

심리학 연구를 위한 실용적인 베이지언 추론의 소개

고 성 룡¹⁾ 주 혜 리²⁾ 오 성 주^{*1)}

¹⁾서울대학교 심리학과

²⁾서울대학교 인지과정협동과정

베이지언 처리는 근래 심리학에서 유용한 도구로 자리 잡아 가고 있다. 이러한 흐름은 심리학에서 인지처리를 보는 관점의 변화를 일으키고 있으며 영가설 검증에 기반을 둔 전통적인 통계 처리 방식과 경쟁하며 뿌리를 내리고 있다. 이 논문에서 통계 처리에 관한 두 관점인 빈도 주의와 베이지언 주의의 논쟁에서 벗어나 베이지언 처리 방식을 하나의 실용적인 처리 도구라는 관점에서, 심리학 연구의 주된 측정치인 반응비율과 반응시간을 베이지언 추론에서 어떻게 분석하는지 소개하고자 한다.

주요어 : 베이지언 추론, 반응 비율, 반응 시간, 신호탐지이론, 고정 시간, 반응분포 추정

† 교신저자: 오성주, 서울대학교 사회과학대학 심리학과, (151-742) 서울시 관악구 관악로 1
Tel: 02-880-6430, E-mail: songjoo@snu.ac.kr

심리학에서 획득한 자료를 통계적으로 처리할 때, p-value에 기반한 영가설(null-hypothesis significance testing, NHST)을 검증하는 방법이 전통적으로 널리 쓰여 왔다. 그렇지만 이 방법으로 도출된 결과는 해석상의 문제가 많아 이를 개정해야 한다는 움직임이 꾸준히 제기되었다. 최근 들어 이러한 흐름은 더욱 거세지고 있는데, 실제로 심리학 저명 학술지인 *Psychological Science*는 논문 투고에 대한 지침서에서 2014년 1월부터 투고되는 논문에 대해 p-value 사용을 자제하도록 권고하고 있다. 그 대신, 새로운 방식으로 통계 결과를 보고할 것을 의무화하고 있는데, 여기에는 효과크기(effect sizes), 신뢰구간(confidence interval), 표본 분포(sample distribution), 그리고 메타 분석(meta analysis)과 같은 자료 등이 포함된다(Cumming, 2014; 2012). 이러한 맥락에서 전통적인 통계 방법과 독립적으로 발달해 온 베이지언 추론에 대하여 재검토해 볼 필요가 있는데, 그 이유는 베이지언 추론이 전통적인 통계 방법이 안고 있는 문제점들을 해소할 수 있는 정교한 자료를 제공하므로 훌륭한 대안이 될 수 있기 때문이다. 그러나 베이지언 추론의 정수를 심리학 초심자가 이해하기에는 문턱이 매우 높은 것이 사실이다. 따라서 본 논문은 베이지언 추론에 대한 기본적인 개념을 알기 쉽게 소개하고, 이를 적용한 실제 예들을 제시함으로써 초심자가 베이지언 추론을 활용하는 데 실질적인 도움을 주고자 한다. 즉, 본 논문은 베이지언 추론을 수학적 논쟁의 대상이 아닌 실용적 도구로 보고 교육적인 관점에서 다루는데 그 목적이 있다.

베이지언 추론 배경에 대한 간단한 소개

실생활에서 베이지언 추론

베이지언 통계의 철학과 수학적 이해는 여러 문헌에서 찾아볼 수 있다(예, Gelman, Carlin, Stern & Rubin. 1995, 2003). 실생활에서 베이지언 추론이 어떻게 적용되는지를 이해하기 위해서 먼저 간단한 예를 소개하고자 한다. 미국을 가본 적이 없었던 어떤 사람이 미국인은 키가 크다는 믿음을 가지고 살고 있었다. 어느 날 이 사람이 미국 출장을 가서 어떤 레스토랑에서 이탈리아계의 키가 작은 고객 한 사람을 만났다. 그렇다면 이 사람이 믿음은 어떻게 될까? 아마도 이 사람은 미국 사람들도 다 크지는 않다고 생각이 바뀌게 될 것이다.

베이지언 추론의 수학적 접근

앞에서 살펴 본 일상에서 경험하는 베이지언 추론을 좀 더 객관적으로 기술하기 위해서는 수학으로 옮길 필요가 있으며 그 밑바탕이 되는 것이 확률이다. 즉, 베이지언 추론에서는 우리의 믿음과 불확실성을 확률로 표현하고, 믿음의 변화를 확률 패턴의 변화로 그려낸다. 이 과정에서 핵심은 변화된 확률 분포를 구하는 것인데, 이를 구할 수 있게 하는 식이 조건 확률 사이의 관계를 나타내는 베이지언 정리이다(1).

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)} \quad (1)$$

위의 베이시언 정리에서, P는 확률 분포를 나타내고, H는 생각이나 가설, D는 자료나 경험을 나타낸다. 이 정리에서 좌변은 경험한 뒤, 즉 자료를 얻고 난 뒤의 생각이나 가설에 대한 사후 확률 분포(posterior)를 나타낸다. 우변은 조건 확률의 공식에 따라 다시 풀어쓴 것인데, 사후 확률은 가설이 있을 때 자료를 얻을 확률 분포(우도, likelihood)와 가설이나 생각을 확률로 나타낸 사전 분포(prior)의 곱을 자료의 확률로 나눈 분포로 나타낼 수 있다는 것이다. 가설 대신 모형이라 기술되는 경우도 많으며, 이때는 H 대신 M(model) 기호를 쓴다. 본 논문에서도 H와 M을 구별 없이 사용하였다.

앞서 언급한 예에 대하여 베이시언 추론을 하려면, 먼저 키를 기술하는 표집 분포(sampling distribution) 혹은 우도를 가정해야 한다. 키의 분포는 정규분포로 가정함이 적절할 것이다. 즉, $P(D|H)$ 는 $\text{Normal}(\mu, \text{variance})$ 로 기술한다. 편의상 분산이 10인 정규분포($\text{Normal}(\mu, 10)$)라고 가정하자. 그 다음은 μ 가 어떤 분포에서 나왔는지, 즉 사전 확률분포를 가정해야 한다. 사전 확률분포에 대해서는 논란이 많은데, 본 논문에서는 정보가 없는 사전 확률분포를 주로 썼으나 때로 켈레(conjugate) 분포를 사전분포로 사용하였다. 특정 표집분포에 대해 사전분포로 가정된 분포가 곧 사후분포가 되는 경우에 이 사전분포를 켈레 분포라고 한다. 예를 들어, 표집 분포가 이항분포일 경우 사전 확률분포로 베타분포를 가정하면 사후분포도 이 베타분포가 된다. 즉, 베타분포의 두 파라미터 값만 달라진다. 아래에 소개되는 반응비율 처리에서 베타분포가 사전 확률분포로 쓰인 예들을 보게 될 것이다.

베이시언 정리의 활용

키의 분포를 정규분포로 가정하면, 정규분포의 두 파라미터인 평균과 분산의 사전 확률분포를 가정해야 한다. 위에서 분산은 고정되었다고 가정했기 때문에 평균에 대한 사전 확률분포만 가정하면 된다. 위에서 소개한 믿음은 미국인은 키가 크다는 것인데, 이를 이 평균 파라미터로 나타낼 수 있다. 사전 확률분포로 분산이 고정된 경우 정규분포의 켈레 분포인 정규분포를 가정해본다. 구체적으로 사전 확률분포는 평균이 180cm이고 분산은 10cm인 정규분포($\text{Normal}(180, 10)$)이고, 키가 작은 이태리계 남자가 160cm라고 하자. 이 경우에 평균의 사후분포는 위에서 말한 대로 정규분포가 되는데, 알려진 사후 분포의 평균과 분산을 유도한 식을 따르면(Gelman, Carlin, Stern & Rubin, 1995; 2003 참조), 평균은 사전 분포의 평균(180)과 그 분산의 역수인 정밀도(precision)(1/10)를 곱한 값에 자료 값(160)과 표집분포의 정밀도(1/10)를 곱한 값을 더해 그 값을 사전분포의 정밀도와 표집분포의 정밀도로 나눈 값이 된다. 즉, 식 (2)로 정리되어 사후분포의 평균은 170이 된다.

$$\frac{180*(1/10)+160*(1/10)}{1/10+1/10} \quad (2)$$

사후 확률분포의 정밀도는 사전분포의 정밀도(1/10)와 표집분포의 정밀도(1/10)를 더한 값인 2/10, 즉 1/5 이 되어서 분산은 5가 된다. 만약 키가 작은 이태리계 사람 두 명을 보았고 그들의 키가 각각 160cm와 170cm이었다면, 이 사후 분포의 평균은 사전 분포의 평균(180)과 그 분산의 역수인 정밀도(1/10)를 곱한 값

에 자료1의 값 (160)과 정밀도(1/10)를 곱한 값과 자료2의 값 (170)과 표집분포의 정밀도 (1/10)를 곱한 값을 더하고 이 값을 각 키 값의 가중치 구실을 한 정밀도를 다 더한 값으로 나눈 값은 식 (3)으로 정리되며 170이 된다.

