Contents

Part I

Unit 1: 확률의 기초와 이산 확률 변수

〔구조〕 이 교재의 전체 구조 (Roadmap)

- Unit 1: 확률의 기초와 이산 확률 변수 (Fundamentals & Discrete Random Variables)
  → Chapter 1: 확률 모형의 수립과 공리 (Probability Models & Axioms) 〈- 현재 위치
  - Chapter 2: 조건부 확률과 베이즈 정리
  - Chapter 3: 독립성과 이산 확률 변수
- Unit 2: 일반 확률 변수 (General Random Variables)
- Unit 3: 확률 과정과 극한 정리 (Random Processes & Limit Theorems)

Chapter 1

확률 모형의 수립과 기초 공리

Intro: 불확실한 세상으로의 첫걸음

〔이전 단계와의 연결〕
지금까지 여러분은 $1+1=2$와 같이 결과가 정해져 있는 '결정론적(Deterministic)' 세계에서 수학을 배웠습니다. 하지만 현실은 "내일 비가 올까?", "이 주식이 오를까?"처럼 불확실성으로 가득 차 있습니다. 이제 우리는 확실하지 않은 미래를 숫자로 다루는 법을 배웁니다.

Summary

이 단원의 핵심 개요 (Overview)〕 이 단원에서는 현실 세계의 모호한 문제를 수학적 '집합'으로 번역하는 방법을 배웁니다.

1. 모델링: 일어날 수 있는 모든 일을 표본 공간(Ω)이라는 우주로 정의합니다.
2. 규칙(공리): 확률이 무너지지 않게 지탱하는 3가지 절대 규칙(콜모고로프 공리)을 배웁니다.
3. 도구(셈하기): 경우의 수를 정확히 세서 확률을 계산하는 기술(순열/조합)을 익힙니다.

1.1 필수 용어 사전 (Terminology)

확률론은 '집합의 언어'를 사용합니다. 이 번역표를 머릿속에 넣어두세요.

mainblue!20 기호	용어 (한국어/영어)	직관적 의미
Ω	표본 공간 (Sample Space)	일어날 수 있는 모든 결과의 전체 집합 (우주)
ω	근원 사건 (Outcome)	실험의 가장 작은 결과 단위 (원소)
A, B	사건 (Event)	우리가 관심 있는 특정 결과들의 모임 (부분집합)
$P(A)$	확률 (Probability)	사건 A가 일어날 믿음의 정도 (크기, 무게)
$A \cap B$	교집합 (Intersection)	A 그리고 B가 동시에 일어남
$A \cup B$	합집합 (Union)	A 또는 B가 일어남

1.2 핵심 개념 1: 확률 모형의 수립 (Modeling)

1.2.1 실험과 표본 공간 (Ω)

〔개념〕 표본 공간은 우리가 노는 '운동장'이다

- 한 줄 요약: 실험에서 나올 수 있는 모든 결과를 빠짐없이 모아둔 집합입니다.
- 직관적 비유: 식당의 전체 메뉴판. 내가 무엇을 주문하든, 반드시 메뉴판 안에 있는 것이어야 합니다.
- 기술적 정의: 상호 배타적(Mutually Exclusive)이고 전체를 포괄하는(Collectively Exhaustive) 모든 결과들의 집합 Ω.
- 구체적 예시: 동전을 한 번 던질 때 $\Omega = \{$앞면(H), 뒷면(T)$\}$.

〔주의〕 잠깐! 오해하기 쉬워요: Outcome vs Event 구분하기

- Outcome (근원 사건): 주사위를 던져 '1'이 나오는 것. (더 쪼갤 수 없음)
- Event (사건): 주사위를 던져 '홀수'가 나오는 것 ($\{1, 3, 5\}$). (Outcome들의 묶음)
- 확률론에서는 주로 **Event(집합)**에 확률을 부여합니다.

1.2.2 사건(Event)과 집합 연산

현실의 언어를 집합의 언어로 번역해야 합니다.

- "적어도 한 번 앞면" $\rightarrow \{HHT, HTH, THH, HHH, \ldots\}$
- "A가 일어나지 않음" $\rightarrow A^c$ (여집합)
- "A와 B가 겹치지 않음" $\rightarrow A \cap B = \emptyset$ (배반 사건, Disjoint)

1.3 핵심 개념 2: 확률의 3공리 (The Axioms)

확률은 감으로 찍는 것이 아닙니다. 러시아 수학자 콜모고로프가 만든 3가지 절대 규칙 위에서만 작동합니다.

〔개념〕 콜모고로프의 3공리

1. Non-negativity (비음수성): 확률은 절대 음수가 될 수 없다.
$$P(A) \geq 0$$
2. Normalization (정규화): 전체 우주(모든 가능성)의 확률 합은 1(100%)이다.
$$P(\Omega) = 1$$
3. Additivity (가산성): 서로 겹치지 않는(Disjoint) 사건들의 합집합 확률은 각 확률의 단순 합과 같다.
$$\text{If } A \cap B = \emptyset, \text{ then } P(A \cup B) = P(A) + P(B)$$

〔예제〕 예시: 공리의 적용: 케이크 자르기

- 전체 케이크(Ω)의 크기는 1입니다. (공리 2)
- 케이크 한 조각(A)의 크기는 0보다 큽니다. (공리 1)
- 딸기 조각(A)과 초코 조각(B)이 겹치지 않는다면, 두 조각을 합친 크기는 그냥 두 조각의 무게를 더하면 됩니다. (공리 3)

1.3.1 공리에서 유도된 유용한 도구들

- 포함-배제 원리: 겹치는 게 있다면?

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

(두 번 더해진 교집합 부분을 한 번 빼줘야 함)

- Union Bound (안전빵 부등식): 겹치는지 아닌지 모를 때 최대 확률은?

$$P(A \cup B) \leq P(A) + P(B)$$

(엔지니어링에서 최악의 시나리오를 계산할 때 매우 중요!)

1.4 핵심 개념 3: 셈하기 (Counting)

모든 결과가 동등하게 일어날 때(주사위 등), 확률 계산은 결국 **"분모와 분자를 잘 세는 싸움"**입니다.

$$P(A) = \frac{\text{사건 A의 경우의 수}}{\text{전체 경우의 수}}$$

1.4.1 셈하기 결정 트리 (Decision Tree)

문제를 보자마자 다음 두 질문을 던지세요.

1. 순서가 중요한가? (Order matters?)
   - Yes $\rightarrow$ 순열 (Permutation): 비밀번호 1234와 4321은 다르다.
   - No $\rightarrow$ 조합 (Combination): 로또 번호 1, 5, 10과 10, 5, 1은 같다.
2. 중복을 허용하는가? (Replacement?)
   - 뽑고 다시 넣나(복원), 아니면 버리나(비복원).

〔시나리오〕 스토리 시나리오: 비밀요원 K의 가방 열기 (Counting 시나리오)
비밀요원 K가 3자리 숫자 자물쇠가 달린 가방을 열어야 합니다.

1. **상황 A (순열):** 비밀번호는 0 9 숫자 중 서로 다른 3개로 이루어져 있고, 순서가 중요합니다.
   - 계산: 첫 번째 칸 10개 $\times$ 두 번째 칸 9개 $\times$ 세 번째 칸 8개
   - $10 \times 9 \times 8 = 720$가지.
2. **상황 B (조합):** 사실 자물쇠가 아니라, 10개의 버튼 중 3개를 동시에 누르면 열리는 방식이었습니다. (순서 상관 없음)
   - 계산: 순열(720)에서 순서 섞이는 경우($3! = 3 \times 2 \times 1 = 6$)를 나눠줘야 합니다.
   - $\binom{10}{3} = \frac{720}{6} = 120$가지.

교훈: "순서"가 사라지니 경우의 수가 720에서 120으로 확 줄어듭니다!

1.5 실전 응용: 시스템 안정성 (Union Bound)

상황: 당신은 서버 관리자입니다. 서버 A가 다운될 확률은 5%(0.05), 서버 B가 다운될 확률은 10%(0.10)입니다. 두 서버가 동시에 다운되는지, 서로 영향을 주는지는 복잡해서 정확히 모릅니다.

질문: "적어도 하나의 서버가 다운될 최악의 확률은 얼마인가?"

해결: Union Bound를 사용합니다.

$$P(A \cup B) \leq P(A) + P(B) = 0.05 + 0.10 = 0.15$$

결론: 정확한 상관관계를 몰라도, "어쨌든 전체 시스템 에러율은 15%를 넘지는 않겠군"이라고 보수적으로 판단하고 대비할 수 있습니다. 이것이 공학적 사고입니다.

1.6 초심자가 자주 묻는 질문 (FAQ)

Q1. 왜 확률은 1을 넘을 수 없나요?

확률은 '전체(표본 공간 Ω)'에 대한 '부분(사건 A)'의 비율(무게)입니다. 부분이 전체보다 클 수는 없습니다. 만약 계산 결과가 1.2가 나왔다면, 어딘가(주로 겹치는 부분)를 중복해서 더한 것입니다.

Q2. 언제 곱하고, 언제 더하나요?

- 곱할 때: 단계적으로 일이 진행될 때 ("A 하고, 그리고 나서 B 한다"). 예: 옷 입기 (상의 $\times$ 하의).
- 더할 때: 경우를 나눌 때 ("A인 경우 또는 B인 경우"). 예: 등교 방법 (버스 타는 수 + 지하철 타는 수).

Q3. 주사위 2개를 던질 때 표본 공간을 $\{2, 3, \ldots, 12\}$로 잡으면 안 되나요?

가능은 합니다. 하지만 추천하지 않습니다. 왜냐하면 합이 2가 되는 경우(1,1)와 합이 7이 되는 경우(1,6, 2,5, 3,4...)의 확률이 다르기 때문입니다. 계산을 쉽게 하려면 모든 근원 사건의 확률이 같은 '가장 잘게 쪼개진' 모델($\{(1,1), \ldots, (6,6)\}$)을 쓰는 것이 좋습니다.

1.7 요약 및 다음 단계

Summary

Unit 1-1 핵심 요약〕

1. 모델링: 문제를 집합(Ω)과 부분집합(Event)으로 바꿔라.
2. 공리: 확률의 3원칙(0 이상, 전체는 1, 겹치지 않으면 더하기)을 항상 기억하라.
3. 셈하기: 순서가 있는지(Permutation), 없는지(Combination) 먼저 판단하라.

