NOTES AND EXERCISES, LECTURE 1, MATH 642:592, Spring 2008

1. *Probability space, random variables and expectation.*

We summarize the formal mathematical setting of the course:

(1) All analysis takes place in a probability space. This is a triple $(\Omega, \mathcal{F}, \mathbb{P})$, in which Ω is a non-empty set, $\mathcal{F}$ is a σ-algebra of subsets of Ω called *events*, and $\mathbb{P}$ is a probability measure on $(\Omega, \mathcal{F})$. Recall that a σ-algebra is a collection of sets closed under the operations of taking complements, of countable unions, and of countable intersections. A probability measure is a countably additive (non-negative) measure for which $\mathbb{P}(\Omega) = 1$. In this course, $(\Omega, \mathcal{F}, \mathbb{P})$ is usually the notation for a generic probability space.

(2) σ-algebras. In probability theory, events come bundled naturally as packages of σ-algebras. If $\mathcal{C}$ is some collection of subsets of a set Ω, $\sigma\{C\}$, called the σ-algebra generated by $\mathcal{C}$, is the smallest σ-algebra containing $\mathcal{C}$, or equivalently, the intersection of all σ-algebras of subsets of Ω containing $\mathcal{C}$. A particularly important type is a Borel σ-algebra. If $\mathcal{S}$ is a topological space, $\mathcal{B}(\mathcal{S})$ is the σ-algebra generated by the open sets of $\mathcal{S}$ and it is called the Borel σ-algebra of $\mathcal{S}$. When we use the term Borel sets without specifying $\mathcal{S}$, we generally mean $\mathcal{B}(\mathbb{R})$ or $\mathcal{B}(\mathbb{R}^n)$, whichever is appropriate.

(3) *Random Variables.* A random variable is a Borel measurable map from $(\Omega, \mathcal{F})$ to the extended reals. This means that for every Borel subset of $I\!R$, $X^{-1}(U) = \{\omega\,;\, X(\omega) \in U\}$ belongs to $\mathcal{F}$. Hence if $\mathbb{P}$ is a probability measure on $(\Omega, \mathcal{F})$ we can assign a probability to the event that X takes values in U, for every Borel U. The σ-algebra generated by X is the inverse image of the Borel sets of $\mathbb{R}$ under X:

$$\sigma\{X\} \triangleq X^{-1}\left(\mathcal{B}(\mathbb{R})\right) \triangleq \left\{X^{-1}(U)\,;\, U \in \mathcal{B}(\mathbb{R})\right\}.$$

(The symbol "$\triangleq$" will be used for equalities which are definitions.

When a probability measure $\mathbb{P}$ is given the cumulative distribution function of a random variable X is $F_X(x) = \mathbb{P}\left(X^{-1}((-\infty, x])\right)$. We write this last expression less pedantically as $\mathbb{P}(X \leq x)$.

(4) *Expectation* The expected value (mean value) of X is $\mathbb{E}\left[X\right] \triangleq \int_\Omega X(\omega)\, d\mathbb{P}$, where the integral is in the sense defined in general measure and integration theory. To review briefly: define the indicator function of a subset A of Ω, by

$$\mathbf{1}_A(\omega) \triangleq \begin{cases} 1 & \text{if } \omega \in A; \\ 0 & \text{if not.} \end{cases}$$

We shall use this notation throughout the course. When $A \in \mathcal{F}$, $\mathbb{P}(A)$ makes sense and we define

$$\mathbb{E}\left[\mathbf{1}_A\right] = \int_\Omega \mathbf{1}_A(\omega)\,\mathrm{d}\mathbb{P}(\omega) \triangleq \mathbb{P}(A).$$

A simple random variable is a random variable of the form $\sum_1^n c_i \mathbf{1}_{A_i}(\omega)$, where $A_1, \ldots, A_n$ are in $\mathcal{F}$; they are just random variables taking values in a finite set. The integral (expectation) of a simple random variable is defined to be

$$\mathbb{E}\left[\sum_1^n c_i \mathbf{1}_{A_i}\right] \triangleq \sum_1^n c_i \mathbb{P}(A_i).$$