$$\frac{180*(1/10)+160*(1/10)+170*(1/10)}{1/10+1/10+1/10} \quad (3)$$

이 사후분포의 정밀도는 사전분포의 정밀도 (1/10)와 두 자료의 표집분포 정밀도(2/10)를 더한 값인 3/10이 되고 그 역수인 분산은 3.33이 된다. 이 과정을 좀 쉽게 나타내면, 사후분포의 평균은 사전분포의 평균(180)과 정밀도(1/10)를 곱한 값에 자료들의 평균(165)과 그 자료들의 표집분포의 정밀도(2/10)를 곱한 값을 더하고 이 값을 사후분포의 정밀도(3/10)로 나눈 값은 식 (4)로 정리되어 값이 170이 된다.

$$\frac{180*(1/10)+165*(2/10)}{1/10+2/10} \quad (4)$$

베이저언 추론의 특징

이처럼 베이저언 추론은 그 바탕이 우리의 생각과 경험의 절충에 있으며, 경험의 자료가 많아질수록 사후분포는 더욱 그 자료를 반영하게 된다. 많은 경우 우리가 관심을 갖는 것은, 위의 예와 같은 한 집단의 특징이 아니라 집단 사이의 특징을 비교하는 것이다. 예를 들어, 10대의 남자와 여자의 키 차이에 관심이 있다고 하자. 어떤 사람은 십대 남자 아이와 여자 아이들의 키가 차이가 없을 것이라 생각하나(H1) 다른 사람은 차이가 있을 것이

라 생각할 수 있다(H2). 구체적으로 이 두 사람은 남자들과 여자들의 평균 키에서 차이가 있는지 알아보기로 했다. 그래서 10대 남자 15명과 여자 15명의 키를 측정하여 자료(D)를 얻었다. 이 자료를 바탕으로 두 가설에 대해 베이저언 추론을 하려면, 위의 예에서 했던 것처럼 자료를 기술할 수 있는 표집분포/우도와 사전 확률분포를 가정해야 한다. 표집분포를 정규분포라고 가정하면 이 표집분포의 두 파라미터인 평균과 분산에 대한 사전 확률분포를 가정해야 한다. 이 두 파라미터의 사전 확률분포로는 균등분포를 가정해 보자. 이런 가정을 하면, 평균 파라미터의 사후분포는 t분포를 띄고 분산의 파라미터는 역-감마(inverse-gamma)분포를 띄게 된다(Gelman et al., 2003). 감마분포는 N번 성공할 때 까지 걸리는 시간을 나타내는 연속무선변인의 확률밀도 분포로 잘 알려진 분포인데, 역-감마분포는 이 감마분포를 띄는 무선변인 (X)에 역수 (1/Y)을 넣어 얻을 수 있는 분포이다. 이 분포들에서 직접 표집을 하여 파라미터의 사후분포를 추정하면 된다. 이와 비슷한 예는 아래 반응시간의 비교에서 살펴볼 것이다.

모형의 파라미터 추정은 그 수가 늘어나면 직접 계산이 불가능하거나, 가능한 경우 일지라도 쉽지 않기 때문에 주로 근사치를 구하는 방법에 의존하게 된다. 이에 가장 많이 사용하는 것이 마르코프-연쇄-몬테카를로(Markov chain Monte-Carlo) 방법이며, 이 방법의 도입으로 베이저언 추론이 다시금 관심을 끌게 되었다고 해도 과언이 아니다. 이 방법에 대해서는 여러 문헌에서 상세히 소개되어 있으므로 여기에서는 간단한 개념만 알아보도록 하겠다.

마르코프-연쇄-몬테카를로 방법

도박으로 유명한 모나코의 도시인 몬테카를로가 암시하듯이 이 방법은 무선표집(random sampling)에 바탕을 두어 계산하는 방법을 말한다. 예를 들어, 어떤 구간의 함수의 면적을 계산할 때 우리는 적분을 한다. 하지만 함수가 너무 복잡하면 수식으로 적분을 하는 것이 쉽지 않아 근사치를 계산하게 된다. 평균치 정리는 적분 구간 안에 평균값이 있음을 보장하고, 통계의 큰 수의 법칙은 독립적으로 뽑은 표본의 수가 늘어날수록 표본의 평균이 전 집의 평균에 다가간다는 것을 보장한다. 그래서 이 구간에서 무작위로 되도록 많은 수를 뽑고, 뽑은 수의 함수 값을 모아 평균을 계산한다. 이 평균과 적분 구간의 크기를 곱하면 구하려고 했던 면적의 근사치가 된다. 이것이 몬테카를로 방법의 간단한 예이다.

한 무선표본의 확률이 확률 밀도 함수의 면적으로 표현된다는 점을 알면 왜 이 방법이 중요한지 쉽게 이해할 수 있다. 몬테카를로 방법의 구현을 위해 수를 균등분포, 지수분포, 정규분포와 같은 분포에서 무작위로 생성할 수 있는 방법들이 개발되어 왔다. 보통 몬테카를로 방법에서 표본은 독립적으로 모아지는데, 이러한 표집과는 달리 현재 상태의 값에 기대어 다음 값을 뽑는 방법이 마르코프 연쇄 방법이다. 이 마르코프 연쇄방식의 특징은 한 상태에서 가질 수 있는 모든 다음 상태의 값들의 확률이 정해진다는 점이다. 때문에 마르코프 연쇄에서 상태들의 이행이 주기를 띠지 않고 한 상태에서 모든 다른 상태로 갈 수 있게 관계를 갖는다면 이 연쇄 방법으로 모아진 값들은 어느 정도 표집을 하면 안정된 분포에 이르게 된다.

그런데 실제 문제에서 베이지언의 사후분포는 복잡한 형태가 되기 때문에 사후분포에서 직접 표집하기가 어렵다. 그래서 많이 쓰는 것이 메트로폴리스(Metropolis) 방법이다. 이 방법에서는 무작위로 값을 추천하는 추천 분포(proposing distribution)(예: 정규분포)를 정하고, 이 분포가 추천한 값이 지금의 값보다 사후분포에서 확률 값이 더 크면 표집하고 그렇지 않으면 상대적인 비를 계산하여 이 비가 균등분포에서 무작위로 뽑은 수보다 크면 표집하고 그렇지 않으면 추천된 값을 표집하고 않고 지금 값을 다시 표집하게 된다. 이런 방법을 비대칭의 추천 분포에까지 일반화한 것이 Metropolis-Hastings 방법이고 Gibbs 알고리즘은 이런 방법의 특수한 경우로 볼 수 있다. 아래에 소개하는 Bugs 프로그램도 직접 표집이 힘든 경우에 이 마르코프-연쇄-몬테카를로 방법을 써서 사후분포를 만든다. 이 구체적인 과정들은 아래 반응시간 처리 부분에서 살펴 보도록 하겠다.

베이지언 추론에서 선택과 평가

이 파라미터의 추정을 하고 나서 중요한 과정이 모형의 선택(selection) 혹은 평가(evaluation)이다. 전통적인 빈도주의 관점에서 두 가설 가운데 어느 가설이 적절한지를 가리는 한 방법은 신뢰구간(confidence interval)을 보는 것이다. 이와 비슷하게 베이지언 처리에서도 파라미터의 사후 확률분포에 대한 베이지언 신뢰구간(Bayesian credible interval)을 구해 이를 활용한다. 예를 들어 평균 차이를 비교한다면, 평균 차이 파라미터를 정하고 이 파라미터의 사후분포를 구해서 이 분포의 베이지언 신뢰구간이 0을 포함하는지를 확인한다. 구간이 0을

포함하지 않으면 두 집단의 평균 차이가 없다는 가설은 옳을 확률이 매우 낮음을 뜻하는 것이다. 베이저언 신뢰구간을 구하는 방법은 두 가지를 많이 쓰고 있다. 한 가지는 분위수(quantile)에 바탕을 둔 것으로, 구간이 95%라고 할 때 2.5%의 값과 97.5%의 값을 구간의 하한과 상한으로 정한다. 구체적인 예시는 아래에 다시 언급하겠다. 다른 한 가지 방법은 고클러스밀도(highest probability density)를 찾는 방법이다. 말 그대로 두 값 사이에 밀도가 가장 높은 두 값을 구간의 하한과 상한으로 정하는 방법이다. 주로 정해진 확률을 만족하는 두 값들을 찾아 이 가운데 가장 두 값의 범위가 가장 좁은 값을 하한과 상한으로 정하는 방법이 많이 사용된다. 사후 확률분포가 비대칭적일 때는 두 번째 방법이 보다 유용할 수 있다. 베이저언 신뢰구간의 장점은 전통적인 빈도주의의 신뢰구간이 모수치가 표집분포의 구간에 있을 가능성을 나타내는 것과는 달리 파라미터가 이 범위안의 값이 될 수 있는 확률을 나타내고 있다는 점이다. 즉, 베이저언 신뢰구간은 확률의 정의를 충실히 따름으로써 전통적인 빈도주의의 신뢰구간이 불러일으킬 수 있는 오해나 애매함을 피하고 있다.