〔다음 단원 예고〕

이번 장에서는 모든 사건이 똑같은 확률로 일어난다고 가정하거나, 이미 확률이 주어졌다고 쳤습니다. 하지만 현실에서 내일 비가 올 확률은 어떻게 알 수 있을까요? 새로운 정보(구름이 꼈다)가 들어오면 확률은 어떻게 바뀔까요?

다음 장 Chapter 2: 조건부 확률(Conditional Probability)에서 그 비밀을 밝혀냅니다. "정보가 곧 확률을 바꿉니다."

Chapter 2

조건부 확률과 추론 (Inference)

Intro: 정보는 확률을 바꾼다

〔이전 단계와의 연결〕
Chapter 1에서는 아무런 정보가 없는 상태에서 확률을 계산했습니다. 하지만 현실에서는 끊임없이 **"정보(단서)"**가 들어옵니다. "구름이 끼었다"는 정보를 알았을 때, 비가 올 확률은 변해야 합니다. 이번 장에서는 **정보가 주어졌을 때 확률을 업데이트하는 법**을 배웁니다.

Summary

이 단원의 핵심 개요 (Overview)〕

1. 조건부 확률 (Renormalization): 새로운 정보 B가 주어지면, B가 곧 새로운 우주(분모)가 됩니다.
2. 베이즈 정리 (Inference): 결과를 보고 원인을 역추적하는 '탐정의 도구'를 배웁니다.
3. 독립성 (Independence): 정보가 가치가 있는지(확률에 영향을 주는지) 판단합니다.

2.1 필수 용어 사전

mainblue!20 기호	용어	직관적 의미
$P(A\|B)$	조건부 확률	B가 일어났다는 전제 하에 A가 일어날 확률
$P(A \cap B)$	결합 확률	A와 B가 동시에 일어날 확률 (교집합)
Partition	분할	전체를 겹치지 않게 조각내는 것 (케이크 자르기)
Prior	사전 확률	데이터를 보기 전의 믿음 (P(원인))
Posterior	사후 확률	데이터를 본 후 수정된 믿음 (P(원인|결과))
Independent	독립	B를 아는 것이 A 예측에 도움이 안 됨 ($P(A\|B) = P(A)$)

2.2 핵심 개념 1: 조건부 확률 (Conditional Probability)

2.2.1 표본 공간의 재정의 (Renormalization)

〔개념〕 새로운 정보는 세상을 축소시킨다

- 한 줄 요약: 사건 B가 일어났다면, B 바깥의 세상은 소멸하고 B가 새로운 전체(Ω_{new})가 됩니다.
- 직관적 비유: 지도 어플에서 '서울시'를 검색하면, 지도 화면(표본 공간)이 대한민국 전체에서 서울시로 확대(Zoom-in)되는 것과 같습니다.
- 공식:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

(분모가 1이 아니라 $P(B)$로 바뀜 $\rightarrow$ 이것이 Renormalization!)

〔예제〕 예시: 주사위 던지기

주사위를 던졌는데 "짝수가 나왔다(B)"는 힌트를 받았습니다. 이때 그 숫자가 "2 이하(A)"일 확률은?

1. 원래 세상(Ω): $\{1, 2, 3, 4, 5, 6\}$. 확률은 $2/6 = 1/3$.
2. 바뀐 세상(B): $\{2, 4, 6\}$. 이제 전체는 3개입니다.
3. 겹치는 부분($A \cap B$): $\{2\}$ 하나뿐입니다.
4. 계산: $P(A|B) = \frac{1}{3}$. (3개 중 1개)

2.2.2 곱셈 법칙 (Multiplication Rule)

조건부 확률 식을 살짝 바꾸면, 사건이 **순차적**으로 일어나는 과정을 설명할 수 있습니다.

$$P(A \cap B) = P(B) \cdot P(A|B)$$

해석: "먼저 B가 일어나고, 그 상황 속에서 A가 일어난다."

2.3 핵심 개념 2: 전체 확률과 베이즈 정리 (The Core)

이 부분은 MIT 수업에서 가장 강조하는 **"시스템적 추론(Inference)"**의 핵심입니다.

2.3.1 전체 확률의 정리 (Total Probability Theorem)

복잡한 문제를 해결하는 전략: **"분할 정복(Divide and Conquer)"**

- B가 일어날 확률을 한 번에 구하기 어렵다면, 원인별 시나리오($A_1, A_2, \ldots$)로 쪼개서 계산한 뒤 합칩니다.

$$P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \ldots$$

2.3.2 베이즈 정리 (Bayes' Rule)

시간을 거스르는 추론

- 핵심 사고: 결과를 관측(B)한 뒤, 그 원인이 무엇이었는지(A) 역추적합니다.
- 공식:

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{P(B)}$$

- 구조:

$$\text{사후 확률(Posterior)} = \frac{\text{사전 확률(Prior)} \times \text{우도(Likelihood)}}{\text{전체 확률(Evidence)}}$$

〔시나리오〕 스토리 시나리오: 공장 불량품 추적 시나리오
당신은 품질 관리자입니다.

- 공장 A: 전체 부품의 60% 생산 ($P(A) = 0.6$), 불량률 1% ($P(E|A) = 0.01$)
- 공장 B: 전체 부품의 40% 생산 ($P(B) = 0.4$), 불량률 2% ($P(E|B) = 0.02$)

랜덤하게 뽑은 부품이 불량(Error, E)이었습니다. 이 부품이 공장 A에서 왔을 확률은?
1단계: 전체 불량률 구하기 (분모)

$$P(E) = (0.6 \times 0.01) + (0.4 \times 0.02) = 0.006 + 0.008 = 0.014$$

(전체 중 1.4%가 불량)
2단계: 베이즈 정리 적용 (역추적)

$$P(A|E) = \frac{P(A)P(E|A)}{P(E)} = \frac{0.006}{0.014} = \frac{3}{7} \approx 42.8\%$$

결론: 공장 A가 점유율은 더 높지만(60%), 불량품이 나왔다는 증거(Evidence)를 보고 나니 공장 A일 확률이 42.8%로 줄어들었습니다. (공장 B일 확률이 높아짐)

2.4 핵심 개념 3: 사건의 독립성 (Independence)

〔개념〕 정보의 가치 평가

- 정의: $P(A \cap B) = P(A)P(B)$
- 의미: $P(A|B) = P(A)$. 즉, "B를 알게 되어도 A에 대한 나의 믿음은 변하지 않는다."
- 예: "어제 주식 가격이 올랐다(B)"는 정보는 "오늘 내가 점심에 카레를 먹을지(A)"를 예측하는 데 아무런 도움이 안 됩니다. (독립)

〔주의〕 잠깐! 오해하기 쉬워요: 독립(Independent) vs 배반(Disjoint)
이 둘을 절대 혼동하면 안 됩니다!

- 배반 (Disjoint): A와 B는 겹치지 않음 ($A \cap B = \emptyset$).
  - A가 일어나면 B는 절대 안 일어남.
  - 즉, A는 B에 대해 엄청난 정보를 줌. → 강한 종속성!
- 독립 (Independent): A가 일어나든 말든 B 발생 확률은 그대로임.
- 결론: 배반 사건은 (확률이 0이 아닌 이상) 절대로 독립일 수 없습니다.

2.5 실전 응용: 스팸 필터 (Naive Bayes)

이메일에 "무료"라는 단어가 들어있을 때(B), 이것이 스팸(A)일 확률을 어떻게 구할까요?

1. 과거 데이터를 통해 평소 스팸 올 확률 $P(A)$를 구합니다. (Prior)
2. 스팸 메일 중에서 "무료"라는 단어가 나올 확률 $P(B|A)$를 구합니다. (Likelihood)
3. "무료"라는 단어가 포함된 메일이 도착했습니다. 베이즈 정리를 돌려 $P(A|B)$를 계산합니다.
4. 이 확률이 95%를 넘으면 스팸함으로 보냅니다.

우리가 쓰는 지메일(Gmail) 필터의 기초 원리가 바로 이것입니다.

2.6 초심자가 자주 묻는 질문 (FAQ)

Q1. $P(A|B)$와 $P(B|A)$는 같은 거 아닌가요?
완전히 다릅니다! 이를 **"검사의 오류(Prosecutor's Fallacy)"**라고 합니다.

- $P(\text{죽음}|\text{상어에 물림}) \approx 100\%$ (매우 높음)
- $P(\text{상어에 물림}|\text{죽음}) \approx 0\%$ (대부분의 죽음은 상어와 무관함)

순서를 바꾸면 의미가 완전히 달라집니다.

Q2. 독립성을 직관적으로 어떻게 아나요?
직관은 위험합니다. 반드시 $P(A \cap B) = P(A)P(B)$인지 계산해보고 판단하세요. 특히 "조건부 독립(Conditional Independence)" 상황에서는 직관이 자주 틀립니다.

2.7 요약 및 다음 단계

Summary

Unit 1-2 핵심 요약]

1. 조건부 확률: 정보(B)는 분모를 바꾼다. ($/P(B)$)
2. 전체 확률: 복잡하면 시나리오별로 쪼개서(Σ) 계산해라.
3. 베이즈 정리: $P(\text{원인}|\text{결과})$를 구하려면 순서를 뒤집어라.
4. 독립성: $P(A|B) = P(A)$일 때만 정보다 가치가 없다. (배반과는 다르다!)

[다음 단원 예고]
지금까지는 "앞면/뒷면", "성공/실패" 같은 사건(Event) 중심이었습니다. 하지만 "앞면이 100번 중 몇 번 나왔지?"처럼 숫자로 요약하고 싶을 때가 많습니다.
다음 장 Chapter 3: 이산 확률 변수(Discrete Random Variables)에서는 사건을 숫자로 매핑하는 '확률 변수'와 'PMF'라는 강력한 도구를 만납니다.

Chapter 3

이산 확률 변수와 분포의 세계

Intro: 결과(Outcome)를 숫자(Number)로 바꾸다

〔이전 단계와의 연결〕
Chapter 2까지 우리는 "앞면/뒷면", "성공/실패" 같은 추상적인 사건(Event)을 다뤘습니다. 하지만 데이터 분석이나 공학에서는 "그래서 성공을 몇 번 했는데?", "평균적으로 얼마를 버는데?"와 같이 **숫자**로 요약된 정보가 필요합니다. 이제 우리는 추상적인 사건을 숫자로 매핑(Mapping)하는 법을 배웁니다.

