hence $\sigma(\mathcal{E}) = \sigma(\mathcal{H}^+)$. For a fixed h and C, the continuous function $(1 - (h/C)^n)^+$ of h belongs to $\mathcal{H}^+$, and it increases monotonely to the indicator of $\{h < C\}$. Thus the indicators of all sets in $\mathcal{E}$ belong to $\mathcal{H}^+$. The assumptions about $\mathcal{H}^+$ ensure that the class $\mathcal{B}$ of all sets whose indicator functions belong to $\mathcal{H}^+$ is stable under finite intersections (products), complements (subtract from 1), and increasing countable unions (montone increasing limits). That is, $\mathcal{B}$ is a λ-system, stable under finite intersections, and containing $\mathcal{E}$. It is a sigma-field containing $\mathcal{E}$. Thus $\mathcal{B} \supseteq \sigma(\mathcal{E}) = \sigma(\mathcal{H}^+)$. That is, $\mathcal{H}^+$ contains all indicators of sets in $\sigma(\mathcal{H}^+)$.

Finally, let k be a bounded, nonnegative, $\sigma(\mathcal{H}^+)$-measurable function. From the fact that each of the sets $\{k \geq i/2^n\}$, for $i = 1, \ldots, 4^n$, belongs to the cone $\mathcal{H}^+$, we have $k_n := 2^{-n} \sum_{i=1}^{4^n} \{k \geq i/2^n\} \in \mathcal{H}^+$. The functions k_n increase monotonely to k, which consequently also belongs to $\mathcal{H}^+$. $\square$

<45> **Theorem.** *Let $\mathcal{H}^+$ be a λ-cone of bounded, nonnegative functions, and $\mathcal{G}$ be a subclass of $\mathcal{H}^+$ that is stable under the formation of pointwise products of pairs of functions. Then $\mathcal{H}^+$ contains all bounded, nonnegative, $\sigma(\mathcal{G})$-measurable functions.*

Proof. Let $\mathcal{H}_0^+$ be the smallest λ-cone containing $\mathcal{G}$. From the previous Lemma, it is enough to show that $\mathcal{H}_0^+$ is stable under pairwise products.

Argue as in Theorem <38> for λ-systems of sets. A 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-cone containing $\mathcal{G}$, and hence $\mathcal{H}_1^+ = \mathcal{H}_0^+$. That is, $h_0 g \in \mathcal{H}_0^+$ for all $h_0 \in \mathcal{H}_0^+$ and $g \in \mathcal{G}$. Similarly, the class $\mathcal{H}_2^+ := \{h \in \mathcal{H}_0^+ : h_0 h \in \mathcal{H}_0^+ \text{ for all } h_0 \text{ in } \mathcal{H}_0^+\}$ is a λ-cone. 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. $\square$

<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. $\square$

SOLUTION: Define $\mathcal{H}$ as the collection of all bounded functions in $\mathcal{L}^1(\mu)$ that can be approximated arbitrarily closely by functions from $\mathbb{C}_0$. Check that the class $\mathcal{H}^+$ of nonnegative functions in $\mathcal{H}$ is a λ-cone. Trivially it contains $\mathbb{C}_0^+$, the class of nonnegative members of $\mathbb{C}_0$. 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.

To approximate a general f in $\mathcal{L}^1(\mu)$, first reduce to the case of nonnegative functions by splitting into positive and negative parts. Then invoke Dominated Convergence to find a finite n for which $\mu|f^+ - f^+ \wedge n| < \epsilon$, then approximate $f^+ \wedge n$ by a member of $\mathbb{C}_0^+$. See Problem [26] for the extension of the approximation result to infinite measures. $\square$

12. Problems

[1] Suppose events $A_1, A_2, \ldots$, in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, are independent: meaning that $\mathbb{P}(A_{i_1} A_{i_2} \ldots A_{i_k}) = \mathbb{P}A_{i_1} \mathbb{P}A_{i_2} \ldots \mathbb{P}A_{i_k}$ for all choices of distinct subscripts $i_1, i_2, \ldots, i_k$, all k. Suppose $\sum_{i=1}^{\infty} \mathbb{P}A_i = \infty$.

