

Lecture 3: Integration, convergence theorems, and densities

Lecturer: Yiqiao Zhong

Report an issue

Last update: October 19, 2026

These notes support STAT 709 at the University of Wisconsin–Madison. Use them alongside the cited textbooks; corrections and questions are welcome.

Lecture 2 supplied measurable sets, measures, and random variables. We now build integration with respect to a measure and study the theorems that make the construction useful. The main practical question is:

When may a limit, sum, derivative, or iterated integral be exchanged with integration?

The answer is never “because the expressions look well behaved.” Each exchange requires assumptions that control signs, monotonicity, or magnitude.

For example, an average loss can be written as an integral against an empirical measure, a population risk is an integral against a probability law, and a likelihood calculation may require differentiating an integral. The same notation should handle a finite sum, a continuous density, and a law containing both point masses and a continuous part. We will build this common language first, then ask which assumptions justify each exchange. The shrinking-spike example below returns to Lecture 2’s warning that pointwise limits need not preserve area.

Recap from earlier lectures

From Lecture 2: measures, measurable maps, laws

Reminder only: this block introduces no new exam requirements.

- **Measure space.** A measure assigns countably additive mass to measurable sets. Used in Section 1.
- **Measurability.** Inverse images of measurable target sets must be measurable. Used in Section 1.
- **Law of a random variable.** $P_X(B) = P(X \in B)$ transfers the probability model to the value space. Used in Section 1.
- **Continuity of measures.** Masses converge along increasing sets, and along decreasing sets of initially finite mass. The set-indicator analogue of Section 2’s integral convergence theorems.
- **Product measure.** Rectangle masses multiply; the product sigma-algebra also contains nonrectangular sets. Used in Section 3.

Learning goals

After this lecture, you should be able to:

- interpret integration with respect to Lebesgue, counting, and probability measures;
- determine whether a measurable function is integrable;
- state and apply the monotone convergence theorem (MCT), dominated convergence theorem (DCT), and Fatou’s lemma;
- distinguish Tonelli’s theorem from Fubini’s theorem and use each with the correct assumptions;
- diagnose standard failures of limit and order exchanges;

- recognize an ordinary density or mass function as a Radon–Nikodym derivative; and
- use a likelihood ratio to change the measure under an expectation.

Priority guide

High Core concepts and calculations to master; directly examined.

Mid Supporting concepts useful in other lectures; in scope for examinations.

Low Enrichment supporting understanding; will not be examined.

Further reading Optional self-study, references, extensions, and historical notes.

A star (*) marks optional material; priority labels give the scope of each block. Lecture-specific exemptions and the optional status of Part VI override the general exam guide.

1 Abstract integration

High priority *Definitions and integral rules used throughout statistics*

Throughout this section, $(\Omega, \mathcal{F}, \nu)$ is a measure space. We use

$$\mathbf{1}_A(\omega) = \mathbf{1}\{\omega \in A\}$$

for an indicator function. The defining calculation is

$$\int \mathbf{1}_A d\nu = \nu(A).$$

Integration extends this measure-weighted sum in three steps.

Definition 1 (Integral of a nonnegative simple function). *A nonnegative **simple function** has a representation*

$$\phi = \sum_{i=1}^k a_i \mathbf{1}_{A_i},$$

where $0 \leq a_i < \infty$ and $A_1, \dots, A_k \in \mathcal{F}$ may be taken pairwise disjoint. Define

$$\int \phi d\nu = \sum_{i=1}^k a_i \nu(A_i).$$

We use the convention $0 \cdot \infty = 0$. The value does not depend on the chosen representation; the common-refinement argument in Section 6 explains why.

Low priority *Why do we set zero times infinity equal to zero?*

Zero height contributes no area. In Definition 1, $0\mathbf{1}_A$ is the zero function even when $\nu(A) = \infty$. Its contribution must therefore be $0\nu(A) = 0$. Otherwise, adding a zero term could change the integral of the same simple function.

Zero measure contributes no mass. The nonnegative extension below also gives integral zero to a function that is $+\infty$ on a measurable null set N and zero elsewhere: every finite simple function beneath it has integral zero. This agrees with $\infty \cdot \nu(N) = \infty \cdot 0 = 0$. Countably many null sets still have null union, since $\nu(\bigcup_j N_j) \leq \sum_j \nu(N_j) = 0$, so the same reasoning applies on that union. This supports the null-set intuition; the immediate reason in Definition 1 is the zero-height calculation.

See John K. Hunter's *Measure Theory*, Definition 4.1 (p. 39) and Proposition 4.15 (p. 45). This is an integration convention, not a rule for evaluating an indeterminate product of limits.

High priority *Nonnegative and signed integrals*

Definition 2 (Integral of a nonnegative function). *For a nonnegative measurable function $f : \Omega \rightarrow [0, \infty]$, define*

$$\int f \, d\nu = \sup \left\{ \int \phi \, d\nu : 0 \leq \phi \leq f, \phi \text{ simple and measurable} \right\}.$$

The value may be $+\infty$.

Every nonnegative measurable f can be approximated from below by simple functions:

$$0 \leq \phi_1 \leq \phi_2 \leq \cdots \uparrow f.$$

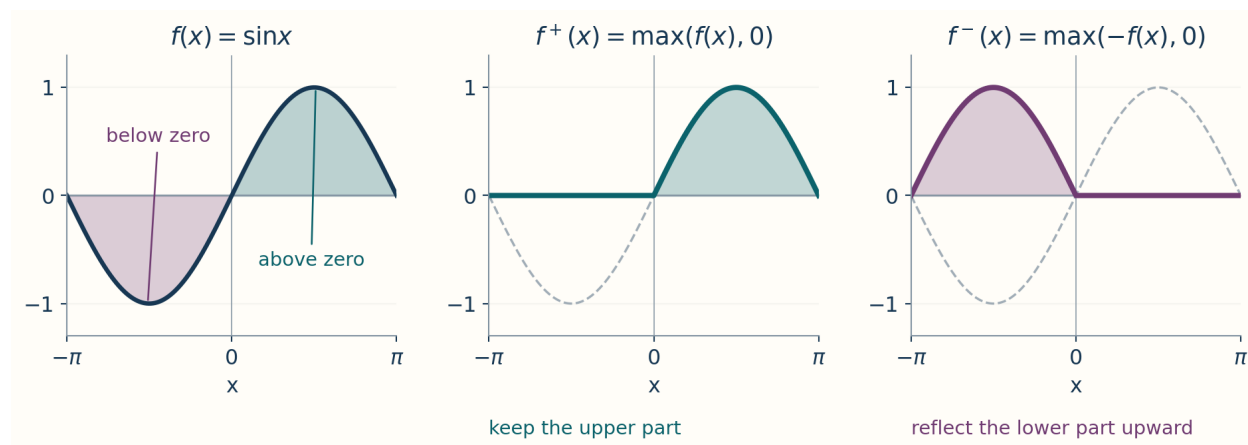
This fact connects the definition to the monotone convergence theorem. An explicit dyadic construction and its verification are given in the final proof details.

For a real-valued measurable function f , define

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

Then

$$f = f^+ - f^-, \quad |f| = f^+ + f^-.$$



A picture of positive and negative parts. For $f(x) = \sin x$ on $[-\pi, \pi]$, f^+ keeps the part above the horizontal axis and sets the rest to zero. The function f^- reflects the part below the axis upward and sets the rest to zero. **Both parts are nonnegative**; the minus sign appears when we reconstruct $f = f^+ - f^-$. Adding the two parts instead gives $|f|$.

Definition 3 (Integrability). *A measurable function f is **integrable** with respect to ν if*

$$\int |f| \, d\nu < \infty.$$

Equivalently, both $\int f^+ \, d\nu$ and $\int f^- \, d\nu$ are finite. In this case,

$$\int f \, d\nu = \int f^+ \, d\nu - \int f^- \, d\nu.$$

Exercise prompt: signed area. On $[0, 1]$ with Lebesgue measure, let $f = 3\mathbf{1}_{[0,1/4]} - 2\mathbf{1}_{(1/4,3/4]} + \mathbf{1}_{(3/4,1]}$. Find its positive and negative parts and compute $\int f$ and $\int |f|$. See Exercise 1 in Section 6 and its complete solution.

