A Biggner's Guide to Probability Theory

Seoncheol Park

2023-07-30

Table of contents

Preface **4**

1 Introduction **5**

2 The Elements of Proability Theory **6**

3 Data and models **7**

4 Field and σ-field **8**

5 Probabilities **9**
5.1 Probability 9

6 Random Variables **11**
6.1 Radon-nikodym derivative 11
6.2 Integration 12
6.3 리만-스틸체스 적분 12
6.4 리만 적분과 르베그 적분 13

7 Convergence **15**
7.1 Definitions 15

8 Uniform Integrability **17**

9 Convergence of Moments **18**
9.1 Almost Sure Convergence 18
9.2 Convergence in Probability 18
9.3 SNU, 2020 Winter, Problem 1 18
9.4 Solution 18
9.5 Convergence in Distribution 19
9.6 SNU, 2019 Summer, Problem 3 19
9.7 Solution 19

10 The Law of Large Numbers **20**

11 Preliminaries **21**
11.1 Moments and Tails 21

12 A Weak Law for Partial Maxima **22**
12.1 SNU, 2017 Winter, Problem 3 22
12.2 Solution 22

13 Summary **23**

References **24**

Preface

확률론은 통계학을 공부하는 데 있어 굉장히 중요한 과목

This is a Quarto book.

To learn more about Quarto books visit https://quarto.org/docs/books.

```
1 + 1
```

```
[1] 2
```

1 Introduction

This is a book created from markdown and executable code.

See Knuth (1984) for additional discussion of literate programming.

```
1 + 1
```

```
[1] 2
```

2 The Elements of Proability Theory

3 Data and models

- Sample space Ω (표본공간)
- Random vector $\mathbf{X} = (X_1, \dots, X_n)$ (확률벡터)
- Outcome of the experiment ω
- $\mathbf{X}(\omega)$: observations or data
- Since it is only $\mathbf{X}$ that we observe, we need only consider its probability distribution. This distribution is assumed to be a member of a family $\mathcal{P}$ of probability distributions on $\mathbb{R}^n$.

4 Field and σ-field

Definition 4.1 (Field)**.** The class $\mathcal{A}$ of subsets of Ω is called a **field** if it contains Ω and is closed under the formulation of complements and finite unions, that is if:

1. $\Omega \in \mathcal{A}$
2. $A \in \mathcal{A} \Longrightarrow A^c \in \mathcal{A}$
3. $A_1, A_2 \in \mathcal{A} \Longrightarrow A_1 \cup A_2 \in \mathcal{A}$

Definition 4.2 (σ-field)**.** The class $\mathcal{F}$ of subsets of Ω is called a σ-**field** if it is a field and if it is closed under the formulation of countable unions, that is if:

4. $A_1, A_2, ... \in \mathcal{F} \Longrightarrow \cup_{n=1}^{\infty} A_n \in \mathcal{F}$

- Recall that the elements of any field or σ-field are called **random events** (or simply **events**).

5 Probabilities

::: {#def-prob}

5.1 Probability

- Let Ω be any set and $\mathcal{A}$ be a field of its subsets. We say that P is a **probability** on the measurable space $(\Omega, \mathcal{A})$ if P is defined for all events $A \in \mathcal{A}$ and satisfies the following axioms.

1. $P(A) \geq 0$ for each $A \in \mathcal{A}$; $P(\Omega) = 1$
2. P is **finitely additive**. That is, for any finite number of pairwise disjoint events $A_1, \dots, A_n \in \mathcal{A}$ we have

$$P\Big(\cup_{i=1}^{n} A_i \Big) = \sum_{i=1}^{n} P(A_i).$$

3. P is continuous at $\emptyset$. That is, for any events $A_1, A_2, \dots, \mathcal{A}$ such that $A_{n+1} \subset A_n$ and $\cap_{n=1}^{\infty} A_n = \emptyset$, it is true that

$$\lim_{n \to \infty} P(A_n) = 0.$$

Note that conditions 2 and 3 are equivalent to the next one 4.