모형을 선택하는 또 다른 과정은 전통적인 빈도주의의 가설검증 방법에 해당되는 베이스 요인(Bayes factor)이다. 베이스 요인은 각각의 가설이 자료를 설명하는 상대적인 비. 즉 $P(D/H1) / P(D/H2)$ 의 값인데, 이 값을 직접 계산하는 것은 많은 경우에 불가능하거나 가능한 경우에도 쉽지 않다. 그래서 근사치를 구해서 많이 쓰는데, 이 근사치를 구하는 일반적인 방법은 되돌아갈 수 있는 마르코프 연쇄 몬테카를로(reversible jump mcmc) 방법(Green, 1995)과 곱공간(product space) 방법

(Lodewyckx, Kim, Lee, Tuerlinckx, Kuppens & Wagenmakers, 2011)이 있다. 하지만 이 방식을 구현 가능한 손쉬운 도구가 아직 마련되지 않은 관계로 아래에서는 베이저언 신뢰구간을 주로 쓰되, 위계 베이저언 분석이 아닌 경우에 한하여 전통적인 분석을 한 뒤 이 정보를 바탕으로 베이스 요인을 구하는 방법을 간략히 살펴보겠다.

베이저언 추론 적용의 실제 예

지금부터는 베이저언 추론을 적용할 수 있는 시스템의 설치를 간략히 소개하고자 한다. 아래의 모든 프로그램은 R에서 Openbugs를 구동할 수 있는 스크립트이고, 결과는 이 스크립트를 R에서 실행한 것을 제시하였다.

Openbugs나 Winbugs 설치

위에서 말했듯이, 베이저언 추론에서 근사치를 구하는 방법 가운데 가장 많이 쓰이는 것이 마르코프-연쇄-몬테카를로(Markov chain Montecarlo) 방식이며, 그 가운데서도 깁스표집(Gibbs sampling)이 베이저언 추론의 부흥을 이끌었다고 볼 수 있다. 이 깁스표집을 바탕으로 베이저언 추론을 할 수 있도록 만든 프로그램이 Bugs(Bayesian inference using Gibbs sampling)이며, 이 Bugs를 윈도우 시스템에서 구현한 것이 winbugs(Lunn, Thomas, Best, & Spiegelhalter, 2000)이고 opensource 전통에서 제작한 프로그램이 openbugs이다. 이 프로그램들을 직접 작성하는 것보다는 R에서 구동하는 것이 편리하므로 아래 (5)와 같은 순서에 따라 R과 openbugs을 이어주는 BRugs 인터페이스를

Openbugs를 이용한 베이지언 추론을 위한 시스템 설치 (5)

1. R 3.01을 설치한다 (<http://cran.r-project.org/bin/windows/base/>)
2. openbugs322를 설치한다 (<http://www.openbugs.net/w/Downloads>)
3. R에서 BRugs 패키지를 설치한다 (R에서 `install.package` 이용)

R에 설치할 것을 권한다.

아래에서는 반응비율과 반응시간을 측정치로 한 구체적인 실험을 간략히 소개한 뒤, 이 실험 결과를 가지고 베이지언 추론을 하는 과정을 소개하고자 한다.

간단한 반응비율 처리

오성주(2014)는 빨강과 초록 대립이 전경 배경 조직화(*figure-ground organization*)에 영향을 미치는지 알아보기 위해 초록 바탕의 빨간 원과 빨강 바탕의 초록 원을 주고 어떤 원이 더 두드러진 도형, 즉 전경처럼 보이는지를 선택하는 과제를 실시하였다. 총 64명(남자 32, 여자 32)이 참가한 실험 결과, 48명(남자 20, 여자 28)이 초록 바탕의 빨간 원을 더 전경답다고 선택했다.

이 자료에 대한 베이지언 추론을 위해서는 먼저 표집분포(*sampling distribution*), 즉 우도에 대한 가정이 필요하다. 이 실험에서 한 피험자가 선택하는 것은 다른 피험자의 선택에 영향을 미치지 않았을 것이다. 그래서 한 피험자가 수행한 한 시행을 베르누이 시행으로 보고 이 피험자들이 수행한 전체 시행을 이항분포로 정리하는 것이 적절해 보인다. 이항분포에서 전체 시행 수는 실험 설계를 통해 결정되는데, 위 예의 경우 전체 피험자 수는 64명이고 빨간 원을 택하는 횟수를 성공 횟수로 보면 이는 48번이다. 이를 바탕으로 성공 확

률을 추정해 볼 수 있다.

베이지언 추론에서 중요한 또 하나의 분포는 사전 확률분포이다. 위 실험에서 파라미터는 한 번의 선택에서 빨간 원을 고를 확률인데, 이에 대한 확률분포가 가정되어야 한다. 이와 관련해서 저자들은 가정한 사전 확률과 사후 확률 분포가 같도록 이항분포의 쥘레(*conjugate*) 분포인 베타분포를 사전 확률 분포로 선택했다. 이 베타분포는 두 파라미터를 갖는데 사전 확률의 영향이 미미하도록 균등 분포와 동일하게 두 파라미터 값을 1로 했다. 이렇게 하여 정해진 위 실험에 대한 베이지언 확률 모형을 Winbugs/openbugs에서 적으면 그림 1과 같다.

그림 1에서 *dbin*은 이항분포를 뜻하고 \sim 는 *y*가 이항분포를 따른다는 것을 나타내며 *dbeta*는 베타분포를 뜻한다. 전체 Winbugs/openbugs 스크립트는 부록 1에 제시하였다. 이를 비롯하여 부록으로 첨부한 모든 스크립트는 크게 세 부분으로 구성되어 있다. 첫 번째는 모형을 설정하는 부분인데, 그림 1과 동일하다. 두 번째 부분은 자료를 기술하되, 성공 횟수가 48번이라는 것을 나타내고 있다. 세 번째

```
model {
  y ~ dbin(theta,64)
  theta ~ dbeta(1,1)
}
```

그림 1. 실험 반응비율의 모형설정

는 MCMC를 돌리는 부분으로, 모형을 컴파 일하고 마르코프 체인의 초기 값을 설정한 다. 이어서 총 11000번을 표집 하는데, 처음 1000번은 버려지고 (burn-in) 10000이 실행된 결과인(modelupdate(10000))를 가지고 theta의 사 후분포를 추정하였다. 이 과정에서 theta 파라 미터의 값을 저장하고(samplesSet(c("theta"))) 이 결과가 보이도록(samplesStats("**")) 하고 있다.

표 1은 위 스크립트를 구동한 결과이다. 첫 번째 열은 파라미터를, 두 번째 열은 파라미 터 사후분포의 평균을, 세 번째 열은 파라미 터 사후분포의 표준편차를, 네 번째는 MCMC 의 표집 수로 나눈 MCMC오차를, 다섯 번째 는 파라미터 사후분포의 2.5% 값을, 여섯 번째는 사후분포의 중앙값을, 일곱 번째는 분 포의 97.5%의 값을, 여덟 번째는 버린 표본의 수를, 그리고 아홉 번째는 사후분포의 표본 수를 나타낸다. 표 1에서 theta의 사후분포는 사전분포 beta(1,1)와 64번의 시행 중 48번 성공한 이항분포 자료를 반영한 beta(1+48, 1+16) 것이며, 이 분포에서 30,000번 표집한 결과를 보여주고 있다.

이 실험 결과가 우연적으로 발생한 것이 아 닌지 알아보는 간단한 방법은 위의 자료와

theta가 0.5인 사후확률 분포를 비교하는 것이 다. 우연의 확률로는 두 원 가운데 어떤 쪽으 로도 선호가 없다고 할 때, 한 원의 선택 비 율을 0.5로 보는 것이 합당하기 때문에 64명 가운데 32명이 빨간 원을 선택하는 결과의 이 상적인 자료를 가정할 수 있다. 그림 2에서 와 같이 우연에 바탕을 둔 성공확률 파라미 터(theta2)와 실험 자료에 바탕을 두고 추정한 성공확률 파라미터(theta1) 간의 차이(diff) 분 포를 얻고, 이 분포가 0을 포함하는지를 근거 로 실험결과가 우연인지 아닌지를 확인할 수 있다. 이 비교 모형에 대한 구체적인 Bugs 프 로그램은 부록 2에 제시하였다.

```

model {
  # Likelihood:
  y1 ~ dbin(theta1,64)
  y2 ~ dbin(theta2,64)
  # Prior distribution:
  theta1 ~ dbeta(1,1)
  theta2 ~ dbeta(1,1)
  diff <- theta1 - theta2
}

```

그림 2. 실험 반응비율 비교

표 1. 빨간 원 선택 시의 파라미터 값에 대한 베이지언 추정 결과

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
theta	0.7423	0.05373	0.0005561	0.6304	0.7453	0.8407	1001	10000

표 2. 빨간 원 선택과 무작위 선택의 비교 처리 결과

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
diff	0.2427	0.08119	0.0004239	0.08035	0.2435	0.3978	1001	30000
theta1	0.7427	0.05318	0.0002577	0.63250	0.7449	0.8394	1001	30000
theta2	0.5000	0.06104	0.0003147	0.37980	0.5000	0.6197	1001	30000

부록 2를 처리한 결과인 표 2를 보면, 실험 자료에 바탕을 둔 파라미터(theta1)의 평균은 0.742이고 베이저언 신뢰구간은 0.63-0.84이며, 가상적인 자료에 바탕을 둔 파라미터(theta2)의 평균은 0.5이고 베이저언 신뢰구간은 0.37-0.62이다. 그리고 이 두 파라미터 차이(diff)의 평균은 0.242이고 베이저언 신뢰구간은 0.080-0.397로 0을 포함하고 있지 않다. 이와 같은 결과는 빨간 원을 고르는 비율이 우연에 바탕을 두고 있다는 가설의 확률이 매우 낮음을 보여준다.