Requiring **absolute integrability** avoids the undefined expression $\infty - \infty$. More generally, a signed integral is well defined as an extended real number if at least one of $\int f^+$ and $\int f^-$ is finite. We reserve “integrable” for $\int |f| < \infty$; a conditionally convergent improper integral is not an integrable signed Lebesgue integral. For integrable f and g and scalars a, b , the integral satisfies

$$\begin{aligned}\int (af + bg) d\nu &= a \int f d\nu + b \int g d\nu, \\ f \leq g \text{ a.e.} &\implies \int f d\nu \leq \int g d\nu, \\ \left| \int f d\nu \right| &\leq \int |f| d\nu.\end{aligned}$$

If $f = g$ ν -almost everywhere, then their integrals agree whenever the integrals are defined.

Almost everywhere and almost surely. A property holds **almost everywhere (a.e.)** with respect to ν if it fails only on a **zero-measure set**, also called a **null set**: a measurable set N with $\nu(N) = 0$. We write ν -a.e. when the measure needs to be explicit. For a probability measure P , the corresponding phrase is **almost surely (a.s.)**. The exceptional set is understood to be contained in a measurable null set; all functions in our integral statements are measurable. Countably many a.e. statements can be enforced simultaneously because a countable union of null sets is null. The same conclusion need not hold for an uncountable family of null sets. For example, every singleton in $[0, 1]$ has Lebesgue measure zero, while their union has measure one.

Mid priority *Lebesgue, counting, and probability measures*

The abstract notation packages several familiar operations.

Lebesgue measure. When $\nu = m$ is **Lebesgue measure** on $\mathbb{R}^d$, write

$$\int_{\mathbb{R}^d} f(x) dx \quad \text{instead of} \quad \int_{\mathbb{R}^d} f dm.$$

If a function is Riemann integrable on a bounded rectangle, its Lebesgue integral has the same value (using the completed Lebesgue measure if needed). A precise proof reference is given in the Riemann comparison below.

Counting measure. On $\mathbb{N}$, let ν_{count} denote **counting measure**. Then

$$\int_{\mathbb{N}} a(k) d\nu_{\text{count}}(k) = \sum_{k=1}^{\infty} a(k)$$

whenever the nonnegative or absolutely convergent expression is defined. Thus the convergence theorems for integrals also provide theorems for interchanging infinite sums and limits.

Probability measure. Preview only: probability measure and CDF notation. We will define expectation in Lecture 4.

For a random variable X on $(\Omega, \mathcal{F}, P)$ and a Borel function h ,

$$\mathbb{E}[h(X)] = \int_{\Omega} h(X(\omega)) dP(\omega) = \int_{\mathbb{R}} h(x) dP_X(x),$$

whenever either side is well defined. The second equality is the **law of the unconscious statistician** or **pushforward identity**. It says that an expectation may be computed either on the original sample space or from the law of X . Its indicator-to-function proof is supplied in the final proof details.

For a CDF F_X , the notation

$$\int h(x) dF_X(x)$$

is shorthand for integration with respect to the probability measure P_X determined by F_X ; it does not mean ordinary integration with respect to the pointwise derivative of F_X .

Instructor's Note *An informal picture: what carries the weight?*

Think of $d\nu$ as specifying where the weight lies. Counting measure puts one unit at each integer, a probability law distributes total weight one, and Lebesgue measure assigns length or volume. The integral combines the function values with those weights. For an empirical measure $P_n = n^{-1} \sum_{i=1}^n \delta_{x_i}$, this gives

$$\int h dP_n = \frac{1}{n} \sum_{i=1}^n h(x_i).$$

Indeed, integrating against δ_{x_i} evaluates h at x_i , and finite linearity combines the point masses. This is a mental picture, not a definition of an infinitesimal $d\nu$. The measurable-function construction above makes it precise even for a mixture of atoms and continuous mass.

For a concrete comparison on $[0, 1]$, take $h(x) = x$ and the measures $D = (\delta_0 + \delta_1)/2$, m , and $M = (\delta_0 + m)/2$. Then $\int h dD = 1/2$, $\int h dm = 1/2$, and $\int h dM = 1/4$: the atom at zero contributes zero, and the uniform continuous part contributes its mass times $\int_0^1 x dx = 1/2$. The same function is weighted differently. A general measure need not be either wholly discrete or wholly described by a density against length.

1.1 Connection between Lebesgue integrals and Riemann integrals

Low priority *Riemann comparison and the fundamental theorem—not examined*

Let m be Lebesgue measure on $[a, b]$ (completed when needed). The abstract integral preserves the familiar calculus rules, while enlarging the class of integrable functions. The following four connections make this precise.

1. **Agreement where the Riemann integral exists.** If f is Riemann integrable on $[a, b]$, it is Lebesgue integrable and

$$\text{Riemann } \int_a^b f(x) dx = \int_{[a,b]} f dm.$$

Indeed, the Lebesgue integral lies between the lower and upper step-function sums, whose difference tends to zero. See [3, Theorem 11.33] for a full proof.

2. **The criterion for Riemann integrability.** A **bounded** function on $[a, b]$ is Riemann integrable if and only if it is **continuous almost everywhere**: its set of *points of discontinuity* has Lebesgue measure zero. Piecewise continuous functions with finitely many pieces satisfy this criterion. A proof is given in Labute's notes on Lebesgue's integrability criterion.
3. **Differentiating an indefinite integral.** If f is Lebesgue integrable on $[a, b]$ and

$$F(x) = \int_a^x f(t) dt,$$

then $F'(x) = f(x)$ for Lebesgue-a.e. $x \in (a, b)$. Thus an integral still recovers the integrand by differentiation, with an a.e. qualification. In fact, F is **absolutely continuous**; this function property is specified below.

4. **Integrating a derivative.** If F is differentiable at **every point** of $[a, b]$ (with one-sided derivatives at endpoints) and F' is Lebesgue integrable, then

$$F(b) - F(a) = \int_a^b F'(t) dt.$$

This is the valid everywhere-differentiability statement in the original notes; see [2, Theorem 7.21]. It must not be weakened to “differentiable almost everywhere” without another condition.

Instructor’s Note *An informal picture: which changes are negligible?*

Changing a measurable function on a null set preserves its Lebesgue integral when defined, but it can destroy Riemann integrability. For example, $\mathbf{1}_{\mathbb{Q} \cap [0,1]}$ equals the continuous zero function a.e. and has Lebesgue integral zero. It is discontinuous at *every* point, since each interval contains both rational and irrational numbers; its lower Riemann sums are zero and its upper sums are one. Therefore “continuous a.e.” and “equal a.e. to a continuous function” are different statements.

The role of absolute continuity. A function F is absolutely continuous on $[a, b]$ if for every $\epsilon > 0$ there is $\delta > 0$ such that every finite disjoint collection of intervals $(a_j, b_j) \subset [a, b]$ with $\sum_j (b_j - a_j) < \delta$ satisfies $\sum_j |F(b_j) - F(a_j)| < \epsilon$. In words, small total interval length forces small total change in F . This function property is distinct from absolute continuity of measures, defined later. For an absolutely continuous F , its derivative exists a.e., is integrable, and $F(x) = F(a) + \int_a^x F'(t) dt$; conversely, every indefinite integral of an integrable function is absolutely continuous. For these statements, including differentiating an indefinite integral above, see Zakon, Chapter 9, Section 2, Theorems 2–3. The **Cantor function**¹ shows why a.e. differentiability alone is insufficient: its derivative is zero a.e., but its total change on $[0, 1]$ is one.

Notation in several dimensions. The expressions $\int f dx$, $\int f(x) dx$, and $\int_{\mathbb{R}^d} f(x) dx$ all mean integration with respect to Lebesgue measure on the indicated domain. Multiple integral signs may make coordinates explicit; changing the order requires the hypotheses in Section 3. For example, $f \geq 0$ Lebesgue-a.e. means $m(\{x : f(x) < 0\}) = 0$. Under a probability measure the same null-set convention is called “almost surely”; neither phrase means “at every point.”

