CS109A Introduction to Data Science
제10강: 베이즈 추론 (Bayes Inference)

강의 요약 및 재구성: Gemini

October 26, 2025

▣ **강의명:** CS109A: 데이터 과학 입문

▣ **주차:** Lecture 10

▣ **교수명:** Pavlos Protopapas, Kevin Rader, Chris Gumb

▣ **목적:** Lecture 10의 핵심 개념 학습

□ 핵심 요약

이 문서의 핵심 요약

이 문서는 통계적 추론의 기본(신뢰구간, 가설검정)을 복습하고, 'Likelihood)' 개념을 소개합니다.

궁극적으로, 이 개념들을 조합하여 '베이즈 추론(Bayesian Inference)'의 핵심 원리를 설명합니다.

베이즈 추론의 핵심은 **'데이터(증거)'를 바탕으로 '사전의 믿음(Prior)'을 '사후의 믿음(Posterior)'으로 갱신하는 것**입니다.

진단 키트 예시를 통해 베이즈 정리의 직관을 배우고, 이를 통계 모델의 모수(θ) 추정에 적용하는 방법을 다룹니다.

Contents

1 핵심 용어 정리

본격적인 학습에 앞서, 이번 강의에서 사용되는 핵심 용어들을 정리합니다.

핵심 용어집 (Glossary)

용어 (한글)	쉬운 설명	용어 (영어)	기호/비고
모수	우리가 알고 싶은 모집단의 실제 값 (e.g., 평균, 회귀계수)	Parameter	θ, β_1, μ
추정량	샘플 데이터를 이용해 계산한 '모수 추정치' (e.g., 표본평균)	Estimate / Estimator	$\hat{\theta}, \hat{\beta}_1, \bar{X}$
표준 오차	추정량이 얼마나 부정확한지(퍼져 있는지) 나타내는 척도	Standard Error (SE)	$\hat{SE}(\hat{\beta}_1)$
신뢰 구간	모수가 포함될 것이라 '신뢰'하는 구간 (점 추정 + 불확실성)	Confidence Interval (CI)	e.g., 95% CI
예측 구간	'새로운 단일 관측치'가 존재할 것이라 '예측'하는 구간	Prediction Interval (PI)	항상 신뢰구간보다 넓음
귀무 가설	우리가 기각하고 싶은 가설 (e.K., "관계가 없다", "차이가 없다")	Null Hypothesis	$H_0 : \beta_1 = 0$
p-값	귀무가설이 맞다고 가정할 때, 현재 데이터를 얻을 확률	p-value	p-value < 0.05
가능도	관찰된 데이터를 기반으로, 특정 모수가 얼마나 '그럴듯한지'	Likelihood	$L(\theta\|\text{data})$
최대가능도추정	가능도를 최대화하는 모수 θ를 찾는 방법	Max Likelihood Est. (MLE)	
사전 확률	데이터를 보기 전, 모수에 대해 갖는 초기 믿음(확률)	Prior Probability	$P(\theta)$, $f(\theta)$
사후 확률	데이터를 본 후, 갱신된 모수에 대한 믿음(확률)	Posterior Probability	$P(\theta\|\text{data})$, $f(\theta\|X)$
민감도	실제 '참'인 것을 '참'이라고 예측할 확률 (True Positive Rate)	Sensitivity	$P(T+\|D+)$
특이도	실제 '거짓'인 것을 '거짓'이라고 예측할 확률 (True Negative Rate)	Specificity	$P(T-\|D-)$

Table 1: 제*10*강의 핵심 통계 용어 정리

2 통계적 추론 복습 (Inference Review)

데이터 분석은 샘플(표본)을 통해 모집단의 특성을 알아내는 ’추론’ 과정입니다. 단순히 모델을 만드는 것을 넘어, 이 모델이 얼마나 신뢰할 수 있는지 평가해야 합니다.

2.1 모집단 모델 vs. 표본 모델

우리가 관심 있는 것은 모집단의 ’진짜’ 관계(Population Model)이지만, 우리가 가진 것은 샘플 데이터로 ’추정’한 모델(Estimated Model)뿐입니다.

- **모집단 모델 (이론):** $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$
  - 우리가 찾고 싶은 진짜 값 β_0, β_1
  - ϵ_i: 측정 불가능한 오차 ($N(0, \sigma^2)$ 가정)
- **표본 모델 (추정):** $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ (e.g., $\hat{y} = 247.4 + 0.5898 \cdot x$)
  - 우리가 가진 데이터로 계산한 추정치 $\hat{\beta}_0, \hat{\beta}_1$
  - 이 값은 샘플이 달라지면 계속 바뀌는 ’하나의 추측’에 불과합니다.

2.2 신뢰 구간 (Confidence Intervals)

점 추정(Point Estimate) $\hat{\beta}_1 = 0.5898$은 하나의 추측일 뿐, 진짜 β_1이 0.6일지, 0.7일지, 아니면 0일지 알 수 없습니다. 이 ’불확실성’을 표현하는 방법이 **신뢰 구간(CI)** 입니다.

