STA 711: Probability & Measure Theory

Robert L. Wolpert

Housekeeping Details

- Introduction
- Lec: Mon/Wed 1:25–2:40pm, in 311 Soc Sci (off stairwell)
- My OH: Mon 4:00–5:00pm, in 211c Old Chem
- TA OH: Tue 4–6pm, in 211a Old Chem
- Web: https://stat.duke.edu/courses/Fall15/sta711/
  Some lectures & midterms will probably change, due to confs & trips.
- HW: Approx 6–10 probs/week; expect to spend 5–10 hrs a week on homework.
  Due each Wed starting 2015-09-02 (9 days from today); returned following Mon.
  BE NEAT. Consider **LaTeX**. Collaborating is encouraged but **DON'T COPY**.
  Seriously, you won't pass the exams if you don't write up your own homework solutions.
- Text: Comments welcome. Other texts listed on class web page.
- Suggested work-flow: Read the chapter, do the problems, talk about them with each other, ask questions in class or office-hours. Solve problems on scratch paper; write up clear concise solution to turn in. Homework and exam scores are based on your success in **communicating a correct answer**. A correct but obscure solution will lose points.
- My role is not to spoon-feed you the book, but rather to add perspective, illustrate and illuminate ideas, offer examples, and help show how the ideas and tools are useful in the theory and application of (especially Bayesian) statistics. In particular, you may need to learn things covered in the book but not in lecture and vice versa.

0 Prologue

This is *mostly* a course about random variables— how to find probabilities they take particular values, or values in certain ranges; how to find their "expectations" (whatever that means), and especially how to find properties of the limits of sequences of random variables, and just what it means for sequences to **have** limits. That turns out to be a very interesting question, with several different answers leading to a rich circle of ideas. You'll also learn a whole new way of thinking about *conditional* probabilities and distributions.

But, before we can do much with random variables, we need to build some background. The first two weeks or so of the course will be more abstract and technical than most of what follows. Some people find it hard and frustrating at first, but it gets easier as you become more familiar with the arguments and approaches we need to take. It's worth it— lots of what you'll do as a Statistician or Economist or Physicist or whatever you want to become will involve thinking carefully about limits of random variables and about conditional probabilities at a depth impossible without this material.

1 Sets and Events

1.1 Motivation

Although it's not a pre-requisite, most students enrolled in this course will have taken an undergraduate calculus-based course in probability theory like Duke's MTH230 = STA230. Such a course teaches about discrete and continuous random variables and their distributions, joint distributions of 2 or 3 RVs, a little about conditional probability and conditional distributions. Most things are done twice: first for discrete RVs (binomial, geometric, Poisson), using sums, and then again a second time for continuous RVs (uniform, normal, exponential), using integrals.

This course instead builds a single coherent (beautiful) structure for one, two, or even infinitely-many random variables that are discrete or continuous or neither. We will be especially concerned with limits of sequences of random variables (we will see there are many sorts of limits to consider) and with conditional distributions, given the values of many (even infinitely-many) other random variables or events.

A recurring theme is application within Bayesian statistics— which we view as simply probability theory on a grand scale, building a joint probability model for all the things we don't know. These might include both the values of parameters (like the probability p of success in a clinical trial of an experimental drug) and observable quantities that we may not yet have observed (for example, the number X of successes in the trial of N subjects). The object is usually to make deductions about the CONDITIONAL DISTRIBUTION of the things we care about, given the things we have observed... like $\mathsf{P}[X \ge 8 \mid p = 0.5, N = 10]$, for predicting outcomes of a future experiment for known value 0.5 of the parameter p, or $\mathsf{P}[\ p > 0.5 \mid X = 8, N = 10\ ]$, for making inference about the parameter p after observing the outcome $X = 8$ successes among $N = 10$ subjects.

1.2 Notation and Basic Mathematical Set-Up

- Ω: Set of possible outcomes of some "experiment"
- ω: One of the outcomes in Ω
  [Idea: nature or fate chooses an ω from Ω; alas she doesn't tell us which one. We just

get hints from observing $X(\omega), Y(\omega)$, ...]

