

품질 및 신뢰성 기법에서 연역 및 귀납 추론에
의한 Conjugate 분포의 적용
Application of Conjugate Distribution using
Deductive and Inductive Reasoning in
Quality and Reliability Tools

최 성 운*

Sung-woon Choi*

Abstract

The paper proposes the guidelines of application and interpretation for quality and reliability methodologies using deductive or inductive reasoning. The research also reviews Bayesian quality and reliability tools by deductive prior function and inductive posterior function.

Keywords: Deductive, Inductive, Conjugate, Bayesian, Prior, Likelihood, Posterior, Quality and Reliability Methodologies

1. 서 론

최근 품질 및 경영혁신 운동에서 6σ 가 전 분야에 걸쳐 활발히 진행되어 큰 효과를 보고 있다. 6σ 에서는 최고경영자의 비전과 CTQ(Critical To Quality)에 의해 도출된 프로젝트를 다양한 기능의 부서팀(CFT : Cross Functional Team)이 공동으로 수행한다. 이 경우 부서간 공통 언어의 수단으로 통계적 분석방법을 사용하고 있으며 특히 MINITAB의 통계패키지를 활용하고 있다.

모집단(Lot, Batch)의 고정된 모수(Parameter)를 미지의 상수인 분포(모형, 이론, 가능성, 정보, 진실, 가설)로 가정하고 랜덤샘플링(Random Sampling)한 샘플(표본, 시료)의 통계량(Statistics)인 증거(사실, 현상)만으로 귀납적 추론(Inductive Reasoning)하는 고전적인 CSA(Confirmatory Statistical Analysis)방법이 있다.

* 경원대학교 산업공학과

그러나 CSA는 고정된 모수의 연역적 가정(Deductive Assumption)하에서 데이터의 요약 정리된 통계량이 이상점(Outlier)과 영향점(Influential Point)에 의해 성능에 크게 차이가 날 수 있다. 즉 샘플링 오차인 유의수준(Significance Level) α 보다 유의확률(Significance Probability) P-Value가 작아 통계적으로 유의적인 결과로 판정된다하더라도 실제 기술적, 실무적 검토를 할 경우 의미가 없을 수 있다. 또한 P-Value에 의한 Fisher 검정과 α 에 의한 Neyman-Pearson 검정에서의 유의성 판정에 따라 모집단 추론결과의 해석이 달라질 수도 있다.[8] 이렇듯 고전적 CSA방법은 분포인 모형의 모수를 추론하는 경우 MLE(Maximum Likelihood Estimator)를 사용하나 수리통계적 유도방법과 추론의 성질에서 UMVE(Unbiased Minimum Variance Estimator)를 만족하지 못하거나 비저항성의 기술통계량에 의해 오도된 모집단 추론판정의 문제가 있을 수 있다.

따라서 미지의(Unknown) 고정상수 모수로 가정된 상황하에서의 통계량만으로 계산된 우도함수(Likelihood Function)를 사용하는 CSA의 단점을 극복하는 방법으로 베이저안 통계적 분석(BSA : Bayesian Statistical Analysis) 방법이 있다. 이 방법은 모집단의 모수에 대한 사전(Prior) 분포함수를 알고 있다고(Known) 연역적으로 가정하고 연역적 추출에 의한 우도(Likelihood) 분포함수와 Joint Normalize된 (Total, Marginal, Expectation) 사후(Posterior) 분포함수로 모수를 귀납적으로 추론하는 방법이다.