위의 자료에서 남녀 성차도 살펴볼 수 있는데, 여자는 32명 가운데 28명이 빨간 원을 골랐고 남자는 32명 가운데 20명이 골랐다. 성별에 따른 전경 선택에 차이가 있는지 알아보기 위한 모형은 그림 2와 거의 동일하며, 전체 스크립트는 부록 3에 제시하였다.

부록 3을 처리한 결과인 표 3을 보면, 여자가 빨간 원을 고른 확률 파라미터의 평균은

0.853이고 신뢰구간은 0.718-0.948이나, 남자는 파라미터의 평균이 0.617이고 신뢰구간은 0.449-0.77이다. 이 두 파라미터의 차이 평균은 0.236이고 베이저언 신뢰구간은 0.032-0.434이다. 신뢰구간에 0이 포함되지 않았으므로 여자와 남자가 빨간 원을 고른 비율에서 차이가 없다는 가설의 확률이 매우 낮음을 알 수 있으며, 그에 따라 남녀가 고른 빨간 원의 비율이 믿을만한 정도로(credibly) 다르다고 결론을 내릴 수 있다.

앞서 설명하였듯이 베이저언 신뢰구간은 분위수(quantile)를 써서 구하기도 하고 고확률밀도(HPD) 구간을 계산하는 방식으로 구하기도 한다. 위의 결과에서 여자들의 빨간 원 선택 파라미터의 평균이 0.853인 것으로 짐작컨대 사후분포가 편포되어 있을 가능성이 높아 사후확률분포를 살펴보았다(그림 3).

그림 3 중 첫 번째 그림에서 분포가 왼쪽으로 꼬리를 늘어뜨리고 있음을 볼 수 있으며,

표 3. 남녀 집단의 파라미터 추정 결과

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
diff	0.2360	0.10180	0.0005925	0.03247	0.2367	0.4346	1001	30000
theta1	0.8533	0.05977	0.0003641	0.71830	0.8599	0.9486	1001	30000
theta2	0.6172	0.08211	0.0004464	0.44930	0.6197	0.7707	1001	30000

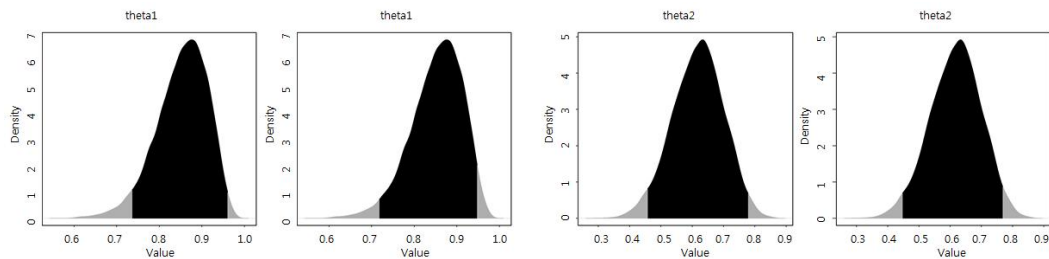


그림 3. 왼쪽부터 theta1의 분위수(quantile) 기반 신뢰구간, theta1의 고확률밀도(HPD) 기반의 신뢰구간, theta2의 분위수 기반 신뢰구간, theta2의 고확률밀도 기반의 신뢰구간을 각각 보여준다.

이와 같이 편포된 경우에는 분위수(quantile) 기반보다 고확률밀도 바탕의 구간이 더 유용할 수 있다. 고확률밀도 바탕의 구간을 계산하기 위해 일단 이 파라미터의 표본 값을 저장해야 한다(부록 3의 $a = \text{buildMCMC}(\text{"theta1"})$). 그리고 이를 바탕으로 95% 고확률밀도 구간(부록 3의 $\text{HPDinterval}(a, 0.95)$)을 구했는데, theta1 의 구간은 0.735-0.959이고 theta2 의 구간은 0.458-0.778로 분위수 기반으로 계산한 구간보다 폭이 조금 좁아졌다. 두 파라미터 차이의 신뢰구간 또한 0.0408-0.440으로 분위수 방법으로 계산한 구간보다 폭이 약간 좁아졌다. 이러한 결과로 미루어 파라미터의 사후분포가 비대칭일 경우 고확률밀도 방식으로 베이저언 신뢰구간을 구한 것이 더 적절하였다고 할 수 있다.

신호탐지이론의 분석

일부 심리학 연구는 반응비율을 좀 더 세분하여 어떤 경우에 어떤 반응을 하는지 분석하기도 한다. 이 같은 분석이 의미 있는 것은 일상생활의 경험에서 쉽게 알 수 있다. 때때로 밤 골목을 걷다가 나무의 그림자를 사람으로 여겨 빨리 걷는 경험을 한 적이 있을 것이다. 혹은 화재 신호인 연기가 있을 경우에만 울려야 하는 화재경보기가 가끔 그렇지 않을 때 울리기도 한다. 이렇듯 특정 경우에 어떤 반응을 했는지 알아볼 필요가 있는 중요한 문제에서, 분석의 바탕이 되는 이론이 신호탐지이론이다(Macmillan & Creelman, 2005; Wickens, 2002; Swets, Dawes, & Monahan, 2000; Swets, 1988). 이 이론에 따르면 우리는 자극/신호와 소음이 섞인 상태에서 반응을 하는데, 이 반응은 신호가 있을 때 신호가 있다고 반응(적

중, Hit), 신호가 있을 때 신호가 없다고 반응(누락, Miss), 신호가 없을 때 신호가 있다고 반응(오경보, False Alarm), 그리고 신호가 없을 때 신호가 없다고 반응(정기각, Correct Rejection)하는 네 유형으로 나뉜다. 이를 바탕으로 얼마나 신호를 잘 구별해내고 있는지를 뜻하는 민감도(d')와 반응을 할 때 어떤 편향이 존재하는지를 나타내는 반응편향(c)을 추정할 수 있다. 이러한 추정은 신호와 소음이 자극의 강도라는 하나의 차원에서 보통 정규 분포를 띄며, 두 분포의 분산은 같다는 가정에서 출발한다. 신호/자극과 소음의 구별은 신호 분포와 소음 분포 평균 사이의 거리(z 값)로 나타낸다. 보통 편향은 편향되지 않은 반응일 경우에 비해서 한쪽으로 쏠렸는지를 나타내는 지표를 쓴다.

$$d' = z(h) - z(f) \quad (6)$$

$$c = -\frac{z(h) + z(f)}{2} \quad (7)$$

식 (6)과 (7)에서 z 는 누적정규분포의 역함수를, f 는 오경보율을, h 는 적중률을 나타낸다. 민감도를 나타내는 (6)은 적중한 비율의 z 값에서 오경보율의 z 값을 뺀 둘 사이의 거리로 나타낼 수 있다. 예를 들어, 적중률이 95%이고 오경보율이 10%였다면 $z(.95)$ 는 1.645가 되고 $z(0.1)$ 은 -1.282가 되며 1.645에서 -1.282를 빼면 약 2.927이 되는데 이 값이 d' 이 된다. 그리고 반응 편향(c)은 (7)에 따라 -0.363이 되어 비편향 기준보다 약간 이완된 기준으로 반응하였음을 알 수 있다. 이 같은 민감도(d')와 반응 편향(c)의 베이저언 추정은 Lee(2008)에서 찾을 수 있다. Lee(2008)의 기본 가정은 민감도와 반

응편향이 정규분포를 따른다는 것이었으며, 민감도와 반응편향이 주어졌을 때 적중률과 오경보율을 계산하고 이 값이 자료와 얼마나 맞는지를 추정하였다.

부록 4에 세 명의 가상 피험자 자료를 바탕으로 민감도와 반응편향을 추정하는 Bugs 스크립트를 제시하였다. 예시된 자료에서 d는 평균이 0이고 정밀도가 0.5(분산이 2)인 정규분포로, c는 평균이 0이고 정밀도가 2인 정규분포로 가정되었다. 이 때 Bugs와 관련해서 알아두어야 할 것은 정규분포가 평균과 정밀도(1/분산)로 정의된다는 점이다. 표집분포는 이항분포로 가정되고, 이항분포의 파라미터는 식 (6)과 (7)에 바탕을 두어 계산된다. 즉, 적중률의 파라미터는 민감도를 2로 나눈 값에서 반응편향을 뺀 z값에 대한 누적정규분포의 값의 비율이 되고, 오반응률의 파라미터는 민감도에 -1/2을 곱한 값에서 반응편향을 빼서 얻은 z값의 비율이 된다. 부록 4에서 H와 F는 실제 얻은 적중률과 오경보율을 나타내고, T와 D는 시행횟수를 나타낸다. 가상 피험자에 대한 추정 결과는 표 4와 같다.

