Probability Theory - Durrett's

Thanh Le

May 2024

This material covers key theorems and concepts of Chapters 1.1-4.3 of Durrett's *Probability: Theory and Examples*, 5^{th} edition.

1 Measure theory backgrounds

A **probability space** is a triple $(\Omega, \mathcal{F}, P)$, where Ω is a set of of "outcomes", $\mathcal{F}$ is a set of "events", and $P : \mathcal{F} \to [0, 1]$ is a function that assigns probabilities to events. We assume that $\mathcal{F}$ is a σ**-field** (or σ**-algebra**), i.e. a (nonempty) collection of subsets of Ω that satisfy

- if $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$, and
- if $A_i \in \mathcal{F}$ is a countable sequence of sets then $\cup_i A_i \in \mathcal{F}$.

We also require that P is a **measure**, i.e. a nonnegative countably additive set function $\mathcal{F} \to \mathbb{R}$.

Let μ be a measure. Then μ is said to be σ**-finite** if there is a sequence of sets $A_n \in \mathcal{A}$ such that $\mu(A_n) < \infty$ and $\cup_n A_n = \Omega$.

A function $f : \Omega \to S$ is said to be a **measurable map** from $(\Omega, \mathcal{F})$ to $(S, \mathcal{S})$ if

$$X^{-1}(B) \equiv \{\omega : X(\omega) \in B\} \in \mathcal{F} \quad \text{for all } B \in \mathcal{S}.$$

If $(S, \mathcal{S}) = (\mathbb{R}^d, \mathcal{R}^d)$ and $d > 1$ then X is called a **random vector**. If $d = 1$, X is called a **random variable**.

Note that if $\mathcal{S}$ is a σ-field, then $\{\{X \in B\} : B \in \mathcal{S}\}$ is a σ-field. It is the smallest σ-field on Ω that makes X a measurable map. It is called the σ**-field generated by** X and denoted by $\sigma(X)$. So,

$$\sigma(X) = \{\{X \in B\} : B \in \mathcal{S}\}.$$

Borel sets: the smallest σ-field containing the open sets.

When the distribution $F(x) = P(X \le x)$ has the form

$$F(x) = \int_{-\infty}^{x} f(y)dy,$$

we say that X has **density function** f.

(Example 1.2.4 - Exponential distribution with rate λ) Density function: $f(x) = \lambda e^{-\lambda x}$ for $x \ge 0$ and 0 otherwise. Distribution function:

$$F(x) = \begin{cases} 0 & x \le 0 \\ 1 - e^{-\lambda x} & x \ge 0 \end{cases}$$

(Example 1.2.5. Standard normal distribution) Density function: $f(x) = (2\pi)^{-1/2} \exp(-x^2/2)$. There is no closed form expression for the distribution function $F(x)$, but for large x, we have useful bound for $F(x)$ (check Theorem 1.2.6.).

We say that X has a **Poisson distribution** with parameter λ if

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

for $k = 0, 1, 2, ...$

(Binomal distribution) `https://en.wikipedia.org/wiki/Binomial_distribution`

(Theorem 1.3.1) If $\{w : X(w) \in A\} \in \mathcal{F}$ for all $A \in \mathcal{A}$ and $\mathcal{A}$ generates $\mathcal{S}$, then X is measurable.

(Theorem 1.3.4) If $X : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ and $f : (S, \mathcal{S}) \to (T, \mathcal{T})$ are measurable maps, then $f(X)$ is a measurable map from $(\Omega, \mathcal{F})$ to $(T, \mathcal{T})$.

It follows from Thm 1.3.4 that if X is a r.v. then so is cX for all $c \in \mathbb{R}$, $X^2, \sin(X)$, etc.

(Notation) $a \wedge b = \min\{a, b\}$; $a \vee b = \max\{a, b\}$; $f^+(x) = f(x) \vee 0$; $f^-(x) = (-f(x)) \vee 0$.

In Chapter 1.4, **integral** is defined for measurable functions only!

We say f is **integrable** if $\int |f| d\mu < \infty$. The **integral** of f is defined by

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu.$$

(This is already step 4 of the 4 steps in defining integral).