(i) Using the inequality $e^{-x} \geq 1 - x$, show that

$$\mathbb{P} \max_{n \leq i \leq m} A_i = 1 - \prod_{n \leq i \leq m} (1 - \mathbb{P}A_i) \geq 1 - \exp\left(- \sum_{n \leq i \leq m} \mathbb{P}A_i \right)$$

(ii) Let m then n tend to infinity, to deduce (via Dominated Convergence) that $\mathbb{P} \limsup_i A_i = 1$. That is, $\mathbb{P}\{A_i \text{ i. o.}\} = 1$.

REMARK. The result gives a converse for the Borel-Cantelli lemma from Example <29>. The next Problem establishes a similar result under weaker assumptions.

[2] Let $A_1, A_2, \ldots$ be events in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Define $X_n = A_1 + \ldots + A_n$ and $\sigma_n = \mathbb{P}X_n$. Suppose $\sigma_n \to \infty$ and $\|X_n/\sigma_n\|_2 \to 1$. *(Compare with the inequality $\|X_n/\sigma_n\|_2 \geq 1$, which follows from Jensen's inequality.)*

(i) Show that

$$\{X_n = 0\} \leq \frac{(k - X_n)(k + 1 - X_n)}{k(k+1)}$$

for each positive integer k.

(ii) By an appropriate choice of k (depending on n) in (i), deduce that $\sum_1^\infty A_i \geq 1$ almost surely.

(iii) Prove that $\sum_m^\infty A_i \geq 1$ almost surely, for each fixed m. Hint: Show that the two convergence assumptions also hold for the sequence $A_m, A_{m+1}, \ldots$.

(iv) Deduce that $\mathbb{P}\{\omega \in A_i \text{ i. o. }\} = 1$.

(v) If $\{B_i\}$ is a sequence of events for which $\sum_i \mathbb{P}B_i = \infty$ and $\mathbb{P}B_i B_j = \mathbb{P}B_i \mathbb{P}B_j$ for $i \neq j$, show that $\mathbb{P}\{\omega \in B_i \text{ i. o. }\} = 1$.

[3] Suppose T is a function from a set $\mathcal{X}$ into a set $\mathcal{Y}$, and suppose that $\mathcal{Y}$ is equipped with a σ-field $\mathcal{B}$. Define $\mathcal{A}$ as the sigma-field of sets of the form $T^{-1}B$, with B in $\mathcal{B}$. Suppose $f \in \mathcal{M}^+(\mathcal{X}, \mathcal{A})$. Show that there exists a $\mathcal{B}\backslash\mathcal{B}[0, \infty]$-measurable function g from $\mathcal{Y}$ into $[0, \infty]$ such that $f(x) = g(T(x))$, for all x in $\mathcal{X}$, by following these steps.

(i) Show that $\mathcal{A}$ is a σ-field on $\mathcal{X}$. (It is called the σ-field generated by the map T. It is often denoted by $\sigma(T)$.)

(ii) Show that $\{f \geq i/2^n\} = T^{-1}B_{i,n}$ for some $B_{i,n}$ in $\mathcal{B}$. Define

$$f_n = 2^{-n} \sum_{i=1}^{4^n} \{f \geq i/2^n\} \qquad \text{and} \qquad g_n = 2^{-n} \sum_{i=1}^{4^n} B_{i,n}.$$

Show that $f_n(x) = g_n(T(x))$ for all x.

(iii) Define $g(y) = \limsup g_n(y)$ for each y in $\mathcal{Y}$. Show that g has the desired property. (Question: Why can't we define $g(y) = \lim g_n(y)$?)

