A User's Guide to Measure Theoretic Probability
Errata and comments

Chapter 2.

page 25, line -3: Upper limit on sum should be 2×4^n

page 34, line -10: case of a probability measure

page 35, line 20–23: A student pointed out that I should also assume f is $\mathcal{A}$-measurable to cover the case where the sigma-field is not μ-complete in the sense of Definition <27>. If the set where convergence fails is $\mathcal{A}$-measurable, the function f could redefined as zero on that negligible set. See the discussion in the following paragraph.
page 43: The whole of Section 2.11 has been changed, with λ-cones being replaced by what I call λ-spaces: vector spaces of bounded real functions that contain the constant functions and that are stable under increasing pointwise limits when the limit function is bounded. The replacement is attached to the end of this list.

page 46, line 2 of Problem [3]: A bit silly to define $\mathcal{A}$ as a sigma-field then ask you to prove it is a sigma-field. Better: Define $\mathcal{A} = \{T^{-1}B : b \in \mathcal{B}\}$.

page 47, line 3 of Problem[12]: $\mu B = \inf\{\mu G : \dots$

page 50, line 4 of Problem [24]: X_I should be X_i

page 51, line 3 of Problem [26]: to ensure $\mu|f|B^c < \epsilon$.

Chapter 3.

page 56, line 14: When I teach this material I usually derive the Radon-Nikodym Theorem directly from Lemma <5>. The proof is essentially the same as the proof of the Lebesgue decomposition.

page 56, line -8: Note that μg_1^2 cannot be zero because $\nu g_1 \neq 0$.

page 61, line 15: The upper bound in inequality <10> should be $2H(P,Q)$.

page 72, line 6: Don't need $\mathcal{E}$ countable. Compare with Example <2.5>.

page 73, line 17: The assertion in part (i) of Problem [12] should be

$$(\mu_1 - \mu_2)^+ B = \sup_\pi \sum_{A \in \pi} ((\mu_1 - \mu_2)AB)^+$$

page 73, line -2: Better hint: first show that $(q - Kp)(\mathfrak{X}_0 - f) \geq 0$.

Chapter 4.

page 82, line 7 and -12: Theorem <1> not <2>. I now think it is better to prove that Theorem by a direct λ-space argument rather than the way it is done on page 82.

page 92, line 14: from Section 2.9

page 96, line -14: Actually the proof wouldn't fail because $\{\tau = i\}$ is always intersected with $\{|S_i| > \epsilon_1 + \epsilon_2\}$.

page 101, line 4: be a consistent.

page 102, line 1: It was misleading to derive Daniell/Kolmogorov from Theorem <49>. That approach obscures the role played by compactness.

page 105, line 17: I forget to mention that the P_i and Q_i are probability measures for Problem [18].

page 106, line 21: I forgot to mention that the X_i's are assumed to be identically distributed in Problem [26]. Without that assumption, the appearance of X_1 in part (iii) would be pure nonsense.

page 109, line 14: There is a lot more known about Problem [14] than I was aware of when writing UGMTP. See for example, the note "The relation between the distance of random variables and their distributions" at the web site http://www.renyi.hu/∼major/probability.html or the two volumes on "Mass Transportation Problems" by Rachev and Rüschendorf (Springer 1998).

page 114, line -9: conditional

Chapter 5.

page 123, line 22: An embarrassing typo here and elsewhere for Kolmogorov's name.

page 124, line 17: Compare with Problem [3.3].

page 124, line -9: should be $\gamma_n \le \gamma_{n+1}$

page 125, line 1: $\alpha(t)$ missing before $h_2(t)$

page 125, line 7: $\alpha(T)$ missing before $h_2(T)$

page 125, line -2: The first $\mathbb{P}$ should be a $\mathbb{Q}$.

page 125, line -1: $\mathcal{B}$-measurable

page 126, line -7: Ω not $\mathcal{X}$

page 132, line 14: Part (iii) should define the conditional density as

$$p_x(y) := \sum\nolimits_i \beta_i p_{1i}(x) p_{2i}(y)/p(x) \qquad \text{on } \{x : p(x) > 0\}$$

page 132, line -2: hence $\{X > t\}$ is $\mathcal{G}$-measurable

page 133, line 6: bound $\mathbb{P}\{|S_i| > \epsilon, \tau = i\}$ by $\beta\mathbb{P}\{|S_N| > \alpha\epsilon, \tau = i\}$.