4. P is σ-additive (countably additive), that is

$$P\Big(\cup_{n=1}^{\infty} A_n \Big) = \sum_{n=1}^{\infty} P(A_n)$$

for any events $A_1, A_2, \dots \in \mathcal{A}$ which are pairwise disjoint.

Example 5.1 (A probability measure which is additive but not σ-additive)**.** Let Ω be the set of all rational numbers r of the unit interval $[0,1]$ and $\mathcal{F}_1$ the class of the subsets of Ω of the form $[a,b]$, $(a,b]$, (a,b) or $[a,b)$ where a and b are rational numbers. Denote by $\mathcal{F}_2$ the class of all finite sums of disjoint sets of $\mathcal{F}_1$. Then $\mathcal{F}_2$ is a field. Let us define the probability measure P as follows:

$$P(A) = b - a, \quad \text{if } A \in \mathcal{F}_1,$$
$$P(B) = \sum_{i=1}^{n} P(A_i), \quad \text{if } B \in \mathcal{F}_2, \text{ that is, } B = \sum_{i=1}^{n} A_i, A_i \in \mathcal{F}_1.$$

Consider two disjoint sets of $\mathcal{F}_2$ say

$$B = \sum_{i=1}^{n} A_i \quad \text{and } B' = \sum_{j=1}^{m} A'_j,$$

where $A_i, A'_j \in \mathcal{F}_1$ and all A_i, A'_j are disjoint. Then $B + B' = \sum_{k=1}^{m+n} C_k$ where either $C_k = A_i$ for some $i = 1, \dots, n$, or $C_k = A'_j$ for some $j = 1, \dots, m$. Moreover,

$$\begin{aligned} P(B+B') &= P\Big(\sum_k C_k\Big) = \sum_k P(C_k) = \sum_{i,j}(P(A_i) + P(A'_j)) \\ &= P(A_i) + \sum_j P(A'_j) = P(B) + P(B'). \end{aligned}$$

and hence P is an additive measure.

Obviously every one-point set $\{r\} \in \mathcal{F}_2$ and $P(\{r\}) = 0$. Since Ω is a countable set and $\Omega = \sum_{i=1}^{\infty} \{r_i\}$, we get

$$P(\Omega) = 1 \neq 0 = \sum_{i=1}^{\infty} P(\{r_i\}).$$

This contradiction shows that P is not σ-additive.

6 Random Variables

6.1 Radon-nikodym derivative

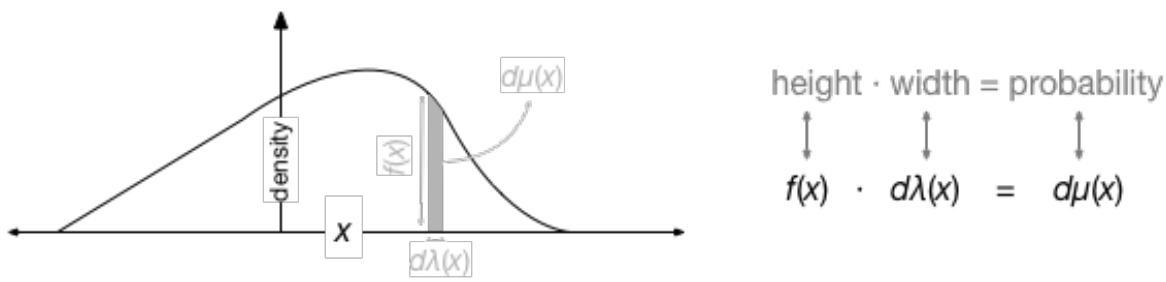

Figure 6.1: Change of measures.

확률측도는 volume element의 일반화라고 볼 수 있다.

- $\mu(x)$: probability measure, interval이나 set of points들을 인풋으로 받고 area/volume에 해당하는 확률(양수)을 아웃풋으로 주는 함수다.
- $\lambda(x)$: reference measure. We often take $\lambda(x)$ as the Lebesgue measure which is essentially just a uniform function over the sample space.