[4] Let $g_1, g_2, \ldots$ be $\mathcal{A}\backslash\mathcal{B}(\mathbb{R})$-measurable functions from $\mathcal{X}$ into $\mathbb{R}$. Show that $\{\limsup_n g_n > t\} = \bigcup_{\substack{r \in \mathbb{Q} \\ r > t}} \bigcap_{m=1}^\infty \bigcup_{i \geq m} \{g_i > r\}$. Deduce, without any appeal to Example <8>, that $\limsup g_n$ is $\mathcal{A}\backslash\mathcal{B}(\overline{\mathbb{R}})$-measurable. Warning: Be careful about

strict inequalities that turn into nonstrict inequalities in the limit—it is possible to have $x_n > x$ for all n and still have $\limsup_n x_n = x$.

[5] Suppose a class of sets $\mathcal{E}$ cannot separate a particular pair of points x, y: for every E in $\mathcal{E}$, either $\{x, y\} \subseteq E$ or $\{x, y\} \subseteq E^c$. Show that $\sigma(\mathcal{E})$ also cannot separate the pair.

[6] A collection of sets $\mathcal{F}_0$ that is stable under finite unions, finite intersections, and complements is called a field. A nonnegative set function μ defined on $\mathcal{F}_0$ is called a finitely additive measure if $\mu\left(\cup_{i \le n} F_i\right) = \sum_{i \le n} \mu F_i$ for every finite collection of disjoint sets in $\mathcal{F}_0$. The set function is said to be countably additive on $\mathcal{F}_0$ if $\mu\left(\cup_{i \in \mathbb{N}} F_i\right) = \sum_{i \in \mathbb{N}} \mu F_i$ for every countable collection of disjoint sets in $\mathcal{F}_0$ *whose union belongs to* $\mathcal{F}$. Suppose $\mu \mathcal{X} < \infty$. Show that μ is countably additive on $\mathcal{F}_0$ if and only if $\mu A_n \downarrow 0$ for every decreasing sequence in $\mathcal{F}_0$ with empty intersection. Hint: For the argument in one direction, consider the union of differences $A_i \backslash A_{i+1}$.

[7] Let $f_1, \ldots, f_n$ be functions in $\mathcal{M}^+(\mathcal{X}, \mathcal{A})$, and let μ be a measure on $\mathcal{A}$. Show that $\mu\left(\vee_i f_i\right) \le \sum_i \mu f_i \le \mu\left(\vee_i f_i\right) + \sum_{i<j} \mu\left(f_i \wedge f_j\right)$ where $\vee$ denotes pointwise maxima of functions and $\wedge$ denotes pointwise minima.

[8] Let μ be a finite measure and f be a measurable function. For each positive integer k, show that $\mu|f|^k < \infty$ if and only if $\sum_{n=1}^{\infty} n^{k-1} \mu\{|f| \ge n\} < \infty$.

[9] Suppose $\nu := T\mu$, the image of the measure μ under the measurable map T. Show that $f \in \mathcal{L}^1(\nu)$ if and only if $f \circ T \in \mathcal{L}^1(\mu)$, in which case $\nu f = \mu(f \circ T)$.

[10] Let $\{h_n\}$, $\{f_n\}$, and $\{g_n\}$ be sequences of μ-integrable functions that converge μ almost everywhere to limits h, f and g. Suppose $h_n(x) \le f_n(x) \le g_n(x)$ for all x. Suppose also that $\mu h_n \to \mu h$ and $\mu g_n \to \mu g$. Adapt the proof of Dominated Convergence to prove that $\mu f_n \to \mu f$.

[11] A collection of sets is called a monotone class if it is stable under unions of increasing sequences and intersections of decreasing sequences. Adapt the argument from Theorem <38> to prove: if a class $\mathcal{E}$ is stable under finite unions and complements then $\sigma(\mathcal{E})$ equals the smallest monotone class containing $\mathcal{E}$.