- A, B, C: Subsets of Ω; A is "true" if nature's $\omega \in A$. Usually UC letters in first half of alphabet, A–M or so.
- Y^X: All functions from a set X to a set Y. Special cases:
  - 2^Ω: All subsets $\{A : A \subseteq \Omega\}$ of Ω ("Power set", often denoted by a spiky $\mathfrak{P}(\Omega)$) $= \{f : \Omega \to \{0,1\}\}$
  - Ω^2: All ordered pairs (ω_1, ω_2) $= \{f : \{1,2\} \to \Omega\}$
- $\mathsf{P}[\ ]$: Probability assignment of numbers $\mathsf{P}[A] \geq 0$ to *some* (maybe not all) subsets A of Ω. The need to limit $\mathsf{P}[\ ]$ to just *some* "events" and not the entire power set 2^Ω is an important distinction of graduate level or "measure theoretic" probability.
- $\mathcal{A}$: Certain collections ("classes") of sets (typ. 1st half of $A - Z$, in $\mathcal{SCRIPT}$).
- X, Y, Z: Random variables, functions $X : \Omega \to \mathbb{E}$, usually to a vector space $\mathbb{E}$ (often $\mathbb{R}$ or $\mathbb{R}^n$). Mostly 2nd half of $A - Z$.
- $\mathsf{E}[X]$: Expectation of SOME (not all!!!) random variables X (why not all?)
- $\{\ \}$ "Slash Oh" ($\emptyset$) is *empty set*, not the Greek letter ϕ (or φ).
- $\omega \in A$: Inclusion ("element of"). $\in$ is not the Greek letter ϵ (or ε).
- $A \subset B$: Subset: means $(\forall \omega \in A)\ \omega \in B$. Same as $A \Rightarrow B$.
- $\mathbb{R} := (-\infty, \infty)$, $\mathbb{R}_+ := (0, \infty)$, $\mathbb{R}_- := (-\infty, 0)$; $\mathbb{C} := \{a + b\mathrm{i}\}$ with $a, b \in \mathbb{R}$ and $\mathrm{i} = \sqrt{-1}$; $\mathbb{N} := \{1, 2, \cdots\}$; $\mathbb{N}_0 := \{0, 1, 2, \cdots\}$; $\mathbb{Z} := \{..., -3, -2, -1, 0, 1, 2, \cdots\}$; $\mathbb{Q} := \{\frac{i}{n} : \ i \in \mathbb{Z},\ n \in \mathbb{N}\}$, the rationals; $\mathbb{Q}_2 := \{i/2^n : \ i, n \in \mathbb{Z}\}$, the *dyadic* rationals.
- $\lfloor x \rfloor := \max\{n \in \mathbb{Z} : \ n \leq x\}$; $\lceil x \rceil := \min\{n \in \mathbb{Z} : \ n \geq x\}$; $\lfloor \pi \rfloor = 3 = \lfloor 3 \rfloor = \lceil 3 \rceil$.

1.3 Four Big Ideas in Probability

1. LLN (Law of Large Numbers):
   If $\{X_i\}$ are Independent Identically-Distributed (IID) RVs with same mean $\mu = \mathsf{E}[X_i]$, and partial sums $S_n := \sum_{i \leq n} X_i$ and sample mean $\bar{X}_n := S_n/n$, then $\bar{X}_n \to \mu$ or, equivalently,
   $$\frac{S_n - n\mu}{n} \to 0$$
   [what does it *mean* for a sequence *random variables* like $\bar{X}_n := \frac{1}{n} S_n$ to "converge" to a constant μ or to a random variable Y??? Or to be independent or identically distributed?]

2. CLT (Central Limit Theorem):
If $\{X_i\}$ are IID with same mean μ and finite variance $\sigma^2 := \mathsf{E}[(X_i-\mu)^2]$, and partial sums $S_n := \sum_{i\le n} X_i$, then $\sqrt{n}(\bar{X}_n - \mu) \Rightarrow \mathsf{No}(0,\sigma^2)$ or, equivalently,

$$Z_n := \frac{S_n - n\mu}{\sqrt{n\sigma^2}} \Rightarrow \mathsf{No}(0,1)$$

[what does it *mean* for a sequence of *distributions* to converge??]

[what happens if $\{X_i\}$ *don't* have finite variances or means?]