이 방법은 알려진 사전분포함수의 확률변수(Random Variable)에 의해 모집단에 관한 정보량이 증대되므로 우도분포함수만을 이용하는 고전적 CSA방법보다 직관적인 해석이 가능하고 우수한 모집단에 대한 사후분포함수를 구할 수 있는 장점이 있다. 그러나 BSA는 사실(Fact)적인 사전분포함수가 감정(Feeling)에 의해 신념의 믿음정도(Degree of Belief)에 영향을 받을 수 있다는 것과 주관적(Subjective) 결정방법의 합리성에 이의를 제기할 수 있다. 또한 사전분포함수와 우도분포함수의 공액분포 (Conjugate Distribution)의 성질을 이용할 경우 사후분포함수의 계산은 간단히 이루어질 수 있으나 그 이외의 경우 계산의 복잡성에 따른 어려움이 있다. 최근 BSA의 효율적 계산을 위해 고속저가 컴퓨터를 이용한 MCMC(Markov Chain Monte Carlo)방법의 연구가 활발히 진행되고 있다. BSA의 기존 연구로는 베이지 추론[1,3,7]과 고찰 논문[5,6]이 있으며 응용분야로 R과 WinBugs 적용[2], 경제경영 적용[4], 실험계획법 적용[9], 프로세스 모니터링 및 제어 적용[10], 생존분석 적용[11,16], 신뢰성분석 적용[12,14,15], 웨이브렛 적용[13] 등이 있다.

따라서 본 연구에서는 CSA와 BSA의 사전, 우도, 사후분포함수에서 사용되는 연역추론과 귀납추론의 차이를 다양한 기법과 이론으로 유형화한다. 또한 식스 시그마 혁신 프로세스에서 적용되는 주요 품질 및 신뢰성 모형에 대해 실무자가 계산이 간단하고 다루기가 용이한 Conjugate 분포를 이용한 베이저안 분석방법을 고찰해 보기로 한다.

2. 품질 및 신뢰성에서 연역추론과 귀납추론의 비교

베이저안 분석은 모집단 모수의 사전분포함수를 연역가정하고 우도분포함수를 연역추론하여 사후분포함수로 모수를 귀납추론하는 복합추론의 방법이다. 따라서 연역추론(Deductive Reasoning)과 귀납추론(Inductive Reasoning)의 차이를 용어, 적용분야 및 기법으로 유형화하면 <표1>과 같다.

<표1> 연역추론과 귀납추론의 차이

<표1.1>	연역추론관점	귀납추론관점
용어	Deductive, Normative, Modeling, Algorithmic, Optimal, Prescriptive, Predictive,	Inductive, Exploratory Heuristics, Trial-and-Error, Ad-Hoc, Hunch, Experience(Expert, Rule), Inference Intuition, Descriptive Observation,
원리	고전물리학, 서양철학, 기독교(Analysis, Certainty, Deterministic)	현대물리학, 동양철학, 불교(Synthesis, Uncertainty, Stochastic, Probabilistic)
특성요인도(Fish Bone Diagram)	특성, 결과(Effect, Consequence, Fish)	요인, 원인(Cause, Bone) 1차요인 2차요인 3차요인
Why-Why	Why : 1차요인	Why : 2차요인
Top-Down	Top	Down
Bottom-Up	Up	Bottom
안전원리	Fail-Safe	Fool-Proof
문장	두괄식	미괄식

