)$ given in the statement of the theorem, we are done. (See also p. 70)□

7 Kolmogorov's Zero-One Law

Kolmogorov's Zero-One Law states that if an event depends only on the tail-behavior of a sequence of random variables then the event will occur with either probability zero or probability one. This is fascinating, although there aren't too many interesting events that depends only on tail-behavior, since they happen either essentially always or basically never.

Definitions

• *Random variable (measurable function)* A random variable or measurable function is a map $f : \Omega \to R$, i.e. a real valued function $f(\omega)$ on Ω such that for every Borel set $B \subset R$, $f^{-1}(B) = \{\omega \in \Omega : f(\omega) \in B\}$ is a measurable subset of Ω. (p. 14)

• *Independent* The events A and B are independent if $P(A \cap B) = P(A)P(B)$ (p. 51)

• *Product measure* If P_1 is a measure on Ω_1 and P_2 is a measure on Ω_2 then we may define a measure P_3 on rectangles in $\Omega_1 \times \Omega_2$ by

$$(\forall A_1 \in \Omega_1), (\forall A_2 \in \Omega_2), P_3(A_1 \times A_2) = P_1(A_1)P_2(A_2)$$

and then extend P_3 in a natural way to the σ-field generated by these rectangles. The measure thus obtained is called the product measure. (See also p. 14)

• *Monotone class* a class that is closed under monotone limits of an increasing or decreasing sequence of sets. (p. 9)

• *Riemann-Stieltjes Integral* Limit as $N \to \infty$ of

$$\Sigma_{j=0}^{N} g(x_j)[F(a_{j+1}^N) - F(a_j^N)]$$

where $-\infty < a_0^N < a_1^N < \cdots < a_N^N < a_{N+1}^N < \infty$ is a partition of the finite interval $[a_0^N, a_{N+1}^N]$ with the limit taken in such a way that $a_0^N \to -\infty, a_{N+1}^N \to \infty$, and the oscillation of g in any and $[a_j^N, a_{j+1}^N]$ goes to 0. In the above g is bounded and continuous and F is a distribution function.

Theorem 3.15(Kolmogorov's Zero-One Law)

If $A \in \mathcal{B}^\infty$ and P is any product measure (not necessarily with identical components) then either $P(A) = 0$ or $P(A) = 1$. In here $\mathcal{B}$ denotes the tail σ-field which contains only events that depend on the tail behavior of the sequence of random variables (See also p. 49)

Proof. The proof depends on showing that A is independent of itself so that

$$P(A) = P(A \cap A) = P(A)P(A) = [P(A)]^2$$

and therefore equals 0 or 1. The proof is elementary. Since $A \in \mathcal{B}^\infty \subset \mathcal{B}^{n+1}$ and P is a product measure, A is independent of $\mathcal{B}_n = \sigma$-field generated by $\{x_j : 1 \leq j \leq n\}$. It is therefore independent of sets in the field $\mathcal{F} = \cup_n B_n$. The class of sets $\mathcal{A}$ that are independent of A is a monotone class. Since it contains the field $\mathcal{F}$ it contains the σ-field $\mathcal{B}$ generated by $\mathcal{F}$. In particular since $A \in \mathcal{B}$, A is independent of itself. (p. 70)$\square$

8 Hahn-Jordan Decomposition

Here we define a signed measure, and then decide that this it is not something we like working with and then prove that it is actually just two normal (unsigned) measures working against each-other. It might have been possible to define two measures working against each other, but what would be the fun in that?

Definition A countably additive *signed measure* is a countably additive measure which allows for sets to have negative measure.

Theorem 4.3(Hahn-Jordan Decomposition)

Given a countably additive signed measure λ on $(\Omega, \mathcal{F})$ it can be written always as

$$\lambda = \mu^+ - \mu^-$$

the difference of two nonnegative measures. Moreover μ^+ and μ^- may be chosen to be orthogonal i.e, there are disjoint sets $\Omega_+, \Omega_- \in \mathcal{F}$ such that $\mu^+(\Omega_-) = \mu^-(\Omega_+) = 0$. In fact Ω_+ and Ω_- can be taken to be subsets of Ω that are respectively totally positive and totally negative for λ. $\mu^\pm$ then become just the restrictions of λ to $\Omega_\pm$. (p. 104)

Proof. Totally positive sets are closed under countable unions, disjoint or not. Let us denote

$$m^+ = sup_A \lambda(A).$$

If $m^+ = 0$ then $\lambda(A) \leq 0$ for all A and we can take $\Omega_+ = \Phi$ and $\Omega_- = \Omega$ which works. Assume that $m^+ > 0$. There exist sets A_n with $\lambda(A) \geq m^+ - \frac{1}{n}$ and therefore totally positive subsets $\overline{A_n}$ of A_n with $\lambda(\overline{A_n}) \geq m^+ - \frac{1}{n}$. Clearly $\Omega_+ = \cup_n A_n$ is totally positive and $\lambda(\Omega_+) = m^+$. It is easy to see that $\Omega = \Omega - \Omega_+$ is totally negative. $\mu^\pm$ can be taken to be the restriction of λ to $\Omega_\pm$. (See also p. 105)□

9 Jensen's Inequality

Jensen's inequality describes the intuitive* and useful behavior of convex functions. That it also holds for expectation values is impressive but not entirely surprising. (*: depending on exactly what inclinations one's intuitions have)

Theorem `(Jensen's Inequality) p.80`

If $\phi(x)$ is a convex function of x, and $g = \mathbb{E}\{f|\Sigma\}$ (see p. 79 for the definition) then

$$\mathbb{E}\{\phi(f(\omega))|\Sigma\} \geq \phi(g(\omega)) \ a.e.$$

and if we take expectations

$$\mathbb{E}[\phi(f)] \geq \mathbb{E}[\phi(g)]$$

(See also p. 80)

Proof. We note that if $f_1 \geq f_2$ then $\mathbb{E}\{f_1|\Sigma\} \geq \mathbb{E}\{f_2|\Sigma\} a.e.$ and consequently $\mathbb{E}\{max(f_1, f_2)|\Sigma\} \geq max(g_1, g_2) a.e.$ where $g_i = \mathbb{E}\{f_i|\Sigma\}$ for $i = 1, 2$. Since we can represent any convex function ϕ as $\phi(x) = \sup_a[ax - \psi(a)]$, limiting ourselves to rational a, we have only a countable set of functions to deal with, and

$$\mathbb{E}\{\phi(f)|\Sigma\} = \mathbb{E}\{\sup_a[af - \psi(a)]|\Sigma\}$$

$$\geq \sup_a[a\mathbb{E}\{f|\Sigma\} - \psi(a)]$$

$$= \sup_a[ag - \psi(a)]$$

$$= \phi(g)$$

a.e. and after taking expectations

$$\mathbb{E}[\phi(f)] \geq \mathbb{E}[\phi(g)].$$