신뢰 구간을 만드는 방법은 크게 두 가지가 있습니다.

1. **부트스트랩 (Bootstrap):** 데이터를 반복 재추출(resampling)하여 $\hat{\beta}_1$의 경험적 분포를 만듦 (컴퓨터 파워)
2. **공식 기반 (Formula-based):** 확률 이론(정규분포, t-분포)에 기반한 공식을 사용 (수학)

2.2.1 공식 기반 신뢰 구간

신뢰 구간은 추정치에서 표준 오차(Standard Error, SE)의 배수만큼 더하고 빼서 만듭니다.

$$95\% \text{ 신뢰 구간} = \hat{\beta}_1 \pm t^* \cdot \hat{SE}(\hat{\beta}_1)$$

- $\hat{SE}(\hat{\beta}_1)$: 추정치 $\hat{\beta}_1$의 불확실성을 나타내는 척도.
- t^*: 신뢰 수준(e.g., 95%)을 결정하는 값. (샘플이 충분히 크면 약 $1.96 \approx 2$)

□ 예제: title

추정치의 불확실성(SE)을 어떻게 줄일 수 있을까요? $SE(\hat{\beta}_1) \approx \frac{\hat{\sigma}_\epsilon}{\sqrt{\sum(x_i - \bar{x})^2}}$ 공식에서 힌트를 얻을 수 있습니다.

- **1. 데이터(n) 늘리기:** 샘플 크기 n이 커지면 분모의 $\sum(\dots)$ 항이 커져 SE가 줄어듭니다. (가장 확실한 방법)

- **2. X의 범위 넓히기:** X 값들이 넓게 퍼져있을수록($Var(X)\uparrow$), 분모가 커져 SE가 줄어듭니다. (좁은 범위의 X로 예측하면 불안정함)
- **3. 노이즈(σ_ϵ) 줄이기:** 데이터가 회귀선 주변에 빽빽하게 모여있을수록($\hat{\sigma}_\epsilon\downarrow$), 분자인 $\hat{\sigma}_\epsilon$이 작아져 SE가 줄어듭니다. (더 정확한 모델)

2.2.2 왜 Z분포가 아닌 t-분포를 사용할까?

- **정규분포 (Z):** 모집단의 실제 노이즈(σ_ϵ)를 **알고 있을 때** 사용합니다.
- **t-분포 (t):** σ_ϵ을 **모르기 때문에** 데이터의 잔차(residual)로 $\hat{\sigma}_\epsilon$를 **추정했을 때** 사용합니다.

σ_ϵ를 추정하는 과정에서 추가적인 불확실성이 발생합니다. t-분포는 이 불확실성을 반영하기 위해 정규분포보다 **'꼬리가 두꺼운(fatter tails)'** 형태를 가집니다. (단, 샘플 크기 n이 50개 이상으로 매우 커지면 t-분포는 정규분포와 거의 동일해집니다.)

2.3 가설 검정 (Hypothesis Testing)

"이 변수가 Y와 정말 관련이 있을까?" (e.g., β_1이 0이 아닐까?)를 통계적으로 답하는 과정입니다.

가설 검정 5단계

1. **가설 설정:**
   - **귀무 가설 (H_0):** 관계가 없다. ($H_0 : \beta_1 = 0$)
   - **대립 가설 (H_A):** 관계가 있다. ($H_A : \beta_1 \neq 0$)
2. **검정 통계량 선택:** t-통계량 $t = \frac{\hat{\beta}_1 - 0}{\hat{SE}(\hat{\beta}_1)}$ (추정치가 0으로부터 표준 오차의 몇 배만큼 떨어져 있는가?)
3. **데이터 수집 및 계산:** 샘플 데이터로 $\hat{\beta}_1$과 $\hat{SE}(\hat{\beta}_1)$을 계산하여 t값 확인.
4. **결정 (p-value):**
   - **p-value란?** H_0가 맞다(즉, $\beta_1 = 0$)고 가정할 때, 우리가 관찰한 t값보다 더 극단적인 값이 나올 확률.
   - p-value가 매우 작으면 (e.g., < 0.05), "H_0가 맞다고 보기엔 너무 희귀한 일이 일어났군. H_0를 기각하자!"
5. **결론 도출:** "p-value가 0.001이므로, X와 Y 사이에는 통계적으로 유의미한 연관성이 있다."

가설 검정을 위한 부트스트랩? → 순열 검정!

신뢰 구간(CI)을 만들 때는 **부트스트랩(Bootstrap)**이 유용합니다. 하지만 가설 검정(p-value)에는 부트스트랩을 쓰면 안 됩니다. (Type I Error 증가 문제)
가설 검정 시에는 **순열 검정(Permutation Test)**을 사용해야 합니다.

- **부트스트랩 (CI용):** H_0와 무관하게 원본 데이터에서 (복원) 재추출.
- **순열 검정 (p-value용):** H_0가 맞다(X, Y 관계 없음)고 가정하고, Y 값의 순서를 무작위로 섞은(shuffle) 후 X와 다시 매칭하여 β_1^*을 계산.

