Notes 2 : Measure-theoretic foundations II

Math 733-734: Theory of Probability *Lecturer: Sebastien Roch*

References: [Wil91, Chapters 4-6, 8], [Dur10, Sections 1.4-1.7, 2.1].

1 Independence

1.1 Definition of independence

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.

DEF 2.1 (Independence) *Sub-σ-algebras $\mathcal{G}_1, \mathcal{G}_2, \ldots$ of $\mathcal{F}$ are* independent *for all $G_i \in \mathcal{G}_i$, $i \geq 1$, and distinct $i_1, \ldots, i_n$ we have*

$$\mathbb{P}[G_{i_1} \cap \cdots \cap G_{i_n}] = \prod_{j=1}^{n} \mathbb{P}[G_{i_j}].$$

Specializing to events and random variables:

DEF 2.2 (Independent RVs) *RVs $X_1, X_2, \ldots$ are* independent *if the σ-algebras $\sigma(X_1), \sigma(X_2), \ldots$ are independent.*

DEF 2.3 (Independent Events) *Events $E_1, E_2, \ldots$ are* independent *if the σ-algebras*

$$\mathcal{E}_i = \{\emptyset, E_i, E_i^c, \Omega\}, \quad i \geq 1,$$

are independent.

The more familiar definitions are the following:

THM 2.4 (Independent RVs: Familiar definition) *RVs X, Y are independent if and only if for all $x, y \in \mathbb{R}$*

$$\mathbb{P}[X \leq x, Y \leq y] = \mathbb{P}[X \leq x]\mathbb{P}[Y \leq y].$$

THM 2.5 (Independent events: Familiar definition) *Events E_1, E_2 are independent if and only if*

$$\mathbb{P}[E_1 \cap E_2] = \mathbb{P}[E_1]\mathbb{P}[E_2].$$

The proofs of these characterizations follows immediately from the following lemma.

LEM 2.6 (Independence and π-systems) *Suppose that $\mathcal{G}$ and $\mathcal{H}$ are sub-σ-algebras and that $\mathcal{I}$ and $\mathcal{J}$ are π-systems such that*

$$\sigma(\mathcal{I}) = \mathcal{G}, \quad \sigma(\mathcal{J}) = \mathcal{H}.$$

Then $\mathcal{G}$ and $\mathcal{H}$ are independent if and only if $\mathcal{I}$ and $\mathcal{J}$ are, i.e.,

$$\mathbb{P}[I \cap J] = \mathbb{P}[I]\mathbb{P}[J], \quad \forall I \in \mathcal{I}, J \in \mathcal{J}.$$

Proof: Suppose $\mathcal{I}$ and $\mathcal{J}$ are independent. For fixed $I \in \mathcal{I}$, the measures $\mathbb{P}[I \cap H]$ and $\mathbb{P}[I]\mathbb{P}[H]$ are equal for $H \in \mathcal{J}$ and have total mass $\mathbb{P}[I] < +\infty$. By the Uniqueness lemma the above measures agree on $\sigma(\mathcal{J}) = \mathcal{H}$.

Repeat the argument. Fix $H \in \mathcal{H}$. Then the measures $\mathbb{P}[G \cap H]$ and $\mathbb{P}[G]\mathbb{P}[H]$ agree on $\mathcal{I}$ and have total mass $\mathbb{P}[H] < +\infty$. Therefore they must agree on $\sigma(\mathcal{I}) = \mathcal{G}$. ■

1.2 Construction of independent sequences

We give a standard construction of an infinite sequence of independent RVs with prescribed distributions.

Let $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1], \mathcal{B}[0,1], \lambda)$ and for $\omega \in \Omega$ consider the binary expansion

$$\omega = 0.\omega_1\omega_2 \ldots .$$

(For dyadic rationals, use the all-1 ending and note that the dyadic rationals have measure 0 by countability.) This construction produces a sequence of independent Bernoulli trials. Indeed, under λ, each bit is Bernoulli$(1/2)$ and any finite collection is independent.

