수리통계학

박 진호

머리말

수리통계학은 통계학과 교과과정의 핵심적인 과목 중의 하나라 할 수 있다. 기초적인 통계학에서 출발하여 다양한 자료분석 방법을 배우기 위한 기본적인 통계 이론을 다루는 과목이 수리통계학이다. 수리통계학에서 다루는 내용은 통계학에서 사용되는 기본적인 언어라 할 수 있다. 기본적인 통계 이론을 이해하지 않고서는 좀더 깊이 있는 자료분석 방법을 다루는 것이 쉽지 않다. 이처럼 수리통계학이 중요한 역할을 하고 있지만 다른 한편으로는 학생들이 많은 어려움을 느끼는 과목이기도 하다. 수리통계학에서 다루는 내용이 수리적이며 구체적인 분석 방법이 아니라 추상적인 이론을 다루기 때문에 어려움을 느끼는 것으로 생각된다.

이 책에서 다루는 내용은 수리통계학에서 다루는 핵심적인 내용만을 간추린 것이다. 통계학을 배우는 과정에 있는 학생들에게 필요한 기초적인 이론만을 포함하고 있다. 학생들의 이해를 돕기 위하여 가능한 예제를 많이 다루도록 하였으며 수리통계학에서 다루는 내용을 다른 과목과 연관 지어 설명하려고 노력하였다. 학생들에게 하나의 정리를 직접 설명하는 것보다 학생들이 이미 알고 있는 내용과 연관 지어 설명할 때 정리의 이해에 많은 도움이 될 것으로 판단되기 때문이다. 이 책을 크게 두 부분으로 나눈다면 주로 확률분포에 대한 여러 가지 내용을 다룬 6장까지와 점추정과 가설검정을 비롯한 통계적 추론을 다룬 7장부터 10장까지로 나눌 수 있다. 저자의 경험에 의하면 학생들이 가장 어려움을 느끼는 부분은 확률변수의 극한을 다룬 6장, 최대가능도추정량의 점근적 성질을 다룬 8.6절과 균일최강력검정에 대한 9.3절이라 할 수 있다. 이 책을 교재로 사용한다면 이점을 참고로 강의를 진행하면 될 것이다.

이 책은 지난 몇 년간 인하대학교에서 사용한 강의 노트에 좀 더 자세한 설명을 첨부한 것이다. 처음 강의에서 사용한 교재는 J.A. Rice가 쓴 *Mathematical Statistics and Data Analysis*이었다. 아마도 이 책에도 그 영향이 남아 있을 것이다. 끝으로 이 책이 발간될 수 있도록 지원을 해준 인하대학교에 감사를 표한다.

2005년 11월

박 진호(jhpark@inha.ac.kr)

차 례

그림 차례

수리통계학에서 다루는 내용

통계학이란 어떤 것을 다루는 것일까? 일반적으로 통계학이라면 흔히 히스토그램이나 평균 또는 표준편차 등을 떠올리게 된다. 이러한 것은 통계학의 한 분야인 자료를 요약 또는 정리하는 방법이라 할 수 있다. 통계학은 이처럼 단순히 자료를 정리하는 단계에서 더 나아가서 자료를 분석하고, 이로부터 유용한 정보를 이끌어내는 방법을 다룬 학문이라 할 수 있다.

예를 들어 제약회사에서 특정한 질병에 대한 치료약의 효과를 알아보려는 경우를 생각하여 보자. 이 제약회사에서 규명하려는 것은 약물의 투여량과 치료효과 사이의 연관성이라 할 수 있다. 어떤 방법으로 약물의 투여량과 치료효과 사이의 연관성을 구할 수 있을까? 연관성을 알아내는 한 가지 방법은 실험을 통해 투여량과 그에 따른 치료효과에 대한 자료를 관측하고 이로부터 필요한 정보를 얻는 것이다.

10명을 대상으로 실험한 결과가 다음과 같다고 하자.

$$(3.0, 45),\ (4.0, 40),\ (4.0, 50),\ (4.5, 45),\ (5.0, 50)$$
$$(5.0, 53),\ (5.5, 60),\ (5.5, 58),\ (6.0, 55),\ (7.0, 45)$$

위에서 괄호 안의 첫 번째 값은 약물 투여량, 두 번째 값은 치료효과를 점수화한 자료를 나타낸다. 이 자료로부터 약물 투여량의 평균은 4.95, 표준편차는 1.14, 치료효과의 평균은 50.1, 표준편차는 6.43임을 알 수 있다. 또 약물 투여량과 치료효과 사이의 표본상관계수는 0.40임을 알 수도 있다. 그러나 우리가 이 자료를 통하여 알고 싶은 것은 위의 10명에 한정된 자료의 요약된 결과는 아니다. 제약회사에서 구하려는 것은 이 질병을 앓고 있는 모든 환자를 대상으로 한 약물 투여량과 치료효과 사이의 연관성이다.

앞에서 실험의 대상이 된 10명은 이 질병을 앓고 있는 환자의 일부분이라 할 수 있다. 우리가 구하려는 것은 10명이라는 일부분의 결과를 이용하여 모든 환자를 대상으로 한 결론을 얻으려는 것이다. 이처럼 주어진 자료를 이용하여 우리가 알고자 하는 대상에 대한 유용한 정보를 이끌어내는 과정을 다루는 것이 통계학이다. 우리가 알고자 하는 대상을 모집단(population)이라 한다. 그림 0.1은 이러한 과정을 나타낸 것이다. 일부분의 자료를 이용하여 모집단에 대한 결론을 얻는 과정은 추측이라는 단계를 거치게 된다. 일부분을 이용하여 전체 대상에 대한 확대 해석을 하기 위해서는 추측을 하여야만 하기 때문이다. 따라서 이러한 추측 과정에서 잘못된 판단을 할 수도 있다. 통계학은 이러한 추측과정을 체계적으로 할 수 있도록 하고 추측과정에서 발생하는 오류의 가능성을 제어하려고 한다.

다른 예로 우리나라 성인 남성의 평균 키를 구하기 위해 성인 남성 100명을 임의로 선택하여

이들의 평균 키를 조사한다고 하자. 우리가 최종적으로 구하려는 것은 이들 100명의 평균 키가 아니라 전체 성인 남성의 평균 키이다. 전체 성인 남성의 평균 키를 추측하는 방법으로 조사한 100명의 평균 키를 이용할 수도 있고 100명 중 첫 번째로 선택된 사람의 키를 이용할 수도 있다.

두 가지 추측방법 중에서 어느 것이 타당할까? 대부분의 사람은 100명의 평균 키를 이용하는 것이 좋을 것이라고 생각할 것이다. 과연 첫 번째 선택된 사람의 키가 아니라 100명의 평균 키를 이용하는 것이 더 좋을까? 어떤 의미에서 더 좋을까? 항상 더 좋다고 주장할 수 있을까? 만약 전체 성인 남성의 평균 키는 172cm인데 우연히도 첫 번째로 선택된 사람의 키는 172cm이고 100명의 평균 키는 170cm라면 어떨까? 이 경우는 100명의 평균 키보다 첫 번째로 선택된 사람의 키가 전체 성인 남성의 평균 키에 가깝게 된다. 따라서 항상 100명의 평균 키가 첫 번째로 선택된 사람의 키보다 전체 성인 남성의 평균 키에 가깝다고 주장할 수는 없다. 그러나 문제는 이 경우처럼 첫 번째로 선택된 사람의 키가 전체 성인 남성의 평균 키에 더 가까울 가능성이 얼마나 될까? 이럴 가능성은 매우 작다. 100명의 평균 키가 전체 성인 남성의 평균 키에 더 가까울 가능성이 훨씬 더 클 것이다. 따라서 첫 번째로 선택된 사람의 키보다는 100명의 평균 키를 이용하는 것이 더 타당할 것이다.

위의 예처럼 추측하는 방법이 여러 가지가 있을 수 있다. 이때 어느 것이 더 타당할 것인지는 가능성을 가지고 얘기할 수밖에 없다. 이러한 가능성이라는 것은 확률을 통해 나타낼 수 있다. 즉 100명의 평균 키가 어느 한 사람의 키보다는 전체 평균 키에 가까울 확률이 높다고 할 수 있다.

통계학은 자료를 분석하여 모집단에 대한 유용한 정보를 이끌어 내는 것을 목표로 한다고 하였다. 일반적으로 자료의 분석방법은 자료의 특징이나 성격에 따라 다를 수 있다. 대표적인 몇 가지 자료분석 방법은 회귀분석, 시계열분석, 범주형자료분석, 다변량분석, 생존분석 등이 있다. 예를 들어 한 변수가 다른 변수의 영향을 받는 종속적인 관계가 있고 영향을 받는 변수가 연속형 변수인 경우에 흔히 사용되는 자료분석 방법이 회귀분석이다. 영향을 받는 변수가 어떤 범주에 속하는지를 나타내는 명목형 변수인 경우는 변수 사이의 연관성을 분석하기 위해 범주형자료분석을 이용한다. 한편 순차적으로 관측되는 자료를 분석하기 위해서는 시계열분석을 이용한다.

통계학의 기본적인 목표인 주어진 자료로부터 유용한 정보를 얻는 한 가지 방법은 자료의 특징이나 성격에 따라 모형화(modeling)하는 것이다. 통계학에서 모형화란 자료분석을 통해 알고자 하는 대상인 모집단에 대한 가정을 하는 것을 말한다. 간단한 예로 모집단의 특징을 파악하기 위해

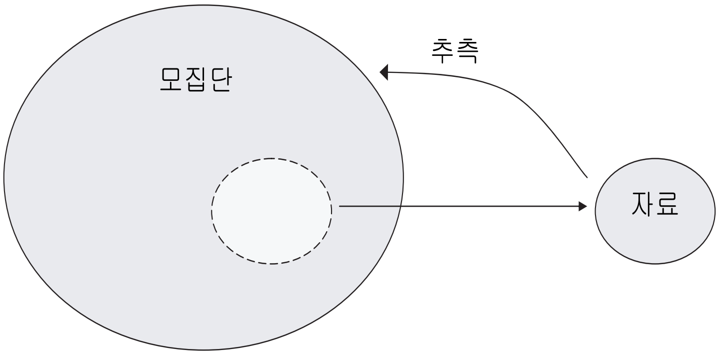

그림 0.1: 통계학 - 자료를 이용하여 알고자 하는 대상에 대한 추측

모집단이 정규분포를 따른다고 가정하는 것도 모형화의 한 방법으로 생각할 수 있다. 또 두 변수 사이의 연관성을 분석하기 위해 두 변수 사이에 선형관계를 가정하는 것도 모형화라 할 수 있다. 이러한 모형화 과정을 통해 모집단은 그 특성을 결정지어주는 몇 개의 값으로 나타낼 수 있다.

예를 들어 모집단이 정규분포를 따른다는 가정을 하는 경우를 생각하여 보자. 정규분포는 평균과 분산에 의하여 결정되므로[1] 모집단이 정규분포를 따른다는 가정 하에서는 평균과 분산에 의하여 모집단의 특성이 결정된다. 이처럼 모집단의 특성을 결정하는 값을 모수(parameter)라 한다. 정규분포에서는 평균과 분산이 모수가 된다. 따라서 통계학은 '자료로부터 유용한 정보를 이끌어 내는 것' 에서 '자료로부터 모집단의 특성을 나타내는 모수에 대한 추측을 하는 것' 으로 단순화할 수 있다. [2]

자료분석을 통해 모집단의 특성을 나타내는 모수의 값을 추측하는 문제에서 중요한 것은 모수를 어떻게 추측할 것인가에 대한 것이다. 또한 추측한 모수의 값이 어떠한 성질을 가지게 될 것인지도 관심사항일 것이다. 이러한 질문에 대한 기본적인 답을 찾는 방법을 다루는 것이 수리통계학이라 할 수 있다. 간단히 말하면 수리통계학은 자료분석에서 사용되는 기본적인 통계학적 방법론을 다루는 분야라고 할 수 있다. 몇 가지 예를 통해 모형화 과정과 모수의 추측 방법을 간단히 알아보도록 하자.

우리나라 성인 남성의 콜레스테롤 수준을 알아보기 위해 n명의 성인남성을 대상으로 조사하였다고 하자. 관측한 n명의 콜레스테롤 수준을 다음과 같다고 하자.

$$X_1, X_2, \cdots, X_n$$

우리가 알고 싶은 것은 조사한 n명에 대한 콜레스테롤 수준은 아니다. 주어진 자료를 분석하여 우리나라 전체 성인 남성의 콜레스테롤 수준의 특성을 파악하려는 것이다. 이 경우 우리가 알고자 하는 대상인 모집단은 우리나라 전체 성인 남성의 콜레스테롤 수준이라 할 수 있다.

주어진 자료를 이용하여 성인 남성의 콜레스테롤 수준이 어떤 특징을 가지고 있는지 알아내는 것은 모집단에 대한 가정(모형화)을 하지 않으면 쉽지 않다. 모집단의 기본적인 형태에 대한 가정이 없이는 주어진 10개의 자료로부터 모집단에 대해 알아낼 수 있는 것은 많지 않다. 모집단이 특정한 확률분포를 가진다는 것과 같은 가정을 했을 때 주어진 자료로부터 모집단에 대한 좀 더 유용한 추측을 할 수 있다. 우리나라 성인 남성의 콜레스테롤 수준에 대한 한 가지 모형화 방법으로 다음과 같이 가정하자.

$$\text{성인 남성의 콜레스테롤 수준} \sim \mathrm{N}(\mu, \sigma^2)$$

즉 성인 남성의 콜레스테롤 수준은 평균이 μ, 분산이 σ^2인 정규분포를 따른다고 가정하자. 모집단이 $\mathrm{N}(\mu, \sigma^2)$이라고 가정했을 때 우리가 알지 못하는 것은 모집단의 확률분포인 정규분포를 결정지어주는 평균 μ와 분산 σ^2이라고 할 수 있다. 이 경우 모집단의 특징을 나타내는 모수는 μ와 σ^2이다. 따라서 자료분석의 주된 목적은 주어진 자료를 이용하여 μ와 σ^2을 추측하는 것이라

[1]정규분포에 관한 51면 참조.

[2]최근의 통계학에서는 모집단이 몇 개의 모수에 의하여 결정된다는 고전적인 모수적 통계학에서 벗어나서 비모수적 통계 분석방법이 많이 활용되고 있다.

할 수 있다. 두 모수의 추측 값을 $\widehat{\mu}$와 $\widehat{\sigma}^2$ 이라고 할 때 성인 남성의 콜레스테롤 수준은 다음과 같은 확률분포를 따른다고 추측할 수 있다.

$$\mathrm{N}(\widehat{\mu}, \widehat{\sigma}^2)$$

여기에서 주된 관심사항은 μ와 σ^2 을 어떻게 추측할 것인가에 관한 것이다. 일반적으로 모집단의 평균과 분산인 μ와 σ^2 을 추측하는 값으로 다음과 같은 자료의 평균과 분산을 이용할 수 있다.

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

모집단의 분산 σ^2 을 추측하는 값으로 표본분산 S^2 을 흔히 사용하지만 이 값이 다른 추측 값보다 좋다고 얘기할 수 있을까? 예를 들어 표본분산 S^2 대신에 $S_0^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ 을 이용하여 분산 σ^2 을 추측할 수도 있다. 이때 σ^2 의 추측 값으로 S^2 와 S_0^2 중에서 어떤 것을 사용하는 것이 더 좋을까? 또는 이보다 더 좋은 추측 방법은 없을까? 어떤 것이 우수한지 판단하기 위해서는 비교기준을 마련해야 할 것이다. 그렇다면 어떤 비교기준을 사용하는 것이 타당할 것인가? 이러한 질문에 대한 답을 얻는 과정에 대한 이론적인 근거를 제공하는 것이 수리통계학에서 주로 다루는 내용이다.

다른 예로 자료분석을 통하여 두 변수 사이의 연관성을 규명하는 경우를 생각하여 보자. 이때 한 변수는 다른 변수에 영향을 받는 경우가 있다. 예를 들어 학생들을 대상으로 텔레비전 시청시간과 시험점수 사이의 연관성을 분석하는 경우 텔레비전 시청시간이 시험점수에 영향을 준다고 할 수 있다. 학생 n명을 대상으로 텔레비전 시청시간과 시험점수를 조사한 자료가 다음과 같다고 하자.

$$(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n).$$

여기서 X_i는 i번째 학생의 텔레비전 시청시간, Y_i는 시험성적을 나타낸다. 우리가 알고자 하는 것은 주어진 자료의 분석을 통하여 텔레비전 시청시간이 시험점수에 어떤 영향을 주는지에 관한 것이다. 자료를 분석하기 위한 한 가지 방법으로 원인이 되는 변수 X와 결과가 되는 변수 Y 사이에 선형관계를 가정할 수 있다. 즉 다음 관계가 성립한다고 가정할 수 있다.

$$Y_i = \alpha + \beta X_i + \epsilon_i$$

여기서 ϵ_i는 평균적으로 0의 값을 가지는 오차항을 나타낸다. 위 모형은 X와 Y 사이에는 평균적으로 절편 α, 기울기 β인 선형관계가 있음을 가정하고 있다. 오차항은 평균적으로 0의 값을 갖는다는 가정과 더불어 오차항에 대한 다음과 같이 가정할 수 있다.

$$\epsilon_i \sim \mathrm{N}(0, \sigma^2)$$

위 가정 하에서 원인이 되는 변수 X_i의 값이 주어지면 결과가 되는 변수 Y 의 확률분포는

다음과 같다.

$$Y_i \sim \mathrm{N}(\alpha + \beta X_i, \sigma^2)$$

이 가정 하에서 모집단의 확률분포를 결정짓는 모수는 α, β, σ^2 이다. 따라서 자료분석을 통하여 추측하고자 하는 것은 α, β, σ^2 에 대한 것이다. 모수에 대한 추측 값으로 흔히 사용되는 $\widehat{\alpha}, \widehat{\beta}, \widehat{\sigma}^2$ 은 다음과 같다.

$$\begin{aligned}
\widehat{\beta} &= \frac{\sum_{i=1}^n (X_i - \overline{X})(Y_i - \overline{X})}{\sum_{i=1}^n (X_i - \overline{X})^2} \\
\widehat{\alpha} &= \overline{Y} - \widehat{\beta}\overline{X} \\
\widehat{\sigma}^2 &= \frac{1}{n-2} \sum_{i=1}^n (Y_i - \widehat{\alpha} - \widehat{\beta} X_i)^2
\end{aligned}$$

수리통계학에서는 이러한 추측 방법이 어떤 기준으로 선택되었으며 추측 값이 어떤 성질을 가지게 되는지에 대한 이론적 근거를 다루게 될 것이다.

텔레비전 시청시간과 시험성적의 연관성을 규명하기 위한 모형에서 중요한 의미를 가지는 것 중의 하나가 기울기 β 의 값이 0 또는 음의 값을 가지는가에 대한 것이다. 기울기가 0이라면 텔레비전 시청시간과 시험성적 사이에는 아무런 연관성이 없다는 것을 의미하며 기울기가 음이라면 텔레비전 시청시간이 늘어나면 시험성적이 나빠진다는 것을 의미한다. 따라서 이 모형에서 기울기에 대한 관심있는 두 가지 가설은 다음과 같다.

$$\text{가설 } 0 : \beta = 0 \quad \text{대} \quad \text{가설 } 1 : \beta < 0$$

주어진 자료를 통하여 위의 두 가지 가설 중에 어떤 가설이 타당한지에 대한 판단하는 방법도 수리통계학에서 다루는 내용이다.

