A Biggner’s Guide to Probability and Extremes

Seoncheol Park

2024-05-30

Table of contents

Preface

확률론은 통계학을 공부하는 데 있어 굉장히 중요한 과목이다. 그러므로 열심히 공부해야 한다.

This is a Quarto book.

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

```
1 + 1
```

```
[1] 2
```

Part I

Intro

1 Introduction

1.1 Probability Theory

- Probability models: random experiment를 묘사하는데 목적이 있음
- Random experiment: 무작위성이 있어 미래에 일어날 결과물을 정확하게 예측할 수 없는 실험
- **Probability space**: 확률론의 기초가 됨, 확률공간의 키가 되는 아이디어는 **stabilization of the relative frequencies**임

우리가 random experiement를 **독립**적으로, 반복적으로 수행한다고 하고 어떤 특정한 **사건(event)** A가 일어나는지 아닌지를 기록한다고 하자. $f_n(A)$를 처음 n개의 독립시행에서 A사건이 일어난 횟수라고 하고, $r_n(A) = f_n(A)/n$이라고 하자. 그러면 이 relative frequency $r_n(A)$는 $n \to \infty$일 때 다음과 같다고 생각하는 것이다(stabilization).

$$r_n(A) \overset{n\to\infty}{\longrightarrow} \text{some real number.}$$

Part II

Probability Theory

2 The Elements of Proability Theory

2.1 Probability Triples

다음은 콜모고로프 가 정리한 수리적 기반의 확률론이다.

Q. 왜 probability triple이 필요한가? Single도 아니고 double도 아니고 왜 triple이어야 하는가?

- **Sample space** Ω (표본공간): 이것은 any non-empty set이면 된다. 예를 들어 uniform distribution일 때 $\Omega = [0, 1]$이 있다.
- $\mathcal{F}$: σ-algebra 또는 σ-field: 이것은 Ω의 subset들의 collection으로 $\emptyset, \Omega$ 등을 포함한다.
- **Probability** P: a mapping from $\mathcal{F}$ to $[0, 1]$ with
  - $P(\emptyset) = 0$
  - $P(\Omega) = 1$
  - P is countably additive, $P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots$

2.2 Field and σ-field

Definition 2.1 (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 2.2 (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. $A_1, A_2, \ldots \in \mathcal{F} \Longrightarrow \cup_{n=1}^{\infty} A_n \in \mathcal{F}$

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

2.3 $\pi - \lambda$ System

Some intuition for $\pi - \lambda$ is that you can take a finite non π-system such as $S = \{\{1,2\},\{2,3\}\}$, and this is not enought to guarantee uniqueness on the σ-algebra generated by S, which includes sets like $\{2\},\{1,2,3\}$. But, at least in the countable case, you can use the π-system property to do disjointification/partitioning on Ω, which finished the proof.

Lemma 2.1 (σ-algebra and π-λ system)**.** A family of sets is a σ-algebra iff it is both π and λ.

2.4 Probabilities

::: {#def-prob}

2.4.1 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 $A_1, \ldots, A_n \in \mathcal{A}$ 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 $A_1, A_2, \ldots, \mathcal{A}$ 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 $A_1, A_2, \ldots \in \mathcal{A}$ which are pairwise disjoint.

Example 2.1 (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 $i = 1, \ldots, n$, or $C_k = A'_j$ for some $j = 1, \ldots, m$. 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.

3 Random Variables

3.1 Random Variables

Definition 3.1 (Random Variables)**.** Given a probability triple $(\Omega, \mathcal{F}, P)$, a **random variable** is a function X from Ω to $\mathbb{R}$, such that

$$\{\omega \in \Omega; X(\omega) \leq x\} \in \mathcal{F}, \quad x \in \mathbb{R}.$$

Q. Random variable을 정의하는데 왜 inverse image를 쓰는가?

Commonly a probability measure P is added to $(\Omega, \mathcal{F})$. Then sets like $\{X \in A\} := \{\omega \in \Omega | X(\omega) \in A\}$ can $= X^{-1}(A)$ be **measured** if they belong to $\mathcal{F}$. 예를 들면 $X : \Omega \to \mathbb{R}$이 확률변수일 때 $X < 1$일 확률을 구하려면 $X^{-1}(-\infty, 1)$이 가측이어야 할 것이다.