(Notation) We define the integral of f over the set E:

$$\int_E f d\mu = \int f \cdot 1_E d\mu$$

(Theorem 1.5.3 - Bounded Convergence Theorem) Let E be a set with $\mu(E) < \infty$. Suppose f_n vanishes on E^c, $|f_n(x)| \leq M$, and $f_n \to f$ in measure. Then

$$\int f d\mu = \lim_{n\to\infty} \int f_n d\mu.$$

(Theorem 1.5.5. Fatou's Lemma): If $f_n \geq 0$ then

$$\liminf_{n\to\infty} \int f_n d\mu \geq \int \big(\liminf_{n\to\infty} f_n\big) d\mu.$$

(Theorem 1.5.7 - Monotone Convergence Theorem) If $f_n \geq 0$ and $f_n \uparrow f$, then $\int f_n d\mu \uparrow \int f d\mu$.

(Theorem 1.5.8 - Dominated Convergence Theorem) If $f_n \to f$ a.e., $|f_n| \leq g$ for all n, and g is integrable, then $\int f_n d\mu \to \int f d\mu$.

(Counting measure) `https://en.wikipedia.org/wiki/Counting_measure`. When Ω is a countable set, $\mathcal{F}$ = all subsets of Ω, and μ is counting measure, then we write $\sum_{i\in\Omega} f(i)$ for $\int f d\mu$.

(Theorem 1.5.1 - Jensen's inequality) Suppose φ is convex, that is,

$$\lambda\varphi(x) + (1-\lambda)\varphi(y) \geq \varphi\big(\lambda x + (1-\lambda)y\big)$$

for all $\lambda \in (0,1)$ and $x, y \in \mathbb{R}$. If μ is a probability measure, and f and $\varphi(f)$ are integrable, then

$$\varphi\Big(\int f d\mu\Big) \leq \int \varphi(f) d\mu.$$

Let $||f||_p = (\int |f|^p d\mu)^{1/p}$ for $1 \leq p < \infty$.

(Theorem 1.5.2 - Holder's inequality) If $p, q \in (1, \infty)$ with $1/p + 1/q = 1$ then

$$\int |fg| d\mu \leq ||f||_p ||g||_q.$$

(Theorem 1.6.4 - Chebyshev's inequality) Suppose $\varphi : \mathbb{R} \to \mathbb{R}$ has $\varphi \geq 0$, let $A \in \mathcal{R}$ and let $i_A = \inf\{\varphi(y) : y \in A\}$. Then

$$i_A P(X \in A) \leq E(\varphi(X); X \in A) \leq E\varphi(X).$$

(Theorem 1.6.7 - Dominated Convergence Theorem) If $X_n \to X$ a.s., $|X_n| \leq Y$ for all n, and $EY < \infty$, then $EX_n \to EX$.

(This is Theorem 1.5.8 rewritten for expectation).

(Theorem 1.6.9 - Change of variables formula) Let X be a random element of $(S, \mathcal{S})$ with distribution μ, i.e, $\mu(A) = P(X \in A)$. If f is a measurable function from $(S, \mathcal{S})$ to $(\mathbb{R}, \mathcal{R})$ so that $f \geq 0$ or $E|f(X)| < \infty$, then

$$Ef(X) = \int_S f(y)\mu(dy).$$

(Theorem 1.7.2 - Fubini's Theorem) If $f \geq 0$ or $\int |f| d\mu < \infty$ then

$$\int_X \int_Y f(x,y)\mu_2(dy)\mu_1(dx) = \int_{X\times Y} f d\mu = \int_Y \int_X f(x,y)\mu_1(dx)\mu_2(dy).$$

2 Law of Large Numbers

Two random variables X and Y are **independent** if for all $C, D \in \mathcal{R}$,

$$P(X \in C, Y \in D) = P(X \in C)P(Y \in D)$$

i.e., the events $A = \{X \in C\}$ and $B = \{Y \in D\}$ are independent.

We say that Y_n converges to Y **in probability** if for all $\epsilon > 0$, $P(|Y_n - Y| > \epsilon) \to 0$ as $n \to \infty$.

Let

$$\Omega_0 \equiv \{w : \lim_{n\to\infty} X_n \text{ exists}\}.$$

If $P(\Omega_0) = 1$, we say that X_n **converges almost surely**.

`https://en.wikipedia.org/wiki/Convergence_of_random_variables`

(Convergence in L^p) Let $p > 0$ be fixed. X_n converges to X in L^p if $E|X_n - X|^p \to 0$ as $n \to \infty$.