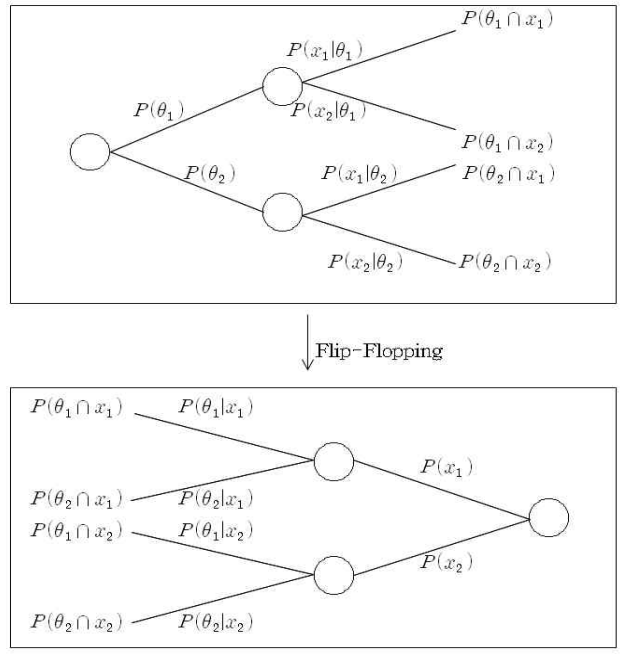
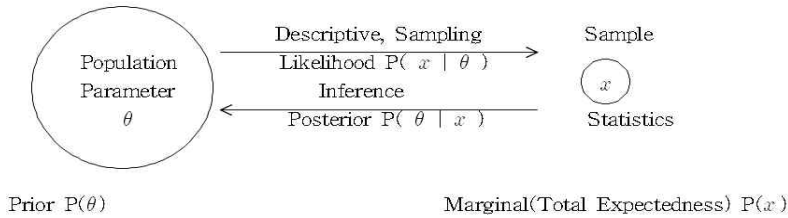
<표1.2>	연역추론관점	귀납추론관점
IF - Then	Then	If
Given - Find	Find	Given
What - If	What	If
안전기법	FTA	FMEA, ETA
나무와 숲	숲(Breadth)	나무(Depth)
수렴과 발산	수렴(Convergence, Zoom-In) ⊙	발산(Divergence, Zoom-Out) ⊖
Knowing - Doing	Knowing	Doing
Premise - Example	Premise(Conclusion, Summary)	Example
정보와 데이터	정보(Information, Truth, Theory, Hypothesis)	데이터(Data, Fact, Evidence, Phenomenon)
전체와 부분	전체(Whole, General)	부분(Part, Specific, Special)

<표1.3>	연역추론관점	귀납추론관점
수학	필요조건(Necessary Condition) 시작 ⊙	충분조건(Sufficient Conditon) 시작 ⊖
Infimum 과 Supremum	Supremum(Least Upper Bound)	Infimum(Greatest Lower Bound)
통계	모집단(Population, Lot, Batch) 모수(Parameter) 분포(Distribution) 가설(Hypothesis)	표본, 시료(Sample) 통계량(Statistics) 데이터, 자료(Data)
베이지안	사전(Prior)확률분포함수 : $P(\theta)$ 우도(Likelihood)확률분포함수 : $P(x \theta)$	사후(Posterior)확률분포함수 : $P(\theta x)$

3. Conjugate 분포를 이용한 베이저안 품질추론

베이저안 분석에서는 사전(Prior) 확률분포함수를 신념의 정도(Degree of Belief)를 주관적 확률(Subjective Probability)로 표시할 경우 Fact인데 Feeling의 영향을 감소하기 위한 방법으로 무정보적(Noninformative, Vague, Flat, Diffused, Let the data speak for themselves), 일양분포(Uniform Distribution)를 사용한다. Laplace에 의한 부적절한(Improper) 사전분포는 pdf아래 면적이 무한대이므로 모수 변환시 불변성(Invariance)이 결여되어 Jeffrey는 Fisher 정보량의 사전분포 방법을 제시하였다.[1]

<표2> 베이저안 Decision Tree



사전(Prior) 확률분포함수에 대한 주관성의 의미부여(Encoding), 합의(Consensus), 선택>Selecting)이 이루어 질 경우 지수족(Exponential Family)의 Conjugate 분포를 활용하면 효율적인 계산이 가능하다. 즉 Posterior = Prior × Likelihood 에서 Conjugate Prior 분포족(Distribution Family)을 선정하게 되면 우도함수가 Neyman-Pearson

Factorization Product에 의해 충분통계량(Sufficient Statistics)으로 표현되어 차원감소(Dimension Reduction), 분석적 조작성(Analytic Tractability), 유연하고 풍부하고 용이한 해석(Flexible, Rich and Easy Interpretation)이 가능하다.

