BAYESIAN STATISTICS

지금까지 배운 통계

베이즈 정리

베이즈 통계

활용

그래서 이번 학기에는?

베이즈 통계란?

이규민

Thomas Bayes
1701-1761

통계학입문, 통계방법론, 수리통계학, ……
배운 내용은?

"… made N repeated performances of the random experiment. Then we can count the number f of times (the frequency) that the event C actually occurred throughout the N performances. The ratio f /N is called the **relative frequency** of the event C in these N experiments. …"(p. 2, Hogg)

"The preceding interpretation of probability is sometimes referred to as the ***relative frequency approach***, and it obviously depends upon the fact that an experiment can be repeated under essentially identical conditions. However, many persons extend probability to other situations by treating it as a ***rational measure of belief***. …… However, since the mathematical properties of probability given in Section 1.3 are consistent with either of these interpretations, the subsequent mathematical development does not depend upon which approach is used." (p. 2, Hogg)

통계학입문, 통계방법론, 수리통계학, ……
배운 내용은?

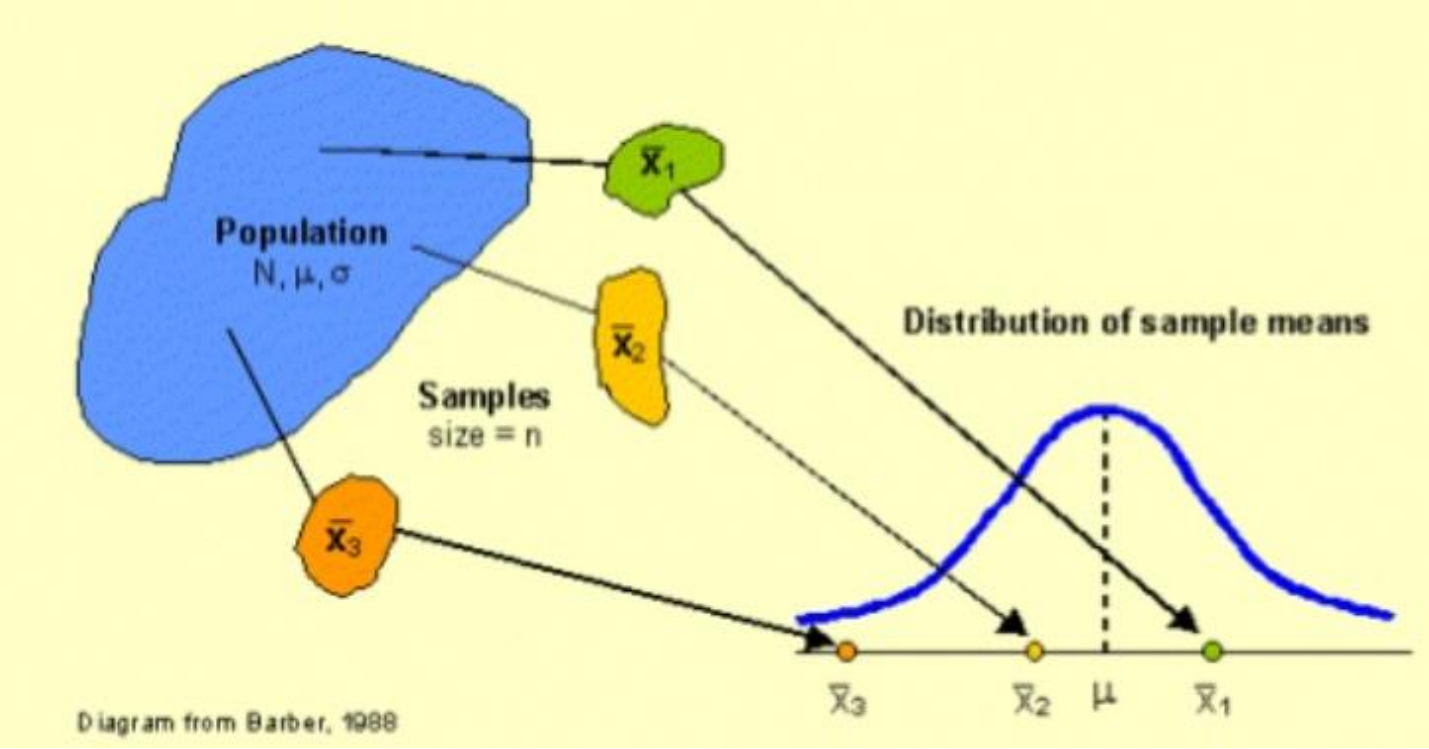
Definition 1.3.1 (Probability). *Let $\mathcal{C}$ be a sample space and let $\mathcal{B}$ be the set of events. Let P be a real-valued function defined on $\mathcal{B}$. Then P is a* **probability set function** *if P satisfies the following three conditions:*