The reference measure $\lambda(x)$ is essentially just a meter-stick that allows us to express the probability measure as a simple function $f(x)$. That is, we represent the probability measure $\mu(x)$ as $f(x)$ by comparing the probability measure to some specified reference measure $\lambda(x)$. This is essentially the intuition that is given by the Radon-Nikodym derivative

$$f(x) = \frac{d\mu(x)}{d\lambda(x)}$$

or equivalently

height = area / width.

Note that we can also represent the same idea by

$$\mu(A) = \int_{A \in X} f(x) d\lambda(x),$$

where $\mu(A)$ is the sum of the probability of events in the set A which is itself a subset of the entire sample space X. Note that when $A = X$ then the integral must equal 1 by definition of probability.

라돈-니코딤 정리는 조건부 확률에 응용된다고 함.

6.2 Integration

6.3 리만-스틸체스 적분

종종 헷갈리는 표현이 기댓값을 다음과 같이 분포함수를 이용해 표현하는 경우가 있다.

$$E(X) = \int x dF(x).$$

우리가 알고 있는 정적분은 x 축을 따라가며 함수값 f(x)가 만드는 면적을 계산한다.

$$\int_a^b f(x)dx.$$

위 식을 더 확장하면 x 대신 임의의 곡선 $g(x)$ 를 적분 변수로 두고 $f(x)$ 를 단순하게 정적분 할 수도 있다.

$$\int_{x=a}^b f(x)dg(x).$$

여기서 $dg(x)$ 는 $g(x)$ 의 미분소(differential)로, $g(x)$ 의 움직임을 결정하는 x 는 단조 증가하거나 감소한다. 위와 같이 리만 적분을 일반화한 정적분을 **리만-스틸체스 적분(Riemann-Stieltjes Integral)**이라 한다. 리만 적분의 정의를 이용해 리만-스틸체스의 적분을 표현할 수도 있다.

$$\int_{x=a}^b f(x)dg(x) = \lim_{N\to\infty} \sum_{n=0}^{N-1} f(t_n)[g(x_{n+1}) - g(x_n)].$$

여기서 x_n 은 정적분을 위해 구간 $[a, b]$ 를 나눈 점, t_n 은 닫힌 세부공간 $[x_n, x_{n+1}]$ 사이에 있는 임의점이다.

6.4 리만 적분과 르베그 적분

여기는 Confused when changing from Lebesgue Integral to Riemann Integral 에 올라왔던 내용을 살펴보기로 한다. 여기서 질문자는 리만 적분을 어떻게 르베그 적분으로 바꾸는지에 대해 관심이 있다.

다음과 같이 확률공간 $(\Omega, \mathcal{F}, P)$에서 정의된 음이 아닌 확률변수 X가 지수분포를 따른다고 하자.

$$P(X < x) = 1 - e^{-\lambda x}.$$

한편, 르베그 적분으로 X의 기댓값을 쓰면 다음과 같다.

$$E[X] = \int_{\{\omega | X(\omega) \geq 0\}} X(\omega) dP(\omega).$$

여기서 질문자는 이것을 리만 적분으로 어떻게 바꾸냐

$$E[X] = \int_0^\infty x\lambda e^{-\lambda x} dx$$

를 물어보고 있다.

답변은 이것이 적분의 문제가 아닌 변수변환의 문제라고 한다.

By definition, given $X : \Omega \to \mathbb{R}$ a random variable, $E[X] = \int_\Omega X$. X defines a measure $\tilde{m}$ in $\mathbb{R}$, called the **push-forward**, by $\tilde{m}(A) = P(X^{-1}(A))$. By definition, this measure is invariant under X, and hence

$$\int_{\mathbb{R}} f d\tilde{m} = \int_\Omega f \circ X dP.$$

The equality follows from the usual arguments (prove for characteristics, simple functions, then use convergence. Recall that $1_A \circ X = 1_{X^{-1}(A)}$).