3. LIL (Law of the Iterated Logarithm):
If $\{X_i\}$ are IID with same mean $\mu := \mathsf{E}[X_i]$ and finite variance $\sigma^2 := \mathsf{E}[(X_i-\mu)^2]$, and partial sums $S_n := \sum_{i\le n} X_i$, then

$$\limsup_n \frac{S_n - n\mu}{\sqrt{2n\sigma^2 \log\log n}} = 1$$

[what is the "lim sup" of a sequence of random variables?]
Note all three of LLN, CLT, LIL describe the convergence of expressions of the form $[S_n - n\mu]/g(n)$ as $n\to\infty$, for functions $g(n)$ that increase at different rates.

4. MCT (Martingale Convergence Theorem):
If X_n is "conditionally constant" in the sense that for every $k \ge 0$ and n ,

$$X_n = \mathsf{E}[X_{n+k} \mid X_1, \ldots, X_n],$$

then *under some conditions* (what conditions? why are they needed?), there exists some limiting random variable X_∞ such that

$$X_n \to X_\infty$$

(what does "$\to$" mean here?) and, for *some* random times $\sigma \le \tau$
(which ones? why just them?), also

$$\mathsf{E}[X_\tau \mid \text{Info up to time } \sigma] = X_\sigma$$

[what does it *mean* to find expectation "given" some "info" ? What *is* "info"?]

1.4 Set Operations & Logical Operations

- **Complement**: A^c = "not A" $= \{w : w \notin A\}$
- **Union** over **arbitrary** index set:

$$\bigcup_\alpha A_\alpha = \{\omega : \omega \in A_\alpha \text{ for at least one } \alpha\}$$

$$A \cup B = \text{"}A \text{ or } B \text{ (or perhaps both)"}$$

[Later we'll see it sometimes matters if the index set has finitely-many, countably-many, or uncountably-many elements; this definition works for all those cases]

- **Intersection** over arbitrary index set:

$$\bigcap_\alpha A_\alpha = \{\omega : \omega \in A_\alpha \text{ for all } \alpha\}$$
$$A \cap B = AB = \text{“both } A \text{ and } B\text{”}$$

- **Set difference**: Those $\omega \in \Omega$ in A but not in B:

$$A \backslash B = A \cap B^c$$

- **Symmetric difference**:

$$\begin{aligned} A \Delta B &= (A \backslash B) \cup (B \backslash A) \\ &= (A \cup B) \backslash (A \cap B) \\ &= \text{“in exactly one of } A, B\text{”} \end{aligned}$$

- Relations:
  - containment: $A \subset B$: “A implies B” ($A \cap B = A$)
  - disjoint: $A \cap B = \emptyset$: “A, B mutually exclusive”
  - equality: $A = B$: “A if-and-only-if B”
- De Morgan's Laws:

$$\Big(\bigcup_\alpha A_\alpha\Big)^c = \bigcap_\alpha (A_\alpha^c) \qquad \Big(\bigcap_\alpha A_\alpha\Big)^c = \bigcup_\alpha (A_\alpha^c)$$

- Countable $\neq$ Infinite (Cantor arg if time allows; note $c = 2^{\aleph_0} > \aleph_0$)
- Define injection, cardinality: $\#A \le \#B$ if exists 1:1 $\phi : A \hookrightarrow B$ (not necessarily a surjection— *i.e.*, *into* but maybe not *onto*.)
- *State* $(\#A \le \#B) \cap (\#B \le \#A) \Rightarrow (\#A = \#B)$, *i.e.*, $\#A \le \#B$ and $\#B \le \#A$ implies there exists 1:1 invertible mapping $\phi : A \leftrightarrow B$
- Convention:
  “i, j, n” (Latin) subscripts $\rightarrow$ *countable* union/intersection/sum/...
  “α, β, γ” (Greek) subscripts $\rightarrow$ *arbitrary* (could be uncountable)
  “Countable” means *finite or countably infinite.*

2 Sets, convergence of sequences of sets, fields

2.1 Convergence

Let $\{A_n\} \subset \mathcal{F}$ be a countable collection of events. In addition to their countable union $\cup_n A_n$ and intersection $\cap_n A_n$, two other combinations of $\{A_n\}$ arise frequently enough to have their own names and notations:

$$\liminf A_n = \text{ All but finitely-many } = \text{ union of intersections } = \bigcup_n \bigcap_{m\ge n} A_m$$
$$\limsup A_n = \text{ Infinitely-many } = \text{ intersection of unions } = \bigcap_n \bigcup_{m\ge n} A_m$$