[12] Let μ be a finite measure on the Borel sigma-field $\mathcal{B}(\mathcal{X})$ of a metric space $\mathcal{X}$. Call a set B ***inner regular*** if $\mu B = \sup\{\mu F : B \supseteq F \text{ closed }\}$ and ***outer regular*** if $\mu B = \inf\{\mu F : B \subseteq G \text{ open }\}$

(i) Prove that the class $\mathcal{B}_0$ of all Borel sets that are both inner and outer regular is a sigma-field. Deduce that every Borel set is inner regular.

(ii) Suppose μ is tight: for each $\epsilon > 0$ there exists a compact K_ϵ such that $\mu K_\epsilon^c < \epsilon$. Show that the F in the definition of inner regularity can then be assumed compact.

(iii) When μ is tight, show that there exists a sequence of disjoint compacts subsets $\{K_i : i \in \mathbb{N}\}$ of $\mathcal{X}$ such that $\mu\left(\cup_i K_i\right)^c = 0$.

[13] Let μ be a finite measure on the Borel sigma-field of a complete, separable metric space $\mathcal{X}$. Show that μ is tight: for each $\epsilon > 0$ there exists a compact K_ϵ such that $\mu K_\epsilon^c < \epsilon$. Hint: For each positive integer n, show that the space $\mathcal{X}$ is a countable

union of closed balls with radius $1/n$. Find a finite family of such balls whose union B_n has μ measure greater than $\mu\mathcal{X} - \epsilon/2^n$. Show that $\cap_n B_n$ is compact, using the total-boundedness characterization of compact subsets of complete metric spaces.

[14] A sequence of random variables $\{X_n\}$ is said to ***converge in probability*** to a random variable X, written $X_n \xrightarrow{\mathbb{P}} X$, if $\mathbb{P}\{|X_n - X| > \epsilon\} \to 0$ for each $\epsilon > 0$.

(i) If $X_n \to X$ almost surely, show that $1 \geq \{|X_n - X| > \epsilon\} \to 0$ almost surely. Deduce via Dominated Convergence that X_n converges in probability to X.

(ii) Give an example of a sequence $\{X_n\}$ that converges to X in probability but not almost surely.

(iii) Suppose $X_n \to X$ in probability. Show that there is an increasing sequence of positive integers $\{n(k)\}$ for which $\sum_k \mathbb{P}\{|X_{n(k)} - X| > 1/k\} < \infty$. Deduce that $X_{n(k)} \to X$ almost surely.

[15] Let f and g be measurable functions on $(\mathcal{X}, \mathcal{A}, \mu)$, and r and s be positive real numbers for which $r^{-1} + s^{-1} = 1$. Show that $\mu|fg| \leq (\mu|f|^r)^{1/r} (\mu|g|^s)^{1/s}$ by arguing as follows. First dispose of the trivial case where one of the factors on the righthand side is 0 or ∞. Then, without loss of generality (why?), assume that $\mu|f|^r = 1 = \mu|g|^s$. Use concavity of the logarithm function to show that $|fg| \leq |f|^r/r + |g|^s/s$, and then integrate with respect to μ. *This result is called the* ***Hölder inequality***.

[16] Generalize the Hölder inequality (Problem [15]) to more than two measurable functions $f_1, \ldots, f_k$, and positive real numbers $r_1, \ldots, r_k$ for which $\sum_i r_i^{-1} = 1$. Show that $\mu|f_1 \ldots f_k| \leq \prod_i (\mu|f_i|^{r_i})^{1/r_i}$.

[17] Let $(\mathcal{X}, \mathcal{A}, \mu)$ be a measure space, f and g be measurable functions, and r be a real number with $r \geq 1$. Define $\|f\|_r = (\mu|f|^r)^{1/r}$. Follow these steps to prove ***Minkowski's inequality***: $\|f + g\|_r \leq \|f\|_r + \|g\|_r$.