Let h be the density of X. We then have, by definition of density, that $\tilde{m}(A) = P(X^{-1}(A)) = \int_A h dm$ for any $A \in \mathcal{B}(\mathbb{R})$, where m is the Lebesgue measure. By **change of variables**, we have

$$\int_{\mathbb{R}} f d\tilde{m} = \int_{\mathbb{R}} f \cdot h dm.$$

Combining these equations,

$$\int_{\mathbb{R}} f \cdot h dm = \int_\Omega f \circ X dP.$$

Taking $f = \text{Id}$ yields

$$\int_{\mathbb{R}} xh(x)dx = \int_{\Omega} XdP = E[X].$$

Taking $f = \text{Id} \cdot \mathbf{1}_I$, where I is some interval (for example, $(0, +\infty)$ as in your case), we have

$$\int_I xh(x)dx = \int_{X^{-1}} XdP,$$

recalling again that $\mathbf{1}_A \circ X = \mathbf{1}_{X^{-1}(A)}$. Since $P(X < 0)$ in your case is 0, this last integral is actually equal to the integral over the whole space, and hence to $E[X]$, which gives your equality.

Definition 6.1 (Integrable Random Variable)**.** Gut (2014) 의 53쪽에 따르면, $E|X| < \infty$ 인 경우 random varible X 가 integrable 하다고 부른다.

Definition 6.2 ($\mathcal{L}^p$)**.** 다음과 같은 확률공간 $(\Omega, \mathcal{F}, P)$ 를 생각하자. $p > 1$ 에 대해, 확률변수 X 가 $E|X|^p < \infty$ 이면 $X \in \mathcal{L}^p$ 라고 하며 다음과 같은 놈 $\|X_p\| = (E|X|^p)^{\frac{1}{p}}$ 를 정의할 수 있다.

7 Convergence

이 장에서는 확률변수의 수렴에 대해 알아본다. $X_1, X_2, \ldots$가 확률변수라고 하자.

- 그러면 만약 이들 n항까지의 합 S_n은 $n \to \infty$일 때 어떻게 될 것인가?
- max$\{X_1, \ldots, X_n\}$은 $n \to \infty$일 때 어떻게 될 것인가?
- 수열의 극한은 어떠할 것인가?
- 수열의 함수는 어디로 수렴할 것인가? 이는 수학에서 적분의 수렴에 대응된다고 한다. (Gut 2014)
- 적분의 극한은 극한의 적분과 같을 것인가?

7.1 Definitions

다음의 정의들은 확률론에서 많이 등장하는 정의들이다. $X_1, X_2, \ldots$를 확률변수열이라 하자.

Definition 7.1 (Almost sure convergence)**.** 확률변수열 X_n은

$$P\{\omega : X_n(\omega) \to X(\omega) \text{ as } n \to \infty\}) = 1$$

을 만족하면 X_n **converges almost surely (a.s.)** to the random variable X as $n \to \infty$라 하고, $X_n \overset{\text{a.s.}}{\to} X$ as $n \to \infty$라 쓴다.

Definition 7.2 (Converge in Probability)**.** 확률변수열 X_n이 임의의 $\varepsilon > 0$에 대해

$$P\{|X_n - X| > \varepsilon) \to 0 \quad \text{as} \quad n \to \infty.$$

을 만족하면 X_n **converges in probability** to the random variable X as $n \to \infty$라 하고, $X_n \overset{\text{p}}{\to} X$ as $n \to \infty$라 쓴다.

Definition 7.3 (Converge in r-mean)**.** 확률변수열 X_n가

$$E|X_n - X|^r \to 0 \quad \text{as} \quad n \to \infty.$$

을 만족하면 X_n **converges in** $r-$mean to the random variable X as $n \to \infty$라 하고, $X_n \xrightarrow{\text{r}} X$ as $n \to \infty$라 쓴다.