2 Convergence theorems

High priority *MCT and DCT: assumptions, conclusions, and calculations*

Theorem 1 (Monotone convergence theorem (MCT)). *Let $f_1, f_2, \dots$ and f be nonnegative measurable functions. If $f_n \uparrow f$ ν -a.e., then*

$$\int f d\nu = \lim_{n \rightarrow \infty} \int f_n d\nu,$$

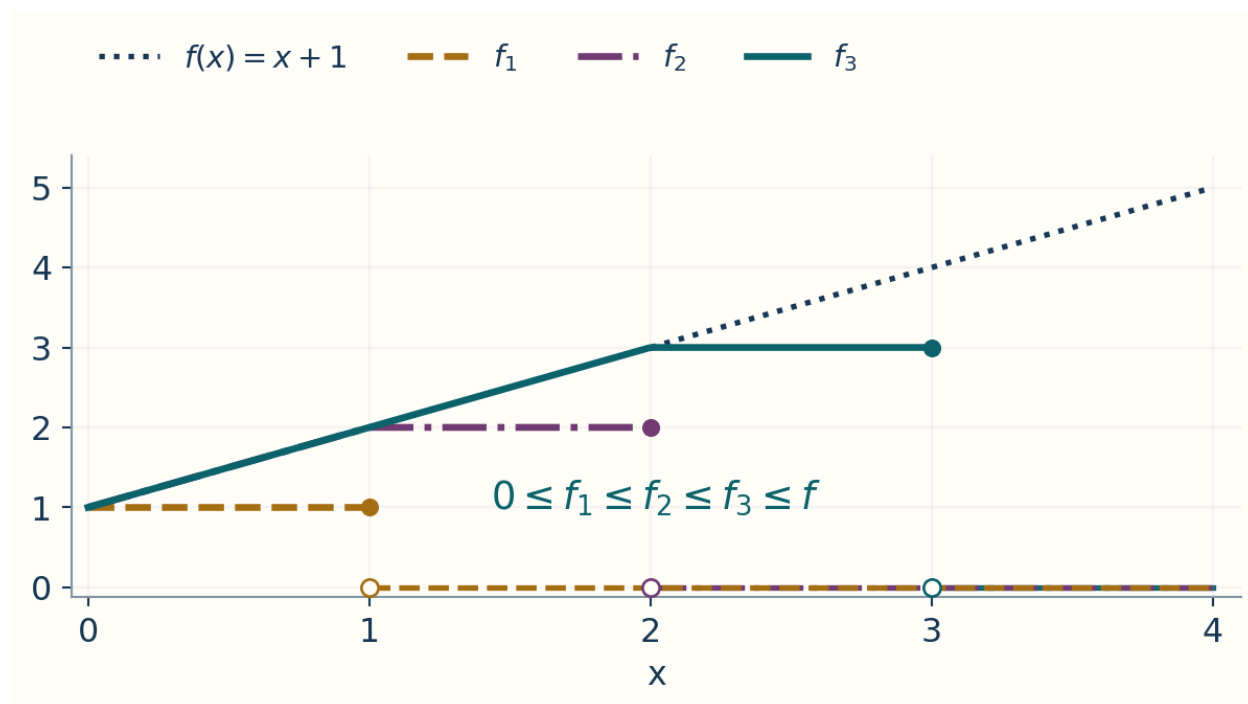
where either side may equal $+\infty$.

¹ The Cantor function is a continuous, nondecreasing function on $[0, 1]$ that rises from zero to one while remaining constant on every interval removed in the middle-thirds construction of the Cantor set. Those removed intervals have total length one, so its derivative is zero a.e.; the function is not absolutely continuous.

MCT requires **nonnegativity and monotone increase**, but it does not require an integrable upper bound. A canonical application is truncation:

$$f_n = \min(f, n)\mathbf{1}_{E_n} \uparrow f$$

for a nonnegative measurable f and measurable increasing sets $E_n \uparrow \Omega$. No finiteness of $\nu(E_n)$ is needed for this pointwise construction; the integrals may be infinite. For a full proof of MCT, see [1, Theorem 1.5.7].



A picture of truncation. For $\Omega = [0, \infty)$, take $f(x) = x + 1$ and $E_n = [0, n]$. Then $f_n(x) = \min(x + 1, n)\mathbf{1}_{[0, n]}(x)$ raises the height cap and enlarges the visible domain as n grows. At each fixed x , neither change lowers the value: $f_n(x) \leq f_{n+1}(x)$, and eventually $f_n(x) = x + 1$. The first three curves make this pointwise ordering visible. The same reasoning applies to the general truncation above: both nonnegative factors increase, and each point eventually belongs to E_n .

A picture of monotone increase. On $[0, 1]$, the functions $u_n(x) = (1 - 2^{-n})x$ increase to $u(x) = x$. Their integrals are $(1 - 2^{-n})/2$, increasing to $1/2 = \int_0^1 x dx$. Each new graph adds nonnegative area beneath the same limiting graph. This is one illustration of MCT, not a proof for all measurable functions.

Example 1 (Interchanging a nonnegative sum and an integral). If $g_1, g_2, \dots \geq 0$ are measurable, then the partial sums $s_n = \sum_{k=1}^n g_k$ increase to $\sum_{k=1}^\infty g_k$. MCT gives

$$\int \sum_{k=1}^\infty g_k d\nu = \sum_{k=1}^\infty \int g_k d\nu,$$

allowing the value $+\infty$.

Exercise prompt: a geometric series under the integral. On $[0, 1]$ with Lebesgue measure, set $g_k(x) = x^{k-1}(1 - x)$ for $k \geq 1$. Evaluate $\int_0^1 \sum_{k \geq 1} g_k(x) dx$ by exchanging sum and integral, and justify the exchange. See the application in Exercise 3 in Section 6 and its complete solution.

Theorem 2 (Dominated convergence theorem (DCT)). *Let $f_1, f_2, \dots$ and f be real-valued measurable functions such that $f_n \rightarrow f$ ν -a.e. If there is an integrable function g satisfying*

$$|f_n| \leq g \quad \nu\text{-a.e. for every } n,$$

then f and every f_n are integrable and

$$\lim_{n \rightarrow \infty} \int f_n d\nu = \int f d\nu.$$

In fact, $\int |f_n - f| d\nu \rightarrow 0$.

The **dominating function** (or **integrable envelope**) must be **independent of n** and integrable on the whole domain. A reliable DCT calculation has three explicit lines:

pointwise limit: $f_n \rightarrow f$ a.e., domination: $ f_n \leq g$ with $\int g < \infty$, conclusion: $\int f_n \rightarrow \int f$.
--

The L^1 **conclusion**² follows by applying the integral-convergence statement to $|f_n - f|$, whose envelope is $2g$ outside one null set. For a full DCT proof, see [1, Theorem 1.5.8]; the optional proof architecture explains the strategy.

Example 2 (A direct DCT calculation). *On $[0, 1]$, let*

$$f_n(x) = \frac{x^n}{1+x}.$$

For $0 \leq x < 1$, $f_n(x) \rightarrow 0$; the value at $x = 1$ is irrelevant to an a.e. statement. Also $0 \leq f_n(x) \leq 1$, and the constant function 1 is integrable on $[0, 1]$. Hence DCT yields