앞에서 설명한 것처럼 모형화를 통해 우리가 알고자 하는 대상인 모집단은 몇 개의 모수에 의하여 결정된다. 따라서 자료분석은 주어진 자료를 통하여 모수의 값을 추측을 하거나 또는 모수에 대한 가설 중 어떤 것이 타당한지 판단하는 것이라 볼 수 있다. 이러한 과정을 통계학에서는 모집단 또는 모수에 대한 추론(inference)이라 한다.

제 1 장

확률

1.1 확률의 성질

일반적인 사회현상에서는 어떤 행위나 실험의 결과가 하나로 결정되지 않고 여러 가능한 결과 중의 하나가 일어나는 경우를 흔히 볼 수 있다. 간단한 예로 주사위를 던질 때 어떤 결과가 나올지는 미리 예측할 수 없다. 우리가 알 수 있는 것은 1부터 6까지 결과 중의 하나가 나타난다는 것이다. 다른 예로 어떤 학생의 현재 키가 145cm일 때 1년 후 이 학생의 키를 예측하는 문제를 생각하여 보자. 이 학생의 연령에서는 평균적으로 1년에 5cm 성장한다면 이 학생도 1년 후 150cm 정도일 것으로 예측할 수 있다. 그러나 1년 후 이 학생의 키를 정확히 예측할 수는 없다. 148cm가 될 수도 있고 152.5cm가 될 수도 있다.

이처럼 여러 가능한 결과 중의 하나가 일어나는 현상을 확률이란 개념을 통해 설명할 수 있다. **확률**(probability)이란 여러 가능한 결과 중에서 일부분이 일어날 가능성을 0과 1 사이의 값으로 나타낸 것을 말한다. 예를 들어 동전을 던질 때 가능한 결과는 '앞면' 또는 '뒷면'의 두 가지가 있다. 이때 앞면이 나올 확률이 0.5라는 것은 앞면이 나올 가능성과 뒷면이 나올 가능성이 같다는 것을 의미한다.

확률에서 사용되는 몇 가지 용어를 알아보도록 하자. 여러 가능한 결과 중의 하나가 일어나도록 하는 행위를 **실험** 또는 **시행**이라 한다. 이러한 실험에서 나타날 수 있는 모든 결과를 모아놓은 집합을 **표본공간**(sample space)이라 하며, 표본공간의 일부분(부분집합)을 **사건**(event)이라 한다. 표본공간은 Ω 또는 S라는 기호를 사용하여 나타내며 사건은 A, B 또는 E 등과 같이 주로 영어 알파벳 대문자를 사용하여 나타낸다. 한편 사건 A가 일어날 확률을 $P(A)$ 또는 $\Pr(A)$로 나타낸다.

예를 들어 동전을 두 번 던지는 실험에서 '앞면'을 H, '뒷면'을 T라 할 때 표본공간 Ω는 다음과 같이 나타낼 수 있다.

$$\Omega = \{\ HH, HT, TH, TT\}$$

이때 앞면이 1회 이상 나올 사건 A는 다음과 같이 나타낼 수 있으며

$$A = \{HH, HT, TH\}$$

사건 A가 발생할 확률은 $P(A) = \frac{3}{4}$ 이다.

확률은 세 가지 구성요소로 이루어져있다고 볼 수 있다. 먼저 실험에서 나타날 수 있는 모든 결과를 모아놓은 집합인 **표본공간**, 결과의 일부분인 (표본공간의 부분집합인) **사건**, 사건이 일어날 가능성을 0과 1 사이의 값으로 나타낸 **확률**의 세 가지로 이루어져 있다. 즉 확률의 세 가지 구성요소는 다음과 같이 나타낼 수 있다.

$$(\Omega,\ \mathcal{A},\ P)$$

표본공간의 부분집합 위에서 정의되어 있는 확률은 다음 공리 (axiom) 를 만족하게 된다.

확률의 공리

(i) 모든 사건 A에 대하여 $0 \le P(A) \le 1$ 이다.

(ii) 표본공간 Ω에 대하여 $P(\Omega) = 1$ 이다.

(iii) 사건 $A_1, A_2, \ldots$ 이 서로 배반 (disjoint), 즉 서로 다른 i와 j에 대하여 $A_i \cap A_j = \emptyset$ 일 때 다음이 성립한다.

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

확률에 대한 두 번째와 세 번째 공리를 이용하면 $P(\emptyset) = 0$ 임을 보일 수 있다. $A_1 = \Omega$, $A_2 = A_3 = \cdots = \emptyset$ 이라 할 때 이 사건들은 서로 배반이므로 다음이 성립한다.

$$P(\Omega) = P(\Omega) + \sum_{i=2}^{\infty} P(\emptyset)$$

한편 $P(\Omega) = 1$ 이므로 $P(\emptyset) = 0$ 임을 알 수 있다.

세 번째 공리는 서로 배반인 사건에 대해여 합집합의 확률은 각 확률의 합이라는 것이다. 이 공리는 서로 배반인 유한개의 사건에 대해서도 성립함을 보일 수 있다. 세 번째 공리에서 $A_{n+1} = A_{n+2} = \cdots = \emptyset$ 이라 할 때 $\cup_{i=1}^{\infty} A_i = \cup_{i=1}^{n} A_i$ 이므로 다음이 성립한다.

$$\begin{aligned} P(\cup_{i=1}^{n} A_i) &= P(\bigcup_{i=1}^{\infty} A_i) \\ &= \sum_{i=1}^{n} P(A_i) + \sum_{i=n+1}^{\infty} P(\emptyset) \end{aligned}$$

위 식에서 $P(\emptyset) = 0$ 이므로 서로 배반인 유한개의 사건 $A_1, A_2, \ldots, A_n$ 에 대하여 다음 식이

성립함을 알 수 있다.

$$P(\bigcup_{i=1}^{n} A_i) \;=\; \sum_{i=1}^{n} P(A_i)$$

확률에 대한 공리를 이용하면 다음 정리를 얻을 수 있다.

정리 1.1 (i) 사건 A의 여집합 A^c에 대하여 $P(A^c) = 1 - P(A)$이다.

(ii) 임의의 두 사건 A와 B에 대하여 $P(A \cup B) = P(A) + P(B) - P(A \cap B)$이다.

(iii) $A \subset B$일 때 $P(A) \leq P(B)$이다.

[증명]

(i) A와 A^c는 서로 배반이고 $A \cup A^c = \Omega$이므로 $P(\Omega) = P(A) + P(A^c)$이다. 따라서 $P(A^c) = 1 - P(A)$이다.

(ii) $A \cup B = A \cup (A^c \cap B)$이고 A와 $A^c \cap B$는 서로 배반이므로 다음이 성립한다.

$$P(A \cup B) = P(A) + P(A^c \cap B) \tag{1.1}$$

한편 $B = (A \cap B) \cup (A^c \cap B)$이고 $A \cap B$와 $A^c \cap B$는 서로 배반이므로 다음이 성립한다.

$$P(B) = P(A \cap B) + P(A^c \cap B) \tag{1.2}$$

위의 두 식 (1.1)과 (1.2)로부터 $P(A \cup B) = P(A) + P(B) - P(A \cap B)$이 성립함을 보일 수 있다.

(iii) A가 B의 부분집합일 때 $B = A \cup (A^c \cap B)$이므로 $P(B) = P(A) + P(A^c \cap B)$이 성립한다. 한편 $P(A^c \cap B) \geq 0$이므로 $P(A) \leq P(B)$임을 보일 수 있다.

■

1.2 경우의 수

어떤 사건의 확률을 구하기 위해 자주 이용되는 것이 경우의 수이다. 일반적인 사건 A의 확률은 실험을 수 없이 반복할 때 사건 A가 발생하는 비율의 극한이라 할 수 있다. 즉 사건 A의 확률은 다음과 같이 정의할 수 있다.

$$P(A) = \text{실험을 반복할 때 } A\text{가 발생하는 비율의 극한} \tag{1.3}$$

가능한 결과의 수가 유한하며 모든 결과들이 같은 가능성을 가지면 위의 극한은 총 가능한 결과의 수 중에서 A에 속하는 결과의 수의 비율임을 보일 수 있다. 따라서 표본공간에 유한개의 가능한 결과만이 있고 모든 결과들이 같은 가능성을 가지고 발생할 때 사건 A의 확률은 다음과 같다.

$$P(A) = \frac{A\text{에 속하는 결과의 수}}{\text{총 가능한 결과의 수}} \tag{1.4}$$

예를 들어 동전을 두 번 던지는 실험에서 총 가능한 결과는 HH, HT, TH, TT의 네 가지이며 앞면이 한 번 이상 나오는 결과는 HH, HT, TH의 세 가지이므로 앞면이 한 번 이상 나올 확률은 $\frac{3}{4}$이다.

식 (1.4)에서 총 가능한 결과의 수 또는 A에 속하는 결과의 수를 구하기 위해서는 경우의 수를 구해야 한다. 경우의 수를 구하기 위해 이용되는 기본적인 두 가지 규칙은 다음과 같다.

합의 법칙

두 사건 A와 B가 일어나는 경우의 수가 각각 m과 n이며 A와 B는 동시에 일어나지 않을 때, 사건 A 또는 B가 일어나는 경우의 수는 $m+n$이다.

곱의 법칙

두 사건 A와 B가 일어나는 경우의 수가 각각 m과 n일 때, 두 사건 A와 B가 동시에 혹은 잇달아 일어나는 경우의 수는 mn이다.

곱의 법칙을 확장하여 여러 차례에 걸쳐 반복적으로 선택하는 경우의 수도 구할 수 있다. 즉 m_1개의 가능한 방법 중 하나를 선택하고, m_2개의 가능한 방법 중 하나를 선택하고, ..., m_k개의 가능한 방법 중 하나를 선택하는 경우의 수는 $m_1 m_2 \cdots m_k$이다.

위의 두 가지 법칙을 활용하여 경우의 수를 구하는 특별한 예가 순열과 조합이다. 먼저 서로 다른 n개의 원소 중에서 두 개를 선택하여 배열하는 경우의 수를 구하여 보자. 처음 원소를 선택할 때는 n가지의 가능한 경우가 있고, 두 번째 원소를 선택할 때는 처음 선택된 원소를 제외한 $(n-1)$개의 원소 중에서 하나를 선택하므로 $(n-1)$가지의 가능한 경우가 있다. 곱의 법칙에 의하여 총 가능한 경우의 수는 $n(n-1)$이다. 한편 서로 다른 n개의 원소 중에서 k개를 선택하는 경우는 처음 선택할 때는 n가지, 두 번째 선택할 때는 $n-1$가지, 마지막으로 k번째 원소를 선택할 때는 $(n-k+1)$개의 원소 중 하나를 선택하므로 $(n-k+1)$가지의 가능한 경우가 있다. 곱의 법칙에 의하여 총 가능한 경우의 수는 $n(n-1)\cdots(n-k+1)$이다.

예를 들어 1부터 10까지 숫자가 쓰여 있는 카드에서 3장의 카드를 뽑아 배열하는 경우의 수는 $10 \times 9 \times 8 = 720$이다. 이때 주의할 사항은 3장의 카드가 3, 7, 6의 순서로 뽑힌 경우와 6, 7, 3의 순서로 뽑힌 경우는 서로 다른 것으로 간주한다는 것이다. 선택된 3장의 카드만 본다면 같은 경우라 할 수 있지만 뽑힌 순서를 함께 고려한다면 위의 두 가지 경우는 서로 다른 것이라 할 수 있다. 이처럼 서로 다른 n개의 원소 중에서 k개를 선택하여 배열하는 경우의 수를 순열 (permutation)이라 한다.

순열 : n개의 서로 다른 원소 중에서 k개를 선택하여 배열하는 경우의 수는 $n(n-1)\cdots(n-k+1)$ 이다. 이 값을 흔히 ${}_n\mathrm{P}_k$로 나타낸다.

서로 다른 n개의 원소 중에서 k개를 선택할 때 배열하는 순서를 고려하지 않은 경우의 수를 구하여 보자. 앞에서 설명한 순열과 차이점은 배열하는 순서를 고려하지 않는다는 것이다. 예를 들어 1부터 10까지 숫자가 쓰여 있는 카드에서 3장의 카드를 선택할 때 3, 7, 6의 순서로 뽑힌 경우와 6, 7, 3의 순서로 뽑힌 경우는 순서를 고려하지 않으면 같은 것으로 볼 수 있다. 서로 다른 n개의 원소 중에서 k개를 선택할 때 뽑힌 순서를 고려하면 총 가능한 경우의 수는 $n(n-1)\cdots(n-k+1)$ 이다. 이때 선택된 k개의 원소를 재배열하여 얻을 수 있는 경우의 수는 $k! = k(k-1)\cdots 1$이고, 순서를 고려하지 않으면 재배열하여 얻는 $k!$가지가 모두 구별이 되지 않는 하나의 경우로 볼 수 있다. 예를 들어 10장의 카드 중 선택된 카드가 3, 6, 7인 경우에 순서를 고려하면 아래의 3!가지 경우가 모두 다른 경우로 볼 수 있지만 순서를 고려하지 않으면 하나의 경우로 볼 수 있다.

$$\{3,6,7\},\ \{3,7,6\},\ \{6,3,7\}$$
$$\{6,7,3\},\ \{7,3,6\},\ \{7,6,3\}$$

이처럼 서로 다른 n개의 원소 중에서 k개를 선택할 때 뽑힌 순서를 고려하지 않는 경우의 수를 조합(combination)이라 한다. 조합에서는 선택된 k개의 원소를 재배열하여 얻는 $k!$가지가 구별이 되지 않는 한 가지 경우이므로 경우의 수는 $\dfrac{n(n-1)\cdots(n-k+1)}{k!}$ 이다.

조합 : n개의 서로 다른 원소 중에서 k개를 선택하는 경우의 수는 $\dfrac{n(n-1)\cdots(n-k+1)}{k(k-1)\cdots 1}$ 이다. 이 값을 흔히 ${}_n\mathrm{C}_k$ 또는 $\dbinom{n}{k}$로 나타낸다.

예 1.1 (a) 10명의 학생 중에서 1명은 회장으로 다른 1명은 부회장으로 선출하는 경우의 수를 구하여라.

(b) 10명의 학생 중에서 2명의 학생 대표를 선출하는 경우의 수를 구하여라.

[풀이]

(a) 10명 중에서 1명은 회장으로 다른 1명은 부회장으로 선출하는 경우의 수는 10개의 서로 원소 중에서 2개를 선택하여 배열하는 경우의 수와 같다. 선택된 2명 중에서 한 명은 회장 다른 한 명은 부회장으로 선출하므로 선택되는 배열에 따라 결과가 달라진다고 할 수 있기 때문이다. 따라서 총 가능한 경우의 수는 ${}_{10}\mathrm{P}_2 = 90$이다.

(b) 단지 학생 대표 2명을 선택하므로 선택되는 순서에 따라 결과가 달라지지는 않는다. 따라서 총 가능한 경우의 수는 조합인 ${}_{10}\mathrm{C}_2 = 45$ 이다.

■

예 1.2 (a) 'MOUSE' 를 재배열하여 얻을 수 있는 경우의 수를 구하여라.

(b) 'EXAMPLE' 을 재배열하여 얻을 수 있는 경우의 수를 구하여라.

(c) 'STATISTICS' 을 재배열하여 얻을 수 있는 경우의 수를 구하여라.

[풀이]

(a) 'MOUSE' 는 5개의 서로 다른 알파벳으로 이루어져 있으므로 이를 재배열하는 경우의 수는 서로 다른 5개의 원소 중에서 5개를 선택하여 배열하는 방법의 수이다. 따라서 경우의 수는 $5! = 120$ 이다.

(b) 'EXAMPLE' 은 7개의 알파벳으로 구성되어 있지만 이 중에서 알파벳 'E' 는 두 번 반복된다. 같은 알파벳이 없다면 재배열하는 경우의 수는 $7!$ 일 것이다. 그러나 문자 'EXAMPLE' 은 같은 알파벳 'E' 가 두 번 있으며, 이들의 순서를 맞바꾸어 배열하여도 구별할 수 없다. 'E' 의 순서를 맞바꾸는 경우의 수는 $2!$ 이므로 경우의 수는 $\frac{7!}{2!} = 2520$ 이다.

(c) 10개의 알파벳이 서로 다르면 경우의 수는 $10!$ 일 것이다. 그러나 'STATISTICS' 에는 'S' 3개, 'T' 3개, 'I' 2개가 있고 같은 문자끼리 재배열하는 경우의 수는 $3! \cdot 3! \cdot 2!$ 이다. 같은 알파벳의 위치를 바꾸어 배열하여도 구별할 수 없으므로 가능한 경우의 수는 $\frac{10!}{3! \cdot 3! \cdot 2!} = 50400$ 이다.

■

예 1.3 그림 1.1과 같은 도로에서 지점 A 에서 출발하여 최단 경로를 통해 지점 B 에 도착하려 한다.

(a) 지점 A 에서 지점 B 까지 최단 경로의 수를 구하여라.

(b) 지점 A 에서 지점 B 까지 가능한 최단 경로 중에서 임의로 하나를 선택할 때 지점 C 를 거쳐갈 확률을 구하여라.

[풀이]

(a) A에서 B까지 최단 경로로 이동하기 위해서는 9단계를 거쳐야 한다. 9번의 단계 중 반드시 오른쪽으로 5번, 위로 4번 이동하여야 한다. 오른쪽으로 5번과 위로 4번의 단계 중에서 어떤 순서로 이동하느냐에 따라 경로가 달라지게 된다. 오른쪽으로 1번 이동하는 것을 'R' 위로 1번 이동하는 것을 'U' 로 나타낼 때 최단경로는 'R' 5개와 'U' 4개를 배열하는 방법으로 나타낼 수 있다. 따라서 최단 경로의 수는 'RRRRRUUUU' 를 재배열하여 얻을 수 있는 경우의 수와 같으므로 $\frac{9!}{5!4!} = 126$이다.

(b) 최단 경로 중 C를 거치는 경우의 수를 구하여보자. A에서 출발하여 최단 경로를 통해 C에 도달하기 위해서는 오른쪽으로 2번, 위로 2번 이동하여야 한다. 따라서 A에서 출발하여 C에 도달하는 경우의 수는 $\frac{4!}{2!2!}$이다. 또 C에서 출발하여 최단 경로를 통해 B에 도달하기 위해서는 오른쪽으로 3번, 위로 2번 이동하여야 하므로 가능한 경우의 수는 $\frac{5!}{3!2!}$이다. 곱의 법칙에 의하여 C를 거쳐 A에서 B까지 이동하는 최단 경로의 수는 $\frac{4!}{2!2!} \cdot \frac{5!}{3!2!} = 60$이다. 모든 가능한 최단 경로 중에서 임의로 하나를 선택하므로 C를 거쳐서 갈 확률은 다음과 같다.

$$\frac{60}{126}$$

■

예 1.4 8명의 남학생과 6명의 여학생으로 이루어진 그룹에서 4명을 선택할 때 선택된 4명 중에서 남학생이 3명일 확률을 구하여라.