Definition 7.4 (Converge in Distribution)**.** $C(F_X) = \{x : F_X(x) \text{ is continuous at } x\} =$ the continuity set of F_X라 하자. 확률변수열 X_n가

$$F_{X_n}(x) \to F_X(x) \text{as} \quad n \to \infty, \quad \forall x \in C(F_X).$$

을 만족하면 X_n **converges in distribution** to the random variable X as $n \to \infty$라 하고, $X_n \xrightarrow{\text{d}} X$ as $n \to \infty$라 쓴다.

다음과 같이 정의할 수도 있다고 한다. 확률변수열 X_n가 모든 $h \in C_B$에 대해

$$Eh(X_n) \to Eh(X) \quad \text{as} \quad n \to \infty.$$

을 만족하면 X_n **converges in distribution** to the random variable X as $n \to \infty$라 한다.

이 두개의 정의가 동치라는 증명이 Gut (2014) 의 Theorem 5.6.1에 있다.

때때로 $X_n \xrightarrow{\text{d}} \mathcal{N}(0,1)$처럼 쓰기도 한다.

Distributional convergence is often called weak convergence in these more general settings. (Gut 2014)

Definition 7.5 (Converge Weakly)**.** 이는 (Durrett 2019) 의 3.2에 나온다. 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 $F_n(x) = P(X_n \leq x)$ converges weakly.

다음은 Gut (2014) 의 5.8.1에 나오는 vague convergence이다. Vague convergence의 limiting random variable이 **proper**하지 않아도 된다는 점이 distributional convergence와의 차이점이다.

Definition 7.6 (Converge Vaguely)**.** A sequence of distribution functions $\{F_n, n \geq 1\}$ **converges vaguely** to the **pseudo-distribution function** H if, for every finite interval $I = (a, b] \subset \mathbb{R}$, where $a, b \in C(H)$,

$$F_n(I) \to H(I) \quad \text{as} \quad n \to \infty.$$

Notation: $F_n \xrightarrow{\text{v}} H$ as $n \to \infty$.

8 Uniform Integrability

Converge in probability가 mean convergence를 imply하지 않는다는 사실로부터, 그러면 어떤 조건이 있을 때 converge in probability 하면 mean convergence를 보장하는지 궁금할 수 있다. **Uniform integrability** 조건이 추가되면 그러함이 알려져 있다. (Gut 2014)

Definition 8.1 (Uniform Integrability)**.** A sequence $X_1, X_2, \ldots$ is called **uniformly integrable** iff

$$E|X_n|I\{|X_n| > a\} \to 0 \quad \text{as} \quad a \to \infty \quad \text{uniformly in } n.$$

분포함수를 이용해 다른 방법으로 정의할 수도 있다. $X_1, X_2, \ldots$ is unifomly integrable iff

$$\int_{|x|>a} |x| dF_{X_n}(x) \to 0 \quad \text{as} \quad a \to \infty \quad \text{uniformly in } n.$$

Remark. $X_1, X_2, \ldots$이 유한한 평균을 갖고 있다는 뜻은 $E|X_n|I\{|X_n| > a\} \to 0$ as $a \to \infty$ for every n을 의미한다. 즉 convergent integrals의 tail이 0으로 수렴해야 하는 것이다. Uniformly integragle은 the contributions in the tails of the integrals tend to 0 **uniformly** for all members of the sequence임을 뜻한다. (Gut 2014)

9 Convergence of Moments

위키의 설명에 따르면 $X_n \overset{L^r}{\to} X$ 이면 $\lim_{n\to\infty} E[|X_n|^r] = E[|X|^r]$ 이 성립한다고 한다. 그러나 일반적인 moment의 convergence에 대해서는 잘 알지 못한다. 여기서는 uniformly integrablility를 추가해 기존 확률변수의 수렴과 moment convergence 사이의 관계에 대해 알아본다.