To get two independent uniform RVs, consider the following construction:

$$\begin{aligned} U_1 &= 0.\omega_1\omega_3\omega_5 \ldots \\ U_2 &= 0.\omega_2\omega_4\omega_6 \ldots \end{aligned}$$

Let $\mathcal{A}_1$ (resp. $\mathcal{A}_2$) be the π-system consisting of all finite intersections of events of the form $\{\omega_i \in H\}$ for odd i (resp. even i). By Lemma 2.6, the σ-fields $\sigma(\mathcal{A}_1)$ and $\sigma(\mathcal{A}_2)$ are independent.

More generally, let

$$\begin{aligned} V_1 &= 0.\omega_1\omega_3\omega_6 \ldots \\ V_2 &= 0.\omega_2\omega_5\omega_9 \ldots \\ V_3 &= 0.\omega_4\omega_8\omega_{13} \ldots \\ \vdots &= \ddots \end{aligned}$$

i.e., fill up the array diagonally. By the argument above, the V_i's are independent and Bernoulli$(1/2)$.

Finally let μ_n, $n \geq 1$, be a sequence of probability distributions with distribution functions F_n, $n \geq 1$. For each n, define

$$X_n(\omega) = \inf\{x \,:\, F_n(x) \geq V_n(\omega)\}$$

By the Skorokhod Representation result from the previous lecture, X_n has distribution function F_n and:

DEF 2.7 (IID Rvs) *A sequence of independent RVs $(X_n)_n$ as above is independent and identically distributed (IID) if $F_n = F$ for some n.*

Alternatively, we have the following more general result.

THM 2.8 (Kolmogorov's extension theorem) *Suppose we are given probability measures μ_n on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ that are* consistent*, i.e.,*

$$\mu_{n+1}((a_1, b_1] \times \cdots \times (a_n, b_n] \times \mathbb{R}) = \mu_n((a_1, b_1] \times \cdots \times (a_n, b_n]).$$

Then there exists a unique probability measure $\mathbb{P}$ on $(\mathbb{R}^{\mathbb{N}}, \mathcal{R}^{\mathbb{N}})$ with

$$\mathbb{P}[\omega \,:\, \omega_i \in (a_i, b_i], 1 \leq i \leq n] = \mu_n((a_1, b_1] \times \cdots \times (a_n, b_n]).$$

Here $\mathcal{R}^{\mathbb{N}}$ is the product σ-algebra*, i.e., the σ-algebra generated by finite-dimensional rectangles.*

1.3 Kolmogorov's 0-1 law

In this section, we discuss a first non-trivial result about independent sequences.

DEF 2.9 (Tail σ-algebra) *Let $X_1, X_2, \ldots$ be RVs on $(\Omega, \mathcal{F}, \mathbb{P})$. Define*

$$\mathcal{T}_n = \sigma(X_{n+1}, X_{n+2}, \ldots), \quad \mathcal{T} = \bigcap_{n\geq 1} \mathcal{T}_n.$$

As an intersection of σ-algebras, $\mathcal{T}$ is a σ-algebra. It is called the tail σ-algebra *of the sequence $(X_n)_n$.*

Intuitively, an event is in the tail if changing a finite number of values does not affect its occurence.

EX 2.10 *If $S_n = \sum_{k\leq n} X_k$, then*

$$\{\lim_n S_n \text{ exists}\} \in \mathcal{T},$$

$$\{\limsup_n n^{-1} S_n > 0\} \in \mathcal{T},$$

but

$$\{\limsup_n S_n > 0\} \notin \mathcal{T}.$$

THM 2.11 (Kolmogorov's 0-1 law) *Let $(X_n)_n$ be a sequence of independent RVs with tail σ-algebra $\mathcal{T}$. Then $\mathcal{T}$ is $\mathbb{P}$-trivial, i.e., for all $A \in \mathcal{T}$ we have $\mathbb{P}[A] = 0$ or 1.*