Always $(\liminf) \subset (\limsup)$ (why?); sometimes, but not always, they coincide.
Some examples, with $\Omega = \mathbb{N}$:

$$A_n = n, n+1, ... \qquad \limsup A_n = \emptyset,\ \liminf A_n = \emptyset$$
$$A_n = 1, 2, ..., n \qquad \limsup A_n = \mathbb{N}, \liminf A_n = \mathbb{N}$$
$$A_{2n} = \text{Evens}, A_{2n+1} = \text{Odds}: \quad \limsup A_n = \mathbb{N}, \liminf A_n = \emptyset$$

The terms "lim sup" and "lim inf" are also the names of operations on sequences of *numbers* or real-valued *functions* $\{a_n\}$:

$$\liminf_{n\to\infty} a_n := \sup_{n\to\infty} \left[\inf_{m\ge n} a_m\right] = \lim_{n\to\infty} \left[\inf_{m\ge n} a_m\right]$$
$$\limsup_{n\to\infty} a_n := \inf_{n\to\infty} \left[\sup_{m\ge n} a_m\right] = \lim_{n\to\infty} \left[\sup_{m\ge n} a_m\right]$$

Always $\liminf a_n \le \limsup a_n$ (why?). The lim inf and lim sup coincide if and only if the sequence $\{a_n\}$ converges, and in that case their common value is $\lim_{n\to\infty} a_n$.

The set-based and numerical meanings of lim inf and lim sup are related, of course. Let $\{A_n\} \subset \mathcal{F}$ be a collection of events and let $a_n := \mathbf{1}_{\{A_n\}}$ be their indicator functions, equal to one for $\omega \in A_n$ and zero elsewhere. Then

$$\liminf_{n\to\infty} \mathbf{1}_{\{A_n\}} = \sup_{n<\infty} \inf_{m\ge n} \mathbf{1}_{\{A_m\}} = \sup_{n<\infty} \mathbf{1}_{\{\cap_{m\ge n} A_m\}} = \mathbf{1}_{\{\cup_{n<\infty}\cap_{m\ge n} A_m\}} = \mathbf{1}_{\{\liminf_{n\to\infty} A_m\}}$$
$$\limsup_{n\to\infty} \mathbf{1}_{\{A_n\}} = \inf_{n<\infty} \sup_{m\ge n} \mathbf{1}_{\{A_m\}} = \inf_{n<\infty} \mathbf{1}_{\{\cup_{m\ge n} A_m\}} = \mathbf{1}_{\{\cap_{n<\infty}\cup_{m\ge n} A_m\}} = \mathbf{1}_{\{\limsup_{n\to\infty} A_m\}}$$

Thus, the lim sup and lim inf of indicator functions of events are the indicators of the lim sups and lim infs of those events, respectively. The *event* that a sequence X_n of functions on Ω converges (pointwise) to a limiting function X is:

$$\{\omega : X_n(\omega) \to X(\omega)\} = \{\omega : \limsup_{n\to\infty} |X_n(\omega) - X(\omega)| = 0\}$$
$$= \bigcap_{k<\infty} \bigcup_{n<\infty} \bigcap_{m\ge n} \{\omega : |X_m(\omega) - X(\omega)| < 1/k\}$$

or, with the limit unspecified, the Cauchy criterion give

$$\begin{aligned}\{\omega : X_n(\omega) \text{ converges}\} &= \{\omega : \limsup_{n\to\infty} X_n(\omega) - \liminf_{n\to\infty} X_n(\omega) = 0\} \\ &= \bigcap_{k<\infty} \bigcup_{n<\infty} \bigcap_{m\geq n} \{\omega : |X_n(\omega) - X_m(\omega)| < 1/k\}.\end{aligned}$$

2.2 Fields and σ-Fields

Not every subset A of Ω will be an "event" whose probability is well-defined, if Ω is uncountable, but we will need to show that some specific sets *are* events, and that some combinations of events (like unions $A \cup B$) will generate events. Here are some tools to help us do that. Think of $\mathcal{A}$ in this section as "the collection of subsets $A \subset \Omega$ to which we can assign a probability $\mathsf{P}[A]$".

A collection $\mathcal{A}$ of sets is a *field* if:

(i) $\Omega \in \mathcal{A}$

(ii) $\mathcal{A}$ is closed under complementation, *i.e.*, $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$

(iii) $\mathcal{A}$ is closed under *finite* unions, *i.e.*, $A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}$.