Summary

이 단원의 핵심 개요 (Overview)〕

1. 확률 변수 (RV): 현실의 결과를 숫자로 바꿔주는 '함수'를 정의합니다.
2. PMF와 CDF: 확률이 어떻게 분포해 있는지 그래프(질량/누적)로 그립니다.
3. 기댓값과 분산: 전체 데이터를 '중심'과 '퍼짐'이라는 두 개의 숫자로 요약합니다.
4. 분포 동물원 (The Zoo): 이항, 기하, 포아송 등 상황별로 꺼내 쓰는 표준 모델을 배웁니다.

3.1 필수 용어 사전

mainblue!20 기호	용어	직관적 의미
X	확률 변수 (Random Variable)	결과를 숫자로 바꿔주는 함수 ($HH \to 2$)
$p_X(k)$	확률 질량 함수 (PMF)	숫자 k가 나올 확률의 높이 (무게)
$F_X(x)$	누적 분포 함수 (CDF)	x까지 쌓아 올린 확률의 합
$E[X]$	기댓값 (Expectation)	확률을 무게로 쳤을 때의 무게 중심 (평균)
$\mathrm{var}(X)$	분산 (Variance)	중심으로부터 데이터가 얼마나 퍼져 있는가

3.2 핵심 개념 1: 확률 변수는 '함수'다

3.2.1 확률 변수의 정의 (Mapping)

〔개념〕 Outcome → Number 변환기

- 한 줄 요약: 세상의 모든 결과(Outcome)에 꼬리표(숫자)를 붙이는 규칙입니다.
- 직관적 비유: 자판기입니다. 버튼(Outcome: 앞면, 앞면)을 누르면 음료수(Number: 2)가 나옵니다.
- 정의: 표본 공간 Ω의 각 원소를 실수 $\mathbb{R}$로 대응시키는 함수 $X(\omega)$.

〔예제〕 예시: 동전 2개 던지기

- 표본 공간 $\Omega = \{HH, HT, TH, TT\}$ (4가지 결과)
- 확률 변수 X 정의: "앞면의 개수"
- 매핑 과정:
  - $X(HH) = 2$
  - $X(HT) = 1, X(TH) = 1$
  - $X(TT) = 0$
- 이제 우리는 $\{HH, \ldots\}$가 아니라 숫자 $0, 1, 2$를 다룹니다.

3.2.2 PMF (확률 질량 함수) vs CDF (누적 분포 함수)

- PMF ($p_X(k)$): "정확히 k일 확률은?" (막대 그래프의 높이)

$$\sum p_X(k) = 1$$

- CDF ($F_X(x)$): "x 이하일 확률은?" (계단식으로 쌓이는 그래프)

$$F_X(x) = P(X \leq x) = \sum_{k \leq x} p_X(k)$$

3.3 핵심 개념 2: 데이터를 요약하는 기술

3.3.1 기댓값 (Expectation, $E[X]$)

기댓값은 단순한 '산술 평균(N빵)'이 아닙니다. **확률 가중치가 반영된 무게 중심**입니다.

- 공식: $E[X] = \sum x \cdot p_X(x)$
- 비유: 시소(See-saw)의 받침점. 확률(무게)이 큰 쪽으로 중심이 이동합니다.

기댓값의 선형성 (Linearity of Expectation) 복잡한 문제를 푸는 마법의 열쇠입니다.

$$E[X + Y] = E[X] + E[Y]$$

중요: X와 Y가 독립이 아니어도(상관 있어도) 무조건 성립합니다.

- 활용 팁: 전체 $E[X]$를 구하기 어려우면, 문제를 아주 작은 단위(지시 변수, Indicator)로 쪼개서 각각 기댓값을 구한 뒤 더하세요.

3.3.2 분산 (Variance, var(X))

- 개념: 데이터가 기댓값 주위에 모여 있나(안정적), 퍼져 있나(Risk)?
- 계산 꿀팁: 정의대로 계산하면 복잡합니다. 아래 공식을 쓰세요.

$$\text{var}(X) = E[X^2] - (E[X])^2$$

(제곱의 평균 - 평균의 제곱)

3.4 핵심 개념 3: 주요 이산 분포 (Story Matching)

공식을 외우기 전에, 어떤 **상황(Story)**에서 이 분포를 쓰는지 파악해야 합니다.

〔시나리오〕 스토리 시나리오: 1. 베르누이 이항 분포 (Binomial)
"동전 던지기의 반복"

- 상황: 성공 확률 p인 실험을 n번 독립적으로 반복했을 때, 성공 횟수 X.
- 예시: 고객 100명(n)에게 메일을 보냈을 때, 클릭($p = 0.1$)한 사람의 수.
- 기댓값: np (100명 × 0.1 = 10명)

〔시나리오〕 스토리 시나리오: 2. 기하 분포 (Geometric)
"성공할 때까지 도전!" (존버 정신)

- 상황: 성공 확률 p인 실험을 성공할 때까지 계속할 때, 시도 횟수 X.
- 예시: 소개팅에서 마음에 드는 사람($p = 0.01$)을 만날 때까지 나가는 횟수.
- 특징 (Memoryless): "지금까지 10번 차였다고 해서, 11번째 확률이 높아지는 건 아니다." (슬프지만 수학적 사실)

〔시나리오〕 스토리 시나리오: 3. 포아송 분포 (Poisson)
"매우 드문 사건의 카운팅"

- 상황: 이항 분포에서 횟수 $n \to \infty$, 확률 $p \to 0$인 극한 상황.
- 예시:
  - 책 한 페이지의 오타 수 (글자 수는 많고, 오타 확률은 낮음)
  - 1시간 동안 콜센터에 걸려오는 전화 수
- 특징: 평균(λ)과 분산(λ)이 같습니다.

3.5 실전 응용: 게임 아이템 뽑기 (Gacha)

당신은 게임 기획자입니다. 전설 아이템이 나올 확률은 $1\%(p = 0.01)$입니다.
Q1. 유저가 전설템을 먹을 때까지 평균 몇 번 뽑아야 할까? (Geometric)

$$E[X] = \frac{1}{p} = \frac{1}{0.01} = 100\text{번}$$

→ "평균 100번은 시도해야 나옵니다."
Q2. 유저가 50번 뽑기를 했을 때, 전설템을 1개도 못 먹을 확률은? (Binomial)

- 50번 시도($n = 50$), 성공 횟수 $k = 0$
- $P(X = 0) = \binom{50}{0}(0.01)^0(0.99)^{50} \approx 0.605$

→ "60.5%의 유저는 50번을 뽑아도 꽝입니다."
Q3. 서버에 1분당 평균 5명의 접속자가 온다. 1분 동안 아무도 안 올 확률은? (Poisson)

- 평균 $\lambda = 5$
- $P(X = 0) = e^{-5}\frac{5^0}{0!} = e^{-5} \approx 0.0067$

→ "0.67%, 즉 거의 일어나지 않는 일입니다."

3.6 확장: 변수가 2개일 때 (Joint PMF)

X(키)와 Y(몸무게)처럼 두 변수를 동시에 고려할 때는, 2차원 표(Grid)를 상상하세요.

- Joint PMF: 각 칸 (x, y)의 확률. $\sum\sum p_{X,Y}(x, y) = 1$.
- Marginal PMF: X만 알고 싶으면, Y 축을 따라 확률을 다 더해버리면(압축) 됩니다.

3.7 요약 및 다음 단계

Summary

Unit 1-3 핵심 요약]

1. 확률 변수: 결과를 숫자로 바꾸는 함수.
2. 기댓값: 선형성($E[X + Y] = E[X] + E[Y]$)을 적극 활용해라.
3. 패턴 매칭:
   - 횟수 고정 + 성공 수 → 이항 분포
   - 성공할 때까지 → 기하 분포
   - 아주 드문 사건 → 포아송 분포

[다음 단원 예고]
지금까지는 1, 2, 3처럼 뚝뚝 끊어지는 '이산(Discrete)' 데이터만 다뤘습니다. 하지만 시간, 온도, 속도처럼 연속적으로 변하는 데이터는 어떻게 다룰까요?
Unit 2에서는 $\sum$(더하기)가 $\int$(적분)으로 바뀌는 연속 확률 변수(Continuous Random Variables)의 세계로 넘어갑니다.

Part Ⅱ

Unit 2: 일반 확률 변수

Chapter 4

연속 확률 변수: 레고에서 유체로

Intro: 더하기($\sum$)에서 적분($\int$)으로

〔이전 단계와의 연결〕
Unit 1에서는 주사위 눈이나 동전 개수처럼 **하나하나 셀 수 있는(Countable)** '레고 블록' 같은 데이터를 다뤘습니다. 이때는 확률을 블록의 높이로 생각하고 단순히 더했습니다($\sum$).

하지만 현실의 시간, 온도, 속도, 전압은 뚝뚝 끊어져 있지 않고 **연속적으로 흐르는 유체(Fluid)**와 같습니다. 물의 양을 잴 때 블록 개수를 세지 않고 '부피'를 측정하듯이, 이제 우리는 확률을 **'구간의 면적(Area)'**으로 다루는 미적분의 세계로 들어갑니다.

Summary

이 단원의 핵심 개요 (Overview)〕

1. PDF (밀도): 점의 확률은 0입니다. 이제 '밀도'를 적분해야 확률이 됩니다.

2. 분포 동물원 (The Zoo): 균등(Uniform), 지수(Exponential), 정규(Normal) 분포의 탄생 배경을 배웁니다.

3. 표준화 (Standardization): 복잡한 정규 분포를 Z-score로 변환하여 쉽게 계산하는 법을 익힙니다.

4. 혼합 추론 (Mixed Bayes): 연속된 신호(전압)를 보고 이산적 원인(0 or 1)을 맞추는 통신의 기초를 다룹니다.

4.1 필수 용어 사전

mainblue!20 기호	용어	직관적 의미
PDF	확률 밀도 함수 ($f_X(x)$)	특정 지점에서의 확률의 **농도(높이)**. (확률 아님!)
CDF	누적 분포 함수 ($F_X(x)$)	$-\infty$부터 x까지 그래프 아래의 **면적** (진짜 확률)
Uniform	균등 분포	"정보가 없다." 모든 구간의 확률 밀도가 평평함.
Exponential	지수 분포	"대기 시간." 사건이 터질 때까지 걸리는 시간.
Normal	정규 분포 (Gaussian)	"오차의 합." 자연계의 잡음(Noise)을 설명하는 종 모양.
Z	표준 정규 분포	평균 0, 분산 1로 통일된 기준 정규 분포.