Proof: Let $\mathcal{X}_n = \sigma(X_1, \ldots, X_n)$. Note that $\mathcal{X}_n$ and $\mathcal{T}_n$ are independent. Moreover, since $\mathcal{T} \subseteq \mathcal{T}_n$ we have that $\mathcal{X}_n$ is independent of $\mathcal{T}$. Now let

$$\mathcal{X}_\infty = \sigma(X_n, n \geq 1).$$

Note that

$$\mathcal{K}_\infty = \bigcup_{n\geq 1} \mathcal{X}_n,$$

is a π-system generating $\mathcal{X}_\infty$. Therefore, by Lemma 2.6, $\mathcal{X}_\infty$ is independent of $\mathcal{T}$. But $\mathcal{T} \subseteq \mathcal{X}_\infty$ and therefore $\mathcal{T}$ is independent of itself! Hence if $A \in \mathcal{T}$,

$$\mathbb{P}[A] = \mathbb{P}[A \cap A] = \mathbb{P}[A]^2,$$

which can occur only if $\mathbb{P}[A] \in \{0, 1\}$. ■

2 Integration and Expectation

2.1 Construction of the integral

Let (S, Σ, μ) be a measure space. We denote by $\mathbb{1}_A$ the indicator of A, i.e.,

$$\mathbb{1}_A(s) = \begin{cases} 1, & \text{if } s \in A \\ 0, & \text{o.w.} \end{cases}$$

DEF 2.12 (Simple functions) *A* simple function *is a function of the form*

$$f = \sum_{k=1}^m a_k \mathbb{1}_{A_k},$$

where $a_k \in [0, +\infty]$ *and* $A_k \in \Sigma$ *for all* k. *We denote the set of all such functions by* SF^+. *We define the* integral of f *by*

$$\mu(f) = \sum_{k=1}^{m} a_k \mu(A_k) \leq +\infty.$$

The following is (somewhat tedious but) immediate. (Exercise.)

PROP 2.13 (Properties of simple functions) *Let* $f, g \in \mathrm{SF}^+$.

1. *If* $\mu(f \neq g) = 0$, *then* $\mu(f) = \mu(g)$. *(Hint: Rewrite* f *and* g *over the same disjoint sets.)*
2. *For all* $c \geq 0$, $f + g, cf \in \mathrm{SF}^+$ *and*

   $$\mu(f + g) = \mu(f) + \mu(g), \quad \mu(cf) = c\mu(f).$$

   (Hint: This one is obvious by definition.)
3. *If* $f \leq g$ *then* $\mu(f) \leq \mu(g)$. *(Hint: Show that* $g - f \in \mathrm{SF}^+$ *and use linearity.)*

The main definition and theorem of integration theory follows.

DEF 2.14 (Non-negative functions) *Let* $f \in (\mathrm{m}\Sigma)^+$. *Then the* integral of f *is defined by*

$$\mu(f) = \sup\{\mu(h) \,:\, h \in \mathrm{SF}^+,\ h \leq f\}.$$

THM 2.15 (Monotone convergence theorem) *If* $f_n \in (\mathrm{m}\Sigma)^+$, $n \geq 1$, *with* $f_n \uparrow f$ *then*

$$\mu(f_n) \uparrow \mu(f).$$

Many theorems in integration follow from the monotone convergence theorem. In that context, the following approximation is useful.

DEF 2.16 (Staircase function) *For* $f \in (\mathrm{m}\Sigma)^+$ *and* $r \geq 1$, *the* r-th staircase function $\alpha^{(r)}$ *is*

$$\alpha^{(r)}(x) = \begin{cases} 0, & \text{if } x = 0, \\ (i-1)2^{-r}, & \text{if } (i-1)2^{-r} < x \leq i2^{-r} \leq r, \\ r, & \text{if } x > r, \end{cases}$$

We let $f^{(r)} = \alpha^{(r)}(f)$. *Note that* $f^{(r)} \in \mathrm{SF}^+$ *and* $f^{(r)} \uparrow f$.

Using the previous definition, we get for example the following properties. (Exercise.)

PROP 2.17 (Properties of non-negative functions) *Let $f, g \in (\mathrm{m}\Sigma)^+$.*

1. *If $\mu(f \neq g) = 0$, then $\mu(f) = \mu(g)$.*
2. *For all $c \geq 0$, $f + g, cf \in (\mathrm{m}\Sigma)^+$ and*
$$\mu(f+g) = \mu(f) + \mu(g), \quad \mu(cf) = c\mu(f).$$
3. *If $f \leq g$ then $\mu(f) \leq \mu(g)$.*

2.2 Definition and properties of expectations

We can now define expectations. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. For a function f, let f^+ and f^- be the positive and negative parts of f, i.e.,

$$f^+(s) = \max\{f(s), 0\}, \quad f^-(s) = \max\{-f(s), 0\}.$$

DEF 2.18 (Expectation) *If $X \geq 0$ is a RV then we define the* expectation of X, *$\mathbb{E}[X]$, as the integral of X over $\mathbb{P}$. In general, if*

$$\mathbb{E}|X| = \mathbb{E}[X^+] + \mathbb{E}[X^-] < +\infty,$$

we let

$$\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-].$$

We denote the set of all such RVs by $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$.