(i) From the inequality $|x + y|^r \leq |2x|^r + |2y|^r$ deduce that $\|f + g\|_r < \infty$ if $\|f\|_r < \infty$ and $\|g\|_r < \infty$.

(ii) Dispose of trivial cases, such as $\|f\|_r = 0$ or $\|f\|_r = \infty$.

(iii) For arbitrary positive constants c and d argue by convexity that

$$\left(\frac{|f| + |g|}{c + d}\right)^r \leq \frac{c}{c + d}\left(\frac{|f|}{c}\right)^r + \frac{d}{c + d}\left(\frac{|g|}{d}\right)^r$$

(iv) Integrate, then choose $c = \|f\|_r$ and $d = \|g\|_r$ to complete the proof.

[18] For f in $\mathcal{L}^1(\mu)$ define $\|f\|_1 = \mu|f|$. Let $\{f_n\}$ be a Cauchy sequence in $\mathcal{L}^1(\mu)$, that is, $\|f_n - f_m\|_1 \to 0$ as $\min(m, n) \to \infty$. Show that there exists an f in $\mathcal{L}^1(\mu)$ for which $\|f_n - f\|_1 \to 0$, by following these steps.

(i) Find an increasing sequence $\{n(k)\}$ such that $\sum_{k=1}^{\infty} \|f_{n(k)} - f_{n(k+1)}\|_1 < \infty$. Deduce that the function $H := \sum_{k=1}^{\infty} |f_{n(k)} - f_{n(k+1)}|$ is integrable.

(ii) Show that there exists a real-valued, measurable function f for which

$$H \geq |f_{n(k)}(x) - f(x)| \to 0 \qquad \text{as } k \to \infty, \text{ for } \mu \text{ almost all } x.$$

Deduce that $\| f_{n(k)} - f \|_1 \to 0$ as $k \to \infty$.

(iii) Show that f belongs to $\mathcal{L}^1(\mu)$ and $\| f_n - f \|_1 \to 0$ as $n \to \infty$.

[19] Let $\{f_n\}$ be a Cauchy sequence in $\mathcal{L}^p(\mathcal{X}, \mathcal{A}, \mu)$, that is, $\| f_n - f_m \|_p \to 0$ as $\min(m, n) \to \infty$. Show that there exists a function f in $\mathcal{L}^p(\mathcal{X}, \mathcal{A}, \mu)$ for which $\| f_n - f \|_p \to 0$, by following these steps.

(i) Find an increasing sequence $\{n(k)\}$ such that $C := \sum_{k=1}^{\infty} \| f_{n(k)} - f_{n(k+1)} \|_p < \infty$. Define $H_\infty = \lim_{N\to\infty} H_N$, where $H_N = \sum_{k=1}^{N} | f_{n(k)} - f_{n(k+1)} |$ for $1 \le N < \infty$. Use the triangle inequality to show that $\mu H_N^p \le C^p$ for all finite N. Then use Monotone Convergence to deduce that $\mu H_\infty^p \le C^p$.

(ii) Show that there exists a real-valued, measurable function f for which $f_{n(k)}(x) \to f(x)$ as $k \to \infty$, a.e. $[\mu]$.

(iii) Show that $| f_{n(k)} - f | \le \sum_{i=k}^{\infty} | f_{n(i)} - f_{n(i+1)} | \le H_\infty$ a.e. $[\mu]$. Use Dominated Convergence to deduce that $\| f_{n(k)} - f \|_p \to 0$ as $k \to \infty$.

(iv) Deduce from (iii) that f belongs to $\mathcal{L}^p(\mathcal{X}, \mathcal{A}, \mu)$ and $\| f_n - f \|_p \to 0$ as $n \to \infty$.

[20] For each random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ define

$$\|X\|_\infty := \inf\{c \in [0, \infty] : |X| \le c \text{ almost surely}\}.$$

Let $L^\infty := L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ denote the set of equivalence classes of real-valued random variables with $\|X\|_\infty < \infty$. Show that $\| \cdot \|_\infty$ is a norm on L^∞, which is a vector space, complete under the metric defined by $\|X\|_\infty$.