2.4 모델 가정과 추론 방법의 선택

공식 기반 추론(t-분포)은 빠르고 편리하지만, 강력한 **모델 가정(Assumptions)**이 필요합니다.

- **L**inearity (선형성): 관계는 선형이다.
- **I**ndependence (독립성): 오차(residual)들은 서로 독립이다. (가장 중요)
- **N**ormality (정규성): 오차는 정규분포를 따른다. (n이 크면(e.g., >50) 덜 중요)
- **E**qual Variance (등분산성): 오차의 분산은 X 값에 상관없이 일정하다. (중요)

사례: 공식(statsmodels) vs. 부트스트랩

강의에서 주택 가격 데이터($n = 592$)를 분석한 결과, 잔차 그림에서 X가 커질수록 분산이 커지는 '부채꼴 모양(fanning out)'이 나타났습니다. 이는 **등분산성(E) 가정이 위반**되었음을 의미합니다.
이때 두 방법으로 구한 β_1 (sqft)의 95% CI는 다음과 같습니다.

- **공식 (statsmodels):** (0.544, 0.636) $\rightarrow$ 가정이 깨져서 **잘못된(너무 좁은)** 구간.
- **부트스트랩:** (0.487, 0.705) $\rightarrow$ 등분산성 가정이 필요 없으므로, 이 상황에서 **더 신뢰할 수 있는** 구간.

결론: 모델 가정이 깨졌을 때는 부트스트랩 같은 비모수적(non-parametric) 방법이 더 강건(robust)합니다.

3 신뢰 구간 vs. 예측 구간 (CI vs. PI)

통계적 추론에서 가장 혼동하기 쉬운 두 가지 '구간'이 있습니다. 강의 중 두 번째 퀴즈가 이 차이점을 정확히 짚었습니다.

- **신뢰 구간 (Confidence Interval):**
  - **질문:** "2860 평방피트(x)를 가진 **집단의 평균 가격**은 어디쯤 있을까?"
  - **의미:** 회귀선 자체의 불확실성. (e.g., $\hat{y} \pm t^* \cdot \mathrm{SE(fit)}$)
- **예측 구간 (Prediction Interval):**
  - **질문:** "2860 평방피트(x)를 가진 **새로운 집 한 채**의 가격은 어디쯤 있을까?"
  - **의미:** 회귀선의 불확실성 + 개별 데이터의 고유한 변동성(노이즈).

예측 구간(PI)은 항상 더 넓습니다

예측 구간은 **두 가지 불확실성**을 모두 고려해야 합니다.

1. **모델(회귀선)의 불확실성:** $\hat{\beta}_0, \hat{\beta}_1$ 추정치의 불확실성. (신뢰구간이 다루는 영역)
2. **개별 데이터의 불확실성 (Irreducible Error, $\hat{\sigma}_\epsilon$):** 아무리 모델이 완벽해도, 개별 주택 가격은 '진짜' 평균선 주위에 흩어져 있습니다.

따라서 예측 구간의 표준 오차는 두 불확실성을 모두 합산합니다.

$$\mathrm{PI} = \hat{y} \pm t^* \cdot \sqrt{(\mathrm{SE(fit)})^2 + \hat{\sigma}_\epsilon^2}$$

(강의 퀴즈의 D번 보기 $247.4 + 0.5898(2860) \pm 2\sqrt{0.023^2 + \hat{\sigma}^2}$ 와 동일한 형태)

4 가능도 (Likelihood)

베이즈 추론을 이해하기 위한 핵심 구성요소인 '가능도'에 대해 알아봅니다.

4.1 정의: PDF를 뒤집어 생각하기

지금까지 우리는 확률 밀도 함수(PDF)를 $P(\text{data}|\theta)$로 생각했습니다. (e.g., $\mu = 0, \sigma = 1$일 때, $x = 2$가 나올 확률은?)

가능도 (Likelihood)는 이 관점을 뒤집습니다.

- **데이터(x)를 '고정된 관찰 값'으로 봅니다.**
- **모수(θ)를 '변수'로 봅니다.**
- **질문:** "우리가 **관찰한 데이터** x가 있을 때, 어떤 θ 값이 이 데이터를 가장 '그럴듯하게(likely)' 설명하는가?"

$$L(\theta|\text{data}) = P(\text{data}|\theta)$$

수학적 형태는 PDF와 같지만, θ에 대한 함수라는 점이 다릅니다.

4.2 최대 가능도 추정 (Maximum Likelihood Estimation, MLE)

가능도 $L(\theta|\text{data})$를 최대로 만드는 θ 값을 찾는 것을 **MLE**라고 부릅니다.