The monotone-convergence theorem implies the following results. (Exercise.) We first need a definition.

DEF 2.19 (Convergence almost sure) *We say that $X_n \to X$ almost surely (a.s.) if*

$$\mathbb{P}[X_n \to X] = 1.$$

PROP 2.20 *Let X, Y, X_n, $n \geq 1$, be RVs on $(\Omega, \mathcal{F}, \mathbb{P})$.*

1. *(MON) If $0 \leq X_n \uparrow X$, then $\mathbb{E}[X_n] \uparrow \mathbb{E}[X] \leq +\infty$.*
2. *(FATOU) If $X_n \geq 0$, then $\mathbb{E}[\liminf_n X_n] \leq \liminf_n \mathbb{E}[X_n]$.*

3. *(DOM) If $|X_n| \leq Y$, $n \geq 1$, with $\mathbb{E}[Y] < +\infty$ and $X_n \to X$ a.s., then*

$$\mathbb{E}|X_n - X| \to 0,$$

and, hence,

$$\mathbb{E}[X_n] \to \mathbb{E}[X].$$

(Indeed,

$$\begin{aligned} |\mathbb{E}[X_n] - \mathbb{E}[X]| &= |\mathbb{E}[X_n - X]| \\ &= |\mathbb{E}[(X_n - X)^+] - \mathbb{E}[(X_n - X)^-]| \\ &\leq \mathbb{E}[(X_n - X)^+] + \mathbb{E}[(X_n - X)^-] \\ &= \mathbb{E}|X_n - X|.) \end{aligned}$$

4. *(SCHEFFE) If $X_n \to X$ a.s. and $\mathbb{E}|X_n| \to \mathbb{E}|X|$ then*

$$\mathbb{E}|X_n - X| \to 0.$$

5. *(BDD) If $X_n \to X$ a.s. and $|X_n| \leq K < +\infty$ for all n then*

$$\mathbb{E}|X_n - X| \to 0.$$

Proof: We only prove (FATOU). To use (MON) we write the $\liminf$ as an increasing limit. Letting $Z_k = \inf_{n \geq k} X_n$, we have

$$\liminf_n X_n = \uparrow \lim_k Z_k,$$

so that by (MON)

$$\mathbb{E}[\liminf_n X_n] = \uparrow \lim_k \mathbb{E}[Z_k].$$

For $n \geq k$ we have $X_n \geq Z_k$ so that $\mathbb{E}[X_n] \geq \mathbb{E}[Z_k]$ hence

$$\mathbb{E}[Z_k] \leq \inf_{n \geq k} \mathbb{E}[X_n].$$

Hence

$$\mathbb{E}[\liminf_n X_n] \leq \uparrow \lim_k \inf_{n \geq k} \mathbb{E}[X_n].$$

■

The following results are well-known.

DEF 2.21 (Space $\mathcal{L}^2$) *We denote the set of all RVs X with $\mathbb{E}[X^2] < +\infty$ by* $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$.

THM 2.22 (Cauchy-Schwarz inequality) *If $X, Y \in \mathcal{L}^2$ and $XY \in \mathcal{L}^1$ then*

$$\mathbb{E}|XY| \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}.$$

THM 2.23 (Jensen's inequality) *Let $h : G \to \mathbb{R}$ be a convex function on an open interval G such that $\mathbb{P}[X \in G] = 1$ and $X, h(X) \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ then*

$$\mathbb{E}[h(X)] \geq h(\mathbb{E}[X]).$$

2.3 Computing expected values

The following result is useful for computing expectations.

THM 2.24 (Change-of-variables formula) *Let X be a RV with law $\mathcal{L}$. If $f : \mathbb{R} \to \mathbb{R}$ is such that $f \geq 0$ or $\mathbb{E}|f(X)| < +\infty$ then*

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}} f(y)\mathcal{L}(\mathrm{d}y).$$

Proof: We use the standard machinery.