베이지안 의사결정에서 주어진 사전(Prior)확률분포함수 $P(\theta)$ 와 우도(Likelihood) 확률분포함수 $P(x|\theta)$ 에 의한 주변(Marginal, Total, Expectedress) 확률분포함수 $p(x)$ 와 사후(Posterior, 확률분포함수 $P(\theta|x)$)를 Decision Tree의 Flip-Flopping에 의해서 구하는 방법은 <표2>와 같다.

모불량률, 모결점률 (모고장률) 추정 등의 계수 이산형 품질 기법과 MTBF, 모평균, 모분산추정 등의 계량 연속형 품질기법에 대한 Prior, Likelihood, Posterior의 Conjugate 분포관계는 <표3>과 같다.[1]

<표3>에서 Likelihood의 모수(Parameter)는 Prior의 확률변수(Random Variable)가 되고 Prior의 모수는 사전에 인지된 것으로 가정한다. 따라서 사용자는 Likelihood와 지수를 중심으로 같은 분포족(Distribution Family)의 Conjugate Prior 분포를 선정한다.

<표3> Conjugate 분포의 품질 적용

Purpose	Prior	Likelihood	Posterior
1. 계수이산형			
1.1 모불량률 추정	베타분포	이항, 음이항 분포	베타분포
1.2 모결점률추정 (모고장률)	감마분포	포아송분포	감마분포
2. 계량연속형			
2.1 MTBF 추정	감마분포	지수분포	감마분포
2.2 모평균 추정 σ Known	정규분포	정규분포	정규분포
2.3 모분산 추정 μ Known	역감마분포	정규분포	역감마분포
	감마분포	정규분포	감마분포
2.4 모평균, 모분산 동시추정	감마분포	정규분포	감마분포
	감마분포	정규분포	정규분포

<표3>에서 실무자가 고전적 모불량률 추정 시 Likelihood에 의한 이항분포만을 사용하기 때문에 $\hat{P} = x/n$ 를 쉽게 구할 수 있다. 그러나 Conjugate Prior 분포를 사용할 경우 베타분포를 통계패키지에 의한 적합도 검정에 의해 적용이 가능하더라도 해석에 어려움이 있다. 따라서 Conjugate Bayesian 분석에서는 분포의 모수에 대한 이해를 선행하지 않고 단지 모수의 정보량을 늘리거나 계산의 효율성을 위해 도입될 경우 실무에서는 해석의 어려움으로 외면당할 수 있다.

모불량률 추정시 Conjugate Prior 분포로 Beta(α, β)를 이용할 경우 $P = (\alpha + x) / (\alpha + \beta + n)$ 으로 $0 < t < 1$ 의 Bound Binding 구간에서 $\alpha + \beta$ 개중 α 개가 고장나는 것으로 해석되며 $\alpha = \beta$ 인 경우 좌우대칭, $\alpha > \beta$ 인 경우 Short Tail, Left Skew 분포, $\alpha < \beta$ 인 경우 Long Tail, Right Skew 분포로 적용된다. Left Skew는 직

릴 신뢰도 모형처럼 가장 신뢰도가 작게 연결된 (Weakest Link) 부품에 적용되며 Right Skew는 최대충격을 견디는 신뢰도 모형에 적용된다.

또한 모고장률 추정시 $\text{Gamma}(\alpha, \beta)$ (α : 형상모수, β : 척도모수)의 Conjugate 사전 분포의 의미는 α 개가 $1/\beta$ 시간 생존하는 확률로 해석된다.