3.2 Radon-nikodym derivative

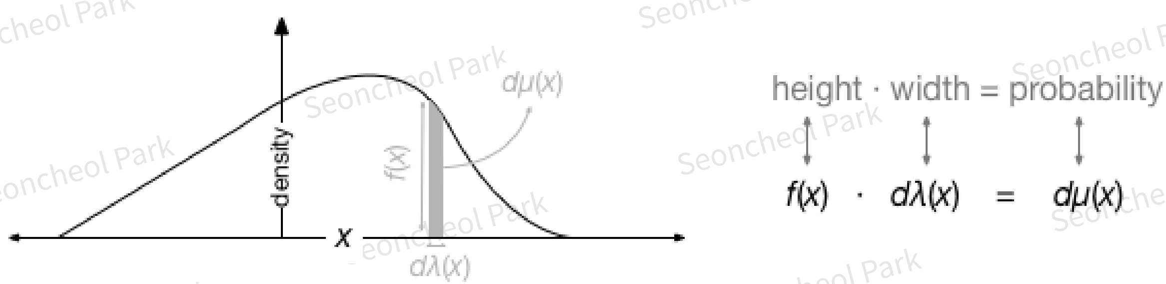

Figure 3.1: Change of measures.

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.

3.3 Integration

3.4 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 x는 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_n은 정적분을 위해 구간 $[a, b]$를 나눈 점, t_n은 닫힌 세부공간 $[x_n, x_{n+1}]$사이에 있는 임의점이다.

3.5 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 $\tilde{m}(A) = P(X^{-1}(A)) = \int_A hdm$ for any $A \in \mathcal{B}(\mathbb{R})$, where m is the Lebesgue measure. By **change of variables**, we have

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

Combining these equations,

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

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

Example 3.1. Given a probability measure P and sample space Ω, it is true that

$$\int_\Omega dP = 1.$$

Because

$$\int_\Omega dP = P(\Omega) = 1.$$

More generally

$$\int_A dP = \int_\Omega 1_A dP = P(A), \quad A \in \mathcal{F}.$$

Definition 3.3 ($\mathcal{L}^p$)**.** 다음과 같은 확률공간 $(\Omega, \mathcal{F}, P)$를 생각하자. $p > 1$에 대해, 확률변수 X가 $E|X|^p < \infty$이면 $X \in \mathcal{L}^p$라고 하며 다음과 같은 놈 $\|X_p\| = (E|X|^p)^{\frac{1}{p}}$를 정의할 수 있다.

4 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?

4.1 Definitions

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

Definition 4.1 (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, $X_n \stackrel{\text{a.s.}}{\to} X$ as $n \to \infty$라 쓴다.

Definition 4.2 (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, $X_n \stackrel{\text{p}}{\to} X$ as $n \to \infty$라 쓴다.

Definition 4.3 (Converge in r-mean)**.** 확률변수열 X_n 가

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

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 4.4 (Converge in Distribution)**.** $C(F_X) = \{x : F_X(x) \text{ is continuous at } x\} =$ the continuity set of F_X 라 하자. 확률변수열 X_n 가

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

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_n 가 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 4.5 (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 $F_n(x) = P(X_n \le x)$ converges weakly.

(Polansky 2011)의 4장에서는 converges weakly를 converge in distribution을 정의할 때 쓴 random variable의 sequence를 생략한 채 $\{F_n\}_{n=1}^{\infty}$ 와 F 로만 정의한 것으로 보았다. 또한 converges weakly를 $F_n \rightsquigarrow F$ 로 표기하기도 하였다.

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

Definition 4.6 (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 $n \to \infty$.

4.2 Bounded in Probability

(Polansky 2011) Convergence of distribution과 관련된 중요한 질문 중 하나는 limiting distribution이 valid distribution function이냐는 것이다. 이 말인 즉슨 sequence of distribution functions $\{F_n\}_{n=1}^{\infty}$가

$$\lim_{n\to\infty} F_n(x) = F(x), \quad \forall x \in C(F) \quad \text{for some fct } F(x)$$

일 때 $F(x)$가 distribution functions여야 할 필요가 있는가? 여기에서 $F(x) \in [0,1]$, $F(x)$가 non-decreasing, right continuity 등은 미적분학 등의 내용을 이용해 보일 수 있으므로

$$\lim_{x\to\infty} F(x) = 1, \qquad \lim_{x\to-\infty} F(x) = 0$$

을 만족하면 F가 valid distribution function이 될 것이다.