Chapter 6.

page 139, line 12: should be the "natural filtration"

page 139, line -4: might not be well defined for unbounded Z; better to stick with bounded, nonnegative Z and justify passage to the limit in particular cases

page 145, line -7: Kolmogorov maximal inequality

page 149, line -13: on $\mathbb{N}_0$

page 151, line 10: almost surely on D^C

page 162, line -15: Tricky problem. Hint: Show that τ is $\mathcal{G}$-measurable, where $\mathcal{G} = \sigma(X_i : i \in \mathbb{N})$ with $X_i = Z_{\tau \wedge i}$.

Chapter 7.

page 170, line -10: Too cute. I should have stuck with the usual definition of $\|\cdot\|_{\text{BL}}$ and put up with the extra factors of 2.

page 171, line 29: I would now use bounded, Lipshitz functions with the outer expectations.

page 175, line 1: Many font problems in the diagram. A $\geq$ is missing before the PG in row two and before the Pg in row three. A $\leq$ is missing before the PF in row two and before the Pf in row three. In rows four and six the X should be an $\mathcal{X}$.

page 186, line 19: subnets not subsets

page 187, line -3: $\Delta(P_n, P) \to 0$

page 188, line 9: $\mathcal{Y} \times \mathcal{Z}$ not $\mathcal{Y} \otimes \mathcal{Z}$

page 189, line -10: Let $\{\xi_i : i \in \mathbb{N}_0\}$

page 189, line -8: The assertion of Problem [17] is false. Compare with $\mathbb{P}\{X_n = \psi^{-1}(n^2)\} = 1/n = 1 - \mathbb{P}\{X_n = 0\}$.

Chapter 9.

page 218, line -1: replace Δ by D

page 220, line 6: In the first printing, some parts of the pictures on this and the next page became invisible.

Chapter 10.

page 239, line -16: and we can find

page 245, line 2: reworking Yurinskii's

page 246, line 4: into the constant C_0

page 247, line -4: the $N(0, \sigma^2 I_k)$ density

Chapter 12.

page 276, line -14: the indicator functions give the wrong value when the x_i's are all < 0. It was intended that the max should run over the set $\{i : 1 \leq i \leq n, i \neq j\}$.

page 276, line -8: $x_j \mapsto f(L - S)$

Appendix D.
See Carter and Pollard (2004, Ann. Statistics) for a cleaner version.

Appendix E.

page 334, line 5: The Theorem as stated is wrong. See the attached corrections from a Stochastic Calculus course that I taught.

page 337, line 11: $\widetilde{\mathcal{F}}_s^B \subseteq \sigma(\mathcal{N} \cup \mathcal{F}_s^B) \subseteq \sigma(\mathcal{N} \cup \mathcal{F}_t^B)$

2.11 Generating classes of functions

Theorem <38> is often used as the starting point for proving facts about measurable functions. One first invokes the Theorem to establish a property for sets in a sigma-field, then one extends by taking limits of simple functions to $\mathcal{M}^+$ and beyond, using Monotone Convergence and linearity arguments. Sometimes it is simpler to invoke an analog of the λ-system property for classes of functions.

<2.43> **Definition.** *Let $\mathcal{H}$ be a set of bounded, real-valued functions on a set $\mathcal{X}$. Call $\mathcal{H}$ a **$\boldsymbol{\lambda}$-space** if:*

(i) $\mathcal{H}$ is a vector space

(ii) each constant function belongs to $\mathcal{H}$;

(iii) if $\{h_n\}$ is an increasing sequence of functions in $\mathcal{H}$ whose pointwise limit h is bounded then $h \in \mathcal{H}$.