By mathematical induction, (iii) implies $\mathcal{A}$ is closed under all finite unions. Together (ii) and (iii) imply that $\mathcal{A}$ is also closed under finite intersections (why?). Finite intersections won't be enough to guarantee that sets like "X_n converges" will be events, and (iii) does *not* imply that countable unions are included in $\mathcal{A}$. For that we need stronger hypotheses:

A collection of subsets $\mathcal{A}$ of Ω is a *σ-field* (or σ-algebra or Borel field) if it satisfies the stronger conditions

(i) $\Omega \in \mathcal{A}$

(ii) $\mathcal{A}$ is closed under complementation, *i.e.*, $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$

(iii) $\mathcal{A}$ is closed under *countable* unions, *i.e.*, $\{A_i\} \subset \mathcal{A} \Rightarrow \cup_{i=1}^{\infty} A_i \in \mathcal{A}$.

Evidently every σ-field is also a field, but the converse is false. For example, for any infinite set Ω the collection $\mathcal{A} = \{$ Finite and co-finite sets $\}$ is a field but not a σ-field. Note also that the condition is only on *countable* unions, and that closure may fail for *arbitrary* unions.

2.3 Probability Assignments

Probabilities are numbers between zero and one that quantify "how likely" events are to occur. Three classical interpretations of probability are:

Symmetry: If exactly one of $k \in \mathbb{N}$ different events A_i will occur, and if each is as likely as another, then $\mathsf{P}[A_i] = 1/k$ for each. For example, the probability of rolling 11 with a pair of fair dice is $2/36 = 1/18 \approx 0.556$; the probability of drawing two queens in a row from a well-shuffled deck of 52 cards is $\binom{4}{2}/\binom{52}{2} = 1/221 \approx 0.04525$.

Frequency: If an event A may be replicated independently over and over, then $\mathsf{P}[A] = \lim_{n\to\infty} \frac{1}{n}\#\{\text{Times } A \text{ occurs in } n \text{ tries}\}$.

Degree of Belief: If you are indifferent between a "game" in which you win \$1 if A occurs and zero if not, and a game in which you win \$1 if a blue ball is drawn from a well-mixed urn containing $100p\%$ blue balls and the rest white ones, then *your* (subjective) probability of A is p.

These are listed in increasing order of applicability— the first applies only to events goverened by symmetry (so "heads or tails" might count but "rain or shine" wouldn't), while the second applies only to events that could (in principle) be replicated indefinitely (so "green smooth peas from a cross of yellow smooth and green smooth" would count, but "Duke beats Carolina in football this year" wouldn't). They agree in situations where they all apply. In each case they satisfy some "rules", like $\mathsf{P}(\Omega) = 1$ and $0 \le \mathsf{P}[A] \le 1$ and $\mathsf{P}[A \cup B] = \mathsf{P}[A] + \mathsf{P}[B]$ if $A \cap B = \emptyset$. Let's codify the rules and start looking at their consequences.

Probability Spaces

A *Probability Space* is a triplet $(\Omega, \mathcal{F}, \mathsf{P})$ of a nonempty set Ω, a σ-field $\mathcal{F} \subset 2^\Omega$, and a *probability measure* $\mathsf{P} : \mathcal{F} \to \mathbb{R}$ with the three properties:

(i) $\mathsf{P}(A) \ge 0$

(ii) σ-additive[1], *i.e.*, if $\{A_i\} \subset \mathcal{F}$ are disjoint then

$$\mathsf{P}\Big(\bigcup_i A_i\Big) = \sum_i \mathsf{P}(A_i)$$

(iii) $\mathsf{P}(\Omega) = 1$.

Other important kinds of (non-Probability) measures P include:

- Finite positive measure: replace (iii) with: $\mathsf{P}(\Omega) < \infty$;
- σ-finite positive measure: replace (iii) with: $\Omega = \bigcup_i A_i$ for some countable collection $\{A_i\} \subset \mathcal{F}$ with each $\mathsf{P}(A_i) < \infty$.