1. For $f = \mathbb{1}_B$ with $B \in \mathcal{B}$,

$$\mathbb{E}[\mathbb{1}_B(X)] = \mathcal{L}(B) = \int_{\mathbb{R}} \mathbb{1}_B(y)\mathcal{L}(\mathrm{d}y).$$

2. If $f = \sum_{k=1}^m a_k \mathbb{1}_{A_k}$ is a simple function, then by (LIN)

$$\mathbb{E}[f(X)] = \sum_{k=1}^m a_k \mathbb{E}[\mathbb{1}_{A_k}(X)] = \sum_{k=1}^m a_k \int_{\mathbb{R}} \mathbb{1}_{A_k}(y)\mathcal{L}(\mathrm{d}y) = \int_{\mathbb{R}} f(y)\mathcal{L}(\mathrm{d}y).$$

3. Let $f \geq 0$ and approximate f by a sequence $\{f_n\}$ of increasing simple functions. By (MON)

$$\mathbb{E}[f(X)] = \lim_n \mathbb{E}[f_n(X)] = \lim_n \int_{\mathbb{R}} f_n(y)\mathcal{L}(\mathrm{d}y) = \int_{\mathbb{R}} f(y)\mathcal{L}(\mathrm{d}y).$$

4. Finally, assume that f is such that $\mathbb{E}|f(X)| < +\infty$. Then by (LIN)

$$\begin{aligned}
\mathbb{E}[f(X)] &= \mathbb{E}[f^+(X)] - \mathbb{E}[f^-(X)] \\
&= \int_{\mathbb{R}} f^+(y)\mathcal{L}(\mathrm{d}y) - \int_{\mathbb{R}} f^-(y)\mathcal{L}(\mathrm{d}y) \\
&= \int_{\mathbb{R}} f(y)\mathcal{L}(\mathrm{d}y).
\end{aligned}$$

■

2.4 Fubini's theorem

DEF 2.25 (Product measure) *Let (S_1, Σ_1) and (S_2, Σ_2) be measure spaces. Let $S = S_1 \times S_2$ be the Cartesian product of S_1 and S_2. For $i = 1, 2$, let $\pi_i : S \to S_i$ be the projection on the i-th coordinate, i.e.,*

$$\pi_i(s_1, s_2) = s_i.$$

The product σ-algebra $\Sigma = \Sigma_1 \times \Sigma_2$ *is defined as*

$$\Sigma = \sigma(\pi_1, \pi_2),$$

i.e., it is the smallest σ-algebra that makes coordinate maps measurable. It is generated by sets of the form

$$\pi_1^{-1}(B_1) = B_1 \times S_2, \quad \pi_2^{-1}(B_2) = S_1 \times B_2, \quad B_1 \in \Sigma_1, B_2 \in \Sigma_2.$$

We now define the product measure and state the celebrated Fubini's theorem. (A proof is sketched in the appendix below.)

THM 2.26 (Fubini's theorem) *For $F \in \Sigma$, let $f = \mathbb{1}_F$ and define*

$$\mu(F) \equiv \int_{S_1} I_1^f(s_1)\mu_1(\mathrm{d}s_1) = \int_{S_2} I_2^f(s_2)\mu_2(\mathrm{d}s_2),$$

where

$$I_1^f(s_1) \equiv \int_{S_2} f(s_1, s_2)\mu_2(\mathrm{d}s_2) \in \mathrm{b}\Sigma_1, \quad I_2^f(s_2) \equiv \int_{S_1} f(s_1, s_2)\mu_1(\mathrm{d}s_1) \in \mathrm{b}\Sigma_2.$$

(The equality and inclusions above are part of the statement.) The set function μ is a measure on (S, Σ) called the product measure *of μ_1 and μ_2 and we write $\mu = \mu_1 \times \mu_2$ and*

$$(S, \Sigma, \mu) = (S_1, \Sigma_1, \mu_1) \times (S_2, \Sigma_2, \mu_2).$$

Moreover μ is the unique measure on (S, Σ) for which

$$\mu(A_1 \times A_2) = \mu(A_1)\mu(A_2), \quad A_i \in \Sigma_i.$$

If $f \in (\mathrm{m}\Sigma)^+$ then

$$\mu(f) = \int_{S_1} I_1^f(s_1)\mu_1(\mathrm{d}s_1) = \int_{S_2} I_2^f(s_2)\mu_2(\mathrm{d}s_2),$$

where I_1^f, I_2^f are defined as before (i.e., as the sup *over bounded functions below). The same is valid if $f \in \mathrm{m}\Sigma$ and $\mu(|f|) < +\infty$.*

Some applications of Fubini's theorem follow.