The sigma-field properties of λ-spaces are slightly harder to establish than their λ-system analogs, but the reward of more streamlined proofs will make the extra, one-time effort worthwhile. First we need an analog of the fact that a λ-system that is stable under finite intersections is also a sigma-field. Remember that $\sigma(\mathcal{H})$ is the smallest σ-field on $\mathcal{X}$ for which each h in $\mathcal{H}$ is $\sigma(\mathcal{H})\backslash\mathcal{B}(\mathbb{R})$-measurable. It is the σ-field generated by the collection of sets $\{h \in B\}$ with $h \in \mathcal{H}$ and $B \in \mathcal{B}(\mathbb{R})$. It is also generated by

$$\mathcal{E}_{\mathcal{H}} := \{\, \{h < c\} : h \in \mathcal{H},\, c \in \mathbb{R}\}.$$

<2.44> **Lemma.** *If a λ-space $\mathcal{H}$ is stable under the formation of pointwise products of pairs of functions then it consists of all bounded, $\sigma(\mathcal{H})$-measurable functions.*

Proof By definition, every function in $\mathcal{H}$ is $\sigma(\mathcal{H})$-measurable. The proof that every bounded, $\sigma(\mathcal{H})$-measurable function belongs to $\mathcal{H}$ will follow from the following four facts:

(a) $\mathcal{H}$ is stable under uniform limits

(b) if h_1 and h_2 are in $\mathcal{H}$ then so are $h_1 \vee h_2$ and $h_1 \wedge h_2$

(c) the collection of sets $\mathcal{A}_0 := \{A \in \mathcal{A} : A \in \mathcal{H}\}$ is a σ-field

(d) $\mathcal{E}_{\mathcal{H}} \subseteq \mathcal{A}_0$ and hence $\sigma(\mathcal{H}) = \sigma(\mathcal{E}_{\mathcal{H}}) \subseteq \mathcal{A}_0$

For suppose g is a bounded, $\sigma(\mathcal{H})$-measurable function. With no loss of generality (or by means of some linear rescaling) we may assume that $0 \le g \le 1$. For each real c, the (indicator function of the) $\sigma(\mathcal{H})$-measurable set $\{g \ge c\}$ belongs to $\mathcal{H}$, by virtue of (d) and (c). The vector space property of $\mathcal{H}$ ensures that the simple function $g_n := 2^{-n}\sum_{i=1}^{2^n}\{g \ge i/2^n\}$ also belongs to $\mathcal{H}$. Stability of $\mathcal{H}$ under uniform limits then implies that $g \in \mathcal{H}$.

Proof of (a).
Suppose $h_n \to h$ uniformly, with $h_n \in \mathcal{H}$. Write δ_n for 2^{-n}. With no loss of generality we may suppose $h_n + \delta_n \ge h \ge h_n - \delta_n$ for all n. Notice that

$$h_n + 3\delta_n = h_n + \delta_n + \delta_{n-1} \ge h + \delta_{n-1} \ge h_{n-1}.$$

the functions $g_n := h_n + 3(\delta_1 + \cdots + \delta_n)$ all belong to $\mathcal{H}$, and $g_n \uparrow h + 3$. It follows that $h + 3 \in \mathcal{H}$, and hence, $h \in \mathcal{H}$.

Proof of (b).
It is enough if we show that $h^+ \in \mathcal{H}$ for each h in $\mathcal{H}$, because $h_1 \vee h_2 = h_1 + (h_2 - h_1)^+$ and $-\big(h_1 \wedge h_2\big) = (-h_1) \vee (-h_2)$. Suppose $c \le h \le d$, for constants c and d. First note that, for every polynomial $p(y) = a_0 + a_1 y \cdots + a_m y^m$, we have

$$p(h) = a_0 + a_1 h + \cdots + a_m h^m \in \mathcal{H},$$

because the constant function a_0 and each of the powers h^k belong to the vector space $\mathcal{H}$. By a minor extension of the Weierstrass approximation result from Problem [25], the continuous function $y \mapsto y^+$ can be uniformly approximated by a polynomial on the interval $[c, d]$. That is, there exists a sequence of polynomials p_n such that $\sup_{c \le y \le d} |p_n(y) - y^+| \to 0$ as $n \to \infty$. In particular, h^+ is a uniform limit of $p_n(h)$, so that $h^+ \in \mathcal{H}$ by virtue of (a).