A family of random variables $X_i, i \in I$ with $EX_i^2 < \infty$ is said to be **uncorrelated** if we have

$$E(X_iX_j) = EX_iEX_j$$

whenever $i \neq j$. (Note here that pairwise is enough. Also, note that being independent implies being uncorrelated).

(Theorem 2.2.3 - L^2 weak law) Let $X_1, X_2, \ldots$ be uncorrelated random variables with $EX_i = \mu$ and $var(X_i) \leq C < \infty$. If $S_n = X_1 + \cdots + X_n$ then as $n \to \infty$, $S_n/n \to \mu$ in L^2 and in probability.

(Theorem 2.2.14 - Weak Law of Large Numbers) Let $X_1, X_2, \ldots$ be i.i.d. with $E|X_i| < \infty$. Let $S_n = X_1 + \cdots + X_n$ and let $\mu = EX_1$. Then $S_n/n \to \mu$ in probability.

(Note: there is a more general version (see Theorem 2.2.12), but this one is more commonly seen. It's more easy to remember, too).

(Lemma 2.2.13) If $Y \geq 0$ and $p > 0$ then $E(Y^p) = \int_0^\infty py^{p-1}P(Y > y)dy$.

(Build-up for Chapter 2.3) If A_n is a sequence of subsets of Ω, we let

$$\limsup A_n = \lim_{m\to\infty} \cup_{n=m}^{\infty} A_n = \{\omega \text{ that are in infinitely many } A_n\}$$

(the limit exists since the sequence is decreasing in m).

It's common to write $\limsup A_n = \{\omega : \omega \in A_n \text{ i.o.}\}$, where i.o. stands for **infinitely often**. An example which illustrates the use of this notation is "$X_n \to 0$ a.s. iff for all $\epsilon > 0$, $P(|X_n| > \epsilon \text{ i.o.}) = 0$."

(Theorem 2.3.1 - Borel-Cantelli lemma) If $\sum_{n=1}^{\infty} P(A_n) < \infty$ then

$$P(A_n \text{ i.o.}) = 0.$$

In other words, if the sum of the probabilities of the events $\{A_n\}$ is finite, then the probability that infinitely many of them occur is 0.

(Theorem 2.3.2) $X_n \to X$ in probability iff for every subsequence $X_{n(m)}$ there exists a further subsequence $X_{n(m_k)}$ that converges almost surely to X.

(Theorem 2.3.7 - The second Borel-Cantelli lemma) If the events A_n are independent then $\sum P(A_n) = \infty$ implies $P(A_n \text{ i.o.}) = 1$.

(Theorem 2.4.1 - Strong law of large numbers) Let $X_1, X_2, \ldots$ be pairwise independent identically distributed random variables with $E|X_i| < \infty$. Let $EX_i = \mu$ and $S_n = X_1 + \cdots + X_n$. Then $S_n/n \to \mu$ a.s. as $n \to \infty$.

Note that the strong law holds whenever EX_i exists. We have **Theorem 2.4.5**: Let $X_1, X_2, \ldots$ be i.i.d. with $EX_i^+ = \infty$ and $EX_i^- < \infty$. If $S_n = X_1 + \cdots + X_n$ then $S_n/n \to \infty$ a.s..

3 Central Limit Theorems

A sequence of distribution functions is said to **converge weakly** to a limit F (written $F_n \Rightarrow F$) if $F_n(y) \to F(y)$ for all y that are continuity points of F. A sequence of random variables X_n is said to **converge weakly** or **converge in distribution** to a limit X_∞ (written $X_n \Rightarrow X_\infty$) if their distribution functions $F_n(x) = P(X_n \le x)$ converges weakly.

(Theorem 3.2.8) If $F_n \Rightarrow F_\infty$ then there are random variables Y_n, $1 \le n \le \infty$, with distribution F_n so that $Y_n \to Y_\infty$ a.s..

(Theorem 3.2.9) $X_n \Rightarrow X_\infty$ iff for every bounded continuous function g, we have $Eg(X_n) \to Eg(X_\infty)$.

(Theorem 3.3.2) If X_1 and X_2 are independent and have ch.f.'s φ_1 and φ_2 then $X_1 + X_2$ has ch.f. $\varphi_1(t)\varphi_2(t)$.