1. $P(C) \geq 0$, *for all* $C \in \mathcal{B}$.
2. $P(\mathcal{C}) = 1$.
3. *If* $\{C_n\}$ *is a sequence of events in* $\mathcal{B}$ *and* $C_m \cap C_n = \phi$ *for all* $m \neq n$, *then*

$$P\left(\bigcup_{n=1}^{\infty} C_n\right) = \sum_{n=1}^{\infty} P(C_n).$$

통계학입문, 통계방법론, 수리통계학, ... 배운 내용은?

Optimization (mle, etc) → error bound (asymptotic dist., bootstrap, etc)



통계학입문, 통계방법론, 수리통계학, ... 배운 내용은?

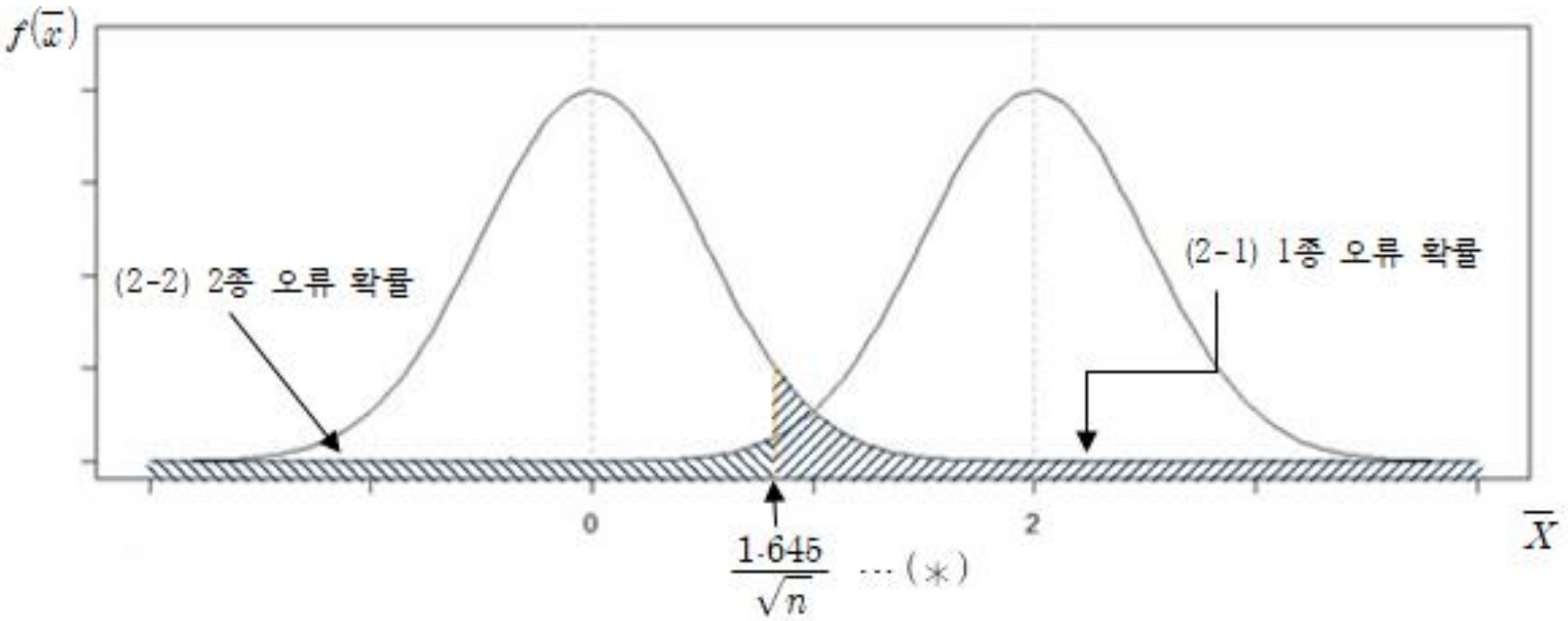
Confidence interval

“Once the sample is drawn, the realized value of the confidence interval is (l, u), an interval of real numbers. Either the interval (l, u) traps θ or it does not. One way of thinking of a confidence interval is in terms of Bernoulli trials with probability of success 1 − α. If one makes, say, M independent confidence intervals over a period of time, then one would expect to have (1 − α)M successful confidence intervals (those that trap θ) over this period of time. Hence one feels (1 − α)100% confident that the true value of θ lies in the interval (l, u).” (p.215, Hogg)