THM 2.27 *Let X and Y be independent RVs with respective laws μ and ν. Let f and g be measurable functions such that $f, g \geq 0$ or $\mathbb{E}|f(X)|, \mathbb{E}|g(Y)| < +\infty$. Then*

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)].$$

Proof: From Fubini's theorem and the change-of-variables formula,

$$\begin{aligned} \mathbb{E}[f(X)g(Y)] &= \int_{\mathbb{R}^2} f(x)g(y)(\mu \times \nu)(\mathrm{d}x \times \mathrm{d}y) \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x)g(y)\mu(\mathrm{d}x) \right) \nu(\mathrm{d}y) \\ &= \int_{\mathbb{R}} (g(y)\mathbb{E}[f(X)])\, \nu(\mathrm{d}y) \\ &= \mathbb{E}[f(X)]\mathbb{E}[g(Y)]. \end{aligned}$$

■

DEF 2.28 (Density) *Let X be a RV with law μ. We say that X has density f_X if for all $B \in \mathcal{B}(\mathbb{R})$*

$$\mu(B) = \mathbb{P}[X \in B] = \int_B f_X(x)\lambda(\mathrm{d}x).$$

THM 2.29 (Convolution) *Let X and Y be independent RVs with distribution functions F and G. Then the distribution function H of $X + Y$ is*

$$H(z) = \int F(z - y)\mathrm{d}G(y).$$

This is called the convolution *of F and G. Moreover, if X and Y have densities f and g, then $X + Y$ has density*

$$h(z) = \int f(z - y)g(y)\mathrm{d}y.$$

Proof: From Fubini's theorem, denoting the laws of X and Y by μ and ν,

$$\begin{aligned}
\mathbb{P}[X+Y\leq z] &= \int\int \mathbb{1}_{\{x+y\leq z\}}\mu(\mathrm{d}x)\nu(\mathrm{d}y)\\
&= \int F(z-y)\nu(\mathrm{d}y)\\
&= \int F(z-y)\mathrm{d}G(y)\\
&= \int \left(\int_{-\infty}^{z} f(x-y)\mathrm{d}x\right)\mathrm{d}G(y)\\
&= \int_{-\infty}^{z} \left(\int f(x-y)\mathrm{d}G(y)\right)\mathrm{d}x\\
&= \int_{-\infty}^{z} \left(\int f(x-y)g(y)\mathrm{d}y\right)\mathrm{d}x.
\end{aligned}$$

■

Further reading

More background on measure theory [Dur10, Appendix A].

References

[Dur10] Rick Durrett. *Probability: theory and examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2010.

[Wil91] David Williams. *Probability with martingales*. Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 1991.

A Proof of Fubini's Theorem

We need a more powerful variant of the standard machinery used in Theorem 2.24.

THM 2.30 (Monotone-class theorem) *Let $\mathcal{H}$ be a class of bounded functions from a set S to $\mathbb{R}$ satisfying:*

1. *$\mathcal{H}$ is a vector space over $\mathbb{R}$.*
2. *The constant 1 is an element of $\mathcal{H}$.*

3. *If $(f_n)_n$ is a sequence of non-negative functions in $\mathcal{H}$ such that $f_n \uparrow f$ where f is a bounded function on S, then $f \in \mathcal{H}$.*

Then if $\mathcal{H}$ contains the indicator function of every set in some π-system $\mathcal{I}$, then $\mathcal{H}$ contains every bounded $\sigma(\mathcal{I})$-measurable function on S.

The proof is omitted.

We begin with two lemmas (which are proved below).

LEM 2.31 *Let $\mathcal{H}$ denote the class of functions $f : S \to \mathbb{R}$ which are in $\mathrm{b}\Sigma$ and are such that*

1. *for each $s_1 \in S_1$, the map $s_2 \mapsto f(s_1, s_2)$ is Σ_2-measurable on S_2,*
2. *for each $s_2 \in S_2$, the map $s_1 \mapsto f(s_1, s_2)$ is Σ_1-measurable on S_1.*

Then $\mathcal{H} = \mathrm{b}\Sigma$.