[1] It's obvious we'll want *finite* additivity, so $\mathsf{P}[A \cup B] = \mathsf{P}[A] + \mathsf{P}[B]$ for disjoint A, B, but less obvious we'll want *countable* additivity. We'll need that to make any strong statements about limits of random variables. If we're ready to assume finite additivity, then the further assumption of *countable* additivity is equivalent to "continuity", to the assertion that if $B_{n+1} \subset B_n$ and $\cap_n B_n = \emptyset$ then $\mathsf{P}[B_n] \to 0$.

- Signed measure: replace (i) with: $\mathsf{P}(A) \in \mathbb{R}$, replace (iii) with[2]: $\Omega = \bigcup_i A_i$, for some countable collection of sets $A_i \in \mathcal{F}$ with $\sum_i |\mathsf{P}(A_i)| < \infty$
- Complex measure: replace (i) with: $\mathsf{P}(A) \in \mathbb{C}$, replace (iii) with: $\Omega = \bigcup_i A_i$, for some countable collection of sets $A_i \in \mathcal{F}$ with $\sum_i |\mathsf{P}(A_i)| < \infty$

Properties of Measures

- Inclusion/Exclusion rule: $\mathsf{P}[A \cup B] = \mathsf{P}[A] + \mathsf{P}[B] - \mathsf{P}[A \cap B]$. More generally,
$$\mathsf{P}\big(\cup A_i\big) = \sum_i \mathsf{P}(A_i) - \sum_{i<j} \mathsf{P}(A_i \cap A_j) + \sum_{i<j<k} \mathsf{P}(A_i \cap A_j \cap A_k) - \dots$$
- Subadditivity:
$$\mathsf{P}\big(\cup A_i\big) \le \sum_i \mathsf{P}(A_i) \qquad \text{(even if not disjoint)}$$
- Continuity:
$$A_n \subset A_{n+1} \Rightarrow \mathsf{P}\big(\cup A_n\big) = \lim \mathsf{P}(A_n)$$
$$B_n \supset B_{n+1} \Rightarrow \mathsf{P}\big(\cap B_n\big) = \lim \mathsf{P}(B_n)$$
- Fatou's Lemma: $\mathsf{E}[\liminf X_n] \le \liminf \mathsf{E}[X_n]$, so (with $X_n = 1_{A_n}$),
$$\mathsf{P}(\liminf A_n) \le \liminf \mathsf{P}(A_n) \le \limsup \mathsf{P}(A_n) \le \mathsf{P}(\limsup A_n)$$
- Distribution Functions (DFs): For $\Omega \subset \mathbb{R}$, the function $F(x) := \mathsf{P}\{(-\infty, x]\}$ satisfies
  - $x < y \Rightarrow F(x) \le F(y)$;
  - $F(x) = F(x+) := \lim\{F(y) : \ y \searrow x\}$;
  - $F(-\infty) := \lim\{F(x) : \ x \searrow -\infty\} = 0$, $F(\infty) := \lim\{F(x) : \ x \nearrow \infty\} = 1$.

  and, for $-\infty < a < b < \infty$, $\mathsf{P}(a, b] = F(b) - F(a)$.

Three special cases admitting simple Probability Measure constructions:

i: Discrete: Countable $\Omega = \{\omega_i\}$, sequence $p_i \ge 0$ with $\sum_i p_i = 1$, and $\mathcal{F} = 2^\Omega$;

ii: Continuous: $\Omega \subseteq \mathbb{R}^n$: $\mathsf{P}[A] = \int_A f(x)\,dx$ for some $f(x) \ge 0$ with $\int_{\mathbb{R}} f(x)\,dx = 1$;

iii: General 1-d: Set $\mathsf{P}(-\infty, b] := F(b)$ on $\mathcal{P} := \big\{(-\infty, b],\ b \in \mathbb{R}\big\}$, for some DF $F(x)$, extend somehow to $\mathcal{B} = \sigma(\mathcal{P})$. We'll see how next week!

Sketch some counter-examples to illustrate what can go wrong when the rules are violated—*e.g.*, try to make uniform distribution on integers or Lebesgue on rationals.

Last edited: August 26, 2015

[2]This is a bit of a simplification. What is actually needed is $\sum_i |\mathsf{P}|(A_i) < \infty$, where $|\mathsf{P}|(A)$ is the σ-finite measure defined by $|\mathsf{P}|(A) := \sup\{|\mathsf{P}(A \cap B)| + |\mathsf{P}(A \cap B^c|,\ B \in \mathcal{F}\} < \infty$. Something similar is needed for complex measures below.