"I am not at all sure that the 'confidence' [in confidence interval] is not a 'confidence trick.’” (Bowley, to Neyman)

통계학입문, 통계방법론, 수리통계학, ……
배운 내용은?

H0: θ=0 vs H1: θ=2



베이즈 정리

$$\begin{aligned} P(C) &= P(C_1)P(C|C_1) + P(C_2)P(C|C_2) + \cdots + P(C_k)P(C|C_k) \\ &= \sum_{i=1}^{k} P(C_i)P(C|C_i). \end{aligned}$$

→ Law of total probability

$$P(C_j|C) = \frac{P(C \cap C_j)}{P(C)} = \frac{P(C_j)P(C|C_j)}{\sum_{i=1}^{k} P(C_i)P(C|C_i)}$$

→ Bayes' theorem

그래서 베이지안 추론은 뭘까?

Likelihood prior

$$p(\theta|X) = \frac{p(X|\theta)p(\theta)}{p(X)} = \frac{p(X|\theta)p(\theta)}{\int p(X|\theta)p(\theta)d\theta} \propto p(X|\theta)p(\theta)$$

posterior

그럼 predict 는?
→ Posterior predictive distribution 으로!

$$p(\tilde{X}|X) = \int p(\tilde{X}|\theta, X)p(\theta|X)d\theta$$

이렇게 좋은데 왜 안 썼을까?

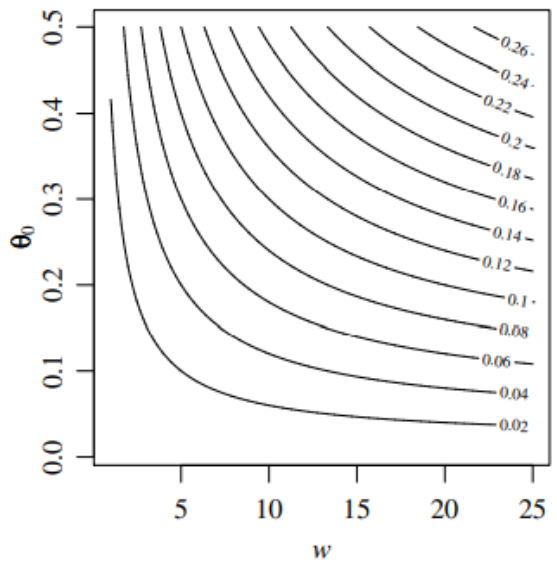
적당한 prior …?

Uninformative 일 때 어떻게 줄까?

→ Jeffrey's prior: $p(\theta) \propto \sqrt{I(\theta)}$

→ Reference prior

→ 결국은 Sensitivity Analysis!