Definition 4.7 (Bounded in Probability). Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of random variables. The sequence is **bounded in probability** if for every $\varepsilon > 0$, $\exists x_\varepsilon \in \mathbb{R}$ and $n_\varepsilon \in \mathbb{N}$ such that $P(|X_n| \leq x_\varepsilon) > 1 - \varepsilon$ for all $n > n_\varepsilon$.

Example 4.1 (Bounded in Probability가 아닌 확률변수열). Consider the situation that $\{X_n\}_{n=1}^{\infty}$ is a sequence of random variables sucht that the distribution function of X_n is given by

$$F_n(x) = \begin{cases} 0, & x < 0 \\ 1 - p_n, & 0 \leq x < n \\ 1, & x \geq n \end{cases},$$

where $\{p_n\}_{n=1}^{\infty}$ is a sequence of real numbers such that

$$\lim_{n\to\infty} p_n = p.$$

- $p = 0$이면 bounded in probability

- 그러나 $p > 0$이면 we set a value of ε such that $0 < \varepsilon < p$. Let x be a positive real value. For any $n > x$ we have the property that $P(|X_m| \le x) = 1 - p \le 1 - \varepsilon$ for all $m > n$. Therefore, it is not possible to find the value of x required in the definition of bounded in probability.

Theorem 4.1 (Bounded in Probability는 Limiting distribution이 valid인 것과 동치)**.** Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of random variables where X_n has distribution function F_n for all $n \in \mathbb{N}$. Suppose that $F_n \rightsquigarrow F$ as $n \to \infty$ where F may or may not be a vaild distribution function. Then,

$$\lim_{x \to \infty} F(x) = 1, \quad \lim_{x \to -\infty} F(x) = 0$$

if and only if the sequence $\{X_n\}_{n=1}^{\infty}$ is bounded in probability.

4.3 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 4.8 (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)

4.4 Convergence of Moments

위키의 설명에 따르면 $X_n \xrightarrow{L^r} X$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)

4.5 Almost Sure Convergence

4.6 Convergence in Probability

4.7 Convergence in Distribution

5 The Law of Large Numbers

6 Preliminaries

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

6.1 Moments and Tails

Proposition 6.1 (Moments and Tails)**.**

1. Let $r > 0$. Suppose that X is a non-negative random variable. Then

$$EX^r < \infty \implies x^r P(X > x) \to 0 \quad \text{as } x \to \infty,$$

   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

$$\left| E \sum_{k=1}^{n} X_k I\{|X_k| \le a\} \right| \le nE|X|I\{|X| > a\}.$$

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 $EY = 0$.

7 A Weak Law for Partial Maxima

Part III

Extremes

8 Extreme Value Theory

9 Multivariate Extreme Value Theory

9.1 Pseudo-Polar Transforms

Kiriliouk et al. (2016) describe a **pseudo-polar** representation of bivariate data as a means to explore right-tail extremal dependency between the variables.

Let (X_i, Y_i) (real values) or (U_i, V_i) (as probabilities) for $i = 1, \ldots, n$ be a bivariate sample of size n. When such data are transformed into a **unit-Pareto** scale by

$$\widehat{X}_i^* = \frac{n}{n+1-R_{X,i}}, \quad \widehat{Y}_i^* = \frac{n}{n+1-R_{Y,i}},$$

where R is `rank()`, then letting each **component sum** or **pseudo-polar radius** be defined as

$$\widehat{S}_i = \widehat{X}_i^* + \widehat{Y}_i^*,$$

and each respective **pseudo-polar angle** be defined as

$$\widehat{W}_i = \frac{\widehat{X}_i^*}{\widehat{X}_i^* + \widehat{Y}_i^*} = \frac{\widehat{X}_i^*}{\widehat{S}_i}$$

a **pseudo-polar representation** is available for study.

A scatter plot of $\widehat{W}_i$ (horizontal) versus $\widehat{S}_i$ (vertical) will depict a **pseudo-polar plot** of the data.