[풀이] 전체 14명 중에서 4명을 선택하는 경우의 수는 $\binom{14}{4}$이다. 선택된 4명 중에서 남학생이 3명인 사건은 남학생 8명 중 3명을 선택하고 여학생 6명 중 1명을 선택하는 경우이므로

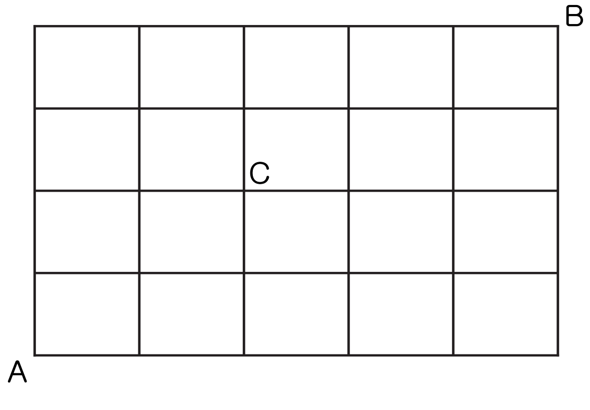

그림 1.1: 최단 경로의 수

경우의 수는 $\binom{8}{3} \cdot \binom{6}{1}$ 이다. 따라서 구하는 확률은 다음과 같다.

$$\frac{\dbinom{8}{3} \cdot \dbinom{6}{1}}{\dbinom{14}{4}} = \frac{48}{143}$$

■

1.3 조건부확률

특정한 사건의 확률을 구할 때 다른 사건에 대한 정보가 주어지는 경우가 있다. 예를 들어 동전을 두 번 던지는 실험에서 앞면이 한 번 이상 나왔다는 정보가 주어질 때 앞면이 2회 나올 확률은 구하는 경우를 생각하여 보자. 어떤 정보도 주어지지 않았다면 앞면이 2회 나올 확률은 $\frac{1}{4}$ 일 것이다. 주어진 정보 하에서도 앞면이 2회 나올 확률은 여전히 $\frac{1}{4}$ 이라고 할 수 있을까? 동전을 두 번 던지는 실험에서 나타날 수 있는 결과는 $\{HH, HT, TH, TT\}$ 의 네 가지가 있고 아무런 정보도 없다면 앞면이 2회 나오는 경우는 $\{HH\}$ 인 한 가지이므로 확률은 $\frac{1}{4}$ 이다. 그렇지만 앞면이 한 번 이상 나왔다는 정보가 주어졌다면 가능한 결과는 $\{HH, HT, TH\}$ 의 세 가지만이 가능하다. 그 중에서 앞면이 2회 나오는 경우는 한 가지이므로 주어진 정보 하에서 앞면이 2회 나올 확률은 $\frac{1}{3}$ 이라 할 수 있다.

다른 사건에 대한 정보가 주어졌을 때 어느 사건의 확률을 **조건부확률**(conditional probability)이라 한다. 조건부확률은 다른 사건에 대한 정보를 이용하여 확률을 구하므로 다른 사건에 대한 정보 없이 구한 확률과 달라질 수 있다. 조건부확률에 대해 좀더 알아보기 위해 다음 예를 생각하여 보자.

어느 반의 학생 30명을 성별과 혈액형에 따라 구분하면 다음과 같다고 한다.

	O형	A형	B형	AB형	계
남학생	5	5	3	3	16
여학생	3	4	5	2	14
계	8	9	8	5	30

이 반에서 임의로 학생을 한 명 선택할 때, 이 학생의 혈액형이 O형일 확률과 선택된 학생이 남학생이라는 정보가 주어졌을 때 혈액형이 O형일 확률을 구해보도록 하자. 전체 학생 30명 중에서 8명이 O형이므로 임의로 선택된 학생이 O형일 확률은 $\frac{8}{30}$ 이다. 만약 선택된 학생이 남학생이라는 정보가 주어졌다면 확률을 계산할 때 30명 전체를 고려할 필요는 없다. 학생의 범위를 남학생으로만 한정하면 16명 중에서 5명이 O형이므로 주어진 조건 하에서는 확률이 $\frac{5}{16}$ 이다. 여기서 분모는 남학생의 수이고 분자는 혈액형이 O형인 남학생의 수이다. 한편 위 확률은

다음과 같이 나타낼 수 있다.

$$\frac{5}{16} = \frac{5/30}{16/30}$$

위 식에서 분모인 $\frac{16}{30}$은 선택된 학생이 남학생일 확률이고 분자인 $\frac{5}{30}$은 선택된 학생이 남학생이며 혈액형이 O형일 확률이다. 즉 분모는 주어진 조건에 해당하는 사건의 확률이고 분자는 주어진 조건과 구하려는 사건을 동시에 만족하는 부분의 확률이다.

위의 예를 바탕으로 조건부확률은 다음과 같이 정의할 수 있다.

정의 1.1 $P(B) \neq 0$인 사건 B에 대하여 B가 발생했다는 정보가 주어졌을 때 사건 A가 발생할 조건부확률을 $P(A|B)$로 나타내며, $P(A|B)$는 다음과 같다.

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

예 1.5 동전을 3번 던지는 실험에서 앞면이 1회 이상 나왔다는 정보가 주어졌을 때 앞면이 2회 이상 나올 확률을 구하여라.

[풀이] 사건 A와 B를 다음과 같이 정의할 때

$$\begin{aligned} A &= \text{앞면이 2회 이상인 사건} \\ B &= \text{앞면이 1회 이상인 사건} \end{aligned}$$

$P(B) = \frac{7}{8},\ P(A \cap B) = P(A) = \frac{4}{8}$이므로 구하는 조건부확률은 다음과 같다.

$$P(A|B) = \frac{(4/8)}{(7/8)} = \frac{4}{7}$$

■

한편 조건부확률 $P(B|A)$와 $P(A)$를 이용하여 사건 $A \cap B$의 확률을 다음과 같이 구할 수도 있다.

$$P(A \cap B) = P(B|A)P(A) \tag{1.5}$$

예를 들어 52장의 카드에서 2장을 선택할 때 2장 모두 스페이드일 확률을 구해보도록 하자. 전체 경우의 수는 $\binom{52}{2}$ 이고, 2장 모두 스페이드인 경우의 수는 $\binom{13}{2}$ 이므로 구하는 확률은 다음과 같다.

$$\frac{\binom{13}{2}}{\binom{52}{2}} = \frac{(13 \times 12)/2}{(52 \times 51)/2}$$

이 확률은 식 (1.5)를 이용하여 구할 수도 있다. 사건 A와 B를 다음과 같다고 하자.

$$\begin{aligned} A &= \text{첫 번째 카드가 스페이드인 사건} \\ B &= \text{두 번째 카드가 스페이드인 사건} \end{aligned}$$

첫 번째 카드를 선택할 때 52장의 카드 중에서 스페이드는 13장이므로 사건 A의 확률은 $\frac{13}{52}$ 이다. 다음으로 첫 번째 카드가 스페이드인 조건 하에서 두 번째 카드가 스페이드일 조건부확률을 구하도록 하자. 첫 번째 선택된 카드가 스페이드라면 두 번째 카드를 선택할 때 남아있는 51장의 카드 중에서 스페이드는 12장이다. 따라서 조건부확률은 $P(B|A) = \frac{12}{51}$ 이다. 선택된 두 장 모두 스페이드인 사건은 $A \cap B$ 이고, 이 확률은 다음과 같다.

$$P(A \cap B) = P(A)P(B|A) = \frac{13}{52} \times \frac{12}{51}$$

이는 앞에서 구한 확률과 같음을 알 수 있다.

조건부확률을 이용하여 다음 정리를 얻을 수 있다.

정리 1.2 [베이즈 정리] $B_1, B_2, \ldots, B_n$ 이 서로 배반이고 $\cup_{i=1}^n B_i = \Omega$ 일 때, 다음이 성립한다.

(i) $P(A) = \displaystyle\sum_{i=1}^{n} P(A|B_i)P(B_i)$

(ii) $P(B_k|A) = \dfrac{P(A|B_k)P(B_k)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)}$

[증명]

(i) $A \cap B_1, A \cap B_2, \ldots, A \cap B_n$ 이 서로 배반이고 $A = \cup_{i=1}^n (A \cap B_i)$ 이므로 다음이 성립한다.

$$\begin{aligned} P(A) &= P\Big((A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_n)\Big) \\ &= \sum_{i=1}^{n} P(A \cap B_i) \end{aligned}$$

조건부확률의 정의에 의하여 $P(A \cap B_i) = P(A|B_i)P(B_i)$ 이므로 사건 A의 확률은 다음과 같다.

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$

(ii) $P(B_k|A) = \dfrac{P(B_k \cap A)}{P(A)}$ 이므로 (i) 에 의하여 다음이 성립한다.

$$\begin{aligned} P(B_k|A) &= \frac{P(B_k \cap A)}{P(A)} \\ &= \frac{P(A|B_k)P(B_k)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)} \end{aligned}$$

■

예 1.6 첫 번째 항아리에 3개의 흰색 공과 2개의 검정색 공이 들어있으며 두 번째 항아리에는 1개의 흰색 공과 2개의 검정색 공이 들어있다. 첫 번째 항아리에서 공 1개를 임의로 선택하여 두 번째 항아리에 옮긴 후 두 번째 항아리에서 공 하나를 임으로 선택한다고 한다.

(a) 두 번째 항아리에서 꺼낸 공이 흰색일 확률을 구하여라.

(b) 두 번째 항아리에서 꺼낸 공이 흰색일 때 첫 번째 항아리에서 꺼낸 공이 흰색이었을 확률을 구하여라.

[풀이]

(a) 사건 A와 $B_1,\ B_2$를 다음과 같이 정의하자.

A = 두 번째 항아리에서 꺼낸 공이 흰색인 사건
B_1 = 첫 번째 항아리에서 꺼낸 공이 흰색인 사건
B_2 = 첫 번째 항아리에서 꺼낸 공이 검정색인 사건

$B_1 \cap B_2 = \emptyset$ 이고 $B_1 \cup B_2 = \Omega$ 이므로 베이즈 정리에 의하여 사건 A의 확률은 다음과 같다.

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2)$$

$P(B_1)=\frac{3}{5},\ P(B_2)=\frac{2}{5}$ 이고 $P(A|B_1)=\frac{2}{4},\ P(A|B_2)=\frac{1}{4}$ 이므로 두 번째 항아리에서 꺼낸 공이 흰색일 확률은 다음과 같다.

$$P(A)=P(A|B_1)P(B_1)+P(A|B_2)P(B_2)=\frac{2}{4}\cdot\frac{3}{5}+\frac{1}{4}\cdot\frac{2}{5}=\frac{8}{20}$$

(b) 두 번째 항아리에서 꺼낸 공이 흰색이었을 때 첫 번째 항아리에서 꺼낸 공이 흰색이었을 확률은 $P(B_1|A)$ 이다. 베이즈 정리에 의하여 $P(B_1|A)$ 는 다음과 같다.

$$P(B_1|A)=\frac{P(A|B_1)P(B_1)}{P(A)}=\frac{2/4\cdot 3/5}{8/20}=\frac{6}{8}$$

■

예 1.7 컴퓨터 C 에서 컴퓨터 D 로 '0' 또는 '1' 의 정보를 전송한다. 전송 중 발생할 수 있는 오류 때문에 컴퓨터 C 에서 '0' 을 전송하였을 때 D 에서 '1' 로 잘못 인식할 확률이 0.05, 컴퓨터 C 에서 '1' 을 전송하였을 때 D 에서 '0' 으로 잘못 인식할 확률이 0.03이라고 한다. 또한 컴퓨터 C 에서 '0' 을 전송할 확률은 0.6이며 '1' 을 전송할 확률은 0.4라고 한다.

(a) 컴퓨터 C 에서 정보를 전송하였을 때 컴퓨터 D 에서 '0' 으로 인식할 확률을 구하여라.

(b) 컴퓨터 D 에서 '0' 으로 인식하였을 때 컴퓨터 C 에서 전송한 정보가 '0' 이었을 확률을 구하여라.

[풀이]

(a) 사건 A 와 $B_1,\ B_2$ 를 다음과 같이 정의하자.

A = 컴퓨터 C 에서 전송한 정보를 컴퓨터 D 에서 '0' 으로 인식하는 사건
B_1 = 컴퓨터 C 에서 '0' 을 전송하는 사건
B_2 = 컴퓨터 C 에서 '1' 을 전송하는 사건

두 사건 B_1 과 B_2 는 서로 배반이고 $B_1\cup B_2=\Omega$ 이므로 베이즈 정리에 의하여 $P(A)$ 는 다음과 같다.

$$P(A)=P(A|B_1)P(B_1)+P(A|B_2)P(B_2)$$

주어진 조건에 의하여 $P(B_1) = 0.6$, $P(B_2) = 0.4$이고 $P(A|B_1) = 0.95$, $P(A|B_2) = 0.03$이므로 $P(A)$는 다음과 같다.

$$\begin{aligned} P(A) &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) \\ &= 0.95 \cdot 0.6 + 0.03 \cdot 0.4 \ = \ 0.582 \end{aligned}$$

(b) 컴퓨터 D에서 '0'으로 인식하였을 때 컴퓨터 C에서 전송한 정보가 '0'이었을 확률 $P(B_1|A)$는 다음과 같다.

$$P(B_1|A) \ = \ \frac{P(A|B_1)P(B_1)}{P(A)} \ = \ \frac{0.95 \cdot 0.6}{0.95 \cdot 0.6 + 0.03 \cdot 0.4} \ = \ 0.9794$$

■

1.4 독립

일반적으로 두 사건은 서로 연관성이 있는 경우가 많다. 예를 들어 동전을 두 번 던지는 실험에서 A는 앞면이 2회인 사건, B는 앞면이 1회 이상인 사건이라 할 때 사건 A가 발생했다는 정보가 주어지면 사건 B는 반드시 발생하였다고 할 수 있다. 이처럼 두 사건은 연관성을 가지고 있는 경우가 있다.

두 사건의 연관성은 조건부확률을 통해 살펴볼 수 있다. 조건부확률 $P(A|B)$는 사건 B에 대한 정보가 주어진 상황에서 사건 A의 확률을 나타낸다. 만약 두 사건 A와 B 사이에 연관성이 없다면 사건 B에 대한 정보는 A의 확률에 아무런 영향을 주지 못할 것이다. 따라서 조건부확률 $P(A|B)$는 $P(A)$와 같은 값을 가질 것이다. 반대로 두 사건 사이에 어떤 연관성이 있다면 사건 B에 대한 정보는 A에 대한 정보를 일정 부분 포함한다고 할 수 있으므로 조건부확률 $P(A|B)$는 $P(A)$와 다른 값을 가질 것이다.

두 사건 사이에 서로 연관성이 없을 때 독립이라 한다. 일반적으로 독립의 정의는 다음과 같이 할 수 있다.

정의 1.2 두 사건 A와 B에 대하여 다음이 성립할 때

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

A와 B는 서로 **독립**(independent)이라 한다.

위 정의로부터 두 사건 A와 B가 서로 독립일 때 다음이 성립함을 알 수 있다.

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

마찬가지로 $P(B|A) = P(B)$이다. 이 식의 의미는 사건 B에 대한 정보가 A의 확률에 영향을 주지 않는다는 것을 의미한다. 반대로 $P(B|A) = P(B)$이면 $P(B \cap A) = P(A) \cdot P(B)$임을 알 수 있다.

예 1.8 52장의 카드 중에서 임의로 한 장을 꺼내는 실험에서 선택된 카드가 스페이드일 사건을 A, 에이스일 사건을 B, 스페이드 에이스일 사건을 C라 할 때 다음 물음에 답하여라.

(a) A와 B는 서로 독립인지 확인하여라.

(b) A와 C는 서로 독립인지 확인하여라.

[풀이]

(a) 52장의 카드 중에서 스페이드는 13장이므로 A의 확률은 $\frac{13}{52}$이다. 한편 사건 B가 발생하였다는 정보가 주어지면 선택된 카드는 에이스임을 의미한다. 에이스 4장의 카드 중에서 스페이드는 1장이므로 조건부확률은 $P(A|B) = \frac{1}{4}$이고 이 조건부확률은 $P(A)$와 같다. 따라서 A와 B는 서로 독립이다. 또한 A와 B의 확률은 각각 다음과 같고

$$P(A) = \frac{13}{52},\ P(B) = \frac{4}{52}$$

스페이드이며 에이스인 카드는 52장 중에서 오직 한 장이므로 $A \cap B$의 확률은 다음과 같다.

$$P(A \cap B) = \frac{1}{52}$$

따라서 $P(A \cap B) = P(A) \cdot P(B)$임을 확인할 수도 있다.

(b) 사건 C가 발생하였다는 정보가 주어지면 선택된 카드는 스페이드 에이스임을 나타낸다. 따라서 주어진 정보로부터 선택된 카드는 스페이드임을 알 수 있다. 즉 조건부확률 $P(A|C) = 1$이고, 이 확률은 사건 A의 확률 $P(A) = \frac{13}{52}$과 다른 값이다. 따라서 A와 C는 서로 독립이 아님을 알 수 있다. 또한 A, C, $A \cap C$의 확률은 다음과 같다.

$$P(A) = \frac{13}{52},\ P(C) = \frac{1}{52},\ P(A \cap C) = \frac{1}{52}$$

이로부터 $P(A \cap C) \neq P(A) \cdot P(C)$임을 확인할 수도 있다.

■

두 사건의 독립을 확장하여 n개의 사건에 대한 독립의 정의는 다음과 같이 할 수 있다.

정의 1.3 $A_1, A_2, \ldots, A_n$에서 임의로 선택한 $k(k = 1, 2, \ldots, n)$개의 모임 $A_{i_1}, A_{i_2}, \ldots, A_{i_k}$에 대하여 다음이 성립할 때 $A_1, A_2, \ldots, A_n$은 서로 독립이라 한다.

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdots P(A_{i_k})$$

위의 정의에서 한 가지 주의할 사항은 n개의 사건의 독립을 다음과 같이 정의해서는 안된다.

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1) \cdot P(A_2) \cdots P(A_n) \tag{1.6}$$

예를 들어 $A_n = \emptyset$이라 하면 위 식의 좌변과 우변 모두 0으로 같게 된다. 따라서 위의 식 (1.6)을 독립의 정의로 생각하면 $A_1, A_2, \ldots, A_{n-1}$에 상관없이 $A_1, A_2, \ldots, A_n$은 항상 독립임을 의미하게 된다. n개의 사건의 독립은 모든 가능한 사건의 조합 $A_{i_1}, A_{i_2}, \ldots, A_{i_k}$에 대하여 다음이 성립하

여야 한다.

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdots P(A_{i_k})$$

연습문제

1. 경우의 수를 구하여라.

   a) 'STATEMENT'를 재배열하여 만들 수 있는 서로 다른 배열의 수.

   b) 총 6개의 서로 다른 선물을 10명에게 나누어줄 때 어느 한 명도 2개 이상의 선물을 받지 않도록 나누어주는 경우의 수.

   c) 같은 색의 공 20개를 서로 다른 4개의 항아리에 분배할 때 4개의 항아리에 최소한 각각 1, 2, 3, 4개의 공이 분배되도록 하는 경우의 수.

   d) 남녀 4명씩 8명을 일렬로 배열할 때 같은 성별끼리는 이웃하지 않도록 하는 경우의 수.

   e) A,B,C,D 네 명을 포함한 10명의 사람을 일렬로 배열할 때 A와 B는 이웃하고 C와 D는 옆에 이웃하지 않도록 배열하는 경우의 수.

   f) 52장의 카드 중에서 5장의 카드를 선택할 때 풀하우스(숫자가 a, a, a, b, b)에 해당하는 경우의 수 (카드의 뽑힌 순서는 무시).