When the c_i's are distinct and the A_i's are disjoint, $A_i = \{X = c_i\}$ and so $\mathbb{E}\left[X\right] = \sum c_i \mathbb{P}(X = c_i)$, which is the definition of expected value for discrete random variables in elementary probability theory. For any nonnegative measurable random variable X, the integral is defined as

$$\mathbb{E}\left[X\right] = \int X \mathrm{d}\mathbb{P} \triangleq \sup\left\{\mathbb{E}\left[Y\right] ; Y \text{ is simple, } Y \leq X.\right\}$$

For a general X, $\mathbb{E}\left[X\right] = \mathbb{E}\left[X^+\right] - \mathbb{E}\left[X^-\right]$, where X^+ and X^- are the positive and negative parts of X and it is assumed at most one of $\mathbb{E}\left[X^+\right]$ and $\mathbb{E}\left[X^-\right]$ is infinite.

We assume familiarity with the big three theorems on limits and integrals: the monotone convergence theorem, the dominated convergence theorem, and Fatou's lemma.

A random variable X for which $\mathbb{E}\left[|X|\right] < \infty$ is said to be *integrable*; sometimes we write instead $X \in L^1(\mathbb{P})$. Similarly, $X \in L^p(\mathbb{P})$ means $\mathbb{E}\left[|X|^p\right] < \infty$. X is *square integrable* if $\mathbb{E}\left[X^2\right] < \infty$.

It is a theorem that

$$\mathbb{E}\left[g(X)\right] = \int_{-\infty}^{\infty} g(x) \mathrm{d}F_X(x),$$

whenever the integral and the expectation make sense, even if the exectation is ∞ or $-\infty$.

Since, for postive M, $\mathbf{1}_{\{|X| \geq M\}} \leq (|X|/M)\mathbf{1}_{\{|X| \geq M\}} \leq |X|/M$, we get the very useful Markov's inequality,

$$\mathbb{P}\left(|X| \geq M\right) = \mathbb{E}\left[\mathbf{1}_{\{|X| \geq M\}}\right] \leq \frac{\mathbb{E}\left[|X|\right]}{M}$$

References. Rogers and Williams, Volume I, section II.1) (Rogers and Williams will hereafter be abbreviated, RW, or R and W.) Also Chapter 1 of probability theory notes available at `http://www.rutgers.edu/ ocone/problect.html`. (This reference is hereafter abbreviated by O.)

2. *Complete probability spaces and completion.* Technical issues often require completeness. A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is *complete* if whenever $B \subset A$ and $A \in \mathcal{F}$, it follows that $B \in \mathcal{F}$ also. If a probability space is not complete it can alway be extended to a complete space. Given an arbitrary $(\Omega, \mathcal{F}, \mathbb{P})$, define

$$\mathcal{N}^{\mathbb{P}} \triangleq \{B \subset \Omega\,;\, B \subset A, A \in \mathcal{F},\ \mathbb{P}(A)=0\}\,.$$

Exercise 1. a) Let $\mathcal{F}^{\mathbb{P}} \triangleq \Big\{B \subset \Omega\,;\,\text{there exist } A \in \mathcal{F},\ N \in \mathcal{N}^{\mathbb{P}} \text{ s.t. } B = A \cup N\Big\}$.