[21] Let $\{X_t : t \in T\}$ be a collection of $\overline{\mathbb{R}}$-valued random variables with possibly uncountable index set T. Complete the following argument to show that there exists a countable subset T_0 of T such that the random variable $X = \sup_{t\in T_0} X_t$ has the properties

(a) $X \ge X_t$ almost surely, for each $t \in T$

(b) if $Y \ge X_t$ almost surely, for each $t \in T$, then $Y \ge X$ almost surely

(The random variable X is called the ***essential supremum*** of the family. It is denoted by $\operatorname{ess\,sup}_{t\in T} X_t$. Part (b) shows that it is, unique up to an almost sure equivalence.)

(i) Show that properties (a) and (b) are unaffected by a monotone, one-to-one transformation such as $x \mapsto x/(1 + |x|)$. Deduce that there is no loss of generality in assuming $|X_t| \le 1$ for all t.

(ii) Let $\delta = \sup\{\mathbb{P} \sup_{t\in S} X_t :$ countable $S \subseteq T\}$. Choose countable T_n such that $\mathbb{P} \sup_{t\in T_n} X_t \ge \delta - 1/n$. Let $T_0 = \cup_n T_n$. Show that $\mathbb{P} \sup_{t\in T_0} X_t = \delta$.

(iii) Suppose $t \notin T_0$. From the inequality $\delta \ge \mathbb{P}(X_t \vee X) \ge \mathbb{P}X = \delta$ deduce that $X \ge X_t$ almost surely.

(iv) For a Y as in assertion (b), show that $Y \ge \sup_{t\in T_0} X_t = X$ almost surely.

[22] Let Ψ be a convex, increasing function for which $\Psi(0) = 0$ and $\Psi(x) \to \infty$ as $x \to \infty$. (For example, $\Psi(x)$ could equal x^p for some fixed $p \ge 1$, or $\exp(x) - 1$ or $\exp(x^2) - 1$.) Define $\mathcal{L}^\Psi(\mathcal{X}, \mathcal{A}, \mu)$ to be the set of all real-valued measurable functions on $\mathcal{X}$ for which $\mu\Psi(|f|/c_0) < \infty$ for some positive real c_0. Define

$\|f\|_\Psi := \inf\{c > 0 : \mu\Psi(|f|/c) \le 1\}$, with the convention that the infimum of an empty set equals $+\infty$. For each f, g in $\mathcal{L}^\Psi(\mathcal{X}, \mathcal{A}, \mu)$ and each real t prove the following assertions.

(i) $\|f\|_\Psi < \infty$. Hint: Apply Dominated Convergence to $\mu\Psi(|f|/c)$.

(ii) $f+g \in \mathcal{L}^\Psi(\mathcal{X}, \mathcal{A}, \mu)$ and the triangle inequality holds: $\|f+g\|_\Psi \le \|f\|_\Psi + \|g\|_\Psi$. Hint: If $c > \|f\|_\Psi$ and $d > \|g\|_\Psi$, deduce that

$$\Psi\left(\frac{|f+g|}{c+d}\right) \le \frac{c}{c+d}\Psi\left(\frac{|f|}{c}\right) + \frac{d}{c+d}\Psi\left(\frac{|g|}{d}\right),$$

by convexity of Ψ.

(iii) $tf \in \mathcal{L}^\Psi(\mathcal{X}, \mathcal{A}, \mu)$ and $\|tf\|_\Psi = |t|\,\|f\|_\Psi$.

REMARK. $\|\cdot\|_\Psi$ is called an Orlicz "norm"—to make it a true norm one should work with equivalence classes of functions equal μ almost everywhere. The L^p norms correspond to the special case $\Psi(x) = x^p$, for some $p \ge 1$.