표 4에서 c[1]은 첫 번째 가상 피험자의 반응편향을 d[1]은 첫 번째 가상 피험자의 민감도를 나타낸다. 첫 번째 가상피험자의 자료는

자극이 나오는 20시행에서 18을 맞추었고(적중률, 18/20), 자극이 나오지 않는 20시행에서 두 시행을 틀렸다(오경보율, 2/20). 이 피험자의 민감도를 식 (3)으로 계산하면 2.563이 나오는데, 베이지언 방법으로 추정된 사후분포는 평균이 2.323 표준편차가 0.4878인 분포가 된다(그림 4).

위와 같은 방식을 적용한 연구로 김윤현, 고성룡, Hazan(2012)은 우리말 화자들이 낱말 내의 위치에 따라 /b/와 /v/를 어떻게 지각하는지 살펴보았다. 이들은 피험자가 자극을 듣고 /b/ 또는 /v/로 판단한 반응을 신호탐지이론에 따라 분석하였는데, /b/음을 신호로 /v/음을 소음으로 간주하여 각 피험자의 조건에 따른 민

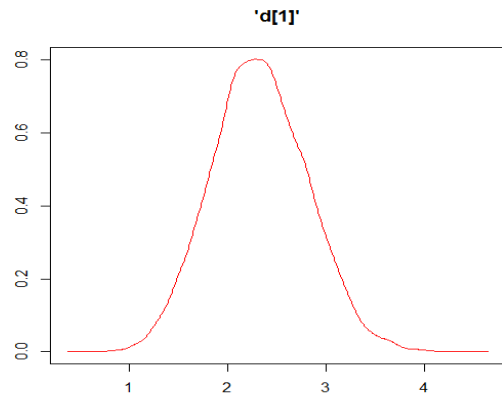


그림 4. 가상 피험자 1의 민감도 사후분포

표 4. 가상 피험자 세 명 자료의 민감도와 반응편향 추정

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
c[1]	0.001114	0.2452	0.002155	-0.48020	0.002171	0.48320	1001	51000
c[2]	-0.265900	0.2211	0.002129	-0.70480	-0.263100	0.15940	1001	51000
c[3]	-0.394400	0.2037	0.001925	-0.79340	-0.393000	0.00322	1001	51000
d[1]	2.323000	0.4878	0.004319	1.39900	2.311000	3.29000	1001	51000
d[2]	1.784000	0.4506	0.004081	0.93130	1.777000	2.69000	1001	51000
d[3]	0.782700	0.4078	0.003790	-0.01317	0.778800	1.59700	1001	51000

감도와 반응편향을 베이저언 프로그램 Lee (2008)을 사용하여 추정하였다. 연구 결과, 민감도의 경우 표적음이 어두에 제시된 조건과 낱말 중간에 제시된 조건 간에 차이가 있었으나(평균, 1.60 vs. 1.29), 반응편향에서는 차이가 없었다(평균, 0.05 vs. 0.15). 이 차이를 검증할 때 개인별로 추정한 민감도 값을 써서 전통적인 집단 간 t 검증을 할 수도 있으나, 더 나아가 베이저언 추론단계 피험자 수와 얻은 t값을 바탕으로 베이스 요인을 계산할 수도 있다 (<http://pcl.missouri.edu/node/32> 참조). 단, 개인의 민감도를 위계모형을 써서 추정했다면 각 개인들의 민감도 파라미터들 사이에 상관관계가 생기므로 위 방법으로 베이스 요인을 구하는 것이 적절하지 않다는 점에 주의해야 한다.

반응시간의 평균처리

Oh(2011)는 흥미로운 착시를 발견했는데, 일정 거리에서 안경을 관찰할 때 안경 다리나 렌즈보다 앞에 놓여 있거나 뒤에 놓여 있는 것으로 애매하게 보이는 현상이었다. 이 연구에서 또 한 가지 흥미로운 점은 안경을 보기 이전에 어떤 애매한 도형(Necker 육면체 vs. 토끼-오리 역전 도형)을 보았느냐에 따라 차이가 있었다는 것이다. 즉, 토끼-오리 역전 도형을 본 집단보다 Necker 육면체를 본 집단에서 더 많은 피험자가 위와 같은 착시를 경험했다고

보고하였다. 연구자는 애매성의 유사성이 이 차이를 가져왔을 것으로 추론하였는데, 즉 토끼-오리 역전 도형은 내용이 애매하지만 Necker 육면체는 안경 역전처럼 깊이가 애매하기 때문이다.

본 연구에서 이 Oh(2011)의 연구 결과를 사전에 경험한 도형에 따라 두 집단에서 역전에 걸린 시간을 다시 분석하였다. 역전 자체를 경험하는 것이 쉬운 일이 아니기 때문에 보고된 전체 자료의 시간을 비교하는 것은 역전을 더 많이 경험한 집단에 불리하였을 수 있다. 즉, 예를 들어, 토끼-오리 사전 관찰 집단에서 애매성을 60초 동안 경험하지 못한 관찰자는 자료 자체가 처리되지 않고, Necker 육면체 사전 집단에서 40초쯤에서 안경의 깊이 애매성을 경험한 관찰자는 자료가 유효하여 이 두 관찰자의 직접 비교가 통계적으로 어려울 수 있다. 그에 따라 연구자는 각 집단에서 역전을 경험했다고 보고한 반응시간들을 순서대로 나열해 보고한 수가 적은 집단의 수만큼을 써서 두 집단을 비교했다. 이 두 집단의 평균반응시간을 추정하는 Bugs 스크립트가 부록 5이다. 표집분포는 정규분포로, 표집분포의 평균 파라미터에 대한 사전 확률분포는 정보가 없는 균등분포로 가정되었으며 이 사전분포의 범위는 합리적으로 보이는 0부터 60초로 설정하였다. 표집분포의 정밀도에 대한 사전분포 역시 동일한 균등분포로 가정되었으며, 추정

표 5. Necker 육면체 집단과 토끼-오리 도형을 본 집단의 평균 반응시간 추정

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
diff	14.930000	4.492000	2.396e-02	6.094000	14.950000	23.860000	1001	30000
mu1	6.697000	1.018000	6.274e-03	4.685000	6.696000	8.711000	1001	30000
mu2	21.630000	4.373000	2.272e-02	12.940000	21.610000	30.300000	1001	30000

된 결과는 표 5와 같다.

결과를 살펴보면, Necker 도형을 본 집단에서 평균 파라미터의 평균은 6.697, 신뢰구간은 4.685-8.71초였고, 토끼-오리 도형을 본 집단에서 평균 파라미터의 평균은 21.630, 신뢰구간은 12.94-30.3초였다. 그리고 이 두 평균 차이 파라미터(diff)의 평균은 14.930이고 구간은 6.094-23.86으로 0을 포함하지 않았다. 위 결과는 두 집단의 반응시간에 차이가 없다는 가설의 확률이 매우 낮음을 보여주며, 이를 바탕으로 두 집단의 반응시간 평균이 믿을만하게 다르다고 결론을 내릴 수 있다.

최근 Kruschke(2013)는 두 집단의 평균 비교에 대한 베이지언 프로그램(BEST, Bayesian estimation supersedes the t test)을 만들어 소개했다. 이 프로그램의 특징은 극단 값들을 잘 반영하기 위해 반응시간의 표집분포로 정규분포가 아닌 t 분포를 가정하였다는 점이다. 위의 자료를 처리할 수 있는 BEST 프로그램을 부록 6에 제시하였다. t 분포를 표집 분포로 가정하기 때문에 t의 파라미터인 자유도에 대한 사전분포가 가정되어야 하는데, 이 분포로 지수(exponential)분포를 가정하였다. 이 지수분포 파라미터 값을 0.1로 지정하여 스크립트를 구동한 결과가 표 6이다.

이 결과에서도 두 집단의 평균 차이에 대한 베이지언 신뢰구간은 0.84-22.06으로 0을 포함하지 않아 기존 방식과 동일하게 두 집단의 반응 시간 평균이 다르다는 결론에 도달하게

된다. 현재까지는 반응시간의 평균 처리와 관련하여 위의 두 방법 가운데 어느 것을 사용해도 무관하다고 여겨지나, 추후 t 분포가 더 적절한 경우들이 밝혀지면 BEST의 활용이 필요할 수 있다.