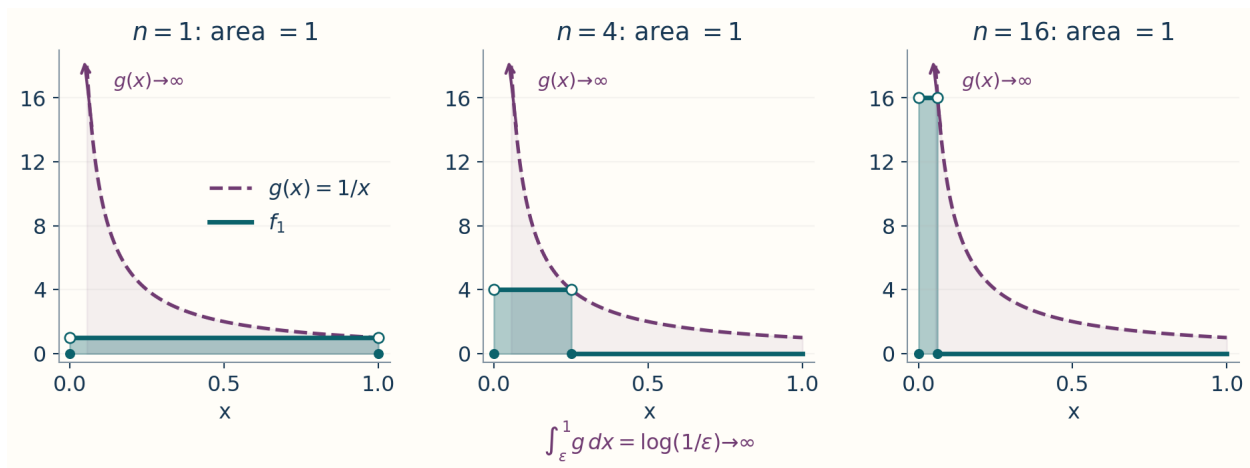
$$\lim_{n \rightarrow \infty} \int_0^1 \frac{x^n}{1+x} dx = 0.$$

Exercise prompt: find one envelope. For $f_n(x) = (1 + x/n)^n e^{-2x}$ on $[0, \infty)$ with Lebesgue measure, identify the pointwise limit and an integrable envelope independent of n , then determine $\lim_n \int_0^\infty f_n(x) dx$. See Exercise 2(4) in Section 6 and its complete solution.

Choosing the theorem.

- If $0 \leq f_n \uparrow f$, first try MCT.
- If $f_n \rightarrow f$ and there is one integrable envelope g , use DCT.
- If the functions are nonnegative but neither monotone nor dominated, Fatou's lemma may still give a one-sided bound.
- If none of the assumptions can be verified, do not exchange the limit and integral without another argument.

² An $L^1(\nu)$ function is a measurable function h with $\int |h| d\nu < \infty$. Convergence in L^1 means $\|f_n - f\|_1 := \int |f_n - f| d\nu \rightarrow 0$: the total absolute error, measured using ν , tends to zero. Taking absolute values prevents positive and negative errors from canceling. This is stronger than convergence of the integrals, since $|\int f_n d\nu - \int f d\nu| \leq \int |f_n - f| d\nu$. The displayed DCT conclusion states this property; the notation does not add a new hypothesis.



Example 3 (Why pointwise convergence is insufficient). On $[0, 1]$, let

$$f_n(x) = n\mathbf{1}_{(0, 1/n)}(x).$$

Then $f_n(x) \rightarrow 0$ for every $x \in [0, 1]$, but

$$\int_0^1 f_n(x) dx = 1.$$

A common dominating function does exist: set $g(0) = 0$ and $g(x) = 1/x$ for $0 < x \leq 1$. On the support of f_n , $x < 1/n$ implies $f_n(x) = n < 1/x = g(x)$; elsewhere $f_n = 0 \leq g$. But

$$\int_0^1 g(x) dx = \lim_{\epsilon \downarrow 0} \int_{\epsilon}^1 \frac{dx}{x} = \lim_{\epsilon \downarrow 0} \log(1/\epsilon) = \infty.$$

Thus this envelope cannot be used in **DCT**. **MCT fails for a different reason**: the sequence is not increasing. For example, at $x = 1/5$, the values for $n = 1, 4, 16$ are $1, 4, 0$.

In fact, there is no integrable envelope, even if domination is required only a.e. for each n . The non-integrability of our chosen g alone does not rule out a different envelope. To do so, discard the countable union of the exceptional null sets. For $0 < x < 1/4$ choose $n = \lfloor 1/(2x) \rfloor$; then $x < 1/n$ and $n \geq 1/(4x)$. Any such envelope must therefore satisfy $g(x) \geq 1/(4x)$ a.e. on $(0, 1/4)$, which has infinite integral. Here the unequal integrals establish failure of the exchange; failure of a theorem's assumptions alone would only mean that theorem cannot justify it.

Mid priority Fatou's lemma and differentiation under the integral

Theorem 3 (Fatou's lemma). If $f_1, f_2, \dots$ are nonnegative measurable functions, then

$$\int \liminf_{n \rightarrow \infty} f_n d\nu \leq \liminf_{n \rightarrow \infty} \int f_n d\nu.$$

To remember the direction, mass may escape in the limit but cannot appear from nowhere under non-negative integration. For $f_n = \mathbf{1}_{[n, n+1]}$ on $\mathbb{R}$,

$$\int \liminf_n f_n dx = 0 < 1 = \liminf_n \int f_n dx.$$

This strict inequality also shows why DCT fails: the natural envelope $\sup_n f_n$ has infinite integral.

A short proof exposes Fatou's connection to MCT. Define

$$h_m = \inf_{n \geq m} f_n.$$

Then $0 \leq h_m \uparrow \liminf_n f_n$. By MCT and monotonicity,

$$\int \liminf_n f_n d\nu = \lim_m \int h_m d\nu \leq \lim_m \inf_{n \geq m} \int f_n d\nu.$$

Theorem 4 (Differentiation under the integral sign). *Let $I \subset \mathbb{R}$ be an open interval, let f be real valued, and suppose $f(\cdot, \theta)$ is measurable for every $\theta \in I$. Assume:*

1. $f(\cdot, \theta_0)$ is integrable for some $\theta_0 \in I$;
2. on one set of full ν -measure, the map $\theta \mapsto f(\omega, \theta)$ is differentiable on I ; and
3. there is an integrable g such that, on that same full-measure set, $|\partial_\theta f(\omega, \theta)| \leq g(\omega)$ for every $\theta \in I$.

Then

$$F(\theta) = \int f(\omega, \theta) d\nu(\omega)$$

is differentiable on I and

$$F'(\theta) = \int \partial_\theta f(\omega, \theta) d\nu(\omega).$$

Define the derivative to be zero off the specified full-measure set; this gives a measurable version, as a pointwise limit of measurable difference quotients. The interval assumption lets the mean value theorem compare $f(\cdot, \theta)$ with the integrable base point. The full difference-quotient proof in Section 6 makes both steps explicit. In statistical models, this is one route to moving derivatives through likelihood integrals, but the boundary and domination conditions must be checked rather than assumed.

Exercise prompt: a Laplace derivative. Let $h \geq 0$ be measurable with $\int_0^\infty (1+x)h(x) dx < \infty$. For $F(\theta) = \int_0^\infty e^{-\theta x} h(x) dx$, $\theta > 0$, justify differentiation under the integral and determine $F'(\theta)$. See Exercise 8 in Section 6 and its complete solution.

Used again in regularity of the score.

3 Tonelli's and Fubini's theorems

High priority *Exchanging the order of integration*

Let $(\Omega_i, \mathcal{F}_i, \nu_i)$, $i = 1, 2$, be σ -finite measure spaces, and let f be $(\mathcal{F}_1 \otimes \mathcal{F}_2)$ -measurable. Recall that σ -finite means each space is a countable union of measurable sets of finite measure; $\mathcal{F}_1 \otimes \mathcal{F}_2$ is generated by measurable rectangles, and the product measure satisfies $(\nu_1 \otimes \nu_2)(A \times B) = \nu_1(A)\nu_2(B)$. These are the product-measure conventions introduced in Lecture 2.

Theorem 5 (Tonelli's theorem). *If $f \geq 0$, then*

$$\begin{aligned} \int_{\Omega_1 \times \Omega_2} f d(\nu_1 \otimes \nu_2) &= \int_{\Omega_2} \left[\int_{\Omega_1} f(\omega_1, \omega_2) d\nu_1(\omega_1) \right] d\nu_2(\omega_2) \\ &= \int_{\Omega_1} \left[\int_{\Omega_2} f(\omega_1, \omega_2) d\nu_2(\omega_2) \right] d\nu_1(\omega_1), \end{aligned}$$

where the common value may be $+\infty$.