2. 항아리에는 검정색 공 2개, 흰색 공 3개, 빨간색 공 4개가 들어 있다. 이 항아리에서 임의로 2개의 공을 선택할 때 다음 물음에 답하여라.

   a) 두 개의 공이 같은 색일 확률을 구하여라.

   b) 두 개의 공이 같은 색이라고 할 때, 모두 흰색일 확률을 구하여라.

3. 항아리 A에는 3개의 흰색 공과 2개의 검정색 공이 들어있고, 항아리 B에는 2개의 흰색 공과 1개의 검정색 공이 들어있다. 먼저 두 항아리에서 임의로 공 하나씩을 선택하여 색을 관찰한 후 원래의 항아리에 집어넣는다. 이때 두 공의 색이 같은 경우에는 다시 항아리 A에서 임의로 하나의 공을 선택하고 색이 다른 경우에는 항아리 B에서 임의로 공을 하나 선택할 때 다음 물음에 답하여라.

   a) 최종적으로 선택된 공이 흰색일 확률을 구하여라.

   b) 최종적으로 선택된 공이 흰색이었을 때 먼저 꺼낸 두 개의 공이 같은 색이었일 확률을 구하여라.

4. 항아리 A에는 흰색 공 3개와 검정색 공 2개가 들어있고, 항아리 B에는 흰색 공 1개와 검정색 공 2개가 들어 있다. 항아리 A에서 임의로 공 2개를 선택하고, 두 개의 색이 같으면 이 중 하나만 항아리 B로 옮긴다. 두 개의 색이 다르면 공 두 개를 모두 항아리 B로 옮긴다. 이후 항아리 B에서 임의로 공 하나를 선택할 때 다음 물음에 답하여라.

   a) 항아리 B에서 선택한 공이 검정색일 확률을 구하여라.

   b) 항아리 B에서 선택한 공이 검정색이었을 때, 항아리 A에서 꺼낸 두 개의 공이 서로 다른 색이었을 확률을 구하여라.

5. 다음과 같은 게임을 한다고 하자. 먼저 10원과 100원짜리 동전을 던져 둘 다 앞면이면 이기고, 둘 다 뒷면이면 진다. 다른 경우에는 앞면이 나온 동전을 선택하여 뒷면이 나올 때까지 던진다고 한다. 이때 3번 안에 뒷면이 나오면 게임을 이기고, 그렇지 못하면 게임을 진다. 이 게임에서 이길 확률을 구하여라. 단, 10원짜리 동전의 앞면이 나올 확률은 p_1 이고, 100원짜리 동전의 앞면이 나올 확률은 p_2 라고 한다.

6. 두 사건 A와 B가 서로 독립일 때 다음 사건이 독립임을 보여라.

   a) A와 B^c

   b) A^c와 B

   c) A^c와 B^c

7. 다음 주장에 대하여 옳고 그름을 판별하고 그 이유를 설명하여라.

   a) $A \subset B$ 이면 $P(B|A) = 1$ 이다.

   b) $A \subset B$ 이면 $P(A|B) \geq P(A)$ 이다.

   c) $P(A|B) > P(A)$ 이면 $P(B|A) > P(B)$ 이다.

   d) $P(A) = P(B) = 0.6$ 일 때 A와 B는 서로 배반일 수 있다.

제 2 장

확률변수와 확률분포

우리가 다루는 자료는 대부분 양적 의미를 가지는 숫자로 구성되어 있다. 이러한 수는 고정된 값이 아니라 상황에 따라 달라질 수 있는 값이다. 예를 들어 임의로 선택된 사람의 키를 나타낸 자료는 사람에 따라 다른 값을 갖게 된다. 키는 여러 가지 가능한 값 중에서 하나의 값을 가지지만 모든 값이 같은 가능성을 가지고 있지는 않을 것이다. 키의 값이 170cm 근처에 있을 가능성이 190cm 근처에 있을 가능성보다는 높을 것이다. 이처럼 자료가 여러 가지 가능한 값 중에서 하나를 가질 때 특정한 값을 가질 확률을 나타낼 필요성이 있다.

1장에서 다룬 확률은 표본공간의 부분집합인 사건 위에 정의되어 있다. 표본공간은 행해지는 실험에 따라 달라지게 된다. 예를 들어 동전을 두 번 던지는 실험에서 표본공간은 $\Omega = \{HH, HT, TH, TT\}$ 이며 주사위를 던지는 실험에서 표본공간은 $\Omega = \{1, 2, 3, 4, 5, 6\}$ 이다. 우리가 다루는 자료는 대부분 숫자로 구성되어 있으므로 여러 가지 가능한 값 중 하나를 가질 확률을 다룰 때 표본공간 대신에 실수에서 정의된 확률을 생각할 필요성이 있다. 실수에서 정의된 확률은 표본공간을 실수로 변환하는 함수를 통해 도입할 수 있다. 이처럼 표본공간을 실수로 바꾸어주는 함수를 확률변수라 한다.

정의 2.1 표본공간에서 정의된 실수로의 함수를 **확률변수**(random variable)라 한다.

확률변수는 흔히 X, Y 또는 Z 처럼 주로 영어의 알파벳 뒷부분의 대문자를 사용하여 나타낸다. 확률변수는 표본공간을 실수로 바꾸어 주는 함수이므로 다음과 같이 나타낼 수 있다.

$$X \ : \ \Omega \to R$$

예를 들어 동전을 2번 던지는 실험에서 표본공간은 $\Omega = \{HH, HT, TH, TT\}$ 이다. 이 실험에서 확률변수 X 를 앞면의 수라 할 때, X 는 표본공간의 각 원소를 그림 2.1과 같이 실수 값 0, 1, 2로 변환하여 준다. 즉 표본공간 $\Omega = \{HH, HT, TH, TT\}$ 위에서 정의된 함수 X 는 다음과 같이 나타낼 수 있다.

$$X(HH) = 2,\ X(HT) = 1,\ X(TH) = 1,\ X(TT) = 0.$$

확률변수에 의하여 우리가 다루는 영역이 표본공간에서 실수로 바뀜에 따라 표본공간에서 정의된 확률도 실수에서 정의된 확률로 변환하여 생각할 수 있다. 실수에서 정의된 확률은 표본공간의 원소가 실수의 어떤 값으로 변환되었는지에 따라 결정된다. 예를 들어 동전을 2번 던지는 실험에서 앞면의 수를 나타내는 확률변수 X는 0, 1, 2 중에서 하나의 값을 가진다. X가 2의 값을 가지는 경우는 표본공간에서 생각하면 $\{HH\}$에 대응되는 사건이므로 $\{X = 2\}$의 확률은 $\{HH\}$가 발생할 확률로 구할 수 있다. 같은 방법으로 확률변수 X에 대한 확률은 다음과 같이 구할 수 있다.

$$\begin{aligned} P(X=2) &= P(HH) = \frac{1}{4} \\ P(X=1) &= P(HT, TH) = \frac{1}{2} \\ P(X=0) &= P(TT) = \frac{1}{4} \end{aligned}$$

예 2.1 앞면이 나올 때까지 동전을 던지는 실험에서 총 동전을 던지는 횟수를 X라 할 때 표본공간의 각 원소가 X에 의하여 어떤 실수 값으로 변환되는지 구하고, X에 의하여 실수의 부분집합에서 정의된 새로운 확률을 구하여라.

[**풀이**] 이 실험에서 나타날 수 있는 결과는 첫 번째 동전의 결과가 앞면인 경우로부터 계속하여 뒷면이 나와 끊임없이 동전을 던지는 경우까지 가능하다. 즉 표본공간은 다음과 같이 나타낼 수 있다.

$$\Omega = \{H,\ TH,\ TTH,\ TTTH,\ \ldots\}$$

앞면이 나올 때까지 동전을 던지는 횟수가 k라는 것은 처음 $k-1$번은 연속하여 뒷면이 나오고 마지막 k번째에서 처음으로 앞면이 나온 것을 나타낸다. 따라서 표본공간의 각

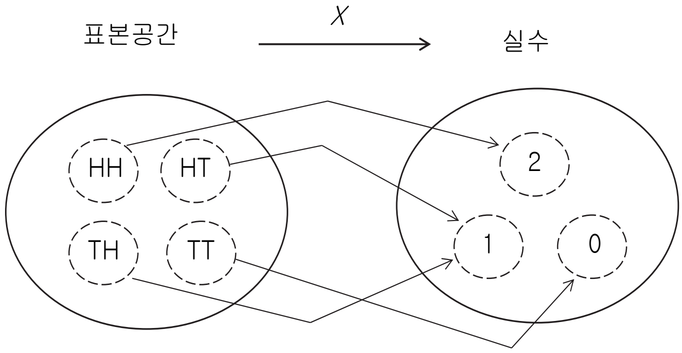

그림 2.1: 동전을 2번 던질 때 나오는 앞면의 수를 나타내는 확률변수 X

원소는 다음과 같은 실수 값에 대응된다.

$$\begin{aligned} X(H) &= 1 \\ X(TH) &= 2 \\ X(TTH) &= 3 \\ X(TTTH) &= 4 \\ &\vdots \end{aligned}$$

이때 확률변수 X 에 의하여 실수에서 정의된 새로운 확률은 다음과 같이 구할 수 있다.

$$\begin{aligned} P(X=1) &= P(H) = \frac{1}{2} \\ P(X=2) &= P(TH) = \frac{1}{4} \\ P(X=3) &= P(TTH) = \frac{1}{8} \\ P(X=4) &= P(TTTH) = \frac{1}{16} \\ &\vdots \end{aligned}$$

■

확률변수에 의하여 우리는 표본공간의 부분집합 위에서 정의된 확률 대신에 실수의 부분집합에서 정의된 확률을 다루게 된다. 확률변수에 대하여 실수의 각 부분집합에서 확률이 어떻게 주어지는지 나타낸 것을 **확률분포**(probability distribution)라 한다. 한편 실수의 부분집합의 형태는 한 값을 원소로 가지는 간단한 경우일 수도 있고 구간으로 주어질 수도 있다. 또한 여러 구간의 합집합 형태의 부분집합도 있을 수 있다. 따라서 확률변수의 확률분포는 가능한 여러 가지 형태의 부분집합에서 확률을 나타내야 한다. 앞에서 다룬 예에서는 확률변수가 몇 개의 값만을 가지는 경우이므로 이 값에서의 확률만 나타내면 되었지만 확률변수가 실수 전체와 같이 큰 범위에서 값을 가진다면 가능한 여러 형태의 부분집합에서 확률을 나타내어야 한다. 예를 들어 다음과 같은 여러 부분집합에 대하여 확률을 나타내어야 한다.

$$3,\ [2,5],\ (-2,0),\ [1,3),\ (-3,2],\ [2,5]\cup(-2,0),\ldots$$

즉 확률변수 X 의 확률분포는 다음과 같은 확률 등에 의하여 결정된다.

$$\begin{gathered} P(X=3),\ P(2\le X\le 5),\ P(-2<X<0), \\ P(1\le X<3),\ P(-3<X\le 2),\ P(X\in[2,5]\cup(-2,0)) \end{gathered}$$

한 가지 알려진 사실은 확률변수의 확률분포는 $(-\infty, x]$[1] 형태를 가지는 구간의 확률에 의하여 결정된다는 것이다. 따라서 확률변수의 확률분포를 나타내기 위해서 여러 가지 가능한 실수의 각 부분집합의 확률을 나타내는 대신에 구간 $(-\infty, x]$의 확률만 나타내면 된다. 확률변수가 구간 $(-\infty, x]$에 속할 확률을 $F(x)$로 나타내며, 이를 **누적분포함수**(cumulative distribution function)라 한다. 즉 누적분포함수 $F(x)$는 다음과 같다.

$$F(x) \ = \ P(X \leq x)$$

구간 $(-\infty, x]$의 확률로 정의된 누적분포함수는 다음과 같은 성질을 가지고 있다. 두 실수 x와 y에 대하여 $x \leq y$일 때, $\{X \leq x\} \subset \{X \leq y\}$이므로 $P(X \leq x) \leq P(X \leq y)$이다. 따라서 x의 값이 커질 때 $F(x)$의 값은 증가함을 알 수 있다. 한편 확률이 가지고 있는 한 가지 성질은 집합 A_n이 A로 가까이 갈 때 A_n의 확률도 A의 확률로 수렴한다는 것이다. 즉 다음이 성립한다.

$$A_n \to A \text{이면} \quad P(A_n) \to P(A)$$

위 사실을 이용하면 누적분포함수가 가지는 몇 가지 성질을 밝힐 수 있다. x의 값이 ∞로 가까이 가면 $\{X \leq x\}$은 전체집합에 접근하게 된다. 따라서 확률은 1에 수렴하게 된다. 반대로 x의 값이 $-\infty$로 가까이 가면 $\{X \leq x\}$은 공집합 접근하게 되므로 확률은 0에 수렴하게 된다. 또한 구간 $(-\infty, x+\frac{1}{n}]$은 n이 커질 때 $(-\infty, x]$로 가까이 간다. 따라서 $P(X \leq x+\frac{1}{n})$은 $P(X \leq x)$로 수렴하게 된다. 이를 누적분포함수로 나타내면 다음과 같다.

$$\lim_{n\to\infty} F\left(x+\frac{1}{n}\right) \ = \ F(x) \tag{2.1}$$

식 (2.1)이 의미하는 것은 x를 우측으로부터 접근할 때 누적분포함수의 값도 $F(x)$로 수렴한다는 것이다. 이러한 성질을 가지는 함수를 우측에서 연속인 함수라 한다.

이번에는 x를 좌측으로부터 접근할 때 누적분포함수의 값이 어떤 값으로 수렴하는지 알아보도록 하자. 한 가지 주의할 사실은 구간 $(-\infty, x-\frac{1}{n}]$은 n이 커질 때 $(-\infty, x]$로 가까이 가지 않고 $(-\infty, x)$로 수렴한다는 사실이다. 구간 $(-\infty, x]$는 x를 포함하지만 $(-\infty, x)$는 x를 포함하지 않는 구간이다. 따라서 다음이 성립한다.

$$P\left(X \leq x-\frac{1}{n}\right) \to P(X < x)$$

이를 누적분포함수로 나타내면 다음과 같다.

$$\lim_{n\to\infty} F\left(x-\frac{1}{n}\right) \ = \ P(X < x) \tag{2.2}$$

일반적으로 $P(X < x) \leq P(X \leq x)$이며 $P(X = x) = 0$일 때는 $P(X < x) = P(X \leq x)$이 된다.

[1]통계학에서는 X와 x를 구분하여 사용한다. X는 확률변수를 나타내고 x는 실수 값을 나타낸다.

따라서 누적분포함수는 다음과 같은 성질을 가지고 있음을 알 수 있다.

(i) $F(x)$는 증가함수이다. 즉 $x \le y$일 때 $F(x) \le F(y)$이다.

(ii) $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$이다.

(iii) $F(x)$는 우측에서 연속인 함수이다. 즉 $\lim_{n \to \infty} F\left(x + \frac{1}{n}\right) = F(x)$이다.

예 2.2 확률변수 X의 누적분포함수 $F(x)$가 다음과 같다고 하자.

$$F(x) = \begin{cases} 0, & x < 0 \text{일 때} \\ x^2, & 0 \le x < \dfrac{1}{4} \text{일 때} \\ \frac{1}{2}, & \dfrac{1}{4} \le x < \dfrac{1}{2} \text{일 때} \\ x, & \dfrac{1}{2} \le x < 1 \text{일 때} \\ 1, & 1 \le x \text{일 때} \end{cases}$$

이때 다음 확률을 구하여라.

(a) $P(X \le \frac{1}{8})$ (b) $P(X < \frac{1}{8})$ (c) $P(X = \frac{1}{8})$ (d) $P(X \le \frac{1}{4})$
(e) $P(X < \frac{1}{4})$ (f) $P(X = \frac{1}{4})$ (g) $P(\frac{1}{4} < X \le \frac{3}{4})$ (h) $P(\frac{1}{4} \le X \le \frac{3}{4})$

[풀이]

(a) 누적분포함수의 정의에 의하여 $P(X \le \frac{1}{8}) = F\left(\frac{1}{8}\right)$이므로 구하는 확률은 다음과 같다.

$$P\left(X \le \frac{1}{8}\right) = F\left(\frac{1}{8}\right) = \left(\frac{1}{8}\right)^2 = \frac{1}{64}$$

(b) 식 (2.2)에 의하여 구하는 확률은 다음과 같다.

$$P\left(X < \frac{1}{8}\right) = \lim_{n \to \infty} F\left(\frac{1}{8} - \frac{1}{n}\right) = \lim_{n \to \infty} \left(\frac{1}{8} - \frac{1}{n}\right)^2 = \frac{1}{64}$$

(c) $\left(-\infty, \frac{1}{8}\right] = \left(-\infty, \frac{1}{8}\right) \cup \left\{\frac{1}{8}\right\}$이므로 $P\left(X \le \frac{1}{8}\right) = P\left(X < \frac{1}{8}\right) + P\left(X = \frac{1}{8}\right)$이다. 따라서 구하는 확률은 다음과 같다.

$$P\left(X = \frac{1}{8}\right) = P\left(X \le \frac{1}{8}\right) - P\left(X < \frac{1}{8}\right) = \frac{1}{64} - \frac{1}{64} = 0$$

(d) $X \le \frac{1}{4}$일 확률은 $P\left(X \le \frac{1}{4}\right) = F\left(\frac{1}{4}\right) = \frac{1}{2}$이다.