눈의 고정시간 분포의 추정

위에서는 주로 반응비율과 반응시간의 평균에 대해 알아보았다. 이제까지의 심리학 연구들은 이 두 변수에 초점을 두었으나 근래 다수 연구자들은 반응시간의 평균만을 분석할 경우 많은 정보를 놓칠 수 있기 때문에 반응시간의 분포 자체를 살펴보아야 한다고 주장하였다. 최근 들어 이루어지고 있는 연구들은 주로 어떤 변인이 낱말 읽기 시간이나 판단시간(Balota & Spieler, 1999; Balota, Yap, Cortese, & Watson, 2008; Yap & Balota, 2007) 혹은 눈의 고정시간(Staub, 2011; Staub, White, Drieghe, Hollway, & Rayner, 2010) 분포에 영향을 미치는지 탐색하고 있다. 사실 분포 분석의 중요성은 오래전부터 강조되어 왔고, Ratcliff(1978)의 확산(diffusion) 모형은 분포를 예측해 왔다. 하지만 분포의 추정은 많은 자료의 수집을 요하기 때문에 실제 연구에 어려움이 많아 연구자들은 중심경향성을 주로 연구하였다. 조금씩 분포 분석이 재조명 되고 있는 것은 반응시간 분포를 다양한 이론적 분포를 통해 효과적으로 추정할 수 있다는 믿음 때문이다. 분

표 6. 지수분포의 파라미터를 0.1로 지정하고 BEST를 구동한 결과

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
diff	0.840	5.441	0.05596	0.8405	10.620	22.06	1001	30000
mu1	8.154	2.629	0.07212	4.2930	7.669	14.66	1001	30000
mu2	19.000	5.253	0.08325	9.1570	18.950	29.60	1001	30000

포의 유형을 정하면 그 유형에서 특정 파라미터 값만 찾으면 되기 때문에 요구되는 표본의 수가 줄어든다. 그에 따라 근래 연구들은 조건의 표본 수가 40개 이상 정도이면 분포를 추정하고자 시도하기도 한다. 분포의 좌우가 서로 비대칭적이며 꼬리를 오른쪽으로 늘어뜨린 모양을 한 반응시간 분포를 잘 기술할 수 있는 이론적인 분포로 감마(gamma), 와이불(Weibull), 로그-정규(log-normal), 지수-정규(ex-Gaussian) 분포들이 사용되고 있으며, 심리학 연구들은 지수-정규분포를 많이 다룬다 (Matzke & Wagenmakers, 2009). 지수-정규분포라는 이름에서 드러나듯이, 이 분포는 지수분포를 밀도함수로 갖는 무선변인과 정규분포를 밀도 함수로 갖는 무선변인의 합으로 이루어진 무선변인의 밀도함수이다. 즉, 이 밀도함수는 독립적인 두 무선변인의 밀도함수에 대한 합성곱(convolution)이다(Matzke & Wagenmakers, 2009). 지수-정규분포는 정규분포의 평균(μ)과 분산 파라미터(σ^2) 및 지수분포의 평균을 나타내는 파라미터(τ)로 정의된다. 최대우도기법을 이용한 이 세 파라미터의 추정법은 이미 도구로 나와 있고(QMPE, Cousineau, Brown, & Heathcote, 2004; Heathcote, Brown & Cousineau, 2004), 실제 연구들에서 활용되었다. 본 논문에서는 마지막으로 위계 베이지언을 이용한 세 파라미터의 추정 과정을 소개하고자 한다.

Koh & Hwang(투고중)은 낱말의 빈도가 안구운동의 고정시간 분포에 미치는 영향을 알아본 황지영(2011)의 연구를 위계 베이지언 방법을 통해 재분석하였다. 이들은 낱말의 판단 시간에 적용한 Matzke & Wagenmakers(2009)의 위계 베이지언 모형을 안구운동 고정시간에 적용하였다. 한 피험자한테서 얻을 수 있는

표본의 수가 많지 않은 경우, 집단 수준을 가정하여 위계 베이지언 추론을 하면 파라미터들의 범위를 축소하여 추정하는 위계 베이지언 방식의 특성 상 좀 더 나은 추정이 가능하다(Rouder, Lu, Speckman, Sun, D., & Jiang, 2005). 안구운동 연구에서도 한 피험자에게 얻을 수 있는 조건별 표본 수가 많은 편이 아니므로 연구자들은 위와 같은 성질을 지닌 위계 베이지언을 써서 추정하는 것이 적절하다고 판단했다. 그에 따라 고빈도 조건과 저빈도 조건에서 얻은 고정시간 분포를 ex-Gaussian으로 가정하고 이 분포의 파라미터들을 구하였다. 부록 8에는 한 어절/낱말에 한 번만 고정된 고정시간인 단일고정시간의 분포 파라미터들을 구하는 Bugs 스크립트를 제시하였다. 부록 7에 3명의 실제 자료를 분석하기 위해 작성한 스크립트를 제시하였다. 고정시간의 표집분포는 위에서 언급한 대로 지수-정규분포가 가정되었고, 이 분포의 세 파라미터에 대한 사전 확률분포는 모두 정규분포로 가정되었으며, 정규분포들의 두 파라미터에 대한 사전 확률분포(hyperprior)는 균등분포가 가정되었다. 균등분포의 두 파라미터 값들은 안구운동 연구에 적절하게 설정하였는데, 평균은 0-0.5 초 사이에서, 표준편차는 0-0.25사이에서 추정되었다. 여기서 눈여겨 볼 점은 정규분포의 정밀도에 대한 사전분포를 감마분포로 쓰지 않고 균등분포에서 표준편차를 무작위로 생성한 뒤 이를 제공한 역수로 추정하였다는 것이다. 부록 9에 있는 프로그램을 돌린 결과가 표 7에 있다.

아래 표에서 $\mu.g$ 은 그룹의 평균, $\sigma.g$ 은 그룹의 표준편차, $\tau.g$ 은 그룹의 지수분포의 평균을 나타내며, 그 외 대괄호가 붙은 파라미터들은 개인 수준의 값들을 의미한다. 앞서

표 7. 지수-정규분포(ex-Gaussian)의 분석 결과

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
mu.g	0.17200	0.028390	0.0003291	0.116600	0.17010	0.23490	1001	51000
mu.i[1]	0.17150	0.008652	0.0002135	0.155900	0.17110	0.18950	1001	51000
mu.i[2]	0.17800	0.016840	0.0006656	0.154400	0.17430	0.22170	1001	51000
mu.i[3]	0.16550	0.009186	0.0002679	0.149500	0.16490	0.18570	1001	51000
sigma.g	0.03084	0.024590	0.0003487	0.004774	0.02523	0.09689	1001	51000
sigma.i[1]	0.01828	0.009078	0.0002692	0.004363	0.01726	0.03895	1001	51000
sigma.i[2]	0.03429	0.013060	0.0004712	0.014690	0.03212	0.06275	1001	51000
sigma.i[3]	0.02343	0.007806	0.0002311	0.010880	0.02255	0.04192	1001	51000
tau.g	0.06705	0.030130	0.0003565	0.017660	0.06353	0.14250	1001	51000
tau.i[1]	0.07271	0.015920	0.0002683	0.047730	0.07070	0.10960	1001	51000
tau.i[2]	0.06457	0.018530	0.0006667	0.021380	0.06549	0.09952	1001	51000
tau.i[3]	0.05344	0.012050	0.0003059	0.030060	0.05337	0.07696	1001	51000

언급한 바와 같이 위계 모형에서 베이스 요인을 추정하는 경우에는 전통적인 검증을 한 뒤 이를 바탕으로 값을 계산하는 방법이 적절하지 않으므로 Koh와 Hwang(투고중)은 곱공간(product space) 방법(Lodewyckx, Kim, Lee, Tuerlinckx, Kuppens & Wagenmakers, 2011)을 사용하여 베이스 요인의 근사치를 구했다. 이 방법에 대한 상세한 설명은 본 논문의 범위를 넘어서므로 다른 기회에 다루도록 하겠다.

맺음말

본 논문은 베이지언 분석방법을 실제 연구에서 활용할 수 있도록 실용적인 측면에 중점을 두어 소개하고자 하였다. 그에 따라 심리학 연구에서 주로 쓰이는 반응비율 처리, 신호탐지이론의 분석, 반응시간 평균 처리, 반응시간의 분포 분석 등의 문제와 관련하여 구체

적인 베이지언 분석 방법을 예와 함께 설명하였고, 특히 프로그램 구동을 위한 Bugs 스크립트를 부록으로 제시하여 연구자들이 응용할 수 있도록 하였다. 그러나 위계 모형 분석의 경우 분포의 파라미터들을 추정하는 예만 언급하여 그 구체적인 내용에 대한 개념적인 설명이 충분하지 못하였으며, 모형 선택 시 베이스 요인을 계산하는 일반적인 방법에 대해서는 아직 손쉬운 도구가 마련되어 있지 않아 제한적인 설명만 가능하였다. 추후 이 같은 점을 보완한 논문과 도구가 제안되기를 기대한다. 끝으로 저자들은 본 논문이 베이지언 추론을 유용한 도구로 인식하고, 나아가 보다 널리 활용하는데 도움이 되기를 바란다.