4. Conjugate 분포를 이용한 베이저안 신뢰성 추론

계수 이산형 샘플링(Attribute Sampling)과 계량 연속형 샘플링(Continuous Sampling)의 Likelihood 분포에 의한 Conjugate Prior, Posterior 분포를 신뢰성 기법에 적용하면 <표4>와 같다.[12]

<표4> Conjugate 분포의 신뢰성 적용

<표4.1> 계수 이산형 우도함수

Purpose	Likelihood	Prior	Posterior
$R(t)$	1. 계수이산형 1.1 이항 샘플링	일양분포	베타분포
		Noninformative and Vague 분포	베타분포
		베타분포	베타분포
$R(t)$	1.2 Pascal 샘플링	베타분포	베타분포
$\lambda(t)$	1.3 Poisson 샘플링	Noninformative 분포	χ^2 분포
		감마분포	감마분포
$R(t)$		감마분포	Negative Log 감마분포

<표4.2> 계량 연속형 우도함수

Likelihood	Purpose	Prior	Posterior
2. 계량연속형 지수샘플링	$\lambda(t)$	Noninformative분포	감마분포
		감마분포	감마분포
	MTTF	Noninformative분포	역감마분포
		역감마분포	역감마분포
	$R(t)$	일양분포	Negative Log 감마분포
		Noninformative분포	Negative Log 감마분포

<표4.1>의 Likelihood 샘플링에서 이항샘플링은 n 개에서 x 개의 고장을 알아보는 경우, Pascal 샘플링은 x 번째 고장이 일어나는 n 개의 샘플을 파악하는 경우 적용된다. <표4.1>의 Poisson 샘플링과 <표4.2>의 지수 샘플링은 역수 관계로 Frequency Domain과 Time Domain의 관계와 같으며 전자의 경우 1시간에 몇 개 고장나는 고장

를 λ 를 후자의 경우 1개가 몇 시간 생존하는 평균고장시간 θ 를 구하는 경우 적용되며 $\theta = 1/\lambda$ 가 된다.

5. 결 론

본 연구에서는 품질 및 신뢰성기법에서 사용되는 연역적 방법과 귀납적 방법을 비교하여 실무자에게 기법의 해석 및 적용을 용이하게 하였다. 또한 연역과 귀납의 복합 추론을 이용하는 Conjugate 베이지안 분포를 이용하며 품질 및 신뢰성 기법에서 적용되는 사전, 우도, 사후 분포간의 관계를 고찰하고 해석방법을 제시하였다.

6. 참 고 문 헌

- [1] 강기훈 외, 베이지안 통계학, 자유아카데미, 2005.
- [2] 김달호, R과 WinBUGS를 이용한 베이지안 통계학, 자유아카데미, 2005.
- [3] 김병휘 외, 베이지안 통계계산, 자유아카데미, 2001.
- [4] 김용대 외, 경제·경영을 위한 베이지안 통계학, 자유아카데미, 2000.
- [5] 김용대 외, “베이지안 통계학의 과거·현재·미래,” 한국통계학회논문집, 8(2001): 47-64.
- [6] 장도석 외, “베이지안 통계학 연구에 대한 소고,” 정보과학연구, (9)(2005) : 133-149.
- [7] 장인식, 베이지안추론, 고려대학교 출판부, 2001.
- [8] 최성운, “샘플링 오차에 의한 품질통계모형의 해석,” 대한안전경영과학회지, 10(2)(2008) : 205-210.
- [9] Chaloner K., Verdinelli I., “Bayesian Experimental Design : A Review,” Statistical Science, 10(3)(1995): 273-304.
- [10] Colosimo B.M., Castillo E.D., Bayesian Process Monitoring, Control and Optimization, Chapman & Hall/CRC, 2007.
- [11] Ibrahim J.G., Chen M., Sinha D., Bayesian Survival Analysis, Springer-Verlag, 2001.
- [12] Martz H.F., Waller R.A., Bayesian Reliability Analysis, John Wiley & Sons, 1982.
- [13] Muller P., Vidakovic B., Bayesian Inference in Wavelet-Based Models, Springer-Verlag, 1999.
- [14] Singpurwalla N.D., Reliability and Risk : A Bayesian Perspective, John Wiley & Sons, 2006.
- [15] Sander P., Badoux R., Bayesian Methods in Reliability, Kluwer Academic Publishers, 1991.
- [16] Woodworth G.G., Biostatistics : A Bayesian Introduction, John Wiley & Sons, 2004.