(e) $X < \frac{1}{4}$일 확률은 다음과 같다.

$$P\left(X < \frac{1}{4}\right) = \lim_{n\to\infty} F\left(\frac{1}{4} - \frac{1}{n}\right) = \lim_{n\to\infty}\left(\frac{1}{4} - \frac{1}{n}\right)^2 = \frac{1}{16}$$

(f) (d)와 (e)로부터 $X = \frac{1}{4}$일 확률은 다음과 같다.

$$P\left(X = \frac{1}{4}\right) \ = \ P\left(X \le \frac{1}{4}\right) - P\left(X < \frac{1}{4}\right) \ = \ \frac{1}{2} - \frac{1}{16} = \frac{7}{16}$$

(g) $\left(-\infty, \frac{3}{4}\right] = \left(-\infty, \frac{1}{4}\right] \cup \left(\frac{1}{4}, \frac{3}{4}\right]$이므로 $P\left(X \le \frac{3}{4}\right) = P\left(X \le \frac{1}{4}\right) + P\left(\frac{1}{4} < X \le \frac{3}{4}\right)$이다. 따라서 구하는 확률은 다음과 같다.

$$\begin{aligned} P\left(\tfrac{1}{4} < X \le \tfrac{3}{4}\right) &= P\left(X \le \frac{3}{4}\right) - P\left(X \le \frac{1}{4}\right) \\ &= \frac{3}{4} - \frac{1}{2} = \frac{1}{4} \end{aligned}$$

(h) $\left(-\infty, \frac{3}{4}\right] = \left(-\infty, \frac{1}{4}\right) \cup \left[\frac{1}{4}, \frac{3}{4}\right]$이므로 $P\left(X \le \frac{3}{4}\right) = P\left(X < \frac{1}{4}\right) + P\left(\frac{1}{4} \le X \le \frac{3}{4}\right)$이다. 따라서 구하는 확률은 다음과 같다.

$$\begin{aligned} P\left(\tfrac{1}{4} \le X \le \tfrac{3}{4}\right) &= P\left(X \le \frac{3}{4}\right) - P\left(X < \frac{1}{4}\right) \\ &= \frac{3}{4} - \frac{1}{16} = \frac{11}{16} \end{aligned}$$

■

앞에서 확률변수의 확률분포는 $(-\infty, x]$의 확률, 즉 누적분포함수에 의하여 결정된다고 하였다. 동전을 두 번 던질 때 앞면의 수를 나타내는 확률변수 X의 경우를 이용하여 이를 확인해보도록 하자. X는 음의 값을 가지지 않으므로 0보다 작은 x에 대하여 누적분포함수의 값은 다음과 같다.

$$F(x) \ = \ 0, \quad x < 0 \text{일 때} \tag{2.3}$$

또 X는 정수 값만 가지므로 0보다 크거나 같고 1보다 작은 x에 대하여 $X \in (-\infty, x]$라는 것은 $X = 0$인 경우 밖에 없다. 따라서 누적분포함수의 값은 다음과 같다.

$$F(x) \ = \ P(X = 0) \ = \ \frac{1}{4}, \quad 0 \le x < 1 \text{일 때} \tag{2.4}$$

이다. 1보다 크거나 같고 2보다 작은 x에 대하여 $X \in (-\infty, x]$라는 것은 $X = 0$ 또는 $X = 1$인 경우이다. 따라서 누적분포함수의 값은 다음과 같다.

$$F(x) \ = \ P(X \in \{0, 1\}) \ = \ \frac{3}{4}, \quad 1 \le x < 2 \text{일 때} \tag{2.5}$$

한편 2보다 크거나 같은 x에 대하여 $X \in (-\infty, x]$라는 것은 $X = 0, 1$ 또는 $X = 2$인 경우이다. 따라서 누적분포함수의 값은 다음과 같다.

$$F(x) \;=\; P(X \in \{0, 1, 2\}) \;=\; 1, \quad 2 \le x \text{일 때} \tag{2.6}$$

반대로 누적분포함수 (2.3), (2.4), (2.5), (2.6) 로부터 다음을 얻을 수 있다.

$$\begin{aligned} P(X = 0) \quad &= P(X \le 0) - P(X < 0) \\ &= F(0) - \lim_{n\to\infty} F\left(X \le -\frac{1}{n}\right) \\ &= \frac{1}{4} - \lim_{n\to\infty} 0 \;=\; \frac{1}{4} \end{aligned}$$

$$\begin{aligned} P(X = 1) \quad &= P(X \le 1) - P(X < 1) \\ &= F(1) - \lim_{n\to\infty} F\left(X \le 1 - \frac{1}{n}\right) \\ &= \frac{3}{4} - \lim_{n\to\infty} \frac{1}{4} \;=\; \frac{1}{2} \end{aligned}$$

$$\begin{aligned} P(X = 2) \quad &= P(X \le 2) - P(X < 2) \\ &= F(2) - \lim_{n\to\infty} F\left(X \le 2 - \frac{1}{n}\right) \\ &= 1 - \lim_{n\to\infty} \frac{3}{4} \;=\; \frac{1}{4} \end{aligned}$$

이는 X의 확률분포가 0, 1, 2에서 각각의 확률이 $\frac{1}{4}$, $\frac{1}{2}$, $\frac{1}{4}$임을 보여주고 있다. 따라서 누적분포함수로부터 X의 확률분포를 구할 수 있다.

확률변수는 몇 가지 가능한 값 중의 하나를 가지는 경우도 있고 어떤 구간에 속하는 무수히 많은 값 중의 하나를 가지는 경우도 있다. 확률변수는 가질 수 있는 값의 범위에 따라 이산형 확률변수와 연속형 확률변수로 구분할 수 있다.

2.1 이산형 확률변수

확률변수가 가질 수 있는 각 값에서의 확률에 의하여 확률분포가 결정될 때 이를 이산형 확률변수라 한다. 그런데 확률변수가 가지는 각 값에서 확률을 구할 수 있는 경우는 확률변수가 가지는 값의 수가 유한개 뿐만이 아니라 무한히 많은 경우도 있을 수 있다. 예를 들어 앞면이 나올 때까지 계속하여 동전을 던지는 실험에서 던지는 총 횟수를 나타내는 확률변수 X를 생각하여 보자.

처음 동전을 던져 앞면이 나오면 $X = 1$이고, 앞면이 100번째에 처음으로 앞면이 나오면 $X = 100$이다. 또한 계속하여 뒷면이 나올 수도 있으므로 확률변수 X가 가질 수 있는 값은 상한이 없다. 이 실험에서 $X = 1$일 확률은 첫 번째 동전의 결과가 앞면일 확률이므로 $\frac{1}{2}$이고,

$X = 100$일 확률은 처음 99번은 뒷면이 나오고 100번째에 앞면이 나올 확률이므로 $\left(\frac{1}{2}\right)^{99} \cdot \frac{1}{2}$ 이다. 이 확률변수는 자연수의 모든 값을 가질 수 있고 각 값에서 확률을 구할 수 있으므로 이산형 확률변수라 할 수 있다. 이처럼 확률변수가 가질 수 있는 각 값에서 확률을 구할 수 있는 경우일지라도 확률변수가 가지는 값의 수는 무한히 많을 수 있다. 단, 무한히 많기는 하지만 동전을 던지는 예처럼 자연수 정도로 많은 정도이다.

확률변수가 가질 수 있는 값이 유한개 또는 무한히 많더라도 셀 수 있을 때(countable) 이를 **이산형 확률변수**(discrete random variable) 라 한다. 어느 집합이 셀 수 있다는 것은 이 집합에 속하는 원소를 하나씩 세어나갈 때 모든 원소를 빠뜨리지 않고 전부 헤아릴 수 있는 경우를 말한다. 예를 들어 자연수의 집합은 1부터 차례대로 끊임없이 세어나가면 모든 자연수를 헤아릴 수 있으므로 자연수는 셀 수 있는 집합이다. 자연수의 집합 외에도 대표적인 셀 수 있는 집합으로 정수와 유리수 등이 있다. 예를 들어 다음과 같은 것이 이산형 확률변수이다.

- 동전을 n번 던져 나오는 앞면의 수
- 앞면이 연속하여 두 번 나올 때까지 동전을 던지는 횟수
- 하루 동안 발생하는 교통사고의 수

동전을 n번 던져 나오는 앞면의 수는 0부터 n 사이의 정수 값을 가지므로 확률변수가 가질 수 있는 값의 수는 유한개인 경우이다. 앞면이 연속하여 두 번 나올 때까지 동전을 던지는 횟수를 나타내는 확률변수는 2이상의 모든 자연수 값을 가질 수 있으며 하루 동안 발생하는 교통사고의 수는 0이상의 모든 정수 값을 가질 수 있다. 위의 두 가지 경우는 확률변수가 가질 수 있는 값의 수는 무한개이지만 셀 수 있는 집합에 속하는 경우이다.

이산형 확률변수가 가질 수 있는 값은 유한개이거나 셀 수 있는 집합에 속하므로 값을 차례로 나열하여 다음과 같이 나타낼 수 있다.

$$x_1,\ x_2,\ x_3, \ldots$$

이산형 확률변수의 확률분포는 각 값에서 확률에 의하여 결정된다. 각 값에서 확률을 나타낸 함수를 **확률질량함수**(probability mass function)라 하며 $p(x)$ 또는 $f(x)$로 나타낸다. 즉 확률질량함수 $p(x_i)$는 X의 값이 x_i일 확률을 나타낸다.

$$p(x_i) \ = \ P(X = x_i)$$

이산형 확률변수 X가 정수 값만을 가지는 경우에는 정수 값을 k로 나타내어 확률질량함수를 $p(k)$로 표현하기도 한다.

예 2.3 동전을 2번 던져 나온 앞면의 수를 X라 할 때 X가 가질 수 있는 값과 확률질량함수를 구하여라.

[풀이] 동전을 2번 던질 때 앞면의 수는 0, 1, 2 중에서 하나의 값을 가진다. 0, 1, 2에서 확률은 0.25, 0.5, 0.25이므로 X의 확률질량함수는 다음과 같다. ■

k	0	1	2
$p(k)$	0.25	0.5	0.25

예 2.4 앞면이 나올 때까지 동전을 던지는 횟수를 X라 할 때 X가 가질 수 있는 값과 확률질량함수를 구하여라.

[풀이] 앞면이 나올 때까지 던지는 횟수는 1이상의 모든 정수 값이 가능하다. 던지는 횟수가 k인 경우는 처음 $k-1$번은 연속하여 뒷면이 나오고 k 번째 처음으로 앞면이 나오는 경우이므로 확률질량함수는 다음과 같다.

$$p(k) \;=\; P(X=k) \;=\; \left(\frac{1}{2}\right)^{k-1}\left(\frac{1}{2}\right) \;=\; \left(\frac{1}{2}\right)^{k}, \quad k=1,2,\ldots$$

■

각 값에서의 확률을 나타내는 확률질량함수는 다음과 같은 성질을 가지고 있다.

(i) $p(x_i) \;\geq\; 0$

(ii) $\sum_i p(x_i) \;=\; 1$

(iii) $P(a \leq X \leq b) \;=\; \sum_{i\,:\,a\leq x_i\leq b} p(x_i)$

$p(x_i)$는 확률을 나타내므로 이 값은 0보다 크거나 같다. $\sum_i p(x_i)$는 모든 가능한 확률의 합이므로 이 값은 1이다. 또한 X가 a이상이고 b이하인 경우는 X가 가질 수 있는 값 $x_1, x_2, x_3, \ldots$ 중에서 a보다 크거나 같고 b보다 작거나 같은 값 중 하나일 때이다. 따라서 X가 a이상이고 b이하일 확률은 다음과 같다.

$$\begin{aligned} P(a \leq X \leq b) &= P(X\text{의 값이 } a\text{이상이고 } b\text{이하인 사건}) \\ &= \sum_{i:a\leq x_i\leq b} p(x_i) \end{aligned}$$

이산형 확률변수의 누적분포함수 $F(x)$는 다음과 같다.

$$F(x) \;=\; P(X \leq x) \;=\; \sum_{i:x_i\leq x} p(x_i)$$

이 함수는 x_i 에서 $p(x_i)$ 만큼 증가하므로 계단 형태의 함수 모양을 가진다. 예를 들어 동전을 두 번 던질 때 앞면의 수인 X 의 누적분포함수는 그림 2.2와 같다.

대표적인 이산형 확률분포로는 베르누이분포, 이항분포, 기하분포, 음이항분포, 초기하분포, 포아송분포 등이 있다.

2.1.1 베르누이분포

실험에서 나타날 수 있는 결과를 두 가지로 분류할 수 있는 경우가 있다. 일반적으로 이러한 실험에서 두 가지 결과를 '성공 (S)' 과 '실패 (F)' 로 나타내게 된다. 이처럼 성공과 실패의 두 가지 중에서 하나가 나타나는 실험을 **베르누이 시행**(Bernoulli trial)이라 한다. 이 실험에서 X 를 시행의 결과가 성공이면 1, 실패면 0의 값을 가지는 확률변수로 정의할 수 있다. 베르누이 시행은 성공이 일어날 확률에 의하여 확률분포가 결정된다. 성공의 확률이 p 라면 실패의 확률은 $1-p$ 가 된다. 이때 X 의 확률분포는 다음과 같이 나타낼 수 있다.

$$X = \begin{cases} 1, & p\text{의 확률로} \\ 0, & 1-p\text{의 확률로} \end{cases}$$

위와 같이 0 또는 1의 값을 가지는 확률변수 X 를 베르누이 확률변수라 하고 기호로 다음과 같이 나타낸다.

$$X \sim \text{Bernoulli}(p)$$

예를 들어 동전을 한 번 던지는 실험에서 X 를 앞면이 나오면 1, 뒷면이 나오면 0의 값을 가진다고 할 때 이 확률변수는 성공 (앞면) 의 확률이 $p = 0.5$ 인 베르누이 확률변수이다. 베르누이 확률변수의 확률질량함수는 다음과 같이 나타낼 수 있다.

$$p(k) = p^k(1-p)^{1-k}, \quad k = 0 \text{ 또는 } 1$$

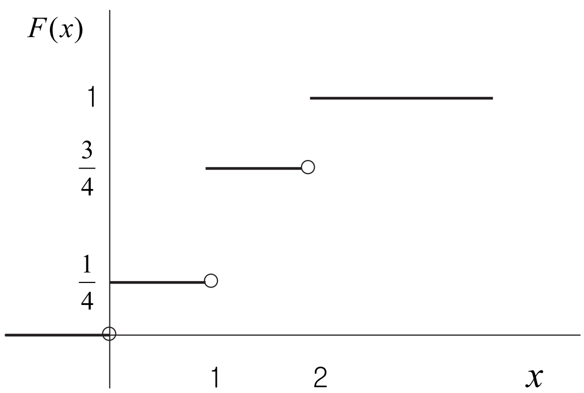

그림 2.2: 동전을 2번 던질 때 앞면의 수에 대한 누적분포함수

베르누이분포는 시행의 결과가 두 가지 이상으로 분류되는 경우에도 이용될 수 있다. 여러 가지 가능한 결과 중에서 특정한 결과가 발생하는지 관심이 있다면 특정한 결과를 성공으로 나머지 결과를 실패로 간주하여 베르누이 시행으로 생각할 수 있다. 예를 들어 주사위를 던지는 실험에서 가능한 결과는 6가지이다. 이때 주사위의 눈이 3의 배수인지에 관심이 있다면 성공을 3의 배수인 3 또는 6의 눈이 나오는 경우로 생각하고 실패를 1, 2, 4, 5의 눈이 나오는 경우로 분류하여 베르누이 시행으로 볼 수 있다. 따라서 주사위의 눈이 3의 배수이면 1의 값을 가지고 다른 눈이 나오면 0의 값을 가지는 확률변수는 성공의 확률 $p = \frac{1}{3}$ 인 베르누이 확률변수라 할 수 있다.

2.1.2 이항분포

동전을 10번 던질 때 나오는 앞면의 수, 주사위를 5번 던질 때 3이나 6의 눈이 나오는 횟수 등에 대한 확률분포로 이용되는 것이 **이항분포**(binomial distribution)이다. 앞의 두 가지 예는 다음과 같은 특징을 가지고 있다. 동전 또는 주사위를 던질 때 나타날 수 있는 결과는 성공이나 실패로 분류할 수 있어서 베르누이 시행이라 할 수 있다. 성공은 동전의 앞면, 주사위의 눈이 3이나 6인 경우라 할 수 있으며 실패는 다른 결과가 나오는 경우라 할 수 있다. 이때 앞면의 수, 3이나 6의 눈이 나오는 횟수는 위의 베르누이 시행을 반복할 때 총 성공의 수가 된다. 또한 시행을 반복할 때 각 시행은 서로 독립이다. 즉 첫 번째 동전의 결과가 앞면이라고 해서 두 번째 동전의 앞면이 나올 확률이 높아진다거나 낮아지지는 않는다.

위에서 동전을 10번 던질 때 나오는 앞면의 수 등은 일반적으로 성공 또는 실패로 분류할 수 있는 베르누이 시행을 독립적으로 n번 반복할 때 나오는 성공의 수라 할 수 있다. 베르누이 시행을 n번 반복하므로 성공의 수를 나타내는 확률변수 X가 가지는 값은 0부터 n까지의 정수 값을 가진다. 각 베르누이 시행에서 성공의 확률이 p일 때 $X = k$일 확률은 다음과 같다.

$$P(X = k) \quad = \quad P(n\text{번 베르누이 시행 중 성공의 수} = k)$$

n번 시행 중에서 성공의 수가 k일 때는 다음과 같은 여러 가지 경우가 있다.

- 처음 k번 시행은 성공이고 나머지 $n-k$번은 시행은 실패
- 처음 시행은 실패, 두 번째 시행부터 $k+1$ 번째 시행까지는 성공이고 나머지 $n-k-1$번은 시행은 실패

  $\vdots$

- 처음 $n-k$번은 시행은 실패이고 마지막 k번 시행은 성공

위의 각 경우에서 성공의 수는 항상 k이다. 따라서 각 확률은 독립적인 베르누이 시행에서 성공이 k번, 실패가 $n-k$번 나올 확률이므로 다음과 같다.

$$p^k(1-p)^{n-k}$$

따라서 성공의 수가 k일 확률은 위 확률에 성공의 수가 k인 경우의 수를 곱하여 주면 된다. 즉 $X=k$일 확률은 다음과 같이 나타낼 수 있다.

$$P(X=k) \;=\; (n\text{번 시행 중 성공의 수가 } k\text{인 경우의 수}) \times p^k(1-p)^{n-k}$$

n번 시행 중에서 성공의 수가 k인 경우의 수는 S가 k번, F가 $n-k$번 있는 문자를 배열하여 얻을 수 있는 경우의 수와 같으므로 $\dfrac{n!}{k!(n-k)!}$ 이다. 따라서 $X=k$일 확률은 다음과 같다.

$$P(X=k) \;=\; \binom{n}{k} p^k(1-p)^{n-k}$$

성공의 확률이 p인 베르누이 시행을 독립적으로 n번 반복할 때 나오는 성공의 수로 정의된 확률변수를 이항 확률변수라 하며 확률질량함수는 다음과 같이 주어진다.

$$p(k) \;=\; \binom{n}{k} p^k(1-p)^{n-k}, \quad k=0,1,\ldots,n.$$

이항분포는 시행 횟수 n과 각 시행에서 성공의 확률 p에 의하여 확률분포가 결정된다. X가 이항분포를 따르는 확률변수일 때 기호로 다음과 같이 나타낸다.

$$X \;\sim\; \mathrm{B}(n,p)$$

$\mathrm{B}(n,p)$ 대신 $\mathrm{Binomial}(n,p)$를 사용하기도 한다.

예를 들어 동전을 10번 던지는 실험에서 X를 앞면의 수라 할 때 이 확률변수는 시행 횟수 $n=10$, 성공(앞면)의 확률 $p=0.5$인 이항분포를 따르게 된다. 그림 2.3은 $n=10$일 때 $p=0.5$와 $p=0.7$인 이항분포의 확률질량함수를 보여주고 있다. $p=0.5$인 경우는 대칭적인 확률분포이지만 $p=0.7$인 경우는 비대칭적인 확률분포임을 알 수 있다.

예 2.5 주사위를 5번 던져 3이나 6이 나오는 횟수를 X라 할 때 X의 확률질량함수와 $X=2$일 확률을 구하여라.

[풀이] 주사위의 눈이 3이나 6이 나올 확률은 $\frac{1}{3}$이므로 X는 시행 횟수 $n=5$, 성공의 확률 $p=\frac{1}{3}$인 이항분포를 따르는 확률변수이다. 따라서 X의 확률질량함수는 다음과 같다.

$$p(k) \;=\; \binom{5}{k}\left(\frac{1}{3}\right)^k\left(\frac{2}{3}\right)^{5-k}, \quad k=0,1,\ldots,5$$

따라서 3이나 6이 나오는 횟수가 2일 확률은 다음과 같다.

$$P(X=2) = \binom{5}{2}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^3 = \frac{80}{243}$$

■

예 2.6 0부터 1 사이의 실수 중에서 임으로 20번 선택하였을 때, 이 중에서 0.7보다 큰 횟수에 대한 확률질량함수를 구하여라.