- **데이터:** $x_1, x_2, \ldots, x_n$ (서로 독립이라 가정)
- **가능도:** $L(\theta) = \prod_{i=1}^{n} P(x_i|\theta)$ (모든 데이터가 동시에 관찰될 확률)

□ **예제: title**

여러 확률을 곱하는 것($\prod$)은 계산이 복잡하고 언더플로우(underflow) 위험이 있습니다. 대신 로그(log)를 취하면 곱셈이 덧셈($\sum$)으로 바뀝니다.

$$\log(L(\theta)) = \sum_{i=1}^{n} \log(P(x_i|\theta))$$

$L(\theta)$를 최대화하는 θ는 $\log(L(\theta))$를 최대화하는 θ와 동일합니다.
예시: $\sigma^2 = 4$인 정규분포에서 [3, 5, 10] (평균 6) 데이터가 관찰됨. $\log(L(\mu|\text{data}))$ 그래프는 $\mu = 6$일 때 정확히 최댓값을 가집니다. 즉, $\hat{\mu}_{MLE} = 6 = \bar{X}$ 입니다.

MLE를 찾는 방법

1. **수학적 방법:** $\log(L(\theta))$를 θ에 대해 미분하여 0이 되는 지점을 찾습니다.
2. **컴퓨터 방법:** 경사 하강법(Gradient Descent)을 사용합니다. (단, 경사 하강법은 '최소화' 알고리즘이므로, $\log(L(\theta))$ 대신 **음의 로그 가능도 (Negative Log-Likelihood)** $-\log(L(\theta))$를 최소화합니다.)

OLS와 MLE의 중요한 관계

선형 회귀(Ordinary Least Squares, OLS)와 MLE는 깊은 관련이 있습니다.
만약 선형 회귀의 오차(ϵ_i)가 **정규분포를 따른다**고 가정한다면, 가능도 $L(\beta_0, \beta_1, \sigma^2|\text{data})$를 계산할 수 있습니다.
이때, **음의 로그 가능도(Negative Log-Likelihood)**를 최소화하는 것은 수학적으로 **오차 제곱합(Sum of Squared Errors, SSE)** $\sum(y_i - (\beta_0 + \beta_1 x_i))^2$을 최소화하는 것과 **정확히 동일**합니다.
결론: OLS는 '오차가 정규분포를 따른다'는 가정 하의 MLE입니다.

5 베이즈 정리 (Bayes' Rule)

베이즈 추론의 기반이 되는 베이즈 정리는 '조건부 확률'을 뒤집는 방법을 제공합니다.

5.1 기본 공식

두 사건 A, B에 대하여 B가 일어났을 때 A가 일어날 확률은 다음과 같습니다.

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

□ **예제: title**

어떤 수업에 STAT 전공자(A)가 40%, CS 전공자(B)가 60%이며, 둘 다 전공하는 학생(A and B)은 20%라고 가정해봅시다.

- **질문 1: STAT 전공자(A) 중에서 CS도 전공(B)할 확률은?**

$$P(B|A) = \frac{P(A \text{ and } B)}{P(A)} = \frac{0.20}{0.40} = 0.50 \quad (50\%)$$

- **질문 2: CS 전공자(B) 중에서 STAT도 전공(A)할 확률은?**

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)} = \frac{0.20}{0.60} = 0.333 \quad (33.3\%)$$

이처럼 $P(A|B)$와 $P(B|A)$는 전혀 다른 값입니다. 베이즈 정리는 한쪽을 알 때 다른 쪽을 계산할 수 있게 해줍니다.

5.2 베이즈 정리의 직관: 믿음의 갱신 (Belief Update)

베이즈 정리는 '증거(Evidence)'를 바탕으로 '사전의 믿음(Prior Belief)'을 '사후의 믿음(Posterior Belief)'으로 갱신하는 논리적 과정입니다.

□ **예제: title**

어떤 사람이 임신했을 확률($D+$)이 30%라고 **'초기에 믿고(Prior)'** 있습니다. 이 사람이 사용한 키트의 정보는 다음과 같습니다.

- **민감도 (Sensitivity):** $P(T+|D+) = 0.97$ (실제 임신 시, '양성'이 나올 확률)
- **특이도 (Specificity):** $P(T-|D-) = 0.99$ ($\rightarrow$ 실제 비임신 시, '양성'이 나올 확률 $P(T+|D-) = 1 - 0.99 = 0.01$)

이 사람이 검사 후 **'양성($T+$)'이라는 증거(Evidence)**를 얻었습니다. 이제 이 사람이 실제 임신($D+$)했을 확률 $P(D+|T+)$은 얼마일까요?
베이즈 정리를 사용합니다.

$$P(D+|T+) = \frac{P(T+|D+)P(D+)}{P(T+)}$$

여기서 $P(T+)$는 양성이 나올 모든 경우의 확률 (Law of Total Probability) 입니다.

$$P(T+) = P(T + |D+)P(D+) + P(T + |D-)P(D-)$$

$$P(T+) = (0.97 \times 0.30) + (0.01 \times 0.70) = 0.291 + 0.007 = 0.298$$

따라서 사후 확률은:

$$P(D + |T+) = \frac{0.291}{0.298} \approx 0.9765 \quad (97.65\%)$$

결론: 30% 였던 '믿음'이 '양성'이라는 '증거'를 통해 97.65% 로 '갱신'되었습니다.

6 베이즈 추론 (Bayesian Inference)

이제 베이즈 정리를 통계적 모수(θ) 추론에 적용합니다.