Theorem 6 (Fubini's theorem). *If*

$$\int_{\Omega_1 \times \Omega_2} |f| d(\nu_1 \otimes \nu_2) < \infty,$$

then the sections are integrable for almost every fixed value of the other coordinate, both iterated integrals are finite, and the two iterated integrals equal the product-space integral.

The operational distinction. Recall the definition of integrability: it requires a finite integral of the absolute value. The distinction between the two theorems is:

	assumption	possible value
Tonelli	$f \geq 0$	$[0, \infty]$
Fubini	$\int f < \infty$	$\mathbb{R}$

For a signed function, checking only that one iterated integral exists is not enough; absolute integrability is the safe condition. For their proofs, see [1, Theorem 1.7.2], which treats both cases under the name Fubini's theorem.

Example 4 (A two-dimensional integral). *For $f(x, y) = e^{-(x+y)}$ on $[0, \infty)^2$, Tonelli applies because $f \geq 0$. Thus*

$$\begin{aligned} \int_0^\infty \int_0^\infty e^{-(x+y)} dx dy &= \int_0^\infty e^{-y} \left(\int_0^\infty e^{-x} dx \right) dy \\ &= 1. \end{aligned}$$

The result also proves that f is integrable, so Fubini applies afterward.

Example 5 (Sums are integrals). *With counting measure on both coordinates, Tonelli says that for $a_{ij} \geq 0$,*

$$\sum_{i=1}^\infty \sum_{j=1}^\infty a_{ij} = \sum_{j=1}^\infty \sum_{i=1}^\infty a_{ij},$$

allowing both sides to be infinite. For signed a_{ij} , the sufficient Fubini condition is

$$\sum_{i=1}^\infty \sum_{j=1}^\infty |a_{ij}| < \infty.$$

Exercise prompt: two ways to slice a region. For $f(x, y) = \mathbf{1}\{0 \leq x \leq y \leq 1\}$, compute its integral on $[0, 1]^2$ in both orders and state which theorem permits the exchange. See Exercise 4 in Section 6 and its complete solution.

Mid priority *What fails without nonnegativity or absolute integrability*

Define an array by

$$a_{ij} = \begin{cases} 1, & j = i, \\ -1, & j = i + 1, \\ 0, & \text{otherwise.} \end{cases}$$

Every row sums to zero, so

$$\sum_{i=1}^\infty \left(\sum_{j=1}^\infty a_{ij} \right) = 0.$$

The first column sums to 1, and every later column sums to zero, so

$$\sum_{j=1}^{\infty} \left(\sum_{i=1}^{\infty} a_{ij} \right) = 1.$$

There is no contradiction with Fubini because

$$\sum_{i,j} |a_{ij}| = \infty.$$

This example is the double-series analogue of conditionally convergent series: changing the order may change the value.

When one coordinate uses counting measure and the other uses Lebesgue measure, the same theorems justify sum-integral exchanges:

$$\int \sum_{k=1}^{\infty} f_k(x) dx = \sum_{k=1}^{\infty} \int f_k(x) dx$$

for nonnegative f_k , or for signed f_k satisfying $\sum_k \int |f_k| < \infty$.

4 Radon–Nikodym derivatives and statistical densities

Mid priority *Absolute continuity, densities, and likelihood ratios*

Definition 4 (Absolute continuity of measures). *Let λ and ν be measures on the same measurable space. We write*

$$\lambda \ll \nu$$

*and say that λ is **absolutely continuous with respect to ν** if*

$$\nu(A) = 0 \implies \lambda(A) = 0 \quad \text{for every } A \in \mathcal{F}.$$

Theorem 7 (Radon–Nikodym theorem). *Let λ and ν be σ -finite measures on $(\Omega, \mathcal{F})$. If $\lambda \ll \nu$, then there exists a nonnegative measurable function f such that*

$$\lambda(A) = \int_A f d\nu \quad \text{for every } A \in \mathcal{F}.$$

*The function is unique ν -a.e. and is called the **Radon–Nikodym derivative**, denoted by*

$$f = \frac{d\lambda}{d\nu}.$$

Density as reweighting. The defining equation says that the Radon–Nikodym (RN) derivative converts reference mass into target mass. On three points a, b, c , let ν assign masses $(1, 2, 3)$ and λ assign masses $(2, 1, 6)$. Their density ratio is $(2, 1/2, 2)$: multiply each reference mass by its ratio to recover the target. For example, on $A = \{a, c\}$, $\lambda(A) = 8 = 2\nu(\{a\}) + 2\nu(\{c\})$. These are finite measures, not normalized probability laws.

For a continuous example on $[0, 1]$, let $\nu = m$ and $\lambda(A) = \int_A 2x dx$. Then $d\lambda/dm = 2x$. The target mass of $A = [1/2, 1]$ is the shaded area $\int_{1/2}^1 2x dx = 3/4$, while the reference mass is its length $m(A) = 1/2$. A density value is a weight per unit reference measure; it is not itself the probability of a point.

For the existence and uniqueness proof, see [1, Theorem A.4.8]. When λ is a probability measure, the derivative depends on the reference measure.

- If ν is Lebesgue measure, $d\lambda/d\nu$ is an ordinary probability density function.
- If ν is counting measure on a countable set, $d\lambda/d\nu$ is a probability mass function.
- If $\nu = Q$ and $\lambda = P$ are probability measures, dP/dQ is a **likelihood ratio** or density ratio.

Changing the measure, not the coordinate. Recall the definition of absolute continuity of measures. Here P and Q are probability laws on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $P \ll Q$, and $X(x) = x$ is the coordinate map. Write $r = dP/dQ$. If h is Borel measurable and $\int |h| dP < \infty$, then

$$\mathbb{E}_P[h(X)] = \int_{\mathbb{R}} h(x) dP(x) = \int_{\mathbb{R}} h(x)r(x) dQ(x) = \mathbb{E}_Q[h(X)r(X)].$$

The same identity holds for nonnegative h , allowing $+\infty$. The indicator-to-function proof also shows $\int |h| dP = \int |h|r dQ$, so integrability transfers exactly. If P, Q are instead measures on an abstract sample space and X is a separate random variable, the weight is $r(\omega)$, not $r(X(\omega))$. This **change-of-measure identity** is used in **importance sampling**³ and likelihood-ratio testing⁴. It also appears throughout information theory.

Example 6 (Gaussian likelihood ratio). Let $P = N(\mu, 1)$ and $Q = N(0, 1)$, with respective Lebesgue densities p and q . Since $q(x) > 0$ for every x , any Q -null set is Lebesgue-null and hence P -null. Thus $P \ll Q$.

To verify that $dP/dQ = p/q$, for every Borel set A we have

$$\int_A \frac{p(x)}{q(x)} dQ(x) = \int_A p(x) dx = P(A).$$

The common Gaussian normalizing constant $1/\sqrt{2\pi}$ cancels, so

$$\frac{dP}{dQ}(x) = \frac{p(x)}{q(x)} = \frac{\exp\{-(x-\mu)^2/2\}}{\exp\{-x^2/2\}} = \exp\left\{\mu x - \frac{\mu^2}{2}\right\}.$$

Consequently,

$$\mathbb{E}_{N(\mu,1)}[h(X)] = \mathbb{E}_{N(0,1)}\left[h(X) \exp\left\{\mu X - \frac{\mu^2}{2}\right\}\right]$$

whenever $h(X)$ is integrable.

Example 7 (One density for a discrete-continuous mixture). Let

$$\mu = \frac{1}{2} \sum_{k=0}^{\infty} p_k \delta_k + \frac{1}{2} h(x) dx,$$

where $p_k \geq 0$, $\sum_k p_k = 1$, and h is a Lebesgue density. This law has neither a pure Lebesgue density nor a pure mass function on $\mathbb{R}$. Define the σ -finite dominating measure

$$\rho = m + \nu_{\text{count}, \mathbb{N}_0}, \quad \nu_{\text{count}, \mathbb{N}_0}(A) = \sum_{k=0}^{\infty} \mathbf{1}\{k \in A\}.$$

Then

$$\frac{d\mu}{d\rho}(x) = \begin{cases} \frac{1}{2}p_k, & x = k \in \mathbb{N}_0, \\ \frac{1}{2}h(x), & x \notin \mathbb{N}_0. \end{cases}$$

The value assigned to h at individual integers does not affect its Lebesgue integral. For the mixture illustration, choose $p_0 = 1$, $p_k = 0$ for $k \geq 1$, and $h = \mathbf{1}_{[0,1]}$. On $[0, 1]$, the reference ρ has length plus atoms at both

³ Importance sampling estimates an expectation under P using draws from another law Q and weights $r = dP/dQ$. For independent draws X_i from Q , the average $n^{-1} \sum_i h(X_i)r(X_i)$ has expectation $\mathbb{E}_P h(X)$ when the weighted variable is integrable; its accuracy also depends on how variable the weights are.