참고문헌

김윤현, 고성룡, Valerie Hazan (2012). 한국어

- 화자의 영어 양순음 /b/와 순치음 /v/ 식별에서 시각 단서의 효과. *말소리와 음성과학*, 4(3), 25-30.
- 오성주 (2014). 빨강-초록 조합에 대한 전경-배경 조직화에서 성차. *인지과학회*, 25(2), 73-90.
- 황지영 (2012). 낱말 빈도와 차폐가 고정시간 분포에 미치는 효과. *서울대학교 석사학위 청구 논문*.
- Koh, S., & Hwang, J. (투고중). A hierarchical Bayesian analysis on eye fixation durations in Korean.
- Balota, D. A. & Spieler D. H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *Journal of Experimental Psychology: General*, 12, 32-55.
- Balota, D. A., Yap, M. J., Cortese, M. J., & Watson, J. M. (2008). Beyond mean response latency: Response time distributional analyses of semantic priming. *Journal of Memory and Language*, 59, 495-523.
- Cumming, G. (2014). The new statistics: Why and How. *Psychological Science*, 25(1), 7-29.
- Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence, intervals and meta-analysis*. New York: Routledge.
- Cousineau, D., Brown, S., & Heathcote, A. (2004) Fitting distributions using Maximum Likelihood: Methods and packages. *Behavior Research, Methods, Instruments & Computers*, 36, 277-290.
- Gelman, A., Carlin, J., Stern, H., & Rubin, D. (2003). *Bayesian data analysis*, Second Edition. Chapman & HALL/CRC.
- Green, P. (1995). Reversible Jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4), 711-732.
- Gelman, A., Carlin, J., Stern, H., & Rubin, D. (1995). *Bayesian data analysis*. Chapman & HALL/CRC.
- Heathcote A, Brown S, & Cousineau D. (2004). QMPE: estimating Lognormal, Wald, and Weibull RT distributions with a parameter-dependent lower bound. *Behav Res Methods Instrum Comput*, 36(2), 277-90.
- Kass, R., & Raftery, A. (1995). Bayes factors. *Journal of the American Statistical Association* 90, 773-795.
- Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General*, 142(2), 573-603.
- Lee, M. D. (2008). BayesSDT: Software for Bayesian inference with signal detection theory. *Behavior Research Methods*, 40, 450-456.
- Lodewyckx, T., Kim, W., Lee, M., Tuerlinckx, F., Kuppens, P. & Wagenmakers, E. (2011). A tutorial on Bayes factor estimation with the product space method. *Journal of Mathematical Psychology*, 55, 331-347.
- Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS-A Bayesian modeling framework: Concepts, structure, and extensibility. *Statistics & Computing*, 10, 325-337.
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide*. New York: Lawrence Erlbaum Associates.
- Matzke, D. & Wagenmakers, E. (2009). Psychological interpretation of the ex-Gaussian

- and shifted Wald parameters: A diffusion model analysis. *Psychonomic Bulletin & Review*, 16, 798-817.
- Oh. (2011). The eyeglass reversal. *Attention, Perception, and Psychophysics*, 73, 1336-1343.
- Ratcliff, R. (1978). *A theory of memory retrieval*. *Psychological Review*, 85, 59-108.
- Rouder, J., Lu, J., Speckman, P., Sun, D., & Jiang, Y. (2005). A hierarchical model for estimating response time distributions. *Psychonomic Bulletin & Review*, 12, 195-223.
- Staub, A. (2011). The effect of lexical predictability on distributions of eye fixation duration. *Psychonomic Bulletin & Review*, 18, 371-376.
- Staub, A., White, S., Drieghe, D., Hollway, E., & Rayner, K. (2010). Distributional effects of word frequency on eye fixation durations. *Journal of Experimental Psychology: Human Perception and Performance*, 36(5), 1280-1293.
- Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Better decisions through science. *Scientific American*, October, 82-87.
- Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science*, 240, 1285-1293.
- Tenenbaum, J., Kemp, C., Griffiths T., & Goodman N. (2011). How to grow a mind: statistics, structure, and abstraction. *Science*, 331, 1279-1285.
- Wickens, T. D. (2002). *Elementary signal detection theory*. New York: Oxford University Press.
- Yap, M., & Balota, D. (2007). Additive and interactive effects on response time distribution in visual word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 274-296.
- 1차원고접수 : 2014. 03. 19.
수정원고접수 : 2014. 08. 18.
최종게재결정 : 2014. 09. 12.

Introduction to the practical using of Bayesian inference for Psychology Research

Sungryong Koh¹⁾

Hyeree Choo²⁾

Songjoo Oh¹⁾

¹⁾Department of Psychology, Seoul National University

²⁾Cognitive Science Program, Seoul National University

Becoming a useful tool in the modern psychology, Bayesian inference is a recent powerful movement to new statistics in order to improve traditional statistics based on the null-hypothesis significance testing (NHST). This tendency substantially challenges the view of cognitive processing and is being widely accepted as a new area of statistics. In this study, the authors introduce Bayesian inference in terms of practical tool beyond the scope of the argument between frequency view and Bayesian view. In addition, the authors present several examples to indicate how to use Bayesian inference for an understanding of the results consisting of response ratio and reaction time that are popular in psychology studies.

Key words : Bayesian inference, response rate, response time, signal detection theory, fixation duration, distribution estimation

부록 1. 반응비율 bugs 스크립트 예

```
library(BRugs)

# Specify the model in BUGS language, but save it as a string in R:
modelString = "
model {
  # Likelihood:
  y ~ dbin(theta,64)
  theta ~ dbeta(1,1)
  # Prior distribution:
}
"

writeLines(modelString,con="model.txt")
modelCheck("model.txt")

#-----#
y = 48
dataList = list(y =y)

# Use BRugs commands to put the data into a file and ship the file to BUGS:
modelData( bugsData( dataList ))

#-----#
# INTIALIZE THE CHAIN.

modelCompile() # BRugs command tells BUGS to compile the model.
modelGenInits() # BRugs command tells BUGS to randomly initialize a chain.

#-----#
# RUN THE CHAINS.

# BRugs tells BUGS to keep a record of the sampled "theta" values:
# R command defines a new variable that specifies an arbitrary chain length:
chainLength = 10000
# BRugs tells BUGS to generate a MCMC chain:
modelUpdate(1000) # burn in
samplesSet(c("theta")) # theta
modelUpdate(chainLength)
samplesStats("*") #summarized results
```

부록 2. 반응비율 비교 bugs 스크립트 예

```
library(BRugs)

# Specify the model in BUGS language, but save it as a string in R:
modelString = "
model {
  # Likelihood:
  y1 ~ dbin(theta1,64)
  y2 ~ dbin(theta2,64)
  # Prior distribution:
  theta1 ~ dbeta(1,1)
  theta2 ~ dbeta(1,1)
  diff <- theta1 - theta2 ##
}
"

writeLines(modelString,con="model2.txt")
modelCheck("model2.txt")

#-----
# THE DATA.
y1 = 48
y2 = 32
dataList = list(y1 =y1, y2=y2)
modelData( bugsData( dataList ))

#-----
# INITIALIZE THE CHAIN.
modelCompile(1) # BRugs command tells BUGS to compile the model.
modelGenInits() # BRugs command tells BUGS to randomly initialize a chain.
#-----
chainLength = 30000
# BRugs tells BUGS to generate a MCMC chain:
modelUpdate(1000) # burn in
samplesSet(c("theta1", "theta2", "diff"))
modelUpdate(chainLength)
samplesStats("*")
####
```

부록 3. 반응비율 비교 bugs 스크립트 예

```
library(BRugs)

modelString = "
model {
  # Likelihood:
  y1 ~ dbin(theta1,32)
  y2 ~ dbin(theta2,32)
  # prior
  theta1 ~ dbeta(1,1)
  theta2 ~ dbeta(1,1)
  diff <- theta1 - theta2 ##
}
"

writeLines(modelString,con="model2.txt")
modelCheck("model2.txt")

#-----
# THE DATA.

y1 = 28
y2 = 20
dataList = list(y1 =y1, y2=y2)

# Use BRugs commands to put the data into a file and ship the file to BUGS:
modelData( bugsData( dataList ))

#-----
# INITIALIZE THE CHAIN.

modelCompile(1) # BRugs command tells BUGS to compile the model.
modelGenInits() # BRugs command tells BUGS to randomly initialize a chain.

#-----
# RUN THE CHAINS.
chainLength = 10000
# BRugs tells BUGS to generate a MCMC chain:
modelUpdate(1000) # burn in
samplesSet(c("theta1", "theta2", "diff"))
modelUpdate(chainLength)
samplesStats("*")

a = buildMCMC("theta1")
HPDinterval(a,0.95)
```

부록 4. 신호탐지이론 분석 bugs 스크립트 예

```
library(BRugs)
# Specify the model in BUGS language, but save it as a string in R:
modelString = "
# BUGS model specification begins ...
model {
  # Likelihood:
  for ( i in 1:N ) {
    d[i] ~ dnorm(0,0.5)
    c[i] ~ dnorm(0,2)

    h[i] <- phi(0.5*d[i]-c[i]) ## phi cumulative function for normal distribution
    f[i] <- phi(-1*0.5*d[i]-c[i])
    H[i] ~ dbin(h[i], T[i]) ##Censoring and Truncation
    F[i] ~ dbin(f[i], D[i])
  }
}“
# Write the modelString to a file, using R commands:
writeLines(modelString,con="model1.txt")
# Use BRugs to send the model.txt file to BUGS, which checks the model syntax:
modelCheck( "model1.txt" )
#-----
#DATA
T= c(20,20,20)
D= c(20,20,20)
H = c(18,18,16)
F = c(2,5,10)
dataList = list(N = 3,H = H, F = F, T=T, D=D)

modelData( bugsData( dataList ) )

#-----
# INITIALIZE THE CHAIN.

modelCompile(2) # BRugs command tells BUGS to compile the model.
modelGenInits() # BRugs command tells BUGS to randomly initialize a chain.

#-----
# RUN THE CHAINS.
chainLength = 11000
modelUpdate(1000) # burn in
samplesSet(c("d", "c"))
modelUpdate(chainLength)
samplesStats("*")
```