4.2 핵심 개념 1: 밀도(Density)는 확률이 아니다

4.2.1 확률 밀도 함수 (PDF, $f_X(x)$)

[Image of Integration Area under Curve]

[개념] 높이가 아니라 '면적'이 확률이다

- 한 줄 요약: PDF 그래프의 y값(높이)은 확률이 아니라 '밀도'이며, 이를 구간으로 **적분(넓이)**해야 비로소 확률이 됩니다.
- 직관적 비유: 금속 막대의 어느 한 점을 바늘로 찔렀을 때의 무게는 0입니다. 하지만 일정 길이(구간)를 잘라내면 무게(확률)가 생깁니다.
- 공식:
$$P(a \le X \le b) = \int_a^b f_X(x)\,dx$$

[주의] 잠깐! 오해하기 쉬워요: 점의 확률은 0이다 ($P(X = x) = 0$)
많은 분들이 가장 혼란스러워하는 부분입니다.

- 친구가 약속 장소에 "정확히 2.00000...초"에 도착할 확률은? 0입니다.
- 수학적으로 **선(Line)의 넓이는 0**이기 때문입니다.
- 따라서 연속 확률 변수에서는 등호의 유무가 중요하지 않습니다.
$$P(X \ge a) = P(X > a)$$

4.3 핵심 개념 2: 주요 연속 분포 (Story Matching)

[시나리오] 스토리 시나리오: 1. 균등 분포 (Uniform): "정보 부재의 공평함"

- 스토리: 버스가 0분에서 10분 사이에 온다는 것만 알고, 다른 정보는 전혀 없을 때.
- 특징: 그래프가 평평한 박스 모양입니다.
- 확률: 전체 구간 길이 분의 내가 원하는 구간 길이 (기하학적 확률).

[시나리오] 스토리 시나리오: 2. 지수 분포 (Exponential): "무기억성(Memoryless)"

- 스토리: 콜센터 전화가 올 때까지 걸리는 대기 시간. (이산 기하 분포의 연속 버전)
- 핵심 성질 (무기억성): "지난 30분 동안 전화가 안 왔다고 해서, 지금 당장 올 확률이 높아지는 건 아니다."
$$P(T > t + s | T > t) = P(T > s)$$
- 과거를 잊어버리고 매 순간 새롭게 시작하는(Fresh Start) 성질입니다.

[시나리오] 스토리 시나리오: 3. 정규 분포 (Normal): "세상의 모든 잡음(Noise)"

- 스토리: 키, 몸무게, 시험 점수, 측정 오차 등 수많은 작은 원인들이 합쳐진 결과.
- 중요성: 중심극한정리(CLT)에 의해, 데이터가 충분히 많으면 세상 만사는 정규 분포(종 모양)를 따르게 됩니다.

4.4 필수 스킬: 정규 분포의 표준화 (Standardization)

정규 분포 곡선 e^{-x^2}은 손으로 적분하는 것이 불가능합니다. 그래서 우리는 모든 정규 분포를 **표준 정규 분포(Z)**로 변환하여 미리 계산된 표(Standard Normal Table)를 사용합니다.

〔단계〕 모든 정규 분포를 $N(0,1)$로 만드는 마법

변수 X가 평균 μ, 표준편차 σ를 가질 때 ($X \sim N(\mu, \sigma^2)$):

1단계: Z-score 변환 (공식 암기 필수)

$$Z = \frac{X - \mu}{\sigma}$$

(의미: 내 점수가 평균에서 표준편차의 몇 배만큼 떨어져 있는가?)

2단계: 표 찾기 변환된 Z값을 이용해 표(CDF)에서 확률 $\Phi(Z)$를 찾습니다.

〔예제〕 예시: 수능 점수 계산

평균이 50점($\mu = 50$), 표준편차가 10점($\sigma = 10$)인 시험에서 70점 이상 받을 확률은?

1. 표준화: $z = \frac{70-50}{10} = 2$. (나는 평균보다 2σ만큼 잘했다.)
2. 확률 확인: 표준 정규 분포 표에서 $P(Z \leq 2) \approx 0.977$.
3. 결론: 상위 약 2.3% $(1 - 0.977)$ 안에 듭니다.

4.5 실전 응용: 디지털 통신과 베이즈 정리

〔상황〕 AI 모델이 이미지를 보고 '고양이(0)'인지 '개(1)'인지 분류하려 합니다.

- **원인(Discrete):** 실제 정답 $K \in \{0, 1\}$.
- **결과(Continuous):** 모델이 뱉어낸 확신 점수(Score) X (0.0 1.0 사이의 실수).

우리는 점수 $X = 0.7$을 관측했을 때, 이것이 실제로 개($K = 1$)일 확률을 알고 싶습니다.

$$P(K = 1|X = 0.7) = \frac{P(K = 1) \cdot f_{X|K}(0.7|1)}{f_X(0.7)}$$

- **이산과 연속의 만남:** 분자는 '이산 확률 $\times$ 연속 밀도'입니다.
- **분모(전체 확률):** 전체 밀도 $f_X(0.7)$은 고양이일 때의 밀도와 개일 때의 밀도를 가중합(Total Probability)하여 구합니다.
- 이 공식이 바로 **모든 머신러닝 분류기(Classifier)**가 작동하는 수학적 원리입니다.

4.6 초심자가 자주 묻는 질문 (FAQ)

Q1. PDF 값(높이)이 1보다 클 수 있나요?

네, 가능합니다! 확률은 1을 넘을 수 없지만, 밀도(Density)는 넘을 수 있습니다.

- 예: 구간 〔0, 0.1〕에 확률 1이 꽉 차 있다면, 높이(밀도)는 10이어야 면적이 1이 됩니다. $(10 \times 0.1 = 1)$

Q2. 왜 적분을 해야 확률이 나오나요?
속도(v)를 적분하면 이동 거리(s)가 나오죠? 마찬가지로 확률의 '변화율(밀도)'을 쌓아야 전체 '양(확률)'이 나오기 때문입니다.

Q3. 정규 분포 식을 외워야 하나요?
복잡한 식($\frac{1}{\sqrt{2\pi}\sigma}e^{-\cdots}$) 자체를 외울 필요는 없습니다. 중요한 것은 그 식을 **Z로 변환하는 방법**과 그래프의 **대칭성(Symmetry)**을 활용하는 능력입니다.

4.7 요약 및 다음 단계

Summary

Unit 2-1 핵심 요약〕

1. 적분 마인드: 점(x)은 0이다. 구간(dx)을 적분해야 확률(면적)이다.
2. 분포 매칭: 랜덤 $\rightarrow$ Uniform / 대기시간 $\rightarrow$ Exponential / 오차합 $\rightarrow$ Normal.
3. 표준화: 정규 분포 문제는 무조건 $Z = \frac{X-\mu}{\sigma}$로 바꿔서 푼다.
4. 혼합 추론: 연속된 데이터(신호)로 이산적 원인(디지털 정보)을 추측한다.

〔다음 단원 예고〕
하나의 변수(키)만으로는 세상을 설명하기 부족합니다. 키(X)와 몸무게(Y)는 서로 관계가 있지 않을까요?
다음 장 Chapter 5: 다변수 연속 확률 변수에서는 2차원 평면 위에서의 적분(이중 적분)과 상관관계(Correlation)를 다룹니다. "적분 기호가 두 개($\int\int$)로 늘어납니다!"

Chapter 5

결합 분포: 평면 위의 미적분

Intro: 1차원 선에서 2차원 평면으로

〔이전 단계와의 연결〕
Chapter 4에서는 변수가 하나(X)인 상황, 즉 수직선 위의 확률 밀도를 적분했습니다. 하지만 현실 세계의 문제는 변수 하나로 설명되지 않습니다. 키(X)와 몸무게(Y), 위도(X)와 경도(Y)처럼 두 변수는 서로 얽혀 있습니다.

이제 우리는 **2차원 평면(x, y)** 위에서 확률을 다룹니다. 선적분($\int$)이 **이중 적분($\iint$)**으로, 그래프의 면적이 **입체의 부피(Volume)**로 확장됩니다.

Summary

이 단원의 핵심 개요 (Overview)〕

1. Joint PDF (지형도): 확률을 3차원 입체의 '부피'로 이해합니다.
2. Marginal PDF (그림자): 입체를 벽면에 투영(Projection)하여 한 변수의 분포만 뽑아냅니다.
3. 독립성 (함정): 수식뿐만 아니라 **'영역의 모양'**이 독립성을 결정한다는 사실을 배웁니다.
4. Bayesian Inference: 미지의 상수(Parameter)를 확률 변수로 취급하여 추정하는 현대 통계의 핵심을 맛봅니다.

5.1 필수 용어 사전

mainblue!20 기호	용어	직관적 의미
$f_{X,Y}(x, y)$	결합 PDF (Joint PDF)	(x, y) 지점에서의 확률 밀도 높이 (지형도)
$f_X(x)$	주변 PDF (Marginal PDF)	Y를 무시하고(적분해서 없애고) X만 본 분포 (그림자)
$\iint_A$	이중 적분	평면의 특정 영역 A 위에 쌓인 확률의 부피 구하기
Support	지지 집합 (정의역)	확률 밀도가 0이 아닌 (x, y)의 영역 (지도상의 땅)
Posterior	사후 확률 ($f_{\Theta\|X}$)	데이터를 본 후 수정된 파라미터의 분포

5.2 핵심 개념 1: 입체로 생각하고 축으로 투영하라

5.2.1 결합 PDF ($f_{X,Y}$): 확률은 부피다

〔개념〕 지형도(Surface)의 부피 구하기

- 한 줄 요약: 평면의 땅(x, y) 위에 확률이라는 흙을 쌓아 올린 산입니다.
- 직관적 비유: 지도 위의 등고선. 특정 지역(A)에 비가 올 확률은 그 지역 위에 떠 있는 구름의 **부피**와 같습니다.
- 공식:

$$P((X,Y) \in A) = \iint_A f_{X,Y}(x,y)\,dx\,dy$$

5.2.2 주변 PDF (f_X): 차원 뭉개기 (Marginalization)

복잡하게 얽힌 두 변수 중 하나만 알고 싶을 때 사용합니다.

- 개념: "나는 Y값은 상관 안 해. X가 x일 확률 밀도만 다 모아줘."
- 직관적 비유 (Projection): 3차원 물체에 Y축 방향에서 빛을 비췄을 때, **X축 벽면에 생기는 그림자**입니다. 입체적인 정보가 납작해집니다.
- 계산: Y가 가질 수 있는 모든 값$(-\infty \sim \infty)$을 적분하여 없앱니다.