For $B = A \cup N$, where $A \in \mathcal{F}$ and $N \in \mathcal{N}^{\mathbb{P}}$, define $\bar{\mathbb{P}}(B) = \mathbb{P}(A)$. Show this definition of $\bar{\mathbb{P}}$ is consistent and $(\Omega, \mathcal{F}^{\mathbb{P}}, \bar{\mathbb{P}})$ is a complete extension of $(\Omega, \mathcal{F}, \mathbb{P})$.

b) R and W (page 94) define completion in a different way. For any $B \in \Omega$ they define inner and outer measures:

$$\mathbb{P}^*(B) \triangleq \inf\{\mathbb{P}(A)\,;\, B \subset A,\ A \in \mathcal{F}\} \quad \textit{and} \quad \mathbb{P}_*(B) \triangleq \sup\left(\mathbb{P}(A)\,;\, A \subset B,\ A \in \mathcal{F}\right\}.$$

Then they define $\tilde{\mathcal{F}}$ to be the σ-algebra of subsets B such that $\mathbb{P}_*(B) = \mathbb{P}^*(B)$, and $\mathbb{P}(\tilde{B}) = \mathbb{P}^*(B)$ for $B \in \tilde{\mathcal{F}}$. Show that

$$\tilde{\mathcal{F}} = \{B \subset \Omega\,;\,\text{there are } A_0, A_1 \text{ in } \mathcal{F} \text{ with } A_0 \subset B \subset A_1 \text{ and } \mathbb{P}(A_1 - A_0) = 0\}\,,$$

and $(\Omega, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ defines the same complete space as the construction in (a).

3. *Uniform integrability.* (References: RW, pp. 113-116; O, Chapter 6, pp 36-40.)

We start off with a simple looking inequality:

$$\mathbb{E}\left[\mathbf{1}_A |X|\right] = \mathbb{E}\left[\mathbf{1}_A |X| \mathbf{1}_{\{|X|<M\}}\right] + \mathbb{E}\left[\mathbf{1}_A |X| \mathbf{1}_{\{|X|\geq M\}}\right] \leq M\mathbb{P}(A) + \mathbb{E}\left[|X| \mathbf{1}_{\{|X|\geq M\}}\right] \tag{1}$$

The first term follows simply because $|X|\mathbf{1}_{\{|X|<M\}} < M$, the second because $\mathbf{1}_A \leq 1$ Uniform integrability is about bounding the second term uniformly in a family of random variables.

Let $\mathcal{X}$ be a family of random variables on a probability space. $\mathcal{X}$ is said to be uniformly integrable (UI) if

$$\lim_{M\to\infty} \sup_{X\in\mathcal{X}} \mathbb{E}\left[|X|\mathbf{1}_{\{|X|\geq M\}}\right] = 0. \tag{2}$$

A simple argument (exercise) shows: if $\mathcal{X}$ is UI, then $\sup_{X\in\mathcal{X}} \mathbb{E}\left[|X|\right] < \infty$.

Theorem 1 *A family $\mathcal{X}$ is uniformly integrable if and only if*

(i) $\sup_{X\in\mathcal{X}} \mathbb{E}\left[|X|\right] < \infty$*;*

(ii) *For every $\epsilon > 0$ there exists a $\delta > 0$ such that $\mathbb{P}(A) < \delta$ implies $\mathbb{E}\left[\mathbf{1}_A|X|\right] < \epsilon$ for all $X \in \mathcal{X}$. Equivalently,*

$$\lim_{\delta\downarrow 0} \sup_{I\!P(A)\leq\delta} \mathbb{E}\left[\mathbf{1}_A|X|\right] = 0 \tag{3}$$

Proof: Assume $\mathcal{X}$ is UI. We have already stated that (i) is true. To prove (ii), let $\epsilon > 0$ and then take M so that $\sup_{X\in\mathcal{X}} \mathbb{E}\left[|X|\mathbf{1}_{|X|\geq M}\right] < \epsilon/2$. Then choose $\delta = \epsilon/(2M)$. By the inequality (1), for every $X \in \mathcal{X}$, $\mathbb{E}\left[\mathbf{1}_A|X|\right] < \epsilon$.

Exercise 2. Prove that (i) and (ii) imply uniform integrability. (Hint: use Markov's inequality, $\mathbb{P}(|X| \geq M) \leq M^{-1}\mathbb{E}\left[|X|\right]$.)