Proof of (c).
The fact that $1 \in \mathcal{H}$ and the stability of $\mathcal{H}$ under monotone limits, differences, and finite products implies that $\mathcal{A}_0$ is a λ-system of sets that is stable under finite intersections, that is, $\mathcal{A}_0$ is a σ-field.

Proof of (d).
Suppose $h \in \mathcal{H}$ and $c \in \mathbb{R}$. By (b), the function

$$h_0 := \big(1 + h - c\big)^+ \wedge 1$$

belongs to $\mathcal{H}$. Notice that $0 \leq h_0 \leq 1$ and $\{h_0 = 1\} = \{h \geq c\}$. As a monotone increasing limit of functions $1-h_0^n$ from $\mathcal{H}$, the (indicator function of the) set $\{h < c\}$ also belongs to $\mathcal{H}$.

□

<2.45> **Theorem.** *Let $\mathcal{G}$ be a set of functions from a λ-space $\mathcal{H}$. If $\mathcal{G}$ is stable under the formation of pointwise products of pairs of functions then $\mathcal{H}$ contains all bounded, $\sigma(\mathcal{G})$-measurable functions.*

Proof Let $\mathcal{H}_0$ be the smallest λ-space containing $\mathcal{G}$. By Lemma <2.44>, it is enough to show that $\mathcal{H}_0$ is stable under pairwise products.

Argue as in Theorem <38> for λ-systems of sets. An almost routine calculation shows that $\mathcal{H}_1 := \{h \in \mathcal{H}_0 : hg \in \mathcal{H}_0 \text{ for all } g \text{ in } \mathcal{G}\ \}$ is a λ-space containing $\mathcal{G}$. The only subtlety lies in showing that $\mathcal{H}_1$ is stable under monotone increasing limits. If $h_n \in \mathcal{H}_1$ and $h_n \uparrow h$ and $g \geq 0$, then $gh_n \uparrow gh$. At points where g is strictly negative, the sequence gh_n would not be increasing. However, we can find a constant C large enough that $g + C \geq 0$ everywhere, and hence gh belongs to $\mathcal{H}_0$ as a monotone inceasing limit of $\mathcal{H}_0$ functions $h_n g + Ch_n - Ch$. It follows that $\mathcal{H}_1 = \mathcal{H}_0$. That is, $h_0 g \in \mathcal{H}_0$ for all $h_0 \in \mathcal{H}_0$ and $g \in \mathcal{G}$.

Similarly, $\mathcal{H}_2 := \{h \in \mathcal{H}_0 : h_0 h \in \mathcal{H}_0 \text{ for all } h_0 \text{ in } \mathcal{H}_0\ \}$ is a λ-space. By the result for $\mathcal{H}_1$ we have $\mathcal{H}_2 \supseteq \mathcal{G}$, and hence $\mathcal{H}_2 = \mathcal{H}_0$. That is, $\mathcal{H}_0$ is stable under products.

□

<2.46> **Exercise.** Let μ be a finite measure on $\mathcal{B}(\mathbb{R}^k)$. Write $\mathbb{C}_0$ for the vector space of all continuous real functions on $\mathbb{R}^k$ with compact support. Suppose f belongs to $\mathcal{L}^1(\mu)$. Show that for each $\epsilon > 0$ there exists a g in $\mathbb{C}_0$ such that $\mu|f - g| < \epsilon$. That is, show that $\mathbb{C}_0$ is dense in $\mathcal{L}^1(\mu)$ under its $\mathcal{L}^1$ norm.

□

Solution: Define $\mathcal{H}$ as the collection of all bounded functions in $\mathcal{L}^1(\mu)$ that can be approximated arbitrarily closely (in $\mathcal{L}^1(\mu)$ norm) by functions from $\mathbb{C}_0$. Check that $\mathcal{H}$ is a λ-space. Trivially it contains $\mathbb{C}_0$. The sigma-field $\sigma(\mathbb{C}_0)$ coincides with the Borel sigma-field. Why? The class $\mathcal{H}$ consists of all bounded, nonnegative Borel measurable functions.

See Problem [26] for the extension of the approximation result to infinite measures.