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$$

5.3 핵심 개념 2: 독립성 (Independence)과 함정

두 변수 X, Y가 독립이라는 것은 $f_{X,Y}(x,y) = f_X(x)f_Y(y)$로 인수분해된다는 뜻입니다. 하지만 **수식만 보면 100% 틀립니다.**

〔주의〕 잠깐! 오해하기 쉬워요: 기하학적 함정 (The Geometry Trap)
수식이 곱하기로 쪼개지더라도, **정의역(Support)의 모양**을 반드시 확인해야 합니다.

- 직사각형 (Rectangle): $a \le x \le b, c \le y \le d$
  - $\to$ 독립일 가능성 있음 (수식 확인 필요).
- 삼각형, 원 등 (Non-Product): $0 \le y \le x \le 1$ 등
  - $\to$ 무조건 종속 (Dependent)!
  - 이유: X값이 변하면 Y가 놀 수 있는 범위(Range)가 변하기 때문입니다. 정보가 섞여 있습니다.

〔예제〕 예시: 삼각형 위에서의 균등 분포
영역 $A = \{(x,y)|0 \le y \le x \le 1\}$ (삼각형) 위에서 균등하게 분포하는 (X,Y)가 있습니다. 면적이 $1/2$이므로 높이(PDF)는 2입니다. $(f_{X,Y}(x,y) = 2)$

1. 주변 PDF 구하기 ($f_X(x)$): x를 고정하고 y에 대해 적분합니다. 이때 y는 0부터 x까지만 존재합니다!

$$f_X(x) = \int_0^x 2\,dy = [2y]_0^x = 2x \quad (0 \le x \le 1)$$

2. 독립성 판별: 정의역이 삼각형이므로 계산해볼 것도 없이 **종속**입니다. (실제로 $f_X(x)f_Y(y)$를 곱해보면 $2 \ne 2x \cdot 2(1-y)$가 되어 성립하지 않습니다.)

5.4 핵심 개념 3: 모수(Parameter)도 확률 변수다

통계학의 패러다임 전환이 일어나는 구간입니다. 동전의 앞면 확률 Θ를 고정된 상수(0.5)가 아니라, **분포를 가진 확률 변수**로 바라봅니다.

〔시나리오〕 스토리 시나리오: 로봇의 위치 추정 (Localization)
로봇이 복도(X)의 어딘가에 있습니다.

1. Prior (사전 확률): 로봇은 처음에 자신이 복도 중앙 쯤($\Theta \approx 5m$)에 있다고 믿습니다. (정규분포)
2. Observation (데이터): 센서로 거리를 쟀더니 $5.2m$가 나왔습니다($X = 5.2$). 하지만 센서에는 오차(Noise)가 있습니다.
3. Posterior (사후 확률): 센서 데이터($5.2m$)와 내 믿음($5m$)을 적절히 섞어서(Update), "아, 나는 실제로는 $5.1m$ 쯤에 있겠구나"라고 믿음을 수정합니다.

〔단계〕 베이지안 추론의 3단계
목표: 관측 데이터 x를 보고 미지의 파라미터 Θ의 분포를 찾아라.

$$f_{\Theta|X}(\theta|x) = \frac{f_\Theta(\theta) \cdot f_{X|\Theta}(x|\theta)}{f_X(x)}$$

1. 분자 계산: Prior(θ 믿음) $\times$ Likelihood(데이터 설명력).
2. 분모 계산 (Normalization): 분자를 θ에 대해 전체 적분하여 상수를 구함 (면적을 1로 맞춤).
3. 추정 (Estimation):
   - MAP: PDF가 가장 높은 봉우리(Mode)를 찍음. ("가장 그럴듯한 값")
   - LMS: PDF의 무게중심(Expectation)을 찍음. ("평균적인 값")

5.5 초심자가 자주 묻는 질문 (FAQ)

Q1. 이중 적분 순서($dxdy$ vs $dydx$)는 어떻게 정하나요?
영역의 모양을 보고 **"화살표를 쏘기 편한 쪽"**으로 정합니다.

- y축 방향(위아래)으로 화살표를 쐈을 때 진입/진출 곡선이 하나라면 dy 먼저.
- x축 방향(좌우)이 편하면 dx 먼저.
- **Tip:** 반드시 그림을 그려야 보입니다!

Q2. 주변 PDF를 구할 때 범위가 헷갈려요.
이게 가장 어렵습니다. $f_X(x)$를 구할 땐, **"x를 상수로 고정했다"**고 생각하세요. 그 고정된 x 선 위에서 y가 어디서부터 어디까지 움직이는지(구간)를 찾아야 합니다. (삼각형 예제에서 y는 0부터 1이 아니라 x까지였습니다!)

5.6 요약 및 다음 단계

Summary

Unit 2-2 핵심 요약]

1. Joint PDF: 확률은 입체의 부피다 ($\int\int$).
2. Marginal PDF: 그림자 놀이다 (한 변수 적분해서 없애기).
3. 독립성 Trap: 직사각형 영역이 아니면 무조건 종속이다.
4. Bayes: 모수(Θ)를 확률 변수로 보고, 데이터를 통해 분포를 업데이트한다.

[다음 단원 예고]
지금까지는 X, Y를 그대로 사용했습니다. 만약 $Z = X + Y$나 $W = X/Y$처럼 변수들을 지지고 볶아서 **새로운 변수**를 만들면, 그 변수의 PDF는 어떻게 될까요?
다음 장 Chapter 6: 파생 변수의 분포(Derived Distributions)에서 미분의 반격, 야코비안(Jacobian)이 등장합니다.

Chapter 6

변환과 관계: 확률 변수를 요리하다

Intro: 관찰을 넘어 조작으로

〔이전 단계와의 연결〕
Chapter 5에서는 주어진 변수 X, Y를 있는 그대로 관찰했습니다. 하지만 공학 현실은 더 복잡합니다. 전압(X)을 제곱해서 전력($Y = X^2$)을 구하거나, 신호(X)와 잡음(Y)이 섞인 수신 값($Z = X + Y$)을 분석해야 합니다.

이제 우리는 확률 변수를 **함수에 넣어 변형**하거나, **서로 더했을 때** 분포가 어떻게 찌그러지고(Distort) 이동하는지 추적하는 기술을 배웁니다.

Summary

이 단원의 핵심 개요 (Overview)〕

1. 파생 분포 (Derived): 입력(X)이 함수(g)를 통과할 때 밀도가 어떻게 변하는가? (야코비안의 마법)
2. 합의 분포 (Sum): 두 변수를 더했을 때의 분포를 구하는 '컨볼루션(Convolution)' 적분.
3. 상관관계 (Correlation): 두 변수가 얼마나 같이 움직이는지 숫자로 요약하기.
4. LIE (반복 기댓값): 어려운 기댓값 문제를 쪼개서 푸는 최강의 무기.

6.1 필수 용어 사전

mainblue!20 기호	용어	직관적 의미
$Y = g(X)$	파생 변수	X를 재료로 요리해서 만든 새로운 변수 Y
CDF Method	2단계 접근법	누적 확률(CDF)을 먼저 구하고 미분하는 안전한 방법
Convolution	합성곱 ($f_X * f_Y$)	두 함수를 뒤집고 밀어서 겹치는 면적을 구하는 것 ($X + Y$)
ρ	상관계수	두 변수의 선형 관계 강도 (-1 $\sim$ 1). (단위 없음)
$E[X\|Y]$	조건부 기댓값	Y에 따라 값이 변하는 **확률 변수** (숫자 아님!)

6.2 핵심 개념 1: 분포 구하기의 정석 (Transformation)

입력 X가 함수 $g(X)$를 통과하면 Y의 분포는 어떻게 될까요? 단순히 X의 확률을 대입하면 안 됩니다. **공간이 왜곡(Stretch Squeeze)**되기 때문입니다.

6.2.1 2단계 접근법 (The 2-Step Method)

MIT에서 권장하는 가장 안전하고 강력한 풀이법입니다.

〔단계〕 CDF 먼저 구하고 미분하라!

1단계: CDF $F_Y(y)$ 구하기 (Accumulate)
Y의 사건을 X의 사건으로 번역합니다.

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \leq g^{-1}(y))$$

2단계: PDF $f_Y(y)$ 구하기 (Differentiate)
위에서 구한 식을 y로 미분합니다.

$$f_Y(y) = \frac{d}{dy} F_Y(y)$$

6.2.2 직관: 야코비안 (The Jacobian Intuition)

만약 $y = g(x)$가 1:1 대응이라면, 공식을 바로 쓸 수 있습니다.

$$f_Y(y) = f_X(x) \cdot \frac{1}{|g'(x)|}$$

〔개념〕 밀가루 반죽 비유

- 밀가루 반죽(X)을 넓게 펴면(g'이 큼), 두께(밀도)는 얇아집니다. $\rightarrow$ 나누기 g'
- 반죽을 좁게 뭉치면(g'이 작음), 두께(밀도)는 두꺼워집니다.
- **야코비안($1/|g'|$):** 공간이 늘어난 만큼 밀도를 줄여주는 **보정 계수**입니다.

6.3 핵심 개념 2: 두 변수를 더하면? (Sum of Independent RVs)

$Z = X + Y$일 때, Z의 PDF는 단순한 합이 아닙니다. **컨볼루션(Convolution)** 적분입니다.

6.3.1 분할 정복 논리

- X가 x로 고정되었다고 칩시다.
- $X + Y = z$가 되려면, Y는 반드시 $z - x$가 되어야 합니다.
- 확률: $f_X(x) \cdot f_Y(z - x)$
- X는 무엇이든 될 수 있으니 모든 x에 대해 더합니다($\int$).

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x)\, dx$$

〔개념〕 Flip and Drag (뒤집고 밀기)
이 적분 식의 기하학적 의미입니다.