6.1 베이즈 추론의 핵심 공식

진단 키트 예시의 구성요소를 통계 모델의 모수(θ)와 데이터(X)로 치환합니다.

- $D+$ (실제 상태) $\to \theta$ (우리가 모르는 실제 모수)
- $T+$ (검사 결과) $\to X$ (우리가 관찰한 데이터)

베이즈 추론 공식

$$f(\theta|X) = \frac{f(X|\theta) \cdot f(\theta)}{f(X)}$$

이 공식의 각 요소를 이해하는 것이 베이즈 추론의 전부입니다.

- $f(\theta|X)$ — **사후 분포 (Posterior Distribution)**
  - **의미:** 데이터를 관찰한 **후**에 갱신된 θ에 대한 믿음의 분포.
  - **결과물:** 이것이 우리가 얻고자 하는 최종 결과입니다.
- $f(X|\theta)$ — **가능도 (Likelihood)**
  - **의미:** 특정 모수 θ를 가정했을 때, 현재 데이터 X가 관찰될 확률.
  - **역할:** 데이터 X가 θ의 어떤 값을 지지하는지 알려주는 '증거'의 역할. (앞서 배운 MLE의 그 가능도입니다.)
- $f(\theta)$ — **사전 분포 (Prior Distribution)**
  - **의미:** 데이터를 관찰하기 **전**에 θ에 대해 우리가 가졌던 초기 믿음.
  - **역할:** 이전 연구나 상식 등, 데이터 외의 정보를 모델에 반영합니다.
- $f(X)$ — **증거 (Evidence) 또는 정규화 상수**
  - **의미:** $f(X) = \int f(X|\theta)f(\theta)d\theta$. (모든 θ에 대해 가능도를 평균낸 값)
  - **역할:** 사후 분포 $f(\theta|X)$의 총합이 1이 되도록 맞춰주는 '상수' 역할. 계산이 복잡해서 종종 생략하고 비례 관계로 표현합니다.

더 간단한 표현: **Posterior $\propto$ Likelihood $\times$ Prior**

6.2 빈도주의 (Frequentist) vs. 베이지안 (Bayesian)

전통적인 통계(빈도주의)와 베이즈 추론은 '확률'과 '모수'를 바라보는 근본적인 관점이 다릅니다.

항목	빈도주의 (Frequentist) (e.g., OLS, MLE)	베이지안 (Bayesian)
모수(θ) 관점	**고정된 상수.** (단지 우리가 모를 뿐)	**확률 변수.** (믿음의 분포를 가짐)
확률의 정의	장기적인 실험에서 발생하는 **빈도**	**주관적 믿음**의 정도 (Belief)
핵심 질문	"θ가 0이라는 H_0를 기각할 수 있는가?"	"데이터 X를 본 후, θ에 대한 믿음은?"
결과물	점 추정치 $\hat{\theta}$와 **신뢰구간(CI)**	**사후 분포** $f(\theta\|X)$ (전체 분포)
구간의 해석	"이 **절차**로 CI를 100번 만들면, 95개는 θ를 포함한다."	"θ가 이 **구간**에 있을 확률(믿음)이 95%이다."

Table 2: 빈도주의와 베이지안 추론의 관점 비교

6.3 이산적 베이즈 추론 예시: 3개의 동전

베이즈 추론이 '믿음을 갱신'하는 과정을 간단한 이산적 예시로 살펴봅니다.

□ 예제: title

주머니 속에 3개의 동전(θ)이 있습니다: $p = 0.1$ (가짜), $p = 0.5$ (공정), $p = 0.9$ (가짜). 이 중 하나를 랜덤하게 뽑아 4번($n = 4$) 던졌더니, **3번의 앞면(H)과 1번의 뒷면(T)**이라는 데이터 (X)가 나왔습니다.

질문: 우리는 3개의 동전 중 어떤 것을 뽑았다고 믿어야 할까요?

1. 사전 분포 (Prior) $f(\theta)$ 설정 데이터를 보기 전, 각 동전을 뽑을 확률은 공평하게 1/3입니다.
$P(p = 0.1) = 1/3, \quad P(p = 0.5) = 1/3, \quad P(p = 0.9) = 1/3$

2. 가능도 (Likelihood) $f(X|\theta)$ 계산 각 동전(θ)을 가정했을 때, '3H 1T' 데이터(X)가 나올 확률 (이항분포)

- $L(p = 0.1) = P(X|p = 0.1) = \binom{4}{3}(0.1)^3(0.9)^1 = 4 \times 0.001 \times 0.9 = \mathbf{0.0036}$
- $L(p = 0.5) = P(X|p = 0.5) = \binom{4}{3}(0.5)^3(0.5)^1 = 4 \times 0.125 \times 0.5 = \mathbf{0.2500}$
- $L(p = 0.9) = P(X|p = 0.9) = \binom{4}{3}(0.9)^3(0.1)^1 = 4 \times 0.729 \times 0.1 = \mathbf{0.2916}$

(데이터 X는 $p = 0.9$일 가능성을 가장 그럴듯하게 봅니다.)