[풀이] 0과 1 사이에서 임의로 실수를 선택하는 실험을 하나의 시행으로 볼 수 있다. 각 시행에서 선택된 수가 0.7보다 큰 경우를 성공으로 간주하면 0.7보다 큰 횟수는 20번의 시행 중에서 성공의 수가 된다. 각 시행에서 성공의 확률은 임의로 선택된 수가 0.7보다 클 확률이므로 0.3이 된다. 따라서 임의로 선택된 20개의 수 중에서 0.7보다 큰 개수 X는 시행 횟수 $n=20$, 성공의 확률 $p=0.3$인 이항분포를 따르는 확률변수이다. 따라서 X의 확률질량함수는 다음과 같다.

$$p(k) = \binom{20}{k} 0.3^k 0.7^{20-k}, \quad k=0,1,\ldots,20$$

■

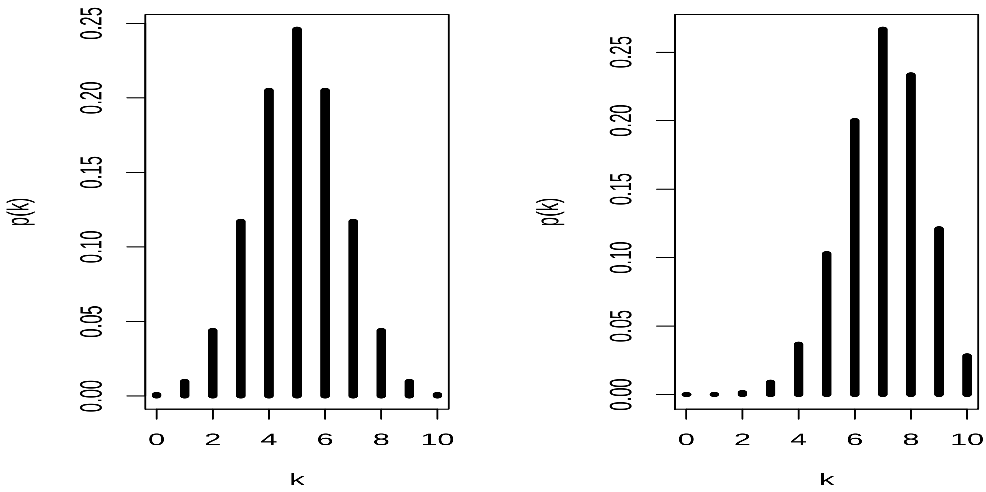

그림 2.3: $n=10$일 때 $p=0.5$와 $p=0.7$인 이항분포의 확률질량함수

2.1.3 기하분포

앞면이 나올 때까지 동전을 던지는 횟수, 주사위의 눈이 3이나 6의 눈이 나올 때까지 주사위를 던지는 횟수 등에 대한 확률분포로 이용되는 것이 **기하분포**(geometric distribution)이다. 앞의 예는 서로 독립인 베르누이 시행을 반복할 때 첫 번째 성공이 나올 때까지 시행 횟수를 나타낸다고 할 수 있다.

첫 번째 성공이 나올 때까지 시행 횟수를 나타내는 확률변수 X가 가지는 값은 1이상의 정수 값을 가지며 $X = k$일 확률은 다음과 같다.

$$\begin{aligned} P(X=k) &= P(k\text{번째 시행에서 첫 번째 성공이 나오는 사건}) \\ &= P(\text{처음 } k-1\text{번 시행은 실패이고 } k\text{번째 시행은 성공}) \\ &= (1-p)^{k-1}p \end{aligned}$$

이처럼 성공의 확률이 p인 베르누이 시행을 독립적으로 반복할 때 첫 번째 성공이 나올 때까지 시행 횟수로 정의된 확률변수 X를 기하 확률변수라 하고 기호로 다음과 같이 나타낸다.

$$X \sim \text{Geometric}(p)$$

예를 들어 동전을 던지는 실험에서 X를 첫 번째 앞면이 나올 때까지 시행 횟수라 할 때 이 확률변수는 성공의 확률 $p = 0.5$인 기하 확률변수이다. 기하 확률변수의 확률질량함수는 다음과 같다.

$$p(k) = (1-p)^{k-1}p, \quad k = 1, 2, \ldots$$

예 2.7 주사위를 던져 3이나 6이 나올 때까지 던지는 시행 횟수를 X라 할 때 X의 확률질량함수와 $X = 5$일 확률을 구하여라.

[풀이] 주사위의 눈이 3이나 6이 나올 확률은 $\frac{1}{3}$이므로 X는 성공(앞면)의 확률 $p = \frac{1}{3}$인 기하분포를 따르는 확률변수이다. 따라서 X의 확률질량함수는 다음과 같다.

$$p(k) = \left(1-\frac{1}{3}\right)^{k-1}\frac{1}{3}, \quad k = 1, 2, \ldots$$

따라서 3이나 6이 나올 때까지 주사위를 던지는 횟수가 5일 확률은 다음과 같다.

$$P(X=5) = \left(1-\frac{1}{3}\right)^{4}\frac{1}{3} = \frac{16}{243}$$

■

2.1.4 음이항분포

기하분포는 베르누이 시행에서 첫 번째 성공이 나올 때까지 시행 횟수에 대한 확률분포이다. 이를 일반화하여 베르누이 시행에서 r번째 성공이 나올 때까지 시행 횟수에 대한 확률분포를 **음이항분포**(negative binomial distribution)라 한다. 예를 들어 동전을 던지는 실험에서 3번째 앞면이 나올 때까지 동전을 던지는 횟수에 대한 확률분포로 이용되는 것이 음이항분포이다.

성공 또는 실패의 결과를 가지는 서로 독립인 베르누이 시행을 반복할 때 r번째 성공이 나올 때까지 총 시행 횟수를 나타내는 확률변수 X가 가지는 값은 r 이상의 정수 값을 가지며 $X = k$일 확률은 다음과 같다.

$$\begin{aligned} P(X = k) &= P(k\text{번째 시행에서 } r\text{번째 성공이 나오는 사건}) \\ &= P(\text{처음 } k-1\text{번 시행에서 성공이 } (r-1)\text{번 나오고} \\ &\qquad \text{마지막 } k\text{번째 시행은 성공인 사건 }) \\ &= P(\text{처음 } k-1\text{번 시행에서 성공이 } (r-1)\text{번인 사건}) \\ &\quad \times P(\text{마지막 } k\text{번째 시행은 성공인 사건}) \end{aligned}$$

위에서 처음 $k-1$번 시행에서 성공의 수는 시행 횟수 $k-1$, 성공의 확률 p인 이항분포를 따르므로 첫 번째 확률은 다음과 같다.

$$P(\mathrm{B}(k-1, p) = r-1) \;=\; \binom{k-1}{r-1} p^{r-1}(1-p)^{k-r}$$

마지막 k번째 시행이 성공일 확률은 p이므로 $X = k$일 확률은 다음과 같다.

$$P(X = k) \;=\; \binom{k-1}{r-1}(1-p)^{k-r}p^r$$

이처럼 성공의 확률이 p인 베르누이 시행을 독립적으로 반복할 때 r번째 성공이 나올 때까지 시행 횟수로 정의된 확률변수 X를 음이항 확률변수라 하고 기호로 다음과 같이 나타낸다.

$$X \;\sim\; \mathrm{NB}(r, p)$$

$\mathrm{NB}(r, p)$ 대신 Negative Binomial(r, p)를 사용하기도 한다. 음이항 확률변수의 확률질량함수는 다음과 같이 나타낼 수 있다.

$$p(k) = \binom{k-1}{r-1}(1-p)^{k-r}p^r, \quad k = r, r+1, \ldots$$

앞에서 살펴본 베르누이, 이항, 기하, 음이항 확률분포는 모두 베르누이 시행과 관련된 확률분포이다. 베르누이분포는 시행의 결과가 성공이나 실패의 두 가지 결과로 분류할 수 있을 때 사용되는 확률분포이고, 이항분포는 성공의 확률이 일정한 베르누이 시행을 n번 독립적으로

반복할 때 성공의 수에 대한 확률분포이다. 기하분포는 베르누이 시행에서 첫 번째 성공이 나올 때까지의 시행 횟수에 대한 확률분포이고, 음이항분포는 r번째 성공이 나올 때까지의 시행 횟수에 대한 확률분포이다.

2.1.5 초기하분포

r개의 원소로 이루어진 하나의 그룹과 $n-r$개의 원소로 이루어진 또 다른 그룹에서 m개를 임의로 선택할 때 첫 번째 그룹에서 나오는 원소의 수에 대한 확률분포를 **초기하분포**(hyper geometric distribution)라 한다.

초기하분포를 따르는 확률변수 X에 대하여 경우의 수를 이용하여 $X = k$일 확률을 구하면 다음과 같다.

$$P(X = k) = \frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}}$$

단, 확률변수가 가지는 값 k의 범위는 $\max(0, m-n+r) \le k \le \min(r, m)$이다. 이와 같이 범위가 한정되는 이유는 두 번째 그룹에는 $n-r$개의 원소만이 있으므로 선택하는 원소의 수 m이 $n-r$보다 크면 첫 번째 그룹에서 최소한 $m-n+r$개의 원소가 선택되어야 하기 때문이다. 마찬가지로 첫 번째 그룹에는 r개의 원소가 있으므로 첫 번째 그룹에서 선택되는 원소의 수가 m이나 r보다 클 수는 없다.

예 2.8 빨간색 구슬 5개와 검정색 구슬 3개가 들어 있는 주머니에서 임의로 4개의 구슬을 꺼낼 때 빨간색 구슬의 수 X의 확률질량함수를 구하여라.

[**풀이**] 빨간색 구슬의 수 X는 첫 번째 그룹에 5개, 두 번째 그룹에 3개의 원소가 있을 때 4개의 원소를 선택하는 초기하 확률분포를 따른다. 즉 X는 $n = 8$, $r = 5$, $m = 4$인 초기하 확률변수이다. 따라서 X의 확률질량함수는 다음과 같다.

$$p(k) = \frac{\binom{5}{k}\binom{3}{4-k}}{\binom{8}{4}}, \quad k = 1, 2, 3, 4$$

■

2.1.6 포아송분포

하루 동안 발생하는 교통사고의 수와 같이 특정한 사건의 발생 수에 대한 확률분포로 흔히 이용되는 것이 **포아송분포**(Poisson distribution)이다. 포아송분포를 따르는 확률변수 X는 음이 아닌 모든 정수 값을 가질 수 있으며 확률질량함수는 다음과 같다.

$$p(k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, \ldots \tag{2.7}$$

단, λ는 포아송분포의 특징을 결정하는 양의 상수 값이다. X가 특정한 사건의 발생 수일 때 λ는 평균적으로 발생하는 사건의 수를 나타낸다. X가 포아송분포를 따르는 확률변수일 때 기호로 다음과 같이 나타낸다.

$$X \sim \text{Poisson}(\lambda)$$

그림 2.4는 $\lambda = 3$인 포아송분포의 확률질량함수를 보여주고 있다. 포아송분포의 확률질량함수는 λ 근처에서 가장 큰 값을 가지고 있음을 알 수 있다.

포아송분포는 다음과 같은 성질을 가지고 있는 특정한 사건의 발생 횟수를 나타내는 확률분포로 주로 이용된다

- 아주 짧은 구간에서는 사건이 2회 이상 발생할 확률은 거의 0에 가까워 무시할 수 있다.
- 아주 짧은 구간에서 사건이 발생할 확률은 구간의 길이에 비례한다.
- 서로 다른 구간에서 발생하는 사건의 수는 서로 독립이다.

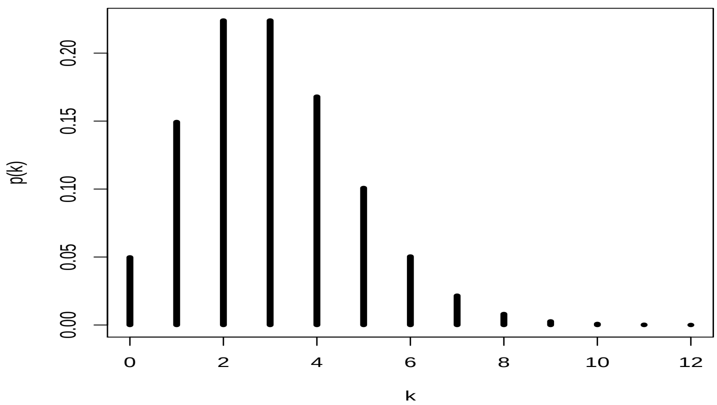

그림 2.4: $\lambda = 3$인 포아송분포의 확률질량함수

일정한 구간에서 발생하는 사건의 수는 그 구간을 아주 짧은 구간으로 나누었을 때 각 구간에서 발생하는 사건의 합으로 생각할 수 있다. 구간을 n개의 같은 간격을 가지는 소 구간으로 나누어 생각하여 보자. 일정한 구간에서 평균적으로 발생하는 사건의 수가 λ이므로 n등분한 구간에서는 평균적으로 $\frac{\lambda}{n}$번 발생한다고 할 수 있다. 한편 n을 충분히 크게 하면 n등분한 구간은 아주 짧아지게 되고 주어진 조건에 따라 아주 짧은 구간에서는 사건이 한 번 발생하거나 발생하지 않는다. n등분한 짧은 구간에서는 기껏해야 사건이 한 번 발생하고 평균적으로는 $\frac{\lambda}{n}$번 발생하므로 이 구간에서 사건이 발생할 확률은 $\frac{\lambda}{n}$라 할 수 있다. 즉 n등분한 소 구간에서 발생하는 사건의 수는 성공의 확률이 $\frac{\lambda}{n}$인 베르누이분포를 따르게 된다. 전체 구간에서 발생하는 사건의 수는 n등분한 구간에서 발생하는 사건의 합이므로 시행 횟수 n, 성공의 확률 $\frac{\lambda}{n}$인 이항분포를 따르게 된다. 등분한 구간이 짧을수록 위의 설명이 타당해지므로 전체 구간에서 발생하는 사건의 수에 대한 확률분포는 $\mathrm{B}\left(n, \frac{\lambda}{n}\right)$의 극한이라 할 수 있다. 이러한 사실을 바탕으로 이항분포의 확률의 극한은 포아송분포의 확률이 된다는 것을 다음과 같이 보일 수 있다.

$\mathrm{B}(n,p)$를 따르는 확률변수 X에 대하여 $p = \frac{\lambda}{n}$일 때 $X = k$일 확률은 다음과 같이 나타낼 수 있다.

$$\begin{aligned} P(X=k) &= \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\ &= \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k}{k!} \cdot \frac{n!}{(n-k)!n^k} \cdot \left(1-\frac{\lambda}{n}\right)^{-k} \left(1-\frac{\lambda}{n}\right)^{n} \end{aligned}$$

위 식에서 n이 커질 때 다음이 성립함을 알 수 있다.

$$\begin{array}{cc} \frac{n!}{(n-k)!n^k} \cdot \left(1-\frac{\lambda}{n}\right)^{-k} & \longrightarrow\ 1 \\ \left(1-\frac{\lambda}{n}\right)^{n} & \longrightarrow\ e^{-\lambda} \end{array}$$

따라서 $X = k$일 확률은 다음 값으로 수렴한다.

$$P(X=k) \longrightarrow \frac{\lambda^k e^{-\lambda}}{k!}$$

이 값은 식 (2.7)에 주어진 포아송분포의 확률과 같음을 알 수 있다.

포아송분포는 다음과 같은 사건의 발생 횟수에 대한 확률분포로 이용된다.

- 하루 동안 걸려오는 전화통화 횟수
- 1년 동안 발생한 지진의 수
- 일정한 기간 동안 발생하는 특정한 사고의 수

포아송분포는 일정기간 동안 발생하는 특정한 사건의 수에 대한 확률분포라 하였다. 한편

구간의 길이가 달라지면 그 구간에서 평균적으로 발생하는 사건의 수는 구간의 길이에 비례하여 변하게 되지만 사건이 발생하는 확률분포의 성격에는 변화가 없으므로 이 구간에서 발생하는 사건의 수도 포아송분포를 따르게 된다. 즉 단위 시간당 발생하는 사건의 수가 Poisson(λ) 분포를 따를 때 t시간 동안 발생 하는 사건의 수의 확률분포는 다음과 같다.

$$\text{Poisson}(\lambda t)$$

예 2.9 어느 사람에게 걸려오는 전화통화 수는 시간당 0.5통인 포아송분포를 따른다고 한다.

(a) 5시간 동안 한 통화도 걸려오지 않을 확률을 구하여라.

(b) 5시간 동안 걸려오는 통화 수가 1일 확률을 구하여라.

[풀이]

(a) 포아송분포의 성질에 의하여 t시간 동안 걸려오는 전화통화 수는 Poisson($0.5t$) 분포를 따르게 된다. 따라서 5시간 동안 한 통화도 걸려오지 않을 확률은 $P(\text{Poisson}(2.5) = 0) = e^{-2.5} = 0.082$이다.

(b) 5시간 동안 걸려오는 통화 수가 1일 확률은 $P(\text{Poisson}(2.5) = 1) = 2.5e^{-2.5} = 0.205$이다.

■

2.2 연속형 확률변수

확률변수가 일정한 구간에 속하는 모든 값을 가질 때 이를 **연속형 확률변수**(continuous random variable)라 한다. 연속형 확률변수는 일정한 구간에 속하는 모든 값을 가지므로 이 구간에서 누적분포함수는 연속적으로 증가한다고 할 수 있다. 이산형 확률변수는 가질 수 있는 각 값에서 확률에 의하여 확률분포가 결정되는 반면 연속형 확률변수는 각 값에서 확률에 의하여 확률분포가 결정되지는 않는다. 연속형 확률변수 X가 a라는 값을 가질 수 있다고 하자. 이때 X가 a라는 값을 가질 확률은 다음과 같다.

$$\begin{aligned} P(X = a) &= P(X \le a) - P(X < a) \\ &= F(a) - \lim_{n \to \infty} F\left(a - \frac{1}{n}\right) \end{aligned}$$

연속형 확률변수의 누적분포함수는 연속이므로 $F(a) = \lim_{n\to\infty} F\left(a - \frac{1}{n}\right)$이다. 즉 연속형 확률변수는 어떤 특정한 값을 가질 확률은 0이 된다. 따라서 연속형 확률변수의 확률분포는 각 값에서의 확률에 의하여 결정되지는 않는다.

이산형 확률분포는 각 값에서의 확률을 나타내는 확률질량함수로 표현할 수 있는 것처럼 연속형 확률변수의 경우도 확률분포를 나타내는 함수가 존재한다. 앞에서 설명한 것처럼 확률변수의 확률분포는 누적분포함수에 의하여 결정되며 일반적으로 연속형 확률변수의 누적분포함수는 미분가능한 함수이다. 연속형 확률변수의 누적분포함수를 미분한 함수를 **확률밀도함수**(probability density function)라 하며 $f(x)$로 나타낸다. 즉 누적분포함수를 $F(x)$라 할 때 확률밀도함수 $f(x)$는 다음과 같다.

$$f(x) \;=\; \frac{d}{dx}F(x)$$

반대로 확률밀도함수를 적분하면 누적분포함수를 얻을 수 있다.

$$F(x) \;=\; \int_{-\infty}^{x} f(t)\;dt$$

따라서 연속형 확률변수의 확률분포는 확률밀도함수로 나타낼 수 있다.

확률밀도함수는 누적분포함수를 미분한 함수이므로 다음과 같은 성질을 가진다.