How can one check uniform integrability? Here is another equivalent condition for uniform integrability.

Theorem 2 *A family $\mathcal{X}$ is uniformly integrable if and only if there exists a nonnegative function ϕ on $[0,\infty)$ such that $\lim_{x\to\infty} x/\phi(x) = 0$ and $\sup_{X\in\mathcal{X}} \mathbb{E}\left[\phi(|X|)\right]$.*

Proof: (if) There is an $\alpha > \infty$ so large that $\phi(x) > 0$ for all $x \geq \alpha$. Thus, for $M \geq \alpha$,

$$|x|\mathbf{1}_{\{|x|\geq M\}} = \phi(|x|)\frac{|x|}{\phi(|x|)}\mathbf{1}_{|x|\geq M} \leq \phi(|x|) \sup_{x\geq M} \frac{x}{\phi(x)}.$$

Therefore $\mathbb{E}\left[|X|\mathbf{1}_{\{|X|\geq M\}}\right] = \sup_{x\geq M} \frac{x}{\phi(x)} \mathbb{E}\left[\phi(|X|)\right]$. Since $\lim_{M\to\infty} \sup_{x\geq M} x/\phi(x) = 0$ and $\sup_{X\in\mathcal{X}} \mathbb{E}\left[\phi(|X|)\right] < \infty$, the uniform integrability of $\mathcal{X}$ follow.

Exercise 3. Assume that $\mathcal{X}$ is uniformly integrable. Show that the following construction produces a convex increasing function ϕ such that $\lim_{x\to\infty} x/\phi(x) = 0$ and $\sup_{X\in\mathcal{X}} \mathbb{E}\left[\phi(|X|)\right] < \infty$.

Set $\phi(0) = 0$. Choose an increasing sequence M_k such that

$$\sup_{X\in\mathcal{X}} \sum_{k=1}^{\infty} k\mathbb{E}\left[|X|\mathbf{1}_{\{|X|\geq M_k\}}\right] < \infty,$$

and define ϕ so that

$$\phi'(x) = \sum_k \left(k - \frac{N_{k+1} - x}{N_{k+1} - N_k}\right)\mathbf{1}_{N_k\leq x<N_{k+1}}.$$

As a consequence of this Theorem, $\mathcal{X}$ is UI if it is $L^p(\mathbb{P})$ bounded for some $p > 1$, that is, if

$$\sup_{X \in \mathcal{X}} \mathbb{E}\left[|X|^p\right] < \infty.$$

This criterion is commonly used for deducing UI.

One thing uniform integrability is good for is that it allows us to interchange limits and expectations without the need to use the Dominated Convergence Theorem. Recall that $X_n \to X$ in probability as $n \to \infty$ means $\lim_{n\to\infty} \mathbb{P}\left(|X_n - X| > \epsilon\right) = 0$ for every $\epsilon > 0$.

Theorem 3 *Suppose that X, and $X_1, X_2, \ldots$ are integrable. Then* $\lim_{n\to\infty} \mathbb{E}\left[|X_n - X|\right] = 0$ *if and only if $X_n \to X$ in probability and $\{X_n\}$ is uniformly integrable.*

Proof: (if). Suppose $\{X_n\}$ is UI and $X_n \to X$ in probability. Then one can check that uniform integrability of $\{X_n\}$ implies that of $\{X_n - X\}$. Fix an arbitrary $\epsilon > 0$. Clearly

$$\begin{aligned} \mathbb{E}\left[|X_n - X|\right] &= \mathbb{E}\left[|X_n - X|\mathbf{1}_{\{|X_n - X| < \epsilon\}}\right] + \mathbb{E}\left[|X_n - X|\mathbf{1}_{\{|X_n - X| \geq \epsilon\}}\right] \\ &\leq \epsilon + \mathbb{E}\left[|X_n - X|\mathbf{1}_{\{|X_n - X| \geq \epsilon\}}\right]. \end{aligned}$$