3. 사후 분포 (Posterior) $\propto$ Likelihood $\times$ Prior

- $P(p = 0.1|X) \propto 0.0036 \times (1/3) = 0.0012$
- $P(p = 0.5|X) \propto 0.2500 \times (1/3) = 0.0833$
- $P(p = 0.9|X) \propto 0.2916 \times (1/3) = 0.0972$

4. 정규화 (Normalize) $\rightarrow$ 총합을 1로 만들기 총합 $= 0.0012 + 0.0833 + 0.0972 = 0.1817$

- $P(p = 0.1|X) = 0.0012/0.1817 \approx \mathbf{0.007}$ **(0.7%)**
- $P(p = 0.5|X) = 0.0833/0.1817 \approx \mathbf{0.458}$ **(45.8%)**
- $P(p = 0.9|X) = 0.0972/0.1817 \approx \mathbf{0.535}$ **(53.5%)**

결론: '3H 1T'라는 데이터를 관찰한 후, 우리의 믿음은 (1/3, 1/3, 1/3)에서 (0.7%, 45.8%, 53.5%)로 **갱신**되었습니다. 이제 우리는 $p = 0.9$ 동전을 뽑았다고 가장 강하게 믿게 되었습니다.

6.4 미리보기: 연속적인 모수 (Normal-Normal 모델)

위 예시는 모수 p가 3개의 값만 가지는 이산적(discrete) 경우였습니다. 실제로는 모수(μ, β_1 등)가 연속적(continuous)인 경우가 많습니다.

- **모델:** $X_i \sim N(\mu, \sigma^2)$ (단, σ^2은 안다고 가정)

- **사전 분포 (Prior):** μ에 대한 초기 믿음. (e.g., $\mu \sim N(\mu_0, \sigma_0^2)$)
- **데이터:** $X = \{x_1, \dots, x_n\}$ (표본 평균 $\bar{X}$)
- **사후 분포 (Posterior):** $f(\mu|X)$는 놀랍게도 또 다른 정규분포가 됩니다.

이때 사후 분포의 평균(우리의 최종 μ 추정치)은 다음과 같습니다.

$$\hat{\mu}_{\text{posterior}} = \frac{(\sigma^2 \cdot \mu_0) + (n\sigma_0^2 \cdot \bar{X})}{\sigma^2 + n\sigma_0^2}$$

이 식은 **사전 믿음(μ_0)과 데이터($\bar{X}$)의 가중 평균**입니다.

데이터가 사전 믿음을 이긴다

위 가중 평균 식에서 샘플 크기 n이 매우 커지면($n \to \infty$) 어떻게 될까요?

- μ_0의 가중치 $\to 0$
- $\bar{X}$의 가중치 $\to 1$

결론: 데이터가 충분히 많아지면, 초기에 어떤 사전 믿음(μ_0)을 가졌든 상관없이 사후 분포는 데이터가 말해주는 값($\bar{X}$)으로 수렴합니다.

7 참고: ‘statsmodels‘ 결과표 분석

강의에서 사용된 ‘statsmodels‘ (Python 라이브러리)의 OLS 회귀분석 결과표를 해석하는 방법입니다.

단순 선형 회귀: ‘price   sqft‘

```
==============================================================================
Dep. Variable:                  price   R-squared:
                         0.519
Model:                            OLS   Adj. R-squared:
                    0.518
Method:                 Least Squares   F-statistic:
                       635.6
Date:                Tue, 03 Oct 2023   Prob (F-statistic):
            9.97e-96
Time:                        22:00:05   Log-Likelihood:
                  -4566.2
No. Observations:                 592   AIC:
                               9136.
Df Residuals:                     590   BIC:
                               9145.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025
                         0.975]
------------------------------------------------------------------------------
Intercept    247.4382     45.388      5.452      0.000     158.296
       336.581
sqft           0.5898      0.023     25.211      0.000       0.544
         0.636
==============================================================================
```

Listing 1: 단순 선형 회귀*(OLS)* 결과

핵심 해석:

- **coef (계수):** ‘Intercept‘($\hat{\beta}_0$) = 247.4, ‘sqft‘($\hat{\beta}_1$) = 0.5898 (평방 피트가 1 증가할 때마다 가격이 0.5898 (천 달러) 증가)
- **std err (표준 오차):** ‘sqft‘의 $\hat{SE}(\hat{\beta}_1) = 0.023$ (0.5898 이라는 추정치의 불확실성 정도)
- **t (t-통계량):** 25.211 = (0.5898 - 0) / 0.023 ($H_0 : \beta_1 = 0$ 으로부터 25 표준오차만큼 떨어져 있음)
- **P>|t| (p-값):** 0.000 ($\beta_1 = 0$ 인데 이런 결과가 나올 확률이 0에 가까움 → H_0 기각)
- **[0.025 0.975] (95% 신뢰구간):** (0.544, 0.636) (모집단의 실제 β_1 이 0.544와 0.636 사이에 있을 것이라 95% 신뢰함)