A density plot of the $\widehat{W}_i$ is a representation of extremal dependence.

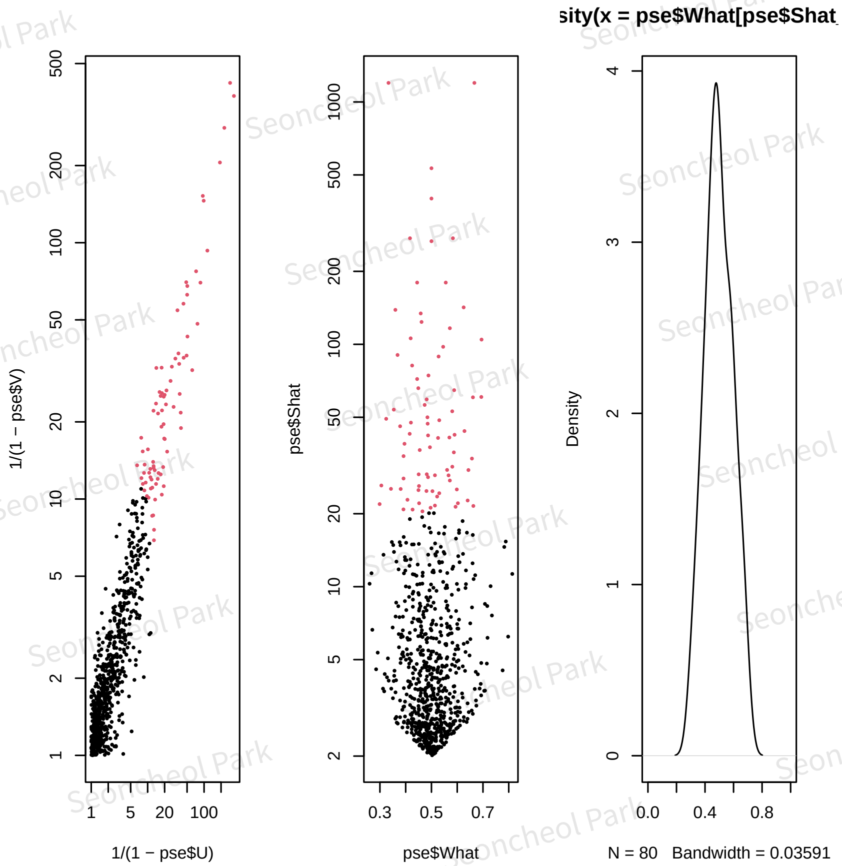

Figure 9.1: Figure: Extremal dependence.

9.2 Clustering Methods in Extremes

- Handbook on Statistics of Extremes 책 발간 예정
- Vector quantization

9.2.1 K-means clustering

- Given obs $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$, find K cluster centroids $\boldsymbol{c}_1, \dots, \boldsymbol{c}_K$ s.t. the avg data-point-to-centroid dist is minimized:

$$(\boldsymbol{c}_1, \dots, \boldsymbol{c}_K) := \arg \min_{\boldsymbol{c}_1, \dots, \boldsymbol{c}_K} \sum_{k=1}^{K}$$

- Estimate the centroids $\boldsymbol{c}_k$ and the cluster membership of each $\boldsymbol{x}_i$ in turns.
  - Given $\hat{\boldsymbol{c}}_1, \ldots, \hat{\boldsymbol{c}}_K$, assign $\boldsymbol{x}_i$ to the cluster k with the closest centroid $\hat{\boldsymbol{c}}_k$.

$$i \in C_k \iff d(\boldsymbol{x}_i, c_k) = \min_{k'} d(\boldsymbol{x}_i, \boldsymbol{c}_{k'})$$

  - Given all $\boldsymbol{x}_i$'s in cluster k, update each $\hat{\boldsymbol{c}}_k$

Q. Choice of $d(\cdot,\cdot)$: + Euclidean $d(\boldsymbol{x},\boldsymbol{y}) = (\boldsymbol{x}-\boldsymbol{y})^T(\boldsymbol{x}-\boldsymbol{y})$ + Then the centroids can be calculated as

$$\hat{\boldsymbol{c}}_k = \arg\min_{\boldsymbol{c}} \sum_{i \in C_k} (\boldsymbol{x}_i - \boldsymbol{c})^T(\boldsymbol{x}_i - \boldsymbol{c}) = \frac{1}{|C_k|} \sum_{i \in C_k}$$