Then define, for $f \in \mathrm{b}\Sigma$,

$$I_1^f(s_1) = \int_{S_2} f(s_1, s_2)\mu_2(\mathrm{d}s_2), \quad I_2^f(s_2) = \int_{S_1} f(s_1, s_2)\mu_1(\mathrm{d}s_1).$$

LEM 2.32 *Let $\mathcal{H}'$ be the class of elements in $\mathrm{b}\Sigma$ such that the following property holds:*

1. *$I_1^f \in \mathrm{b}\Sigma_1$ and $I_2^f \in \mathrm{b}\Sigma_2$,*
2. *we have*
$$\int_{S_1} I_1^f(s_1)\mu_1(\mathrm{d}s_1) = \int_{S_2} I_2^f(s_2)\mu_2(\mathrm{d}s_2).$$

Then $\mathcal{H}' = \mathrm{b}\Sigma$.

We are now ready to prove the two lemmas above.
Proof: We begin with the first lemma. Let

$$\mathcal{I} = \{B_1 \times B_2 \,:\, B_i \in \Sigma_i\}$$

be a π-system generating Σ. Note that if $A \in \mathcal{I}$ then $\mathbb{1}_A \in \mathcal{H}$ since, for fixed s_1, $\mathbb{1}_A$ reduces to an indicator on S_2. The assumptions of the Monotone-class theorem are satisfied by the standard properties of measurable functions. (Note that, for fixed s_1, a sum of measurable functions is measurable, and so is the limit.) Therefore, $\mathcal{H} = \mathrm{b}\Sigma$.

The second lemma follows in the same way. Note that for $A = A_1 \times A_2 \in \mathcal{I}$ and $f = \mathbb{1}_A$

$$I_1^f(s_1) = \mu_2(A_2)\mathbb{1}_{A_1}(s_1), \qquad \int_{S_1} I_1^f(s_1)\mu_1(\mathrm{d}s_1) = \mu_2(A_2)\mu_1(A_1),$$

and similarly interchanging 1 and 2. The assumptions of the Monotone-class theorem are satisfied by (LIN) and (MON). That concludes the proof. ■

Finally, we obtain Fubini's theorem.

THM 2.33 (Fubini's theorem) *For $F \in \Sigma$, let $f = \mathbb{1}_F$ and define*

$$\mu(F) \equiv \int_{S_1} I_1^f(s_1)\mu_1(\mathrm{d}s_1) = \int_{S_2} I_2^f(s_2)\mu_2(\mathrm{d}s_2),$$

where

$$I_1^f(s_1) \equiv \int_{S_2} f(s_1, s_2)\mu_2(\mathrm{d}s_2) \in \mathrm{b}\Sigma_1, \quad I_2^f(s_2) \equiv \int_{S_1} f(s_1, s_2)\mu_1(\mathrm{d}s_1) \in \mathrm{b}\Sigma_2.$$

(The equality and inclusions above are part of the statement.) The set function μ is a measure on (S, Σ) called the product measure *of μ_1 and μ_2 and we write $\mu = \mu_1 \times \mu_2$ and*

$$(S, \Sigma, \mu) = (S_1, \Sigma_1, \mu_1) \times (S_2, \Sigma_2, \mu_2).$$

Moreover μ is the unique measure on (S, Σ) for which

$$\mu(A_1 \times A_2) = \mu(A_1)\mu(A_2), \quad A_i \in \Sigma_i.$$

If $f \in (\mathrm{m}\Sigma)^+$ then

$$\mu(f) = \int_{S_1} I_1^f(s_1)\mu_1(\mathrm{d}s_1) = \int_{S_2} I_2^f(s_2)\mu_2(\mathrm{d}s_2),$$

where I_1^f, I_2^f are defined as before (i.e., as the sup *over bounded functions below). The same is valid if $f \in \mathrm{m}\Sigma$ and $\mu(|f|) < +\infty$.*

Proof: The fact that μ is a measure follows from (LIN) and (MON). The uniqueness follows from the Uniqueness lemma. The second follows from the previous lemma, the staircase approximation and (MON). ■