(i) $f(x) \geq 0$

(ii) $\int_{-\infty}^{\infty} f(x)\;dx \;=\; 1$

(iii) $P(a < X \leq b) \;=\; F(b) - F(a) \;=\; \displaystyle\int_{a}^{b} f(x)\,dx$

누적분포함수는 증가함수이므로 이를 미분한 확률밀도함수는 항상 0보다 크거나 같게 된다. 또한 연속형 확률변수는 어느 한 값을 가질 확률은 0이므로 확률을 구할 때 어떤 한 점이 포함되는지는 중요하지 않다. 예를 들어 아래의 확률은 모두 같게 된다.

$$P(a \leq X \leq b) \;=\; P(a < X \leq b) \;=\; P(a < X < b)$$

이산형 확률변수의 확률질량함수에 대응되는 것이 연속형 확률변수의 경우는 확률밀도함수이다. 한 가지 차이점은 확률질량함수는 어느 점에서의 확률을 나타내지만 확률밀도함수는 그 점에서의 확률을 나타내지는 않는다는 것이다. 그렇다면 확률밀도함수는 어떤 의미를 가지는 것일까? 예를 들어 $f(2) = 4$라는 것은 어떤 뜻일까? 확률질량함수가 어떤 값을 가질 확률을 나타내는 값이라면 확률밀도함수는 어떤 값 근처에 속할 확률을 결정지어주는 값이다.

아주 짧은 구간 $[x, x + \Delta x]$에 대하여 $X \in [x, x + \Delta x]$일 확률은 다음과 같다.

$$P(x \leq X \leq x + \Delta x) \;=\; \int_{x}^{x+\Delta x} f(t)\;dt \;\approx\; f(x)\Delta x$$

그림 2.5에서 볼 수 있는 것과 같이 짧은 구간 $[x, x + \Delta x]$에서 확률밀도함수와 x축으로 둘러싸인 부분의 면적은 근사적으로 $[x, x + \Delta x]$에서 $f(x)$를 높이로 하는 사각형의 면적과 같다. 즉 확률변수 X가 x 근처의 아주 짧은 구간에 속할 확률은 근사적으로 x에서의 확률밀도함수와 그 구간의 길이의 곱으로 주어진다. 이와 같이 확률밀도함수는 그 점에서의 확률은 아니지만 그 점 근처에

있을 확률을 결정지어주는 값이다. 예를 들어 $f(2) = 4$ 일 때 확률변수가 구간 [2, 2.005] 에 속할 확률은 근사적으로 $4 \times 0.005 = 0.02$ 임을 알 수 있다.

예 2.10 X 를 [0, 1] 에서 임의로 선택된 수라 할 때 X 의 확률밀도함수를 구하여라.

[풀이] X 는 임의로 선택된 수이므로 0과 1 사이의 값 x 에 대하여 $X \in [0, x]$ 일 확률은 전체 구간에서 구간 $[0, x]$ 의 길이의 비율인 x 가 될 것이다. 따라서 X 의 누적분포함수는 $F(x) = x, \quad 0 \le x \le 1$ 이며 확률밀도함수는 다음과 같다.

$$f(x) \;=\; \frac{d}{dx}F(x) \;=\; \frac{d}{dx}x \;=\; 1, \quad 0 \le x \le 1$$

■

예 2.11 연속형 확률변수 X 의 확률밀도함수가 $f(x) = 2x, \; 0 \le x \le 1$ 일 때, 다음 물음에 답하여라.

(a) X 의 누적분포함수를 구하여라.

(b) $X \in [0, 0.5]$ 일 확률을 구하여라.

[풀이]

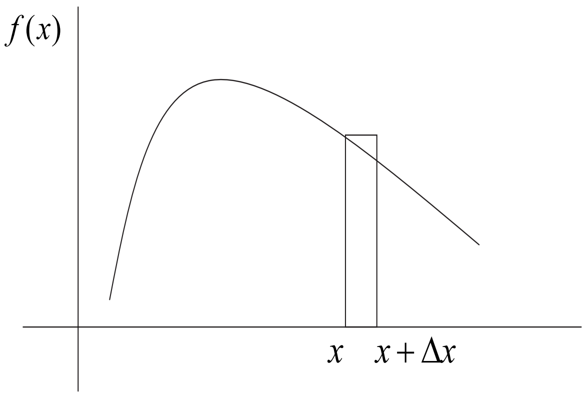

그림 2.5: 구간 $(x, x + \Delta x]$ 에 속할 확률

(a) 주어진 확률밀도함수는 0과 1 사이의 구간에서만 양의 값을 가지며 나머지 영역에서는 0의 값을 가진다. 즉 X는 0과 1 사이의 구간에서만 값을 가지는 확률변수이다. 따라서 0보다 작은 x에 대하여 누적분포함수 $F(x) = 0$이고 1보다 큰 x에 대하여 $F(x) = 1$이다. 또 0과 1 사이의 x에 대하여 누적분포함수는 다음과 같다.

$$F(x) \ = \ \int_{-\infty}^{x} f(t)\ dt \ = \ \int_{0}^{x} 2t\ dt \ = \ x^2$$

앞의 세 가지 경우를 간단히 표현하여 $F(x) = \ x^2, \ 0 \le x \le 1$로 나타내기도 한다.

(b) $P(0 \le X \le 0.5) = \ \int_0^{0.5} 2t\ dt \ = \ 0.25$이다.

■

대표적인 연속형 확률분포로는 균일분포, 지수분포, 감마분포, 정규분포 등이 있다.

2.2.1 균일분포

일정한 구간에서 임의로 선택된 수에 대한 확률분포를 **균일분포**(uniform distribution) 또는 균등분포라 한다. 구간 $[a,\ b]$에서 임의로 선택된 수를 X라 할 때 X는 이 구간에 속하는 모든 값을 가질 가능성이 균일하다고 할 수 있다. 따라서 이 구간에 속한 x에 대하여 $X \in [a, x]$일 확률은 전체 구간의 길이 $[a, b]$ 중 구간 $[a, x]$의 길이의 비율이 될 것이다. 따라서 X의 누적분포함수는 다음과 같고

$$F(x) \ = \ P(a \le X \le x) \ = \ \frac{x-a}{b-a}, \quad a \le x \le b$$

확률밀도함수는 다음과 같다.

$$f(x) \ = \ \frac{1}{(b-a)}, \quad a \le x \le b$$

X가 구간 $[a,\ b]$에서 임의로 선택된 수인 균일분포를 따르는 확률변수일 때 기호로 다음과 같이 나타낸다.

$$X \ \sim \ \mathrm{Uniform}[a, b]$$

$\mathrm{Uniform}[a, b]$ 대신 $\mathrm{U}[a, b]$를 사용하여 나타내기도 한다.

예 2.12 X가 구간 $[2, 5]$에서 임의로 선택된 수일 때, 다음 물음에 답하여라.

(a) X가 4이하일 확률을 구하여라.

(b) X의 확률밀도함수를 구하여라.

[풀이]

(a) $X \le 4$일 확률은 전체 구간 [2, 5]의 길이에서 구간 [2, 4]의 길이의 비율인 $\frac{2}{3}$이다.

(b) X의 확률밀도함수는 $f(x) = \frac{1}{3},\ 2 \le x \le 5$이다.

■

2.2.2 지수분포

일정한 구간에서 발생하는 특정한 사건의 수는 포아송분포를 따른다는 것을 앞에서 설명하였다. 발생한 사건의 수는 이산형 확률분포를 따르는 반면, 첫 번째 사건이 발생할 때까지의 시간은 0 이상의 모든 실수 값을 가질 수 있으므로 연속형 확률분포를 따르게 된다. 구체적으로 포아송분포에서 첫 번째 사건이 일어날 때까지 대기 시간(waiting time)에 대한 확률분포를 구하여 보자. 첫 번째 사건이 발생할 때까지의 시간을 X라 할 때 X의 누적분포함수는 다음과 같이 나타낼 수 있다.

$$\begin{aligned} P(X \le t) &= 1 - P(X > t) \ = \ 1 - P(\text{첫 번째 사건이 발생하는 시간} > t) \\ &= 1 - P([0, t] \text{ 사이에서 발생하는 사건의 수} = 0) \end{aligned}$$

길이가 t인 구간에서 발생하는 사건의 수는 Poisson(λt)를 따르므로 X의 누적분포함수는 다음과 같다.

$$P(X \le t) \ = \ 1 - P(\text{Poisson}(\lambda t) = 0) \ = \ 1 - e^{-\lambda t}$$

따라서 X의 확률밀도함수는 $f(x) = \lambda e^{-\lambda x},\ x > 0$이다. 여기서는 X를 첫 번째 사건이 발생할 때까지의 대기 시간이라 하였지만 사건과 사건 사이의 시간 간격도 지수분포를 따르게 된다.

다음과 같은 확률밀도함수를 가지는 확률분포를 **지수분포**(exponential distribution)라 한다.

$$f(x) \ = \ \lambda e^{-\lambda x}, \quad x \ge 0$$

X가 지수분포를 따르는 확률변수일 때 기호로 다음과 같이 나타낸다.

$$X \ \sim \ \text{Exp}(\lambda)$$

양의 상수인 λ의 의미는 포아송분포에서 단위시간 당 평균적으로 발생하는 사건의 수이므로 시간의 관점에서 본다면 $\frac{1}{\lambda}$은 사건과 사건 사이의 평균적인 시간 간격 또는 첫 번째 사건이 발생할 때까지의 평균 시간이라 할 수 있다. 따라서 지수분포에서 $\frac{1}{\lambda}$은 첫 번째 사건이 발생할 때까지의 평균적인 대기 시간이다. 포아송분포와의 관계에서 알 수 있는 것처럼 지수분포는 주로 어떤 사건이 발생할 때까지의 대기 시간 또는 제품의 수명에 대한 확률분포로 이용된다.

조건부 확률을 이용하여 지수분포에 대한 한 가지 흥미로운 성질을 보일 수 있다. X가 지수분포를 따르는 확률변수일 때 0보다 큰 s와 t에 대하여 다음이 성립함을 알 수 있다.

$$\begin{aligned} P(X > t+s | X > s) &= \frac{P(X > t+s, X > s)}{P(X > s)} = \frac{P(X > t+s)}{P(X > s)} \\ &= \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda t} = P(X > t) \end{aligned}$$

X가 어느 제품의 수명을 나타낸는 경우 위 식은 제품의 수명이 s이상이라는 사실이 주어졌을 때 제품이 추가적으로 t이상 지속될 확률은 제품의 수명이 t이상일 확률과 같다는 사실을 말하고 있다. 즉 앞으로 남은 제품의 수명에 대한 확률분포는 과거 얼마 동안 사용하였는지에 상관없다는 것이다. 이러한 지수분포의 성질을 무기억성(memoryless property)이라 한다. 일반적으로 제품의 수명은 이러한 성질을 만족하지는 않지만 확률분포의 단순함 때문에 제품의 수명에 대한 확률분포로 지수분포를 이용하기도 한다.

예 2.13 어느 사람에게 걸려오는 전화 통화 수는 평균적으로 1시간에 0.5통인 포아송분포를 따른다고 할 때, 다음 물음에 답하여라.

(a) 다음 통화가 걸려올 때까지의 대기 시간의 확률밀도함수를 구하여라.

(b) 다음 통화가 걸려올 때까지의 대기 시간이 1시간 이상일 확률을 구하여라.

[풀이]

(a) 다음 통화가 걸려올 때가지 대기 시간은 $\lambda = 0.5$인 지수분포를 따르므로 확률밀도함수는 $f(x) = 0.5e^{-0.5x},\ x > 0$이다.

(b) 다음 통화가 걸려올 때가지 대기 시간(X)이 1시간 이상일 확률은 다음과 같다.

$$P(X > 1) = \int_1^\infty 0.5e^{-0.5x}\,dx = e^{-0.5} = 0.607$$

■

2.2.3 감마분포

다음과 같은 확률밀도함수를 가지는 확률분포를 **감마분포**(gamma distribution)라 한다.

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1}e^{-\lambda x}, \quad x \geq 0 \tag{2.8}$$

위 확률밀도함수에서 α와 λ는 양의 상수이며 $\Gamma(\alpha)$는 다음과 같이 정의된 값이다.

$$\Gamma(\alpha) = \int_0^\infty u^{\alpha-1}e^{-u}\,du \tag{2.9}$$

X가 감마분포를 따르는 확률변수일 때 기호로 다음과 같이 나타낸다.

$$X \sim \text{Gamma}(\alpha, \lambda)$$

위의 확률밀도함수 (2.8)에서 λ 대신에 $\beta = \frac{1}{\lambda}$를 사용하여 기호로 다음과 같이 나타내기도 한다.

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}, \quad x \geq 0$$

그림 2.6은 감마분포의 확률밀도함수를 보여주고 있다.

한편 식 (2.9)에서 정의한 $\Gamma(\alpha)$에 대하여 다음과 같은 사실을 확인할 수 있다.

- $\Gamma(1) = \int_0^\infty e^{-u}\, du = 1$
- 부분적분을 사용하면 다음과 같은 사실을 알 수 있다.

$$\begin{aligned}\Gamma(\alpha) &= \int_0^\infty u^{\alpha-1} e^{-u}\, du \\ &= (\alpha-1)\int_0^\infty u^{\alpha-2} e^{-u}\, du \\ &= (\alpha-1)\Gamma(\alpha-1)\end{aligned}$$

- $\alpha = n$일 때 위의 사실을 반복하여 사용하면 $\Gamma(n) = (n-1)!$임을 알 수 있다.
- 극좌표를 이용한 적분을 통해 $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$임을 알 수 있다.

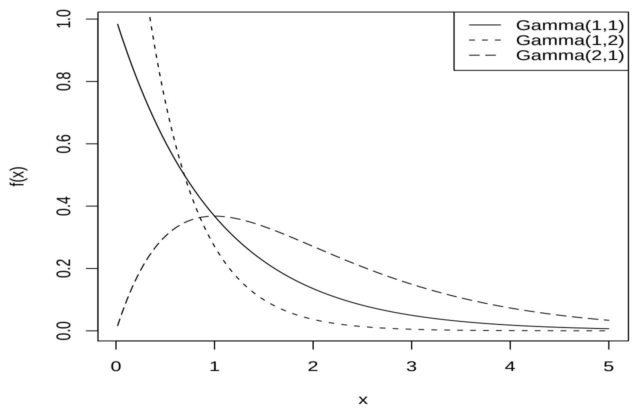

그림 2.6: 감마분포의 확률밀도함수

감마분포는 지수분포를 일반화한 것으로 볼 수 있다. $\alpha = 1$일 때 감마분포의 확률밀도함수는 지수분포의 확률밀도함수가 됨을 쉽게 알 수 있다. 즉 지수분포와 감마분포 사이에는 다음 관계가 있다.

$$\mathrm{Gamma}(1, \lambda) = \mathrm{Exp}(\lambda)$$

또한 감마분포가 지수분포를 일반화한 것이라는 것은 포아송분포와의 관계를 통해서도 확인할 수 있다. 포아송분포에서 첫 번째 사건까지 대기 시간이 지수분포를 따르는 것과 같이 포아송분포에서 n번째 사건까지의 대기 시간이 $\alpha = n$인 감마분포를 따른다는 것을 보일 수 있다. X를 포아송분포에서 n번째 사건까지 대기 시간이라 할 때, 대기 시간이 t보다 클 확률은 다음과 같이 나타낼 수 있다.

$$\begin{aligned} P(X > t) &= P(n\text{번째 사건의 발생 시간 } > t) \\ &= P([0, t] \text{ 사이에서 발생하는 사건의 수} \le n - 1) \end{aligned}$$

한편 구간 $[0, t]$에서 발생하는 사건의 수는 평균이 λt인 포아송분포를 따르므로 $P(X > t)$는 다음과 같다.

$$\begin{aligned} P(X > t) &= P(\mathrm{Poisson}(\lambda t) \le (n-1)) \\ &= \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t} \end{aligned}$$

따라서 X의 누적분포함수는 다음과 같다.

$$F(t) = 1 - \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

이를 미분한 확률밀도함수는 다음과 같다.

$$\begin{aligned} f(t) &= \frac{d}{dt}\left(1 - \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t}\right) \\ &= -\sum_{k=0}^{n-1}\left(\frac{k\lambda^k t^{k-1}}{k!} e^{-\lambda t} - \frac{\lambda^k t^k}{k!} \lambda e^{-\lambda t}\right) \\ &= -\sum_{k=1}^{n-1} \frac{\lambda^k}{(k-1)!} t^{k-1} e^{-\lambda t} + \sum_{k=0}^{n-1} \frac{\lambda^{k+1}}{k!} t^k e^{-\lambda t} \\ &= \frac{\lambda^n}{(n-1)!} t^{n-1} e^{-\lambda t} \end{aligned}$$

즉 포아송분포에서 n번째 사건까지 대기 시간의 확률밀도함수는 $\alpha = n$인 감마분포의 확률밀도함수와 같음을 알 수 있다.

감마분포의 특별한 경우를 카이제곱분포라 한다. 감마분포에서 $\alpha = \frac{n}{2}$ 이고 $\lambda = \frac{1}{2}$ 인 확률분포를 자유도(degree of freedom)가 n 인 카이제곱분포라 한다. 확률변수 X 가 자유도 n 인 카이제고분포를 따를 때 다음과 같이 나타낸다.

$$X \sim \chi_n^2$$

따라서 자유도 n 인 카이제곱분포의 확률밀도함수는 다음과 같다.

$$f(x) = \frac{1}{\Gamma\left(\frac{n}{2}\right)} \left(\frac{1}{2}\right)^{\frac{n}{2}} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, \quad x \geq 0$$

2.2.4 정규분포

다음과 같은 확률밀도함수를 가지는 확률분포를 **정규분포**(normal distribution)라 한다.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

위 확률밀도함수에서 μ 와 σ^2 은 정규분포의 중심 위치와 퍼진 정도를 나타내는 상수이다. X 가 정규분포를 따르는 확률변수일 때 기호로 다음과 같이 나타낸다.

$$X \sim \mathrm{N}(\mu, \sigma^2)$$

$\mu = 0$ 이고 $\sigma^2 = 1$ 인 정규분포를 **표준정규분포**(standard normal distribution)라 한다. 그림 2.7은 표준정규분포의 확률밀도함수를 보여주고 있다. 표준정규분포의 누적분포함수는 일반적으로 $\Phi(x)$ 를 사용하여 나타낸다. 즉 $\Phi(x)$ 는 표준정규분포의 확률밀도함수를 x 까지 적분한 값이다.

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt$$

2.2.5 베타분포

다음과 같은 확률밀도함수를 가지는 확률분포를 **베타분포**(beta distribution)라 한다.

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad 0 \leq x \leq 1$$

위의 확률밀도함수에서 α 와 β 는 확률분포를 결정지어주는 양의 상수이다. X 가 베타분포를 따르는 확률변수일 때 기호로 다음과 같이 나타낸다.

$$X \sim \mathrm{Beta}(\alpha, \beta)$$

2.3 변수변환

X 의 확률분포를 이용하여 $Y = g(X)$ 의 확률분포를 구하는 방법을 알아보도록 하자. 예를 들어 X 의 확률분포를 이용하여 $Y = X^2$ 또는 $Y = e^X$ 등의 확률분포를 구하여 보자.