Since $\lim_{n\to\infty} \mathbb{P}\left(|X_n - X| \geq \epsilon\right) = 0$, by assumption, Theorem 1) says that from uniform integrability of $\{X_n - X\}$, we can conclude that $\mathbb{E}\left[|X_n - X|\mathbf{1}_{\{|X_n - X| \geq \epsilon\}}\right] \to 0$ as $n \to \infty$. Therefore, from the last inequality,

$$\limsup_{n\to\infty} \mathbb{E}\left[|X_n - X|\right] \leq \epsilon.$$

This being true for every $\epsilon > 0$, take $\epsilon \downarrow 0$ to derive $\lim_{n\to\infty} \mathbb{E}\left[|X_n - X|\right] = 0$.

Exercise 4. a) Show that if $\{X_n\}$ is UI and $\mathbb{E}\left[|X|\right] < \infty$, then $\{X_n - X\}$ is UI. (One could as well prove $\{X_n + X\}$ is UI and then apply the result with $-X$ in place of X to get UI of $\{X_n - X\}$.)

b) Prove the "only if" direction in the theorem above.

Exercise 5. Recall that $X_n \to X$ in distribution if $\lim_{n\to\infty} \mathbb{E}\left[h(X_n)\right] = \mathbb{E}\left[h(X)\right]$ for every bounded continuous funtion h. We also write convergence in distribution as $X_n \Longrightarrow X$. Show that if $X_n \Longrightarrow X$ and $\{X_n\}$ is UI, then $\mathbb{E}\left[X_n\right] \to \mathbb{E}\left[X\right]$. (Hint: Consider $h_M(x) = \max\{|x|, M\}$ and write $|x| = |x| - h_M(x) + h_M(x)$. Conversely, show that $X_n|LongrightarrowX$ and $\lim_{n\to\infty} \mathbb{E}\left[|X_n|\right] = E|X|$ imply that $\{X_n\}$ is uniformly integrable. (Source, Ethier and Kurtz, p. 494, *Markov Processes: Characterization and Convergence*, Wiley.)

4. *Signed measures and the Radon-Nikodym theorem.* (Reference: RW, Volume I, pp 98-99.)

Fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let X be an integrable random variable. Define on $\mathcal{F}$ the function

$$\lambda_X(A) \stackrel{\triangle}{=} \mathbb{E}\left[\mathbf{1}_A X\right].$$

One can verify that (i) $\lambda_X(\emptyset) = 0$; (ii) $\sup_{A\in\mathcal{F}} |\lambda_X(A)| < \infty$ (because $|\lambda_X(A)| \leq \mathbb{E}\left[\mathbf{1}_A|X|\right] \leq \mathbb{E}\left[|X|\right] < \infty$ for all A; and (iii) If $A_1, A_2, \ldots$ are disjoint events $\lambda_X(\cup A_i) = \sum_i \lambda(A_i)$, (this follows by use of the Dominated Convergence Theorem). Thus λ_X is countably additive, but it could take on negative values.

In general, a measure λ on $(\Omega, \mathcal{F})$, possibly taking negative as well as positive values, and satisfying conditions (i), (ii), and (iii) (with λ_X being replaced by λ) is called a bounded, signed measure. The previous paragraph thus says that for any integrable random variable, λ_X is a bounded, signed measure.

One can also define, unbounded signed measures; an unbounded signed measure may take on the value $+\infty$ on a set or $-\infty$, but it may not take $+\infty$ on one set and $-\infty$ on another.