다중 선형 회귀: ‘price sqft + dist + ... + type‘

```
... (R-squared: 0.733, Adj. R-squared: 0.729) ...
==============================================================================================

                          coef    std err          t      P>|t|
                                   [0.025      0.975]
----------------------------------------------------------------------------------------------

Intercept           -1949.0670    745.203     -2.615      0.009
   -3412.677    -485.457
type[T.multifamily] -452.2352     77.451     -5.839      0.000
   -604.352    -300.119
type[T.singlefamily] 335.7612     54.642      6.145      0.000
   228.441     443.081
type[T.townhouse]    -76.4372     56.859     -1.344      0.179
   -188.111      35.237
sqft                    0.6411      0.044     14.720      0.000
         0.556       0.727
dist                 -173.5430     20.099     -8.634      0.000
   -213.018    -134.067
beds                  -89.9345     23.532     -3.822      0.000
   -136.152     -43.717
baths                 198.4646     31.332      6.334      0.000
   136.928     260.002
year                    1.2300      0.388      3.169      0.002
         0.468       1.992
==============================================================================================
```

Listing 2: 다중 선형 회귀 *(OLS)* 결과 *(일부)*

핵심 해석 (다중 회귀):

- **계수 해석의 변화:** ‘sqft‘의 계수가 0.5898 → 0.6411로 변경되었습니다. 이는 다른 변수 (dist, beds 등)의 효과를 ’통제’했기 때문입니다.
- **‘beds‘ 계수 (-89.93):** ”다른 모든 변수(sqft, baths, dist 등)가 **동일하게 고정**되어 있다면, 침실 수가 1개 증가할 때 오히려 가격이 89.9 (천 달러) 감소한다.”
- **인과관계 주의:** 이는 상관관계(association)일 뿐, 침실을 늘리면 집값이 떨어진다는 인과관계(causation)를 의미하지 않습니다. (e.g., 같은 크기의 집에 침실만 무리하게 늘리면 가치가 떨어질 수 있음)

8 학습 체크리스트

이 강의를 통해 다음 질문에 답할 수 있는지 확인해 보세요.

최소 학습 목표 (Checklist)

- ☐ 모집단 모델(β_1)과 표본 모델($\hat{\beta}_1$)의 차이를 설명할 수 있는가?
- ☐ 표준 오차(SE)가 무엇인지, 그리고 SE를 줄이는 3가지 방법을 아는가?
- ☐ 왜 Z-분포 대신 t-분포를 사용하며, 둘의 형태상 차이는 무엇인가?
- ☐ p-value가 0.05보다 작다는 것이 (e.g., $H_0 : \beta_1 = 0$일 때) 무엇을 의미하는지 설명할 수 있는가?
- ☐ **신뢰 구간**과 **예측 구간**의 차이를 '질문'과 '불확실성'의 관점에서 설명할 수 있는가?
- ☐ 가능도(Likelihood)가 확률(Probability)과 어떻게 다른지, $L(\theta|X)$로 설명할 수 있는가?
- ☐ OLS(최소제곱법)와 MLE(최대가능도추정)의 관계는 무엇인가? (힌트: 정규분포 가정)
- ☐ 베이즈 정리를 '믿음의 갱신' 과정 (사전확률 → 증거 → 사후확률)으로 설명할 수 있는가?
- ☐ 베이즈 추론 공식의 4가지 요소($f(\theta|X), f(X|\theta), f(\theta), f(X)$)를 각각 명명하고 설명할 수 있는가?
- ☐ 빈도주의자와 베이지안이 '모수(θ)'를 바라보는 근본적인 관점 차이는 무엇인가?

9 FAQ (자주 묻는 질문)

초심자가 흔히 가질 수 있는 질문들을 정리했습니다.

Q: 부트스트랩 샘플 수(B)를 1000번에서 100만 번으로 늘리면 신뢰구간이 더 좁아지나요?

A: 아닙니다. 부트스트랩 샘플 수 B를 늘리는 것은 단지 우리가 추정한 샘플링 분포($\hat{\beta}_1$의 히스토그램)를 더 **'매끄럽게(smoother)'** 만들 뿐입니다. 이는 계산의 정밀도를 높이는 것이지, 원본 추정치의 근본적인 불확실성을 줄여주지 않습니다.
신뢰 구간을 **좁히는(불확실성을 줄이는) 유일한 방법**은 원본 데이터의 샘플 크기 n**을 늘리는 것**입니다. (e.g., 500개 홈 데이터 → 5000개 홈 데이터)

□ 예제: title

A: 좋은 지적입니다. 하지만 '근거 없는 찍기'와는 다릅니다.

1. **정보 사전확률 (Informative Prior):** 과거의 연구 결과나 해당 분야의 전문가적 합의(e.g., "회귀계수가 0에서 10 사이일 것")를 반영합니다.
2. **무정보 사전확률 (Uninformative Prior):** 정말 아는 것이 없을 때, 모든 가능성을 평평하게(uniform) 두는 사전확률을 사용합니다. (e.g., p가 0 1 사이 모든 값일 확률이 동일)
3. **데이터의 힘:** 가장 중요한 점은, **데이터(n)가 충분히 많아지면,** 초기에 어떤 사전확률을 설정했더라도 사후확률(결과)은 **거의 같은 결론으로 수렴**한다는 것입니다. 데이터가 주관적인 믿음을 '압도'합니다.