부록 5. 평균 비교 bugs 스크립트 예

```
library(BRugs)

modelString = "
model {
  # Likelihood:
  for (i in 1:N1){
    x1[i] ~ dnorm(mu1,pre1)
  }
  for (i in 1:N2){
    x2[i] ~ dnorm(mu2,pre2)
  }
  mu1 ~ dunif(0, 60)
  mu2 ~ dunif(0, 60)
  diff <- mu2 - mu1

  pre1<- 1/pow(sigma1, 2)
  sigma1~dunif(0, 60)
  pre2<- 1/pow(sigma2, 2)
  sigma2~dunif(0, 60)
}
"

writeLines(modelString,con="model2.txt")
modelCheck("model2.txt")

#-----
# THE DATA.

# Specify the data in R, using a list format compatible with BUGS:

x1=c(2.87,3.32,3.35,3.84,3.85,4.09,4.19,4.41,4.50,5.40,7.03,8.78,10.00,10.21,11.06,11.06,1
5.91)
x2=c(5.09,6,6.2,6.28,8.32,9.81,10,10,11.78,22.19,23.38,36.87,37.43,39.56,40,46.03,49.22 )
#group = c(rep(1,length(x1)), rep(2,length(x2)))
y = c(x1,x2) ### the number of targets
```

한국심리학회지: 일반

```
N1 = length(x1)
N2 = length(x2)
dataList = list(x1 =x1, x2=x2, N1=N1,N2=N2)

# Use BRugs commands to put the data into a file and ship the file to BUGS:
modelData( bugsData( dataList ))

#-----
# INITIALIZE THE CHAIN.

modelCompile(1) # BRugs command tells BUGS to compile the model.
modelGenInits() # BRugs command tells BUGS to randomly initialize a chain.

#-----
# RUN THE CHAINS.

chainLength = 30000
modelUpdate(1000) # burn in
samplesSet(c("diff", "mu1", "mu2", "pre1", "pre2"))
modelUpdate(chainLength)
samplesStats("*")
```


부록 6. BEST

```

library(BRugs)

# THE MODEL.
modelString = "
model{
  for ( i in 1:Ntotal ){
    y[i] ~ dt(mu[x[i]] , tau[x[i]] , nu)
  }
  for (j in 1:2 ) {
    mu[j] ~ dnorm( muM , muP)
    tau[j]<- 1/pow(sigma[j], 2)
    sigma[j]~dunif(sigmaLow , sigmaHigh )
  }
  nu <- nuMinusOne+ 1
  nuMinusOne ~dexp(0.1)
}"

writeLines(modelString , con="model.txt")
modelCheck("model.txt")

#####

y1=c(4.09,50,2.87,3.84,10,4.5,3.32,35.49,4.19,47.1,11.06,21,11.06,3.85,4.41,5.4,7.03,3.35,8
.78,10.21,15.91)
y2 = c(8.32,6.28,40,5.09,46.03,22.19,36.87,49.22,23.38,9.81,37.43,6,10,6.2,10,39.56,11.78 )

y = c( y1 , y2 ) # combine data into one vector
x = c( rep(1,length(y1)) , rep(2,length(y2)) ) # create group membership code
Ntotal = length(y)
# Specify the data in a list, for later shipment to JAGS:
dataList = list(
  y = y ,
  x = x ,
  Ntotal = Ntotal ,
  muM = mean(y) ,
  muP = 0.000001 * 1/sd(y)^2 ,
  sigmaLow = sd(y) / 1000 ,

```

한국심리학회지: 일반

```
sigmaHigh = sd(y) * 1000

)
modelData( bugsData( dataList ))
#-----
# INITIALIZE THE CHAIN.

modelCompile(1) # BRugs command tells BUGS to compile the model.
modelGenInits() # BRugs command tells BUGS to randomly initialize a chain.

#-----
# RUN THE CHAINS.
chainLength = 10000
modelUpdate(1000) # burn in
samplesSet(c("mu[1]", "mu[2]")) #
modelUpdate(chainLength) #
samplesStats("*") # the summarized results
```

부록 7. ex-Gaussian 파라미터들의 추정

```
library(BRugs)

modelString = "
model {
  for (i in 1:NSUB ){ ## for each individual
    mu.i[i] ~ dnorm(mu.g,lambda.mu.g)I(0,)
    sigma.i[i] ~ dnorm(sigma.g,lambda.sigma.g)I(0,)
    tau.i[i] ~ dnorm(tau.g,lambda.tau.g)I(0,)
    lambda.i[i] <- pow(sigma.i[i],-2)
    rate.i[i] <- 1/tau.i[i]
  }
  for (j in 1:NOBS){
    y[j] ~ dexp(rate.i[sub[j]])I(0,) # rate is 1/tau
    z[j] <- mu.i[sub[j]] + y[j]
    d[j] ~ dnorm(z[j],lambda.i[sub[j]])
  }
  ### group, prior
  ## for means
  mu.g ~ dunif(0, 0.5)
  sigma.g ~ dunif(0, 0.5)
  tau.g ~ dunif(0, 0.5)
  ## sds
  sd.mu.g ~ dunif(0, 0.25)
  sd.sigma.g ~ dunif(0, 0.25)
  sd.tau.g ~ dunif(0, 0.25)
  ### precision
  lambda.mu.g <- pow(sd.mu.g,-2)
  lambda.sigma.g <- pow(sd.sigma.g, -2)
  lambda.tau.g <- pow(sd.tau.g, -2)

}
# ... BUGS model specification ends.
" # close quote to end modelString

# Write the modelString to a file, using R commands:
writeLines(modelString,con="model.p")
```


부록 8. BEST

```

library(BRugs)

# THE MODEL.
modelString = "
model{
  for ( i in 1:Ntotal ){
    y[i] ~ dt(mu[x[i]] , tau[x[i]] , nu)
  }
  for (j in 1:2 ) {
    mu[j] ~ dnorm( muM , muP)
    tau[j]<- 1/pow(sigma[j], 2)
    sigma[j]~dunif(sigmaLow , sigmaHigh )
  }
  nu <- nuMinusOne+ 1
  nuMinusOne ~dexp(0.1)
}"

writeLines(modelString , con="model.txt")
modelCheck("model.txt")

#####

y1=c(4.09,50,2.87,3.84,10,4.5,3.32,35.49,4.19,47.1,11.06,21,11.06,3.85,4.41,5.4,7.03,3.35,8
.78,10.21,15.91)
y2      =      c(8.32,6.28,40,5.09,46.03,22.19,36.87,49.22,23.38,9.81,37.43,6,10,6.2,10,39.
56,11.78 )

y = c( y1 , y2 ) # combine data into one vector
x = c( rep(1,length(y1)) , rep(2,length(y2)) ) # create group membership code
Ntotal = length(y)
# Specify the data in a list, for later shipment to JAGS:
dataList = list(
  y = y ,
  x = x ,
  Ntotal = Ntotal ,
  muM = mean(y) ,
  muP = 0.000001 * 1/sd(y)^2 ,

```

한국심리학회지: 일반

```
sigmaLow = sd(y) / 1000 ,
sigmaHigh = sd(y) * 1000

)
modelData( bugsData( dataList ))
#-----
# INITIALIZE THE CHAIN.

modelCompile(1) # BRugs command tells BUGS to compile the model.
modelGenInits() # BRugs command tells BUGS to randomly initialize a chain.

#-----
# RUN THE CHAINS.
chainLength = 10000
modelUpdate(1000) # burn in
samplesSet(c("mu[1]", "mu[2]")) #
modelUpdate(chainLength) #
samplesStats("*") # the summarized results
```

부록 9. ex-Gaussian 파라미터들의 추정

```
library(BRugs)

modelString = "
model {
  for (i in 1:NSUB ){ ## for each individual
    mu.i[i] ~ dnorm(mu.g,lambda.mu.g)I(0,)
    sigma.i[i] ~ dnorm(sigma.g,lambda.sigma.g)I(0,)
    tau.i[i] ~ dnorm(tau.g,lambda.tau.g)I(0,)
    lambda.i[i] <- pow(sigma.i[i],-2)
    rate.i[i] <- 1/tau.i[i]
  }
  for (j in 1:NOBS){
    y[j] ~ dexp(rate.i[sub[j]])I(0,) # rate is 1/tau
    z[j] <- mu.i[sub[j]] + y[j]
    d[j] ~ dnorm(z[j],lambda.i[sub[j]])
  }
  ### group, prior
  ## for means
  mu.g ~ dunif(0, 0.5)
  sigma.g ~ dunif(0, 0.5)
  tau.g ~ dunif(0, 0.5)
  ## sds
  sd.mu.g ~ dunif(0, 0.25)
  sd.sigma.g ~ dunif(0, 0.25)
  sd.tau.g ~ dunif(0, 0.25)
  ### precision
  lambda.mu.g <- pow(sd.mu.g,-2)
  lambda.sigma.g <- pow(sd.sigma.g, -2)
  lambda.tau.g <- pow(sd.tau.g, -2)
}
# ... BUGS model specification ends.
" # close quote to end modelString

# Write the modelString to a file, using R commands:
writeLines(modelString,con="model.p")
```