⁴ For two specified probability models with densities p and q against a common measure, likelihood-ratio testing compares $p(x)/q(x)$ at the observed data with a threshold. The ratio describes how the models weight the same observation; this lecture supplies the measure-theoretic identity, not a theory of choosing tests.

endpoints; the derivative is $1/2$ at zero and on $(0, 1)$, but zero at one. Thus the target has an atom of mass $1/2$ at zero and a continuous part of mass $1/2$, with no atom at one. To verify the derivative, integrate it over a Borel set A : the atomic part gives $\frac{1}{2} \sum_{k \in A} p_k$, and the Lebesgue part gives $\frac{1}{2} \int_A h(x) dx$. Their sum is $\mu(A)$. The reference measure is σ -finite because its mass on every bounded interval is finite. See Exercise 7 for a one-atom calculation and its complete solution.

Used again in dominated statistical models, likelihoods.

Mid priority Why the reference measure matters

A law concentrated on a curve. Return to Lecture 2's circle example: if U is uniform on $[0, 2\pi)$, then $(\cos U, \sin U)$ lies on the unit circle S with probability one. Its law is not absolutely continuous with respect to plane area: S has area zero. Indeed, annuli of radii $1 - \epsilon$ and $1 + \epsilon$ cover it and have area $4\pi\epsilon \rightarrow 0$. Against arc-length measure on S (the measure assigning each arc its ordinary length), however, this law has the constant density $1/(2\pi)$: an arc of length ℓ has probability $\ell/(2\pi)$. Failure of a density against area does not mean failure against every reference measure.

5 Technical extensions and further reading

Low priority How the convergence theorems fit together—not examined

The standard proof architecture is economical:

$$\text{simple-function construction} \implies \text{MCT} \implies \text{Fatou} \implies \text{DCT}.$$

For DCT, apply Fatou to the nonnegative functions $g + f_n$ and $g - f_n$. One obtains upper and lower bounds that force $\int f_n \rightarrow \int f$. This proof explains why an integrable envelope controls both positive and negative fluctuations.

There is also a useful *reverse Fatou* consequence. For real-valued measurable f_n with $f_n \leq g$ a.e. and integrable g ,

$$\limsup_{n \rightarrow \infty} \int f_n d\nu \leq \int \limsup_{n \rightarrow \infty} f_n d\nu,$$

The integrals are well defined in $[-\infty, \infty)$ because $f_n^+ \leq g^+$; the same holds for the limsup, with no $+\infty - \infty$ subtraction. Apply Fatou to $g - f_n$ and subtract from the finite number $\int g$.

Low priority Change of variables for densities—not examined

Suppose X has Lebesgue density f_X on $\mathbb{R}^k$ and $g : \mathbb{R}^k \rightarrow \mathbb{R}^k$ is Borel measurable. Let $A_1, \dots, A_m$ be pairwise disjoint open sets with $P(X \in \bigcup_j A_j) = 1$. Suppose each restriction $g|_{A_j}$ is a C^1 **diffeomorphism** onto its open image: it is a bijection and both it and its inverse have continuous first partial derivatives.⁵ Its Jacobian is nonsingular (has nonzero determinant). The images may overlap even though the source branches are disjoint. If g_j^{-1} denotes the inverse on $g(A_j)$, then the density of $Y = g(X)$ is

$$f_Y(y) = \sum_{j=1}^m f_X(g_j^{-1}(y)) |\det Dg_j^{-1}(y)| \mathbf{1}_{\{y \in g(A_j)\}}.$$

Each summand is defined to be zero outside $g(A_j)$; do not evaluate g_j^{-1} there. The indicator factors are essential: an inverse branch contributes only on its image. For a Borel set B , split $P(g(X) \in B)$ over the

⁵ A C^1 diffeomorphism is a change of coordinates whose forward and inverse maps are continuously differentiable. The Jacobian is the matrix of first partial derivatives; its determinant describes the local signed volume scaling. The density formula uses the absolute value of the inverse Jacobian determinant.

disjoint A_j and apply the ordinary change-of-variables theorem on each branch. This yields $\int_B f_Y(y) dy$ and proves the formula; the coverage assumption ensures no probability mass has been omitted. This formula is a special, coordinate-based way to compute a pushforward law; the measure-theoretic definition remains $P_Y = P_X \circ g^{-1}$.

Further reading *References and directions for self-study*

The integration and convergence results in this lecture are developed from a probability perspective in [1] and from a mathematical-statistics perspective in [4]. For a fuller real-analysis treatment of Lebesgue integration, absolute continuity, signed measures, and the Radon–Nikodym theorem, see [2].

The author’s freely available January 2019 fifth-edition draft of Durrett contains the numbered results cited here. Proof routes are Sections 1.4–1.5 for integral construction and convergence, Theorem 1.6.9 for pushforward expectations, Theorem 1.7.2 for product integration, and Theorem A.4.8 for RN derivatives.

Useful next topics include:

- **uniform integrability**, which generalizes domination in many expectation-limit problems;⁶
- signed measures and the Hahn–Jordan decomposition;⁷
- conditional expectation as a Radon–Nikodym derivative;⁸
- parameter-dependent integrals with weaker local domination conditions; and
- importance sampling and change-of-measure methods in computation.

Key ideas

- Integration adds values using the mass supplied by a measure.
- MCT, Fatou and DCT control different limit exchanges; check their hypotheses.
- Tonelli uses nonnegativity; Fubini uses absolute integrability.
- A density is a Radon–Nikodym derivative relative to a specified measure.

Looking ahead

Lecture 4 reads an expectation as an integral against a probability law. The convergence and order-exchange theorems here will justify its calculations and inequalities. Continue to Lecture 4.

6 Exercises and complete solutions

In the future HTML version, this section should be rendered as an expandable block. In the PDF, all contents are shown.

High priority *Core practice: constructing an integral*

Exercise 1 (Positive and negative parts). On $([0, 1], \mathcal{B}([0, 1]), m)$, let

$$f(x) = 3\mathbf{1}_{[0, 1/4]}(x) - 2\mathbf{1}_{(1/4, 3/4]}(x) + \mathbf{1}_{(3/4, 1]}(x).$$

⁶ A family $\{X_i\}$ of integrable random variables is uniformly integrable if $\sup_i \mathbb{E}[|X_i| \mathbf{1}\{|X_i| > K\}] \rightarrow 0$ as $K \rightarrow \infty$. Large values then contribute uniformly little to their expectations.

⁷ A signed measure is a countably additive set function that may take positive and negative values. For a finite signed measure, the Hahn decomposition separates positive and negative regions, and the Jordan decomposition writes it as the difference of two finite positive measures concentrated on those respective regions.

⁸ For integrable X and a sub-sigma-algebra $\mathcal{G}$, conditional expectation $\mathbb{E}[X \mid \mathcal{G}]$ is a $\mathcal{G}$ -measurable random variable with the same integral as X over every event in $\mathcal{G}$. It represents an average based only on the information in $\mathcal{G}$.