We are now in the position to show that uniform integrability is the **correct** concept, that is, that a sequence that converges almost surely, in probability, or in distribution, and is uniformly integrable, converges in the mean, that moments converge and that uniform integrability is the minimal additional assumption for this to happen. (Gut 2014)

9.1 Almost Sure Convergence

9.2 Convergence in Probability

9.3 SNU, 2020 Winter, Problem 1

Let $0 < r < \infty$, $X_n \in \mathcal{L}^r$, and $X_n \to X$ in probability. Then, show the following three propositions are equivalent.

(a) $\{|X_n|^r\}$ is uniformly integrable.

(b) $X_n \to X$ in $\mathcal{L}^r$.

(c) $E(|X_n|^r) \to E(|X|^r) < \infty$.

9.4 Solution

Gut (2014) 책에는 almost sure convergence에만 초점을 맞추어 서술하였다. (Theorem 5.5.2) 실제 converge in probability에서도 성립함이 알려져 있다. (Theorem 5.5.4) 이 증명을 하기 위해서는 Fatou's lemma가 필요하다.

9.5 Convergence in Distribution

9.6 SNU, 2019 Summer, Problem 3

Show that, if X_n converges weakly to X and $\{X_n, n \geq 1\}$ is uniformly integrable, then X is integrable and $EX_n \to EX$ as $n \to \infty$.

9.7 Solution

일단 integrable random variable에 대해 모르면 Definition 6.1 를 참고하자.

Gut (2014) 책의 Theorem 5.5.9 참고

10 The Law of Large Numbers

11 Preliminaries

제일 많이 쓰이는 기술은 **truncation**이라고 하는 것으로, 이 방법의 특징은 원래 확률변수열과 asymptotically equivalent 하면서 좀 더 다루기 쉬운 수열을 생각하는 것이다.

11.1 Moments and Tails

Proposition 11.1 (Moments and Tails)**.**

1. Let $r > 0$. Suppose that X is a non-negative random variable. Then
$$EX^r < \infty \quad \Longrightarrow \quad x^r P(X > x) \to 0 \quad \text{as } x \to \infty,$$
but not necessarily conversely.
2. Suppose that $X, X_1, X_2, \ldots$ are i.i.d. random variables with mean 0. Then, for any $a > 0$,
$$EXI\{|X| \le a\} = -EXI\{|X| > a\},$$
and
$$\Big|E\sum_{k=1}^{n} X_k I\{|X_k| \le a\}\Big| \le nE|X|I\{|X| > a\}.$$
3. Let $a > 0$. If X is a random variable with mean 0, then $Y = XI\{|X| \le a\}$ does not in general havemean 0. However, if X is **symmetric**, then $EY = 0$.

12 A Weak Law for Partial Maxima

12.1 SNU, 2017 Winter, Problem 3

Suppose that X_n are i.i.d. r.v.'s with $EX_1 > 0$ and $S_n = X_1 + \cdots + X_n$. Show that $\max_{1 \le k \le n} |X_k|/\sqrt{n} \to 0$ a.s.

(a) $\{|X_n|^r\}$ is uniformly integrable.

(b) $X_n \to X$ in $\mathcal{L}^r$.

(c) $E(|X_n|^r) \to E(|X|^r) < \infty$.

12.2 Solution

Gut (2014) 책에는 almost sure convergence에만 초점을 맞추어 서술하였다. (Theorem 5.5.2) 실제 converge in probability에서도 성립함이 알려져 있다. (Theorem 5.5.4) 이 증명을 하기 위해서는 Fatou's lemma가 필요하다.

13 Summary

In summary, this book has no content whatsoever.

```
1 + 1
```

```
[1] 2
```

References

Durrett, Rick. 2019. Probability: Theory and Examples. 5th ed. Cambridge University Press. https://www.ebook.de/de/product/34699864/rick_duke_university_north_carolina_durrett_probability.html.

Gut, Allan. 2014. Probability: A Graduate Course. 2nd ed. Springer New York.

Knuth, Donald E. 1984. "Literate Programming." Comput. J. 27 (2): 97–111. https://doi.org/10.1093/comjnl/27.2.97.