1. Flip: f_Y 그래프를 y축 대칭시킵니다 ($x \rightarrow -x$).
2. Drag: 그래프를 z만큼 옆으로 밉니다.
3. Area: f_X와 겹치는 부분의 넓이를 구합니다.

6.4 핵심 개념 3: 관계의 척도 (Correlation)

[Image of scatter plots with different correlation coefficients]

6.4.1 공분산(Covariance) vs 상관계수(Correlation)

- **공분산:** 방향(+, -)은 알 수 있지만, 크기가 단위에 의존합니다. (키를 m로 재냐 cm로 재냐에 따라 값이 10000배 차이남)
- **상관계수 (ρ):** 표준편차로 나누어 정규화(-1 ∼ 1)한 값. 단위가 사라집니다.
  - $\rho = 1$: 완벽한 직선 ($Y = X$)
  - $\rho = 0$: 선형 관계 없음 (구름 모양)

[주의] 잠깐! 오해하기 쉬워요: Uncorrelated $\neq$ Independent
가장 많이 틀리는 함정입니다.

- **Uncorrelated ($\rho = 0$):** "직선 관계"가 없다는 뜻입니다.
- **Independent:** "아무런 관계"가 없다는 뜻입니다.
- 반례: X가 $-1 \sim 1$ 사이의 값이고, $Y = X^2$이라고 합시다.
  - X를 알면 Y를 완벽히 알 수 있으므로 **종속(Dependent)**입니다.
  - 하지만 계산해보면 $\rho = 0$이 나옵니다. (2차 곡선 관계이므로 선형성은 0)
- 예외: 정규 분포(Gaussian)인 경우에만 $\rho = 0$이면 독립입니다.

6.5 핵심 개념 4: 통계학의 맥가이버 칼 (LIE)

6.5.1 조건부 기댓값 ($E[X|Y]$)

- $E[X]$는 숫자입니다.
- $E[X|Y]$는 **Y에 대한 함수(확률 변수)**입니다.
- 예: $X =$키, $Y =$성별. $E[X|Y]$는 '남자의 평균 키' 또는 '여자의 평균 키'라는 값을 갖는 변수입니다.

6.5.2 반복 기댓값의 법칙 (Law of Iterated Expectations)

$$E[E[X|Y]] = E[X]$$

"부분의 평균들을 다시 평균 내면 전체 평균이 된다."

[단계] LIE 사용법: Divide and Conquer
복잡한 $E[X]$를 구할 때:

1. 상황을 가르는 변수 Y를 찾습니다. (Divide)
2. 각 상황별 평균 $E[X|Y = y]$를 구합니다. (Local Average)
3. 그 값들의 기댓값을 다시 구합니다. (Global Average)

6.6 실전 응용: 금융 포트폴리오의 위험 관리

〔시나리오〕 스토리 시나리오: 주식 분산 투자의 원리
당신이 주식 A와 B에 절반씩 투자해서 포트폴리오 $Z = 0.5X + 0.5Y$를 만들었습니다.

- 수익률($E[Z]$)은 단순 평균입니다.
- 하지만 **위험(분산, var(Z))**은 다릅니다.

$$\text{var}(Z) = 0.25\text{var}(X) + 0.25\text{var}(Y) + 2(0.5)(0.5)\text{cov}(X, Y)$$

- 만약 두 주식이 **역의 상관관계($\rho < 0$)**라면?
- 공분산 항이 마이너스가 되어 **전체 위험(분산)이 줄어듭니다!**
- 이것이 "계란을 한 바구니에 담지 말라"는 격언의 수학적 증명입니다.

6.7 요약 및 다음 단계

Summary

Unit 2-3 핵심 요약〕

1. 변환: CDF를 먼저 구하고 미분하는 것이 가장 안전하다.
2. 합: 독립 변수의 합은 컨볼루션($\int f_X f_Y$)이다.
3. 상관관계: $\rho = 0$이어도 종속일 수 있다. ($Y = X^2$ 기억하기)
4. LIE: 어려운 평균 문제는 $E[E[X|Y]]$로 쪼개서 푼다.

〔다음 단원 예고〕
지금까지는 한 번의 실험이나 고정된 시점의 확률을 다뤘습니다. 하지만 주식 가격이나 대기 행렬처럼 **시간에 따라 변하는 확률**은 어떻게 다룰까요?
이제 Unit 3 **"확률 과정(Random Processes)"**으로 넘어가 베르누이 프로세스와 포아송 프로세스를 배웁니다. "시간(t)이 변수로 들어옵니다!"

Part III

Unit 3: 확률 과정과 극한 정리

Chapter 7

극한 정리: 무한(∞)이 주는 선물

Intro: 정확한 계산에서 '근사(Approximation)'로

〔이전 단계와의 연결〕
지금까지는 확률 변수의 분포(PMF/PDF)를 정확히 알고 있다고 가정하고, $P(X \leq 3)$ 같은 값을 '정확하게' 계산했습니다. 하지만 현실 세계(빅데이터)에서는 데이터가 너무 많거나 분포를 모르는 경우가 대부분입니다. 이제 우리는 **"데이터가 무한히 많아지면($n \to \infty$), 분포를 몰라도 결과를 예측할 수 있다"**는 강력한 도구를 배웁니다.

Summary

이 단원의 핵심 개요 (Overview)〕

1. 부등식 (Bounds): 분포를 몰라도 평균과 분산만 알면 '최악의 시나리오'를 막을 수 있다. (마르코프, 체비셰프)

2. 대수의 법칙 (LLN): 데이터가 많아지면 통계치(평균)는 결국 진실(True Mean)로 수렴한다.

3. 중심 극한 정리 (CLT): 데이터가 많아지면 합과 평균의 분포는 무조건 **정규 분포(Bell Curve)**가 된다.

7.1 필수 용어 사전

mainblue!20 기호	용어	직관적 의미
Bounds	한계(부등식)	정확한 값은 모르지만, "적어도 이 선은 넘지 않는다"는 상한선
M_n	표본 평균	데이터 n개를 모아서 낸 평균 ($\frac{X_1+\cdots+X_n}{n}$)
Converge	수렴 ($\to$)	n이 커질수록 특정 값이나 분포에 가까워지는 현상
LLN	대수의 법칙	"많이 던지면 결국 확률대로 나온다." (진실의 발견)
CLT	중심 극한 정리	"많이 합치면 종 모양(정규 분포)이 된다." (형태의 발견)

7.2 핵심 개념 1: 분포를 모를 때의 생존법 (부등식)

분포의 모양(PDF)을 몰라도, **평균(μ)**과 **분산(σ^2)**만 안다면 확률의 한계(Limit)를 설정할 수 있습니다. 이것은 리스크 관리(Risk Management)의 기초입니다.

7.2.1 1. 마르코프 부등식 (Markov Inequality)

〔개념〕 평균의 시소 원리

- 조건: 데이터가 음수가 아닐 때 ($X \geq 0$).
- 직관: 평균(무게중심)이 정해져 있다면, 아주 큰 값이 나올 확률은 작아야만 합니다. 그래야 시소가 균형을 잡으니까요.
- 공식:

$$P(X \geq a) \leq \frac{E[X]}{a}$$

〔예제〕 예시: 연봉 분포 미스터리

어떤 회사의 평균 연봉이 5,000만 원이라는 것만 알고 있습니다. 이 회사에서 5억 원(평균의 10배) 이상 받는 사람은 최대 몇 명일까요?

- 정확한 분포는 모릅니다. 하지만 마르코프 부등식에 의해:

$$P(X \geq 5\text{억}) \leq \frac{5\text{천만}}{5\text{억}} = \frac{1}{10} = 10\%$$

- 결론: 고액 연봉자는 아무리 많아봤자 전 직원의 10%를 넘을 수 없습니다. (절대적 상한선)

7.2.2 2. 체비셰프 부등식 (Chebyshev Inequality)

〔개념〕 분산이 알려주는 거리의 제약

- 조건: 모든 분포에 적용 가능. (분산 σ^2을 알 때)
- 직관: 분산(퍼짐)이 작다면, 데이터는 평균 근처에 모여 있어야 합니다. 평균에서 멀리 떨어진 꼬리(Tail) 확률은 급격히 줄어듭니다.
- 공식:

$$P(|X - \mu| \geq c) \leq \frac{\text{var}(X)}{c^2}$$

(거리 c가 멀어질수록 확률은 제곱(c^2)으로 줄어듭니다.)

7.3 핵심 개념 2: 대수의 법칙 (Laws of Large Numbers)

"데이터가 깡패다." 데이터(n)가 많아지면 표본 평균(M_n)은 흔들림을 멈추고 진실(μ)로 수렴합니다.

7.3.1 분산의 축소 (Variance Shrinking)

왜 수렴할까요? 수학적 엔진은 바로 **분산이 줄어들기 때문**입니다.

$$M_n = \frac{X_1 + \cdots + X_n}{n} \implies \text{var}(M_n) = \frac{\sigma^2}{n}$$

- $n \to \infty$이면 분산 $\to 0$이 됩니다.
- 분산이 0이라는 뜻은, 분포가 평균 μ 한 점에 뾰족하게 모인다는 뜻입니다. (불확실성 소멸)

[주의] 잠깐! 오해하기 쉬워요: 약법칙(Weak) vs 강법칙(Strong)

- 약법칙 (WLLN): "오차가 발생할 확률이 0이 된다." (가끔 튀는 놈이 나올 수는 있지만, 그 가능성이 희박해짐)
- 강법칙 (SLLN): "결과값 자체가 100% 확률로 진실에 꽂힌다." (시뮬레이션 그래프가 진동하다가 결국 일직선이 됨)
- 실무적 결론: 엔지니어링에서는 둘 다 "데이터 많으면 평균 믿어도 된다"로 해석해도 무방합니다.

7.4 핵심 개념 3: 중심 극한 정리 (Central Limit Theorem)

이 챕터의 하이라이트입니다. LLN이 "평균으로 간다"는 **위치**를 알려줬다면, CLT는 "**어떤 모양**으로 가는가?"를 알려줍니다.

[개념] 모든 길은 정규 분포로 통한다

원래 데이터가 이항 분포든, 균등 분포든, 지수 분포든 상관없습니다. "독립적인 확률 변수들을 많이 더하면(S_n) 그 합의 분포는 정규 분포(N)에 가까워진다."

7.4.1 CLT 문제 해결 3단계 레시피

복잡한 합의 확률을 구해야 할 때, 다음 절차를 따르세요.

[단계] 표준화(Standardization) 후 근사

상황: $S_n = X_1 + \cdots + X_n$ (합계)의 확률을 구하고 싶다.

1단계: 평균과 분산 계산

- 합의 평균: $E[S_n] = n\mu$
- 합의 분산: $\text{var}(S_n) = n\sigma^2$ (표준편차는 $\sigma\sqrt{n}$)

2단계: 표준화 (Z-score 변환) 정규 분포표를 쓰기 위해 Z로 바꿉니다.