1. Find f^+ , f^- , and $|f|$.
2. Compute $\int f \, dx$ and $\int |f| \, dx$.
3. Verify directly that f is integrable.

Solution. The positive and negative parts are

$$\begin{aligned} f^+(x) &= 3\mathbf{1}_{[0,1/4]}(x) + \mathbf{1}_{(3/4,1]}(x), \\ f^-(x) &= 2\mathbf{1}_{(1/4,3/4]}(x), \end{aligned}$$

and

$$|f(x)| = 3\mathbf{1}_{[0,1/4]}(x) + 2\mathbf{1}_{(1/4,3/4]}(x) + \mathbf{1}_{(3/4,1]}(x).$$

Therefore

$$\int f^+ \, dx = 3 \left(\frac{1}{4} \right) + 1 \left(\frac{1}{4} \right) = 1, \quad \int f^- \, dx = 2 \left(\frac{1}{2} \right) = 1.$$

It follows that

$$\int f \, dx = 1 - 1 = 0, \quad \int |f| \, dx = 1 + 1 = 2 < \infty.$$

The last inequality verifies integrability. □

High priority Core practice: selecting a convergence theorem

Exercise 2 (MCT, DCT, or neither?). For each sequence, identify an applicable theorem and evaluate the limit of the integrals when justified. Use Lebesgue measure in each case.

1. $f_n(x) = \mathbf{1}_{[0,n]}(x)$ on $\mathbb{R}$.
2. $f_n(x) = x^n$ on $[0, 1]$.
3. $f_n(x) = n\mathbf{1}_{(0,1/n)}(x)$ on $[0, 1]$.
4. $f_n(x) = (1 + x/n)^n e^{-2x}$ on $[0, \infty)$.

Solution. For (1), $0 \leq f_n \uparrow 1$ pointwise on $[0, \infty)$ and to zero on $(-\infty, 0)$. MCT gives

$$\lim_n \int_{\mathbb{R}} f_n \, dx = \int_{\mathbb{R}} \mathbf{1}_{[0,\infty)} \, dx = \infty.$$

For (2), $x^n \rightarrow 0$ for $x \in [0, 1)$, and the exceptional point $x = 1$ has measure zero. Since $0 \leq x^n \leq 1$ and 1 is integrable on $[0, 1]$, DCT gives

$$\lim_n \int_0^1 x^n \, dx = 0.$$

For (3), $f_n \rightarrow 0$ pointwise, but $\int_0^1 f_n \, dx = 1$ for every n . The sequence is not increasing, and there is no integrable domination that would permit DCT. Thus neither MCT nor DCT justifies an exchange; in fact the exchange is false.

For (4), $(1 + x/n)^n \rightarrow e^x$, so $f_n(x) \rightarrow e^{-x}$. The inequality $1 + u \leq e^u$ gives

$$0 \leq f_n(x) \leq e^x e^{-2x} = e^{-x},$$

and e^{-x} is integrable on $[0, \infty)$. DCT yields

$$\lim_n \int_0^\infty (1 + x/n)^n e^{-2x} \, dx = \int_0^\infty e^{-x} \, dx = 1.$$

□

High priority Core practice: a nonnegative sum

Exercise 3 (Integral of an infinite series). Let $g_1, g_2, \dots$ be nonnegative measurable functions on $(\Omega, \mathcal{F}, \nu)$. Prove

$$\int \sum_{k=1}^{\infty} g_k \, d\nu = \sum_{k=1}^{\infty} \int g_k \, d\nu.$$

Then apply the result on $[0, 1]$ to $g_k(x) = x^{k-1}(1-x)$.

Solution. The partial sums

$$s_n = \sum_{k=1}^n g_k$$

are nonnegative measurable functions with $s_n \uparrow \sum_{k=1}^{\infty} g_k$. MCT and finite linearity give

$$\int \sum_{k=1}^{\infty} g_k \, d\nu = \lim_{n \rightarrow \infty} \int s_n \, d\nu = \lim_{n \rightarrow \infty} \sum_{k=1}^n \int g_k \, d\nu = \sum_{k=1}^{\infty} \int g_k \, d\nu.$$

For $0 \leq x < 1$,

$$\sum_{k=1}^{\infty} x^{k-1}(1-x) = 1.$$

The equality also holds a.e. on $[0, 1]$; the endpoint $x = 1$ is irrelevant. Therefore

$$1 = \int_0^1 1 \, dx = \sum_{k=1}^{\infty} \int_0^1 x^{k-1}(1-x) \, dx.$$

Indeed, the k th integral equals $1/k - 1/(k+1) = 1/[k(k+1)]$, recovering the telescoping identity $\sum_{k \geq 1} 1/[k(k+1)] = 1$. □

High priority Core practice: Tonelli and order of integration

Exercise 4 (Area above the diagonal). Let $f(x, y) = \mathbf{1}\{0 \leq x \leq y \leq 1\}$.

1. Evaluate $\int_0^1 \int_0^1 f(x, y) \, dx \, dy$.
2. Reverse the order and verify the same value.
3. State precisely why the exchange is valid.

Solution. For fixed $y \in [0, 1]$, the allowed x values form $[0, y]$, so

$$\int_0^1 \int_0^1 f(x, y) \, dx \, dy = \int_0^1 \left(\int_0^y 1 \, dx \right) dy = \int_0^1 y \, dy = \frac{1}{2}.$$

For fixed x , the allowed y values form $[x, 1]$, so

$$\int_0^1 \int_0^1 f(x, y) \, dy \, dx = \int_0^1 \left(\int_x^1 1 \, dy \right) dx = \int_0^1 (1-x) \, dx = \frac{1}{2}.$$

The function is nonnegative and measurable, so Tonelli's theorem permits the exchange. Since its integral is finite, Fubini's theorem also applies. □

Mid priority Supporting practice: Fatou's lemma

Exercise 5 (Escaping mass). On $\mathbb{R}$, let $f_n = \mathbf{1}_{[n, n+1]}$.

1. Compute $\liminf_n f_n(x)$ for every fixed x .
2. Verify the two sides of Fatou's inequality.
3. Explain why DCT cannot be applied using $g = \sup_n f_n$.

Solution. For fixed x , only finitely many intervals $[n, n+1]$ can contain x . Therefore $f_n(x) = 0$ eventually and $\liminf_n f_n(x) = 0$. Hence

$$\int_{\mathbb{R}} \liminf_n f_n dx = 0.$$

On the other hand, $\int_{\mathbb{R}} f_n dx = 1$ for every n , so

$$\liminf_n \int_{\mathbb{R}} f_n dx = 1.$$

Fatou's inequality reads $0 \leq 1$ and is strict.

The envelope is exactly $g = \sup_n f_n = \mathbf{1}_{[1, \infty)}$, so $\int_{\mathbb{R}} g dx = \infty$. It is not an integrable dominating function, and DCT does not apply. $\square$

Mid priority *Supporting practice: failure of Fubini*

Exercise 6 (Different iterated sums). For

$$a_{ij} = \begin{cases} 1, & j = i, \\ -1, & j = i + 1, \\ 0, & \text{otherwise,} \end{cases}$$

compute both iterated sums and $\sum_{i,j} |a_{ij}|$. Identify the failed Fubini assumption.

Solution. For each fixed row i , the only nonzero entries are $a_{ii} = 1$ and $a_{i,i+1} = -1$, so

$$\sum_{j=1}^{\infty} a_{ij} = 0 \quad \text{and} \quad \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_{ij} = 0.$$

In column 1, the sum is 1. For each column $j \geq 2$, the entries $a_{jj} = 1$ and $a_{j-1,j} = -1$ cancel. Hence

$$\sum_{j=1}^{\infty} \sum_{i=1}^{\infty} a_{ij} = 1.$$

Every row contributes two units to the absolute sum:

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} |a_{ij}| = \sum_{i=1}^{\infty} 2 = \infty.$$

The array is signed and not absolutely summable, so neither Tonelli's nonnegativity condition nor Fubini's absolute-integrability condition holds. $\square$

Mid priority *Supporting practice: Radon–Nikodym derivatives*

Exercise 7 (A mixed dominating measure). Let

$$\mu = \frac{1}{2} \delta_0 + \frac{1}{2} \phi(x) dx, \quad \rho = \delta_0 + m,$$

where ϕ is the standard normal density and m is Lebesgue measure.

1. Show that $\mu \ll \rho$.
2. Find $d\mu/d\rho$.
3. Verify that $\int (d\mu/d\rho) d\rho = 1$.

Solution. If $\rho(A) = 0$, then both $\delta_0(A) = 0$ and $m(A) = 0$. Therefore

$$\mu(A) = \frac{1}{2}\delta_0(A) + \frac{1}{2}\int_A \phi(x) dx = 0,$$

so $\mu \ll \rho$.