[23] Define $\|f\|_\Psi$ and $\mathcal{L}^\Psi$ as in Problem [22]. Let $\{f_n\}$ be a Cauchy sequence in $\mathcal{L}^\Psi(\mu)$, that is, $\|f_n - f_m\|_\Psi \to 0$ as $\min(m, n) \to \infty$. Show that there exists an f in $\mathcal{L}^\Psi(\mu)$ for which $\|f_n - f\|_\Psi \to 0$, by following these steps.

(i) Let $\{g_i\}$ be a nonnegative sequence in $\mathcal{L}^\Psi(\mu)$ for which $C := \sum_i \|g_i\|_\Psi < \infty$. Show that the function $G := \sum_i g_i$ is finite almost everywhere and $\|G\|_\Psi \le \sum_i \|g_i\|_\Psi < \infty$. Hint: Use Problem [22] to show that $\mathbb{P}\Psi\left(\sum_{i \le n} g_i/C\right) \le 1$ for each n, then justify a passage to the limit.

(ii) Find an increasing sequence $\{n(k)\}$ such that $\sum_{k=1}^\infty \|f_{n(k)} - f_{n(k+1)}\|_\Psi < \infty$. Deduce that the functions $H_L := \sum_{k=L}^\infty |f_{n(k)} - f_{n(k+1)}|$ satisfy

$$\infty > \|H_1\|_\Psi \ge \|H_2\|_\Psi \ge \ldots \to 0.$$

(iii) Show that there exists a real-valued, measurable function f for which

$$|f_{n(k)}(x) - f(x)| \to 0 \qquad \text{as } k \to \infty, \text{ for } \mu \text{ almost all } x.$$

(iv) Given $\epsilon > 0$, choose L so that $\|H_L\|_\Psi < \epsilon$. For $i > L$, show that

$$\Psi\left(H_L/\epsilon\right) \ge \Psi\left(|f_{n(L)} - f_{n(i)}|/\epsilon\right) \to \Psi\left(|f_{n(L)} - f|/\epsilon\right).$$

Deduce that $\|f_{n(L)} - f\|_\Psi \le \epsilon$.

(v) Show that f belongs to $\mathcal{L}^\Psi(\mu)$ and $\|f_n - f\|_\Psi \to 0$ as $n \to \infty$.

[24] Let Ψ be a convex increasing function with $\Psi(0) = 0$, as in Problem [22]. Let Ψ^{-1} denote its inverse function. If $X_1, \ldots, X_N \in \mathcal{L}^\Psi(\mathcal{X}, \mathcal{A}, \mu)$, show that

$$\mathbb{P}\max_{i \le N} |X_i| \le \Psi^{-1}(N) \max_{i \le N} \|X_i\|_\Psi.$$

Hint: Consider $\Psi(\mathbb{P}\max|X_I|/C)$ with $C > \max_{i \le N} \|X_i\|_\Psi$.

REMARK. Compare with van der Vaart & Wellner (1996, page 96): if also $\limsup_{x,y\to\infty} \Psi(x)\Psi(y)/\Psi(cxy) < \infty$ for some constant $c > 0$ then $\|\max_{i\le N}|X_i|\|_\Psi \le K\Psi^{-1}(N)\max_{i\le N}\|X_i\|_\Psi$ for a constant K depending only on Ψ. See page 105 of their Problems and Complements for related counterexamples.

[25] For each θ in $[0,1]$ let $X_{n,\theta}$ be a random variable with a Binomial(n,θ) distribution. That is, $\mathbb{P}\{X_{n,\theta}=k\} = \binom{n}{k}\theta^k(1-\theta)^{n-k}$ for $k=0,1,\dots,n$. You may assume these elementary facts: $\mathbb{P}X_{n,\theta}=n\theta$ and $\mathbb{P}(X_{n,\theta}-n\theta)^2 = n\theta(1-\theta)$. Let f be a continuous function defined on $[0,1]$.