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

3단계: 근사 (Approximation) 이제 Z_n을 표준 정규 분포 $N(0,1)$로 취급하여 표에서 확률 $\Phi(z)$를 찾습니다.

7.5 실전 응용: 게임 서버 용량 설계 (Capacity Planning)

〔시나리오〕 스토리 시나리오: 넥슨 서버 개발자의 고민
신규 게임 출시를 앞두고 있습니다.

- 유저 1명의 접속 트래픽(X_i)은 평균 10MB(μ), 표준편차 5MB(σ)입니다. (분포는 모름, 아마도 지수 분포?)
- 동시 접속자가 10,000명(n)일 때, 총 트래픽(S_n)이 서버 용량인 100,500MB를 초과하여 서버가 터질 확률은?

해결 과정 (CLT 적용):

1. 기초값: $n = 10000$, $\mu = 10$, $\sigma = 5$.
2. 합의 통계량:
   - 평균 $E[S_n] = 10000 \times 10 = 100,000$ MB
   - 분산 var$(S_n) = 10000 \times 25 \implies$ 표준편차 $\sqrt{250000} = 500$ MB
3. 표준화: 우리가 궁금한 임계값 100,500을 Z로 변환.
$$Z = \frac{100,500 - 100,000}{500} = \frac{500}{500} = 1$$
4. 결론: $P(Z > 1)$을 구하면 됩니다. 정규 분포표에서 $P(Z \leq 1) \approx 0.84$.
$$P(S_n > 100500) \approx 1 - 0.84 = 0.16$$
해석: 서버가 터질 확률이 16%나 됩니다! (위험함). 용량을 늘려야 합니다.

7.6 초심자가 자주 묻는 질문 (FAQ)

Q1. n이 얼마나 커야 정규 분포라고 볼 수 있나요?
보통 통계학 교과서에서는 **$n \geq 30$**이면 충분하다고 봅니다. 하지만 분포가 심하게 치우친(Skewed) 경우라면 데이터가 더 많이 필요할 수 있습니다.

Q2. 왜 $\sqrt{n}$으로 나누나요?
이게 핵심입니다! 합(S_n)은 n에 비례해서 커지지만, 변동성(표준편차)은 $\sqrt{n}$에 비례해서 커집니다. 즉, 신호(Signal, n)가 소음(Noise, $\sqrt{n}$)보다 더 빨리 커지기 때문에, 데이터가 많을수록 비율적으로 안정되는 것입니다.

Q3. 마르코프와 체비셰프 중 뭘 써야 하나요?
정보가 많을수록 더 강력한(Tight) 부등식을 쓸 수 있습니다.

- 평균만 안다 $\rightarrow$ 마르코프 (범위가 넓음)
- 분산까지 안다 $\rightarrow$ 체비셰프 (범위가 좁혀짐)
- 분포 모양(종 모양)까지 안다 $\rightarrow$ 정규 분포 계산 (정확함)

7.7 요약 및 다음 단계

Summary

Unit 3-1 핵심 요약〕

1. 부등식: 분포를 몰라도 평균(μ)과 분산(σ^2)만 있으면 리스크의 상한선을 그을 수 있다.
2. LLN: 데이터가 많으면 표본 평균 $\approx$ 진짜 평균이다. (믿어라!)
3. CLT: 데이터 합계의 분포는 결국 정규 분포(Z)로 수렴한다.
4. 공학적 의의: 복잡한 세상의 현상을 정규 분포라는 단순한 모델로 **근사(Approximation)**하여 해석할 수 있게 해준다.

〔다음 단원 예고〕
지금까지는 $n \to \infty$로 갈 때의 '분포'를 봤습니다. 이제 시간 t가 흐르면서 사건이 발생하는 **'과정(Process)'**을 다룹니다.
다음 장 Chapter 8: 베르누이와 포아송 프로세스에서는 "동전을 영원히 던지는 상황"과 "콜센터 전화가 계속 오는 상황"을 모델링합니다.

Chapter 8

시간 위의 확률: 사건의 흐름

Intro: 사진(Snapshot)에서 영상(Video)으로

〔이전 단계와의 연결〕
지금까지 우리는 주사위를 한 번 던지거나, 시험을 한 번 보는 '정적(Static)'인 상황을 다뤘습니다. 하지만 현실은 시간이 흐릅니다. 1초에 한 번씩 클릭이 발생하고, 불규칙하게 서버 요청이 들어옵니다.

이제 우리는 **시간 축(Time Axis)** 위에 확률 변수를 늘어놓습니다. 이것이 바로 **확률 과정(Random Process)**입니다.

Summary

이 단원의 핵심 개요 (Overview)〕

1. 베르누이 과정 (Discrete): 시계가 똑딱거릴 때마다 동전을 던지는 '디지털' 세상.
2. 포아송 과정 (Continuous): 시간 간격을 0으로 보내서 만든 '아날로그' 세상.
3. Fresh Start (무기억성): "과거는 잊어라." 확률 과정 해석의 가장 강력한 도구.
4. 시스템 결합: 두 개의 확률 흐름을 합치거나(Merge) 쪼개는(Split) 방법.

8.1 필수 용어 사전

mainblue!20 기호	용어	직관적 의미
Slot	시간 슬롯	베르누이 과정의 최소 시간 단위 (1초, 1프레임 등)
λ (Lambda)	도착률 (Arrival Rate)	단위 시간당 평균 발생하는 사건의 수 (예: 5회/초)
Inter-arrival	대기 시간	사건과 사건 사이의 간격 (Wait Time)
Memoryless	무기억성	"지금까지 안 나왔다고 해서 곧 나올 확률이 높지는 않다."
Splitting	분할	하나의 흐름을 확률 p로 두 갈래로 나누는 것

8.2 핵심 개념 1: 디지털 시계의 심장박동 (Bernoulli)

〔개념〕 정의: 매 순간의 동전 던지기
시간 $t = 1, 2, 3, \ldots$ 마다 독립적으로 성공(p) 또는 실패($1 - p$)가 결정됩니다.

- 핵심 성질 (Fresh Start): 매 슬롯은 독립입니다. 과거에 실패가 100번 연속 나왔어도, 이번 슬롯의 성공 확률은 여전히 p입니다. (도박사의 오류 방지)

8.2.1 두 가지 관점: 세느냐(Count), 기다리느냐(Time)?

베르누이 과정을 바라보는 시각은 딱 두 가지입니다.

- View 1: 카운팅 (Counting)
  - 질문: "n번 시도했는데 성공이 몇 번 나왔어?"
  - 분포: **이항 분포 (Binomial)** (n, p)
- View 2: 타이밍 (Timing)
  - 질문: "첫 성공이 나올 때까지 몇 번(T) 기다려야 해?"
  - 분포: **기하 분포 (Geometric)** (p)

$$P(T = k) = (1 - p)^{k-1}p$$

8.3 핵심 개념 2: 빗방울처럼 떨어지는 사건들 (Poisson)

베르누이 과정의 시간 슬롯을 $\Delta \to 0$으로 아주 잘게 쪼개면 **포아송 과정**이 됩니다.

- p는 아주 작아지고, n은 아주 커지지만, 곱(np)은 λt로 일정하게 유지됩니다.
- **예시:** 콜센터 전화, 웹사이트 접속, 방사능 붕괴.

8.3.1 완벽한 대응 (The Rosetta Stone)

이 표를 머릿속에 넣는 것이 이번 단원의 목표입니다. 이산과 연속은 쌍둥이입니다.

subgray 질문 (Perspective)	베르누이 (Discrete)	포아송 (Continuous)
매개변수	확률 p (성공/슬롯)	빈도 λ (회/시간)
Counting (몇 번?) (일정 시간 내 발생 횟수)	이항 분포 (Binomial) $P(K = k) = \binom{n}{k}p^k(1-p)^{n-k}$	포아송 분포 (Poisson PMF) $P(k; \tau) = \frac{(\lambda\tau)^k e^{-\lambda\tau}}{k!}$
Timing (얼마나?) (다음 사건까지 대기)	기하 분포 (Geometric) $P(T = k) = (1-p)^{k-1}p$	지수 분포 (Exponential PDF) $f_T(t) = \lambda e^{-\lambda t}$

[주의] 잠깐! 오해하기 쉬워요: 기하 분포와 지수 분포의 관계

기하 분포의 막대 그래프 폭을 아주 좁게 만들면 지수 분포의 매끄러운 곡선이 됩니다. 둘 다 **"무기억성(Memoryless)"**을 가지는 유일한 분포들입니다.

8.4 핵심 개념 3: 흐름을 합치고 나누기 (System Ops)

여러 개의 확률 흐름(Stream)이 만날 때 어떤 일이 벌어질까요?

1. 병합 (Merging)

- 상황: PC 접속자 흐름(λ_1)과 모바일 접속자 흐름(λ_2)이 서버로 들어옵니다.
- 결과: 합쳐진 흐름도 포아송 과정이며, 속도는 단순 합입니다.

$$\lambda_{total} = \lambda_1 + \lambda_2$$

- 확률: 방금 들어온 요청이 PC일 확률은? $\frac{\lambda_1}{\lambda_1+\lambda_2}$ (속도에 비례)

2. 분할 (Splitting)

- 상황: 전체 트래픽(λ) 중 스팸 메일을 확률 p로 걸러냅니다.
- 결과:
  - 스팸 흐름: 속도 λp인 포아송 과정
  - 정상 흐름: 속도 $\lambda(1-p)$인 포아송 과정
- 직관 파괴 (Independence): 놀랍게도 스팸 흐름과 정상 흐름은 서로 **독립(Independent)**입니다.
- 이유: 전체 개수가 정해진 게 아니라, 매 순간 독립적으로 주사위를 굴려 분류하기 때문입니다.

8.5 실전 응용: 게임 아이템과 서버 관리

[시나리오] 스토리 시나리오: Nexon 게임 기획 및 서버 운영

상황 1: 아이템 강화 (베르누이 - Timing) 강화 성공 확률이 1%($p = 0.01$)입니다. 유저가 성공할 때까지 평균 몇 번 시도해야 할까요?

- 이는 '첫 성공까지의 대기 시간'이므로 **기하 분포**입니다.
- $E[T] = \frac{1}{p} = \frac{1}{0.01} = 100$번.
- "100번 실패했으니 다음엔 되겠지?" $\rightarrow$ 틀렸습니다. 여전히 확률은 1%입니다 (Memoryless).