A signed measure λ is absolutely continuous with respect to $\mathbb{P}$ ($\lambda \ll \mathbb{P}$) if $\mathbb{P}(A) = 0$ implies $\lambda(A) = 0$. Clearly, if X is an integrable random variable, then $\lambda_X \ll \mathbb{P}$. The classic Radon-Nikodym theorem of real analysis implies the converse:

Theorem 4 *If λ is a bounded, signed measure and λ is absolutely continuous with respect to $\mathbb{P}$, then there is an integrable random variable X such that $\lambda = \lambda_X$.*

We omit the proof. We only note that it suffices to treat the case when λ is actually a positive measure, because the Jordan-Hahn decomposition says a signed measure can be expressed as a difference of positive measures. If λ is positive, one considers the set $\mathcal{Y}$ of nonnegative random variables Z such that $\lambda_Z(A) \leq \lambda(A)$ for all $A \in \mathcal{F}$. This set has the property that if Z and Y are both members, so is $\max(Z, Y)$. One can then construct an increasing sequence $\{Z_n\}$ of elements of $\mathcal{Y}$ such that $\mathbb{E}\left[Z_n\right] \uparrow \sup_{Y\in\mathcal{Y}} \mathbb{E}\left[Y\right]$. Then $\lim_{n\to\infty} Z_n$ exists and defines the sought for X. Another popular proof uses the Riesz representation theorem.

Remark: It may be shown that if λ is bounded, then absolute continuity of $\mathbb{P}$ is equivalent to:

$$\forall \;\; \epsilon > 0, \exists \;\; \delta > 0, \text{ such that } \mathbb{P}(A) < \delta \text{ implies } |\lambda(A)| < \epsilon. \tag{4}$$

Observe that condition (ii) of Theorem 1 may be rephrased as saying that $\{\lambda_X\,;\, X \in \mathcal{X}\}$ is a *uniformly* absolutely continuous family of signed measures, in the sense of (4). Thus Theorem 2 implies the equivalence of UI with $L^1(P)$ boundedness of $\mathcal{X}$ plus uniform absolute continuity of $\{\lambda_X\,;\, X \in \mathcal{X}\}$.

5. *Dunford-Pettis theorem.* We will state a fundamental theoretical result that will be important in the proof of the Doob-Meyer decomposition. It is certainly a result worth knowing, and it is not hard to understand the statement. The proof is involved and so students should consider the proof optional. For those that are interested, a detailed exposition of one direction of the proof may be found in another document available from the course web site.

The result, called the Dunford-Pettis compactness criterion, implies that uniform integrability is a necessary and sufficient condition for weak sequential compactness of a family of integrable random variables. The theorem applies to measure spaces more general than probability spaces, but we shall only state its application to probability spaces.

In the discussion, $(\Omega, \mathcal{F}, \mathbb{P})$ stands for a fixed probability space and L^1 for the space of integrable ($E[|X|] < \infty$) random variables on . Let $\{X_n\}$ be a sequence of random variables in L^1. Let X be in L^1, also. We say that $\{X_n\}$ converges weakly to X ($X_n \to X$ (weakly)) if

$$\lim_{n\to\infty} E\left[\xi X_n\right] = E\left[\xi X\right] \qquad \text{for every bounded random variable } \xi.$$

A family $\mathcal{X}$ of random variables in L^1 is said to be *(relatively) weakly sequentially compact* if any sequence contained in $\mathcal{X}$ contains a weakly convergent subsequence.

Theorem 5 *A family $\mathcal{X}$ is weakly sequentially compact if and only if it is uniformly integrable.*

Remark: Recall from real analysis that the dual of L^1 is the set of all linear functionals ℓ on L^1 that are continuous in the sense that $E|X_n - X| \to 0$ implies $\ell(X_n \to \ell(X)$. This dual is isomorphic to the L^∞, the space of bounded random variables, by the identification of $\xi \in L^\infty$ with the linear functional $\ell_\xi(X) = E[\xi X]$. The weak topoolgy on L^1 is the smallest topology making all such linear functions continuous. So weak sequential compactness is a property concerning the weak topology.

This theorem is important because in analysis one is often trying to construct a function (here, random variable) by constructing a sequence of what one hopes are better and better approximations. The Dunford-Pettis criterion allows us to extract a weak limit from the conjectured approximants, if we can prove they are uniform integrable, and this weak limit may give us what we are looking for. This will happen in the case of the Doob-Meyer approximation.