□ **예제: title**

A: 상황에 따라 다릅니다. (It depends.)

- **빈도주의 (Frequentist):** 고전적인 방법이며, 계산이 간단하고 빠릅니다. 객관적인 '절차'를 중시하는 분야(e.g., 의학 임상시험)에서 표준으로 사용됩니다.
- **베이지안 (Bayesian):** 계산이 복잡하지만(최근 10+년간 MCMC 등 컴퓨터 파워로 해결), '사전 지식'을 모델에 통합할 수 있습니다. 데이터가 적거나(n이 작을 때), 모수에 대한 불확실성을 '분포' 자체로 얻고 싶을 때 매우 강력합니다.

현대 데이터 과학에서는 두 가지 접근법을 모두 이해하고, 문제 상황에 맞게 사용하거나 두 결과를 비교 분석하는 경우가 많습니다.

10 한눈에 훑어보기 (1-Page Summary)

신뢰 구간 (Mean) vs. 예측 구간 (Single)

- **신뢰 구간 (CI):** ”2860 sqft 집들의 **평균** 가격은?” (불확실성 1개: 모델)
- **예측 구간 (PI):** ”2860 sqft 집 **한 채**의 가격은?” (불확실성 2개: 모델 + 개별 노이즈 $\hat{\sigma}_\epsilon$)
- $\rightarrow$ PI는 항상 CI보다 넓습니다.

OLS

선형 회귀에서 오차(ϵ)가 **정규분포**를 따른다고 가정하면, **오차 제곱합을 최소화(OLS)** 하는 것은 **가능도를 최대화(MLE)** 하는 것과 수학적으로 동일합니다.

베이즈 정리의 핵심 흐름

사전 믿음 (Prior) ($P(D+) = 30\%$) $+$ **증거 (Evidence)** (테스트 ’양성’) $\rightarrow$ **사후 믿음 (Posterior)** ($P(D+|T+) = 97.7\%$)

베이즈 추론의 심장

$$\underbrace{f(\theta|X)}_{\text{사후 분포}} \propto \underbrace{f(X|\theta)}_{\text{가능도}} \times \underbrace{f(\theta)}_{\text{사전 분포}}$$

(Posterior $\propto$ Likelihood $\times$ Prior)

빈도주의 vs. 베이지안

- **빈도주의자:** 모수(θ)는 **고정된 값**. 나의 추정치($\hat{\theta}$)가 확률적.
- **베이지안:** 모수(θ)는 **확률 변수**(분포). 나의 믿음이 데이터에 따라 갱신됨.

11 부록: 몬티 홀 문제 (Monty Hall Problem)

강의에서 간단히 언급된 몬티 홀 문제는 조건부 확률과 베이즈 정리의 직관을 보여주는 유명한 예시입니다.

문제 상황

1. 3개의 문(Door 1, 2, 3) 뒤에 1개는 자동차(Car), 2개는 염소(Goat)가 있습니다.
2. 당신은 하나의 문(e.g., Door 1)을 선택합니다.
3. 진행자(Monty)는 **당신이 선택하지 않은** 두 개의 문(Door 2, 3) 중에서, **염소가 있는 문 하나**를 열어서 보여줍니다. (e.g., Door 3에 염소가 있다고 열어줌)
4. 진행자가 묻습니다: "처음 선택(Door 1)을 **유지**하시겠습니까, 아니면 남은 문(Door 2)으로 **바꾸**시겠습니까?"

결론: 무조건 바꾸는 것이 유리합니다.

- **유지(Stay)할 경우 승률: 1/3**
- **변경(Switch)할 경우 승률: 2/3**

□ **예제: title**

시나리오 1: 맨 처음에 '자동차'를 선택한 경우 (확률 1/3)

- 당신: Door 1 (Car) 선택
- 진행자: Door 2(Goat) 또는 Door 3(Goat) 중 하나를 열어줌
- 당신: Door 2(Goat) 또는 Door 3(Goat)으로 **바꾼다.** → **패배!**

시나리오 2: 맨 처음에 '염소'를 선택한 경우 (확률 2/3)

- 당신: Door 1 (Goat) 선택
- 진행자: Door 2(Car), Door 3(Goat) 중 **염소가 있는 문(Door 3)을 반드시 열어야 함**
- 당신: 남은 문(Door 2)으로 **바꾼다.** → **승리!**

요약:

- 처음 선택을 '유지'하면, 처음에 자동차를 골랐을 때(1/3)만 이깁니다.
- 처음 선택을 '변경'하면, 처음에 염소를 골랐을 때(2/3) 항상 이깁니다.

진행자의 행동(P(염소 문 열기|내 선택))은 새로운 정보를 제공하며, 이 정보가 조건부 확률을 변경하여 바꾸는 것의 승률을 2/3로 높여줍니다. (100개 문으로 확장하면, 99/100의 확률로 이길 수 있습니다.)