(Theorem 3.3.9) If $F_1, \ldots, F_n$ have ch.f. $\varphi_1, \ldots, \varphi_n$ and $\lambda_i \ge 0$ have $\lambda_1 + \cdots + \lambda_n = 1$ then $\sum_{i=1}^{n} \lambda_i F_i$ has ch.f. $\sum_{i=1}^{n} \lambda_i \varphi_i$.

(Theorem 3.3.18) If $\int |x|^n \mu(dx) < \infty$ then its characteristic function φ has a continuous derivative of order n given by $\varphi^{(n)}(t) = \int (ix)^n e^{itx} \mu(dx)$.

This implies if $E|X|^n < \infty$, then its characteristic function is n times differentiable at 0, and $\varphi^{(n)}(0) = E(iX)^n$.

(Theorem 3.4.1) Let $X_1, X_2, \dots$ be i.i.d. with $EX_i = \mu$, $var(X_i) = \sigma^2 \in (0, \infty)$. If $S_n = X_1 + \cdots + X_n$ then

$$(S_n - n\mu)/\sigma n^{1/2} \to \chi$$

where χ has the standard normal distribution.

(This is the central limit theorem for i.i.d sequences).

(Theorem 3.4.10 - The Lindeberg-Feller theorem) For each n, let $X_{n,m}, 1 \leq m \leq n$, be independent random variables with $EX_{n,m} = 0$. Suppose

(i) $\sum_{m=1}^{n} EX_{n,m}^2 \to \sigma^2 > 0$

(ii) For all $\epsilon > 0$, $\lim_{n\to\infty} \sum_{m=1}^{n} E(|X_{n,m}|^2; |X_{n,m}| > \epsilon) = 0$.

Then, $S_n = X_{n,1} + \cdots + X_{n,n} \Rightarrow \sigma\chi$ as $n \to \infty$.

This is the CLT for triangular arrays, which generalizes the version for i.i.d.. In words, the theorem says that a sum of a large number of small independent effects has approximately a normal distribution.

4 Martingales

Given a probability space $(\Omega, \mathcal{F}_0, P)$, a σ-field $\mathcal{F} \subset \mathcal{F}_\prime$, and a random variable $X \in \mathcal{F}_0$ with $E|X| < \infty$. We define the **conditional expectation of X given $\mathcal{F}$**, $E(X|\mathcal{F})$, to be any random variable Y that has

(i) $Y \in \mathcal{F}$, i.e., Y is $\mathcal{F}$-measurable.

(ii) for all $A \in \mathcal{F}$, $\int_A X dP = \int_A Y dP$.

Actually, Y exists and is unique.

(Theorem 4.1.12) If $\mathcal{F} \subset \mathcal{G}$ and $E(X|\mathcal{G}) \in \mathcal{F}$ then $E(X|\mathcal{F}) = E(X|\mathcal{G})$.

(Theorem 4.1.14) If $X \in \mathcal{F}$ and $E|Y|, E|XY| < \infty$ then

$$E(XY|\mathcal{F}) = XE(Y|\mathcal{F}).$$

Let $\mathcal{F}_n$ be a **filtration**, i.e., an increasing sequence of σ-fields. A sequence X_n is said to be **adapted** to $\mathcal{F}_n$ if $X_n \in \mathcal{F}_n$ (i.e. X_n is $\mathcal{F}_n$-measurable) for all n. If X_n is a sequence with

(i) $E|X_n| < \infty$,

(ii) X_n is adapted to $\mathcal{F}_n$,

(iii) $E(X_{n+1}|\mathcal{F}_n) = X_n$ for all n,

then X is said to be a **martingale** (wrt $\mathcal{F}_n$). If in this definition, $=$ is replaced by $\leq$ or $\geq$, then X is said to be a **supermartingale** or **submartingale**, respectively.

(Theorem 4.2.9) If N is a stopping time and X_n is a supermartingale, then $X_{N \wedge n}$ is a supermartingle.

(Theorem 4.2.11 - Martingale convergence theorem) If X_n is a submartingale with $\sup X_n^+ < \infty$ then as $n \to \infty$, X_n converges a.s. to a limit X with $E|X| < \infty$.

(Theorem 4.2.12) If $X_n \geq 0$ is a supermartingale then as $n \to \infty$, $X_n \to X$ a.s. and $EX \leq EX_0$.