상황 2: 서버 대기열 (포아송 - Counting) 평소에는 초당 10명($\lambda = 10$)이 접속합니다. 1초 동안 접속자가 15명 이상 폭주하여 렉이 걸릴 확률은?

- 이는 '정해진 시간 내 발생 횟수'이므로 **포아송 분포**입니다.
- $P(K \geq 15) = 1 - \sum_{k=0}^{14} \frac{10^k e^{-10}}{k!}$ (보통 근사값이나 표 사용)

8.6 요약 및 다음 단계

Summary

Unit 3-2 핵심 요약]

1. 관점의 전환: 실험(Experiment)에서 흐름(Process)으로.
2. 대응 관계: 이산(베르누이-이항-기하) $\leftrightarrow$ 연속(포아송-포아송-지수).

3. 무기억성: 확률 과정 해석의 핵심 키워드. ”시스템은 매 순간 리셋된다.”

4. 시스템: 병합은 더하고($+$), 분할은 곱한다($\times p$). 분할된 흐름은 독립이다.

〔다음 단원 예고〕
지금까지는 과거가 미래에 영향을 주지 않는(Memoryless) 특수한 상황만 다뤘습니다. 하지만 현실에서는 **”오늘 비가 오면 내일 비가 올 확률이 높다”**처럼 과거 상태가 미래에 영향을 줍니다.
Unit 3의 마지막 장 Chapter 9: 마르코프 체인(Markov Chains)에서는 ”바로 직전의 과거(State)”가 미래를 결정하는 조금 더 현실적인 모델을 배웁니다.

Chapter 9

마르코프 체인: 과거는 잊고 미래로

Intro: 무작위성(Randomness)에 구조(State)를 입히다

[이전 단계와의 연결]
Chapter 8의 베르누이/포아송 과정은 "매 순간이 똑같다(Memoryless)"는 가정하에 진행되었습니다.
하지만 현실은 다릅니다. '맑음' 뒤에는 '맑음'일 확률이 높고, '비' 뒤에는 '흐림'일 확률이 높습니다.
이제 우리는 **"현재의 상태(State)가 미래의 확률을 결정하는"** 조금 더 똑똑한 시스템, 마르코프 체인을 배웁니다.

Summary

이 단원의 핵심 개요 (Overview)]

1. 마르코프 성질: "미래는 오직 현재에만 달려있다." 과거 이력은 깡그리 무시합니다.
2. 전이 행렬 (Map): 상태들 사이를 이동하는 확률 지도를 행렬로 표현합니다.
3. 구조 분석 (Class): 한번 가면 못 돌아오는 길(Transient)과 갇히는 길(Recurrent)을 구분합니다.
4. 정상 상태 (π): 시간이 아주 오래 흐르면 시스템은 어떤 평형 상태에 도달하는가?

9.1 필수 용어 사전

mainblue!20 기호	용어	직관적 의미
State (i)	상태	시스템이 머무를 수 있는 위치 (예: 맑음, 흐림, 비)
p_{ij}	전이 확률	상태 i에서 j로 한 번에 건너갈 확률
P	전이 행렬	모든 p_{ij}를 모아놓은 지도 (행의 합은 1)
Recurrent	재귀 상태	언젠가는 반드시 다시 돌아오는 상태 (집)
Transient	일시적 상태	떠나면 다시는 못 돌아올 수도 있는 상태 (호텔 로비)
π_j	정상 상태 확률	먼 미래($n \to \infty$)에 내가 j에 있을 확률

9.2 핵심 개념 1: 개구리의 점프 (Markov Property)

[Image of Markov Chain state transition diagram with 3 nodes and arrows]

9.2.1 마르코프 성질 (The Markov Property)

[개념] 과거 세탁의 원칙

- 정의: 내일의 상태(X_{n+1})를 예측하는 데 필요한 정보는 오직 오늘(X_n)뿐입니다. 어제(X_{n-1}) 날씨가 어땠는지는 상관없습니다.
- 수식:
$$P(X_{n+1} = j \mid X_n = i, X_{n-1}, \dots) = P(X_{n+1} = j \mid X_n = i) = p_{ij}$$
- 의미: 모든 "역사(History)" 정보를 현재의 "상태(State)" 변수 하나에 압축해 넣어야 모델링이 성공한 것입니다.

9.2.2 전이 행렬 (Transition Matrix P)

확률을 표(Matrix)로 정리합니다.

$$P = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}$$

- **규칙:** 각 행(Row)의 합은 반드시 1이어야 합니다. (어디론가는 가야 하니까요.)
- **n-단계 전이:** n번 점프해서 $i \to j$로 갈 확률은 행렬을 n번 곱한 P^n의 성분과 같습니다. (행렬 연산의 강력함!)

9.3 핵심 개념 2: 구조 파악하기 (Topology)

계산기를 들기 전에, 반드시 **그림(다이어그램)**을 그려서 화살표를 따라가 봐야 합니다.
[Image of Recurrent vs Transient states diagram]

- 재귀 (Recurrent): "개미지옥". 한번 들어가면 그 집합 안에서만 뱅뱅 돌고, 절대 밖으로 못 나옵니다. 무한히 방문합니다.
- 일시적 (Transient): "환승역". 잠시 머물 수는 있지만, 언젠가는 떠나서 다시는 돌아오지 않습니다. 방문 횟수가 유한합니다.
- 흡수 (Absorbing): 나한테서 나로 가는 확률이 100%($p_{ii} = 1$)인 재귀 상태. (블랙홀)

[주의] 잠깐! 오해하기 쉬워요: 접근 가능성(Accessibility) 체크

서로 왔다 갔다 할 수 있는 상태끼리 묶어서 '클래스(Class)'라고 부릅니다. 마르코프 체인 문제를 풀 땐 가장 먼저 **"전체 시스템이 몇 개의 덩어리(Recurrent Classes)로 쪼개져 있는가?"**를 파악해야 합니다.

9.4 핵심 개념 3: 먼 미래의 평형 (Long-run Equilibrium)

시간이 아주 오래 흐르면($n \to \infty$), 내가 어디에 있을지 예측할 수 있을까요?

9.4.1 평형 방정식 (Balance Equations)

수조의 물 높이가 일정하다면, **"들어오는 물 = 나가는 물"**이어야 합니다.

[개념] Global Balance Equation
상태 j에 대해:

$$\pi_j = \sum_k \pi_k p_{kj}$$

(나가는 양) = (들어오는 양의 합)

- π_j: 내가 상태 j에 있을 확률 (다음 턴에 무조건 나가므로 유출량과 같음)
- $\sum \pi_k p_{kj}$: 다른 모든 곳(k)에서 나(j)한테 올 확률의 합.

[단계] 정상 상태(π) 구하는 법

1. **연립 방정식 세우기:** 모든 상태 j에 대해 $\pi_j = \sum \pi_k p_{kj}$를 씁니다.
2. **정규화 조건 추가:** 위 식만으로는 해가 안 나옵니다. 반드시 아래 식을 추가하세요.

$$\sum_j \pi_j = 1$$

3. **풀기:** 연립 방정식을 풀면 유일한 해 π가 나옵니다. (단, 단일 재귀 클래스일 때)

9.5 핵심 개념 4: 첫 스텝 분석법 (First Step Analysis)

MIT 6.431에서 가장 우아하고 강력한 문제 해결 테크닉입니다. 무한 급수를 계산하는 대신, **"딱 한 발자국"**만 생각해서 재귀식을 세웁니다.

문제: 상태 i에서 시작해서 흡수 상태(Goal)까지 가는 평균 시간 μ_i는?

$$\mu_i = 1 + \sum_j p_{ij} \mu_j$$

- 1: 일단 한 발자국 움직임 (시간 1 소모).
- $\sum p_{ij}\mu_j$: 그 다음 도착한 곳(j)에서 겪게 될 평균 시간의 가중 평균.

이 연립 방정식을 풀면 복잡한 미로 찾기의 평균 시간을 아주 쉽게 구할 수 있습니다.

9.6 실전 응용: 게임 유저 이탈 분석 (Churn Analysis)

[Image of User Funnel Markov Chain: Install -> Tutorial -> Play -> Purchase or Churn]

[시나리오] 스토리 시나리오: Nexon 모바일 게임 유저 분석
유저의 상태를 다음과 같이 정의합니다.

- 상태 1: 튜토리얼 (시작)
- 상태 2: 일반 플레이 (Recurrent 같지만 Transient임)
- 상태 3: 구매 (Absorbing Goal) - 우리가 원하는 목표

- 상태 4: 삭제/이탈 (Absorbing Bad) - 원치 않는 결과

질문 1: 튜토리얼을 시작한 유저가 결국 구매(상태 3)에 도달할 확률은?

- 흡수 확률 a_i에 대한 First Step Analysis 방정식을 세워 풉니다.
- $a_1 = p_{12}a_2 + p_{14}a_4$ (여기서 $a_3 = 1, a_4 = 0$)

질문 2: 구매까지 평균 몇 번의 플레이(Click)를 하는가?

- 기대 시간 μ_i 방정식을 세워 풉니다.
- 이 데이터를 바탕으로 "튜토리얼 난이도를 낮춰야 구매 전환율이 오르겠구나" 같은 의사결정을 합니다.

9.7 Course Summary: 확률론의 여정을 마치며

Summary

Unit 3-9 핵심 요약]

1. **모델링:** "상태(State)"를 잘 정의하면 과거를 잊을 수 있다(Markov).
2. **구조:** 한번 갇히면 못 나오는 방(Recurrent)과 지나가는 방(Transient)을 구분해라.
3. **평형:** $\pi P = \pi$. 들어오는 만큼 나간다. (장기적 확률)
4. **첫 스텝 분석:** 무한히 더하지 말고, 재귀 방정식($x = 1 + \sum px$)을 세워라.

[Final Comment]
수고하셨습니다!
우리는 **불확실성을 집합으로 정의**하는 것부터 시작해(Unit 1), **확률 변수라는 함수**로 세상을 모델링하고(Unit 2), 무한한 데이터 속에서 **극한의 법칙**을 발견하고, 마지막으로 **시간에 따른 변화(Process)**까지 다루었습니다(Unit 3).

이제 여러분은 "불확실한 세상"을 두려워하는 것이 아니라, **수학적 모델로 구조화하고 예측할 수 있는 눈**을 가지게 되었습니다. 이 지식은 머신러닝, 통계, 금융, 공학 어디서든 가장 든든한 무기가 될 것입니다.