Q. Choice of K: + Prespecified + Use a scree plot where the obj fct

$$\min_{\boldsymbol{c}_1,\ldots,\boldsymbol{c}_K} \sum_{k=1}^{K} \sum_{i \in C_k} d(\boldsymbol{x}_i, \boldsymbol{c}_k)$$

이러한 K-mean 같이 Euclidean dist를 쓰는 방법은 extreme value에서 통하기 어려움

9.3 Spectral Clustering

- Can detect nonlinear cluster patterns
- Can identify noise clusters

9.4 Clustering the Angluar Components

- $\boldsymbol{Y}$ be multivariate regularly varying with standardized margin (Frechet 등이 해당) Then

$$\frac{\boldsymbol{Y}}{\|\boldsymbol{Y}\|}_{\|\boldsymbol{Y}\|>t} \xrightarrow{d} \Theta, \quad t \to \infty.$$

- Clustering for extremes:
  - Obtain angular compts $\Theta_1, \ldots, \Theta_{k_n}$ from $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$

- Cluster $\Theta_1, \dots, \Theta_{k_n}$ instead.

여기서 Θ는 unit sphere $\{\boldsymbol{x}\|\boldsymbol{x}\| = 1\}$이라는 매우 좋은 space에 놓여 있다. (이때 $\|\cdot\|$은 any norm이나 되지만 $L2$ norm을 쓰기로 한다)

9.5 Max-Linear Models

- Max-linear random vector:

$$\boldsymbol{X} = (X_1, \dots, X_d) = \vee_{i=1,\dots,K} \boldsymbol{b}_i Z_i$$

- Factors $\boldsymbol{b}_1, \dots, \boldsymbol{b}_K \in [0, \infty)^d$
- $Z_1, \dots, Z_k$: i.i.d. Frechet

Then the angular measure Θ consists of point masses at

$$\frac{\boldsymbol{b}_1}{\|\boldsymbol{b}_1\|}, \dots$$

9.5.1 Spherical K-means

- Apply to $\Theta_1, \dots, \Theta_{k_n}$: K-means clustering with choice of distance

$$d(\boldsymbol{x}, \boldsymbol{y}) = 1 - \cos(\boldsymbol{x}, \boldsymbol{y})$$

On the unit sphere $\mathbb{S}_+^{d-1}$,

- $d(\boldsymbol{x}, \boldsymbol{y}) = 1 - \boldsymbol{x}^T \boldsymbol{y}$
- d is equiv to the Euclidean dist

9.5.2 Spherical K-PCs clustering for extremes

- 앞선 방법과 달리 $d(\boldsymbol{x}, \boldsymbol{y}) = 1 - (\boldsymbol{x}^T \boldsymbol{y})^2$을 쓰는 것이 차이점(제곱이 들어감)
- https://academic.oup.com/biomet/article-abstract/110/1/135/6551983?redirected-From=PDF
- arg $\max_{\|\boldsymbol{c}\|_2=1} \boldsymbol{c}^T \Sigma_k \boldsymbol{c}$ 형태가 나옴

- For any spectral measure that can be decomposed into two sub-faces l_1 and l_2, we would like the optimal centroids to satisfy

$$\boldsymbol{c}_1 \in \mathbb{F}_{l_1}, \quad \boldsymbol{c}_2 \in \mathbb{F}_{l_2}$$

- This holds for spheical K-means iff

$$\||l_1| - |l_2|\| \leq 1$$

- This holds for spherical K-PCs always.
- If angular components $\boldsymbol{x}$ and $\boldsymbol{y}$ belongs to different sub-faces, then $\boldsymbol{x}^T\boldsymbol{y}$ close to 0.

9.5.3 Spectral clustering for extremes

- Linear factor model with noise:

$$\boldsymbol{X} = (X_1, \ldots, X_d) = \sum_{i=1}^{K}$$

Part IV

Intro

10 Summary

In summary, this book has no content whatsoever.

```r
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.

Polansky, Alan M. 2011. Introduction to Statistical Limit Theory. CRC Press.