(i) Show that $p_n(\theta)=\mathbb{P}f(X_{n,\theta}/n)$ is a polynomial in θ.

(ii) Suppose $|f|\le M$, for a constant M. For a fixed ϵ, invoke (uniform) continuity to find a $\delta>0$ such that $|f(s)-f(t)|\le\epsilon$ whenever $|s-t|\le\delta$, for all s, t in $[0,1]$. Show that

$$|f(x/n)-f(\theta)| \le \epsilon + 2M\{|(x/n)-\theta|>\delta\} \le \epsilon + \frac{2M|(x/n)-\theta|^2}{\delta^2}.$$

(iii) Deduce that $\sup_{0\le\theta\le 1}|p_n(\theta)-f(\theta)|<2\epsilon$ for n large enough. That is, deduce that $f(\cdot)$ can be uniformly approximated by polynomials over the range $[0,1]$, a result known as the ***Weierstrass approximation theorem***.

[26] Extend the approximation result from Example <46> to the case of an infinite measure μ on $\mathcal{B}(\mathbb{R}^k)$ that gives finite measure to each compact set. Hint: Let B be a closed ball of radius large enough to ensure $\mu|f|B<\epsilon$. Write μ_B for the restriction of μ to B. Invoke the result from the Example to find a g in $\mathbb{C}_0$ such that $\mu_B|f-g|<\epsilon$. Find $\mathbb{C}_0$ functions $1\ge h_i\downarrow B$. Consider approximations gh_i for i large enough.

13. Notes

I recommend Royden (1968) as a good source for measure theory. The books of Ash (1972) and Dudley (1989) are also excellent references, for both measure theory and probability. Dudley's book contains particularly interesting historical notes.

See Hawkins (1979, Chapter 4) to appreciate the subtlety of the idea of a negligible set.

The result from Problem [10] is often attributed to (Pratt 1960), but, as he noted (in his 1966 Acknowledgment of Priority), it is actually much older.

Theorem <38> (the π–λ theorem for generating classes of sets) is often attributed to Dynkin (1960, Section 1.1), although Sierpiński (1928) had earlier proved a slightly stronger result (covering generation of sigma-rings, not just sigma-fields). I adapted the analogous result for classes of functions, Theorem <45>, from Protter (1990, page 7) and Dellacherie & Meyer (1978, page 14). Compare with the "Sierpiński Stability Lemma" for sets, and the "Functional Sierpiński Lemma" presented by Hoffmann-Jørgensen (1994, pages 8, 54, 60).

References

Ash, R. B. (1972), *Real Analysis and Probability*, Academic Press, New York.

Dellacherie, C. & Meyer, P. A. (1978), *Probabilities and Potential*, North-Holland, Amsterdam.

Dudley, R. M. (1989), *Real Analysis and Probability*, Wadsworth, Belmont, Calif.

Dynkin, E. B. (1960), *Theory of Markov Processes*, Pergamon.

Hawkins, T. (1979), *Lebesgue's Theory of Integration: Its Origins and Development*, second edn, Chelsea, New York.

Hoffmann-Jørgensen, J. (1994), *Probability with a View toward Statistics*, Vol. 1, Chapman and Hall, New York.

Oxtoby, J. (1971), *Measure and Category*, Springer-Verlag.

Pratt, J. W. (1960), 'On interchanging limits and integrals', *Annals of Mathematical Statistics* **31**, 74–77. Acknowledgement of priority, *same journal*, vol 37 (1966), page 1407.

Protter, P. (1990), *Stochastic Integration and Differential Equations*, Springer, New York.

Royden, H. L. (1968), *Real Analysis*, second edn, Macmillan, New York.

Sierpiński, W. (1928), 'Un théorème général sur les familles d'ensembles', *Fundamenta Mathematicae* **12**, 206–210.

van der Vaart, A. W. & Wellner, J. A. (1996), *Weak Convergence and Empirical Process: With Applications to Statistics*, Springer-Verlag.