Normalizing constant

분모에 있는 적분이 너무 어렵다!

→Conjugacy prior!
Ex) data: binomial, prior: beta → posterior: beta!!
→MCMC(Markov Chain Monte Carlo)

→Variational inference

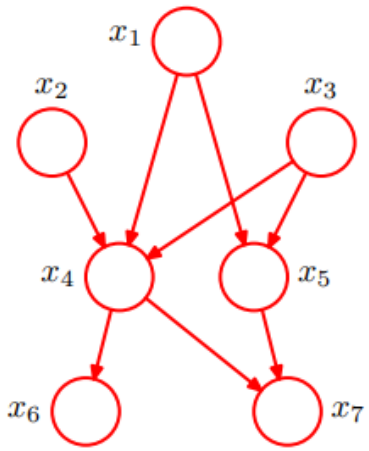
→Normal approximation

이걸 잘 하면……

“Naïve Bayes Classifier”, “Bayesian Neural Nets”, … (The Elements of Statistical Learning)

“Bayesian Linear Regression”, “Bayesian Model Selection”, “Bayesian Network”, “Bayesian PCA”, …
(Pattern Recognition and Machine Learning)

Figure 8.2 Example of a directed acyclic graph describing the joint distribution over variables $x_1, \ldots, x_7$. The corresponding decomposition of the joint distribution is given by (8.4).



is therefore given by

$$p(x_1)p(x_2)p(x_3)p(x_4|x_1, x_2, x_3)p(x_5|x_1, x_3)p(x_6|x_4)p(x_7|x_4, x_5). \qquad (8.4)$$

대선배님들의 격려

"… the only good statistics is Bayesian statistics. Bayesian statistics is not just another technique to be added to our repertoire alongside, for example, multivariate analysis; it is the only method that can ***produce sound inferences and decisions*** *in multivariate, or any other branch of, statistics. It is not just another chapter to add to that elementary text you are writing; it is that text. It follows that the unique direction for mathematical statistics must be along the Bayesian roads."*
– D.V.Lindley ("The Future of Statistics – A Bayesian 21st Century", 1975, Advances in Applied Probability)

"… The bottom line is that we have entered ***an era of massive scientific data collection****, with a demand for answers to large-scale inference problems that lie beyond the scope of classical statistics. In the struggle to find these answers, the statistics profession needs to use* ***both frequenstist and Bayesian ideas****, as well as new combinations of the two. Moreover, I think this is already beginning to happen. . . which was the real point of my examples. …"*
– B.Effron ("Bayesian, Frequentists, and Scientists", 2005, JASA)

지금까지 배운 통계

베이즈 정리

베이즈 통계

활용

그래서 이번 학기에는?

2021 spring curriculum

교재 : *A First Course in Bayesian Statistical Methods*

날짜	메인 세션 (목)	데이터분석 세션 (토)
1 주차 (3/4)	Likelihood and Prior (FCB Chapter 1, 2)	
2 주차 (3/11, 3/13)	One-parameter and Normal model (FCB Chapter 3)	EDA
3 주차 (3/18)	One-parameter and Normal model (FCB Chapter 4)	
4 주차 (3/25, 3/27)	One-parameter and Normal model (FCB Chapter 5)	Modeling I
5 주차 (4/1)	Multivariate Normal model (FCB Chapter 7)	
6 주차 (4/8, 4/10)	Bayesian Hierarchical Models (FCB Chapter 8)	Modeling II
중간고사 휴식기간		
7 주차 (5/6)	MCMC and Diagnosis (FCB Chapter 6, 10)	TBD
8 주차 (5/13)	MCMC and Diagnosis (FCB Chapter 6, 10)	
9 주차 (5/20)	Bayesian Linear Regression (FCB Chapter 9)	TBD
10 주차 (5/27)	Bayesian Linear Regression (FCB Chapter 9)Ⅱ	
11 주차 (6/3)	Final Project	TBD
12 주차 (6/10)	Final Project	
종강 및 기말고사		

감사합니다