A valid derivative is

$$r(x) = \frac{d\mu}{d\rho}(x) = \begin{cases} \frac{1}{2}, & x = 0, \\ \frac{1}{2}\phi(x), & x \neq 0. \end{cases}$$

Indeed, the δ_0 component of ρ reads the value $r(0) = 1/2$, while its Lebesgue component integrates $r(x) = \phi(x)/2$ a.e.; the value at the singleton zero does not affect the Lebesgue integral. Finally,

$$\int r d\rho = \int r d\delta_0 + \int r dm = \frac{1}{2} + \frac{1}{2}\int_{\mathbb{R}} \phi(x) dx = 1.$$

□

Mid priority *Supporting practice: differentiation under the integral*

Exercise 8 (A Laplace transform derivative). Let $h \geq 0$ be measurable and satisfy

$$\int_0^\infty (1+x)h(x) dx < \infty,$$

and define

$$F(\theta) = \int_0^\infty e^{-\theta x} h(x) dx, \quad \theta > 0.$$

Show that

$$F'(\theta) = -\int_0^\infty x e^{-\theta x} h(x) dx.$$

Solution. Fix $\theta > 0$ and choose a neighborhood $I = (\theta/2, 3\theta/2)$. For $t \in I$,

$$\left| \frac{\partial}{\partial t} e^{-tx} h(x) \right| = x e^{-tx} h(x) \leq x h(x).$$

The function $xh(x)$ is integrable by assumption. Also $e^{-tx}h(x)$ is integrable because it is bounded by $h(x)$, which is integrable. The differentiation-under-the-integral theorem therefore gives

$$F'(\theta) = \int_0^\infty \frac{\partial}{\partial \theta} \{e^{-\theta x} h(x)\} dx = -\int_0^\infty x e^{-\theta x} h(x) dx.$$

□

Low priority *Optional practice: many-to-one change of variables—not examined*

Exercise 9 (Squaring a continuous random variable). Let X have a continuous density f_X on $\mathbb{R}$, and let $Y = X^2$. Find the density of Y on $(0, \infty)$. Specialize your answer to $X \sim N(0, 1)$.

Solution. For $y > 0$, the map $g(x) = x^2$ has the two inverse branches $g_1^{-1}(y) = \sqrt{y}$ and $g_2^{-1}(y) = -\sqrt{y}$. Each inverse derivative has absolute value $1/(2\sqrt{y})$. Therefore

$$f_Y(y) = \frac{f_X(\sqrt{y}) + f_X(-\sqrt{y})}{2\sqrt{y}}, \quad y > 0,$$

and $f_Y(y) = 0$ for $y < 0$. There is no atom at zero when X has a continuous density.

For a standard normal X , symmetry gives

$$f_Y(y) = \frac{1}{\sqrt{2\pi y}} e^{-y/2}, \quad y > 0,$$

which is the χ_1^2 density. □

Proof details for linked claims

High priority *Why the integral is well defined*

Representation and approximation. For the simple-function integral, extend two finite disjoint representations by a zero coefficient on the complement, so each is a partition of Ω . Refine the two partitions to their intersections $C_{ij} = A_i \cap B_j$. On a nonempty C_{ij} the two coefficients agree, since they are values of the same function. Finite additivity gives

$$\sum_i a_i \nu(A_i) = \sum_{i,j} a_i \nu(C_{ij}) = \sum_{i,j} b_j \nu(C_{ij}) = \sum_j b_j \nu(B_j).$$

All terms are nonnegative, with $0 \cdot \infty = 0$, so no undefined subtraction occurs.

For a nonnegative measurable f , an explicit simple approximation is

$$\phi_n(\omega) = \begin{cases} 2^{-n} \lfloor 2^n f(\omega) \rfloor, & f(\omega) < n, \\ n, & f(\omega) \geq n. \end{cases}$$

Its finitely many level sets are measurable inverse images of intervals. Refining the dyadic mesh rounds downward by less and increasing the cap cannot lower a value; hence $0 \leq \phi_n \leq \phi_{n+1} \leq f$. At a point with finite f , eventually $f < n$ and $0 \leq f - \phi_n < 2^{-n}$; at a point with $f = \infty$, $\phi_n = n$. Thus $\phi_n \uparrow f$. This proves the approximation assertion without using MCT; applying MCT afterward gives $\int \phi_n \uparrow \int f$.

Mid priority *Pushforward and change-of-measure details*

From the sample space to the law. Let $X : (\Omega, \mathcal{F}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be measurable and $P_X(B) = P(X^{-1}(B))$. For an indicator $h = \mathbf{1}_B$,

$$\int_{\Omega} h(X) dP = P(X \in B) = P_X(B) = \int_{\mathbb{R}} h dP_X.$$

Finite linearity proves the identity for nonnegative simple h . For nonnegative Borel h , take simple $h_n \uparrow h$ and apply MCT on both spaces. Finally apply that result to h^+ and h^- . It gives equality for every well-defined signed integral and, in particular, $\int |h(X)| dP = \int |h| dP_X$. For integrable h , both positive and negative integrals are finite and can be subtracted. This proves the pushforward identity with its conditions.

From one measure to another. Suppose $P \ll Q$ are probability measures on a common measurable space and $r = dP/dQ$. The RN defining identity says $\int \mathbf{1}_A dP = P(A) = \int \mathbf{1}_A r dQ$. By finite linearity it holds for nonnegative simple functions; simple approximation and MCT extend it to every nonnegative measurable h . Apply the result to $|h|$ to obtain

$$\int |h| dP = \int |h| r dQ.$$

If this value is finite, apply the nonnegative identity to h^+ and h^- and subtract. Thus $\int h dP = \int hr dQ$. Setting $h = 1$ also gives $\int r dQ = P(\Omega) = 1$. When the common space is the line, $X(x) = x$ gives the coordinate notation in Section 4. On an abstract space replace h by $h \circ X$; the density remains $r(\omega)$ because it changes the measure on that space.

Mid priority *Differentiation details*

Why the difference quotients are dominated. Under Theorem 4, work on the common measurable full-measure set where all derivatives exist and are bounded by g . For any $\theta \in I$, the segment joining θ_0 and θ lies in I , and the mean value theorem gives

$$|f(\omega, \theta)| \leq |f(\omega, \theta_0)| + |\theta - \theta_0|g(\omega).$$

The right side is integrable, so $F(\theta)$ is finite throughout I . Fix θ and take any nonzero sequence $t_n \rightarrow 0$ with $\theta + t_n \in I$. The measurable difference quotients

$$q_n(\omega) = \frac{f(\omega, \theta + t_n) - f(\omega, \theta)}{t_n}$$

satisfy $|q_n| \leq g$ on the same set and converge there to $\partial_\theta f(\omega, \theta)$. Defining that derivative to be zero off the set gives a measurable version, dominated by g a.e. DCT and linearity now yield

$$\lim_{n \rightarrow \infty} \frac{F(\theta + t_n) - F(\theta)}{t_n} = \lim_{n \rightarrow \infty} \int q_n d\nu = \int \partial_\theta f(\omega, \theta) d\nu.$$

This holds for every such sequence, so the ordinary derivative exists and has the asserted value. Neither the null set nor the envelope changes with θ within I .

References

- [1] Rick Durrett. *Probability: theory and examples*, volume 49. Cambridge university press, 5th edition, 2019.
- [2] W. Rudin. *Real and Complex Analysis*. Higher Mathematics Series. McGraw-Hill Education, 1987.
- [3] Walter Rudin. *Principles of mathematical analysis*. McGraw-Hill, New York, 3 edition, 1976.
- [4] Jun Shao. *Mathematical Statistics*. Springer, 2nd edition, 2003.

What changed

October 19, 2026: Corrected notation, wording and cross-references after the course-wide errata review; refreshed the linked HTML and PDF editions.

October 14, 2026. Added issue-reporting links to the webpage and PDF. The mathematics and priority assignments are unchanged in this update.