2.3.1 이산형 확률변수의 변환

X 가 $x_1, x_2, \ldots$ 중에서 하나의 값을 가지는 이산형 확률변수일 때, 함수 $y = g(x)$ 에 의하여 $x_1, x_2, \ldots$ 는 그림 2.8과 같이 $g(x_1), g(x_2), \ldots$ 로 변환된다. 따라서 $Y = g(X)$ 는 $g(x_1), g(x_2)$, $\ldots$ 중에서 하나의 값을 가지는 이산형 확률변수이다. 사건 $\{Y = y\}$ 는 확률변수 X 를 이용하여 표현할 때 $\{g(X) = y\}$ 로 나타낼 수 있다. 그런데 함수 $y = g(x)$ 는 일대일 함수가 아닐 수 있으므로 $g(x) = y$ 인 x 는 하나 이상 있을 수 있다. $g(x) = y$ 인 x 들의 모임을 $g^{-1}(y)$ 라 하자. 즉 $g^{-1}(y)$ 는 $\{x : g(x) = y\}$ 인 집합을 나타낸다. 이때 $\{g(X) = y\}$ 인 사건은 X 의 값이 $g^{-1}(y)$ 에 속하는 경우를 말한다. 따라서 $g(X) = y$ 일 확률은 X 의 값이 $g^{-1}(y)$ 에 속할 확률이 된다. 이를 요약하면 다음과 같이 정리할 수 있다. X 의 확률질량함수가 $p_X(x)$ 일 때 $Y = g(X)$ 의 확률질량함수 $p_Y(y)$ 는 다음과 같이 구할 수 있다.

$$
\begin{aligned}
p_Y(y) &= P(Y = y) \;=\; P(g(X) = y) \\
&= P(X \in g^{-1}(y)) \\
&= \sum_{x \in g^{-1}(y)} p_X(x)
\end{aligned}
$$

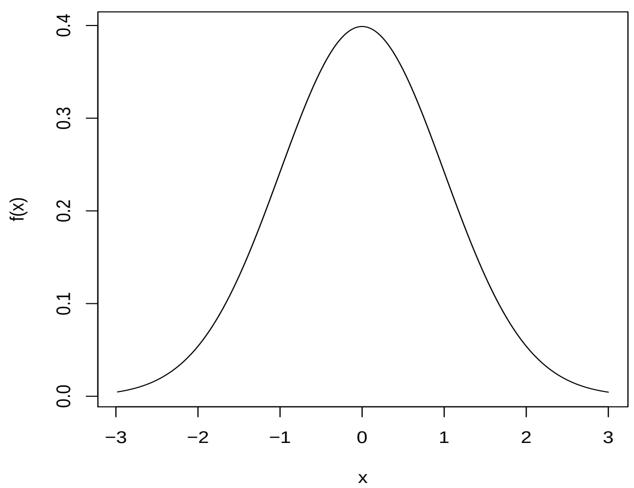

그림 2.7: 표준정규분포의 확률밀도함수

앞에서 설명한 이산형 확률변수의 변환을 예를 통해 알아보도록 하자. 예를 들어 X의 확률질량함수가 다음과 같을 때 $Y = X^2$의 확률질량함수를 구해보도록 하자.

x	-1	0	1	2
$p_X(x)$	0.1	0.2	0.3	0.4

X가 가지는 값은 -1, 0, 1, 2이므로 이를 제곱한 Y는 0, 1, 4의 값을 가진다. $Y = 0$인 사건은 $X = 0$에 대응되므로 $Y = 0$일 확률은 다음과 같다.

$$P(Y = 0) \;=\; P(X = 0) \;=\; 0.2$$

$Y = 1$인 사건은 $X = -1$ 또는 $X = 1$에 대응되므로 $Y = 1$일 확률은 다음과 같다.

$$P(Y = 1) \;=\; P(X = -1) + P(X = 1) \;=\; 0.4$$

이 예에서 $g^{-1}(1)$은 -1과 1을 원소로 가지는 집합이라 할 수 있다. 한편 $Y = 4$인 사건은 $X = 2$인 경우에 대응되므로 $Y = 4$일 확률은 다음과 같다.

$$P(Y = 4) \;=\; P(X = 2) \;=\; 0.4$$

따라서 $Y = X^2$의 확률질량함수는 다음과 같다.

y	0	1	4
$p_Y(y)$	0.2	0.4	0.4

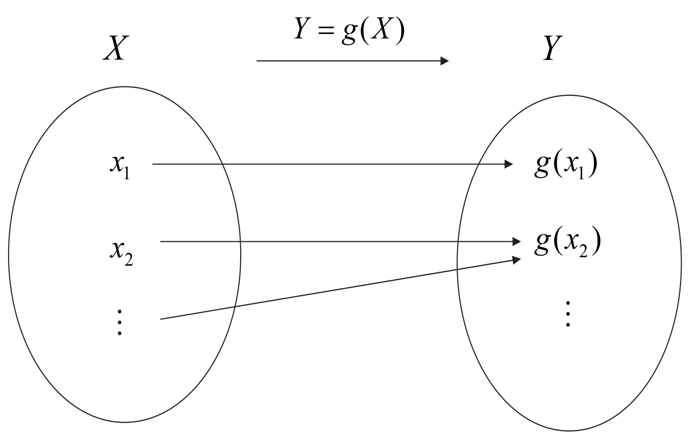

그림 2.8: $Y = g(X)$의 변환

2.3.2 연속형 확률변수의 변환

이산형 확률변수의 변환이 이산형 확률변수가 되는 것처럼 연속형 확률변수 X를 변환한 $Y = g(X)$도 일반적으로 연속형 확률변수가 된다. 연속형 확률변수 $Y = g(X)$의 확률분포를 구하는 방법은 먼저 Y의 누적분포함수를 구하고 이를 미분하여 확률밀도함수를 구할 수 있다. 여기에서는 혼동을 피하기 위하여 X의 확률밀도함수와 누적분포함수를 $f_X(x)$와 $F_X(x)$로 나타내고 Y의 확률밀도함수와 누적분포함수를 $f_Y(y)$와 $F_Y(y)$로 나타내도록 하자.

X의 확률밀도함수 $f_X(x)$로부터 Y의 확률밀도함수 $f_Y(y)$를 구하는 방법을 간단한 형태의 선형변환부터 생각하여 보자.

(i) $Y = 2X$일 때 Y의 누적분포함수를 X의 누적분포함수로 나타내면 다음과 같다.

$$F_Y(y) \;=\; P(Y \le y) \;=\; P(2X \le y) \;=\; P\left(X \le \frac{y}{2}\right) \;=\; F_X\left(\frac{y}{2}\right)$$

X의 누적분포함수 $F_X(\cdot)$를 미분하면 $f_X(\cdot)$이므로 Y의 확률밀도함수는 다음과 같다.

$$f_Y(y) = \frac{d}{dy}F_Y(y) \;=\; \frac{d}{dy}F_X\left(\frac{y}{2}\right) \;=\; \frac{1}{2}f_X\left(\frac{y}{2}\right)$$

(ii) 양의 상수 a에 대하여 $Y = aX+b$의 누적분포함수를 X의 누적분포함수로 나타내면 다음과 같다.

$$F_Y(y) \;=\; P(Y \le y) \;=\; P(aX + b \le y) \;=\; P\left(X \le \frac{y-b}{a}\right) \;=\; F_X\left(\frac{y-b}{a}\right)$$

이를 y에 대하여 미분하여 구할 수 있는 Y의 확률밀도함수는 다음과 같다.

$$f_Y(y) \;=\; \frac{d}{dy}F_Y(y) \;=\; \frac{d}{dy}F_X\left(\frac{y-b}{a}\right) \;=\; \frac{1}{a}f_X\left(\frac{y-b}{a}\right) \tag{2.10}$$

(iii) 음의 상수 a에 대하여 $Y = aX+b$의 누적분포함수를 X의 누적분포함수로 나타내면 다음과 같다.

$$F_Y(y) \;=\; P(Y \le y) \;=\; P(aX + b \le y) \;=\; P\left(X \ge \frac{y-b}{a}\right) \;=\; 1 - \;F_X\left(\frac{y-b}{a}\right)$$

이를 미분한 Y의 확률밀도함수는 다음과 같다.

$$f_Y(y) \;=\; \frac{d}{dy}F_Y(y) \;=\; \frac{d}{dy}\left(1 - \;F_X\left(\frac{y-b}{a}\right)\right) \;=\; -\frac{1}{a}f_X\left(\frac{y-b}{a}\right) \tag{2.11}$$

앞의 선형변환 $Y = aX + b$에 대하여 상수 a가 양수일 때와 음수일 때로 나누어 구한 확률밀도함수는 하나의 식으로 나타낼 수 있다. Y의 확률밀도함수를 나타내는 식 (2.10)과 식 (2.11)의

차이점은 a가 양수일 경우에 $\frac{1}{a}$이고 a가 음수인 경우는 $-\frac{1}{a}$이라는 것이다. 이를 하나로 표현하면 $\left|\frac{1}{a}\right|$로 나타낼 수 있으므로 $Y = aX + b$일 때 Y의 확률 밀도함수는 다음과 같다.

$$f_Y(y) = \left|\frac{1}{a}\right| f_X\left(\frac{y-b}{a}\right) \tag{2.12}$$

예 2.14 X가 $\mathrm{N}(\mu, \sigma^2)$를 따르는 확률변수일 때 $Y = aX + b$의 확률분포를 구하여라. 또한 $Y = \frac{X-\mu}{\sigma}$는 표준정규분포를 따름을 보여라.

[풀이] 위의 식 (2.12)를 이용하면 Y의 확률밀도함수를 구하면 다음과 같다.

$$\begin{aligned} f_Y(y) &= \left|\frac{1}{a}\right| f_X\left(\frac{y-b}{a}\right) \\ &= \frac{1}{|a|}\frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{\left(\frac{y-b}{a}-\mu\right)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi\sigma^2 a^2}}\, e^{-\frac{(y-a\mu-b)^2}{2\sigma^2 a^2}} \end{aligned}$$

이는 정규분포를 나타내는 확률밀도함수에서 μ 대신에 $a\mu + b$, σ^2 대신에 $a^2\sigma^2$을 사용한 식이다. 따라서 $Y = aX + b$는 정규분포 $\mathrm{N}(a\mu + b, a^2\sigma^2)$를 따르는 확률변수이다. 한편 $a = \frac{1}{\sigma}$이고 $b = -\frac{\mu}{\sigma}$일 때 $a\mu + b = 0$, $a^2\sigma^2 = 1$이므로, $Y = \frac{X-\mu}{\sigma}$의 확률분포는 표준정규분포임을 알 수 있다. ■

이번에는 선형변환 대신에 좀더 일반적인 변환 $Y = g(X)$의 확률분포를 구하는 방법을 알아보도록 하자. 먼저 함수 $g(\cdot)$의 역함수 $g^{-1}(\cdot)$가 존재하고 미분가능한 경우를 알아보자. 역함수가 존재한다는 것은 $g(\cdot)$는 단조 증가하거나 단조 감소한다는 것을 의미한다. $g(\cdot)$가 단조증가함수일 때 $Y = g(X)$의 누적분포함수를 X의 누적분포함수로 나타내면 다음과 같다.

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

따라서 Y의 확률밀도함수는 다음과 같다.

$$f_Y(y) \;=\; \frac{d}{dy}F_Y(y) \;=\; f_X(g^{-1}(y)) \cdot \frac{d}{dy}g^{-1}(y) \tag{2.13}$$

한편 $g(\cdot)$가 단조감소함수이면 부등식의 영역 $\{g(X) \le y\}$를 만족하는 X의 영역은 $\{X \ge g^{-1}(y)\}$이다. 따라서 Y의 누적분포함수와 이를 미분한 확률밀도함수는 다음과 같이 구할 수

있다.

$$\begin{aligned} F_Y(y) &= P(Y \le y) \ = \ P(g(X) \le y) \ = \ P(X \ge g^{-1}(y)) \ = \ 1 - F_X(g^{-1}(y)) \\ f_Y(y) &= \frac{d}{dy}F_Y(y) \ = \ -f_X(g^{-1}(y)) \cdot \frac{d}{dy}g^{-1}(y) \end{aligned} \quad (2.14)$$

함수 $g(\cdot)$ 의 역함수가 미분가능할 때 $Y = g(X)$ 의 확률밀도함수를 나타내는 식 (2.13) 과 (2.14) 의 차이점은 단조감소함수일 때 '-1' 이 곱해져 있다는 것이다. 그런데 $g(\cdot)$ 가 단조증가함수일 때 역함수 $g^{-1}(\cdot)$ 도 단조증가함수이다. 따라서 이를 미분한 $\frac{d}{dy}g^{-1}(y)$ 는 양의 값을 가진다. 반대로 $g(\cdot)$ 가 단조감소함수일 때 역함수 $g^{-1}(\cdot)$ 도 단조증가함수이므로 $\frac{d}{dy}g^{-1}(y)$ 는 음의 값을 가진다. 따라서 $g(\cdot)$ 가 단조증가함수일 때와 단조감소함수일 때를 함께 묶어서 다음 정리를 얻을 수 있다.

정리 2.1 $f_X(x)$ 는 연속형 확률변수 X 의 확률밀도함수이며 함수 $g(\cdot)$ 의 역함수가 존재하고 미분가능할 때, $Y = g(X)$ 의 확률밀도함수는 다음과 같다.

$$f_Y(y) \ = \ f_X(g^{-1}(y)) \cdot \left| \frac{d}{dy}g^{-1}(y) \right| \quad (2.15)$$

앞의 정리 2.1에서 한 가지 주의할 사항은 실수 전체에서 $y = g(x)$ 의 역함수가 존재하고 미분가능할 필요는 없다는 것이다. 위의 식 (2.15) 를 사용하여 $Y = g(X)$ 의 확률밀도함수를 구하기 위한 조건은 X 가 값을 가지는 영역에서 $g(\cdot)$ 의 역함수가 존재하고 미분가능하면 된다. 참고로 아래의 예에서 $g(x) = \frac{1}{x}$ 은 $x = 0$ 에서 함수값이 존재하지 않는 경우이고 $X = 0$ 일 확률은 0이다.

예 2.15 X 가 Uniform$[0, 1]$ 를 따르는 확률변수일 때 $Y = \dfrac{1}{X}$ 의 확률밀도함수를 구하여라.

[풀이] 함수 $g(x) = \frac{1}{x}$ 는 영역 $(0, 1]$ 에서 역함수 $g^{-1}(y) = \frac{1}{y}$ 가 존재하고 미분가능한 함수이다. 주어진 함수가 미분가능하지 않은 한 점 $x = 0$ 의 확률은 0이므로 정리 2.1을 적용하여 $Y = \dfrac{1}{X}$ 의 확률밀도함수를 구할 수 있다. X 의 확률밀도함수는 $f_X(x) = 1, \ 0 \le x \le 1$ 이므로 $Y = \dfrac{1}{X}$ 의 확률밀도함수는 다음과 같다.

$$f_Y(y) = f_X(g^{-1}(y)) \cdot \left| \frac{d}{dy}g^{-1}(y) \right| = f_X\left(\frac{1}{y}\right) \cdot \left| \frac{d}{dy}\frac{1}{y} \right| = \frac{1}{y^2}, \quad 1 \le y < \infty$$

위에서 확률변수 X 의 범위는 $[0, 1]$ 이므로 $Y = \dfrac{1}{X}$ 의 범위는 $[1, \infty)$ 로 바뀐다. ■

정리 2.1은 함수 $g(\cdot)$의 역함수가 존재하고 미분가능한 경우만 적용될 수 있으며 일반적인 변환 $Y = g(X)$에 대해서는 어떤 공식에 의하여 Y의 확률밀도함수를 구할 수는 없다. 일반적인 변환에 대해서는 먼저 Y의 누적분포함수를 찾고 이를 이용하여 확률밀도함수를 구하여야 한다. 다음 예는 이러한 과정을 보여주고 있다.

예 2.16 X가 표준정규분포를 따르는 확률변수일 때 $Y = X^2$의 확률분포를 구하여라.

[풀이] X는 표준정규분포를 따르는 확률변수이므로 $f_X(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ 이다. 한편 Y의 누적분포함수를 X의 누적분포함수로 나타내면 다음과 같다.

$$F_Y(y) = P(Y \le y) = P(-\sqrt{y} \le X \le \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$$

이를 미분한 Y의 확률밀도함수는 다음과 같다.

$$\begin{aligned} f_Y(y) &= \frac{d}{dy}F_Y(y) = f_X(\sqrt{y})\frac{1}{2\sqrt{y}} + f_X(-\sqrt{y})\frac{1}{2\sqrt{y}} \\ &= \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y}\frac{1}{\sqrt{y}} = \frac{1}{\sqrt{\pi}}\left(\frac{1}{2}\right)^{\frac{1}{2}} y^{\frac{1}{2}-1}e^{-\frac{1}{2}y}, \quad y > 0 \end{aligned}$$

이는 $\alpha = \frac{1}{2}$, $\lambda = \frac{1}{2}$인 감마분포, 즉 자유도가 1인 카이제곱분포의 확률밀도함수이다. ■

변수변환을 이용하여 다음 두 정리를 얻을 수 있다.

정리 2.2 연속형 확률변수 X의 누적분포함수가 $F(\cdot)$일 때 $Y = F(X)$는 균일분포를 갖는다. 즉 다음이 성립한다.

$$Y = F(X) \ \sim \ \text{Uniform}(0, 1)$$

[증명] 0과 1 사이의 값 y에 대하여 Y의 누적분포함수는 다음과 같이 나타낼 수 있다.

$$F_Y(y) \ = \ P(Y \le y) \ = \ P(F(X) \le y)$$

연속형 확률변수의 누적분포함수인 $F(x)$는 단조증가함수이며 그 역함수가 존재하게 된다. 따라서 사건 $\{F(X) \le y\}$는 $\{X \le F^{-1}(y)\}$와 같고 Y의 누적분포함수는 다음과 같다.

$$F_Y(y) \ = \ P(X \le F^{-1}(y)) \ = \ F(F^{-1}(y)) \ = \ y$$

즉 Y는 Uniform$(0, 1)$을 따른다. ■

위 정리를 이용하면 $X \sim \mathrm{Exp}(\lambda)$ 일 때, X 의 누적분포함수는 $F(x) = 1 - e^{-\lambda x}$ 이므로 $Y = 1 - e^{-\lambda X}$ 의 확률분포는 $\mathrm{Uniform}(0,1)$ 임을 알 수 있다.

정리 2.3 U 가 $\mathrm{Uniform}(0,1)$ 을 따르는 확률변수이고 $F(\cdot)$ 는 연속형 확률분포의 누적분포함수일 때 $Y = F^{-1}(U)$ 는 누적분포함수 $F(\cdot)$ 를 갖는 확률변수이다.

[증명] Y 의 누적분포함수는 다음과 같음을 확인할 수 있다.

$$F_Y(y) = P(Y \le y) = P(F^{-1}(U) \le y) = P(U \le F(y)) = F(y)$$

■

예를 들어 지수분포의 누적분포함수는 $F(x) = 1 - e^{-\lambda x}$ 이고 그 역함수는 $F^{-1}(x) = -\frac{1}{\lambda}\log(1 - x)$ 이다. 따라서 $U \sim \mathrm{Uniform}(0,1)$ 일 때 다음과 같이 정의된 Y 는 $\mathrm{Exp}(\lambda)$ 를 따르는 확률변수이다.

$$Y = -\frac{1}{\lambda}\log(1 - U)$$

위 정리를 활용하면 $\mathrm{Uniform}(0,1)$ 를 따르는 랜덤넘버로부터 다른 확률분포를 따르는 랜덤넘버를 만들 수 있다. 컴퓨터를 이용하여 $\mathrm{Uniform}(0,1)$ 로부터 랜덤넘버를 생성하였을 때, 이 값을 어떤 확률분포의 누적분포함수의 역함수를 통하여 변환하면 이 값은 주어진 확률분포를 따르는 랜덤넘버가 된다. 예를 들어 $\mathrm{Uniform}(0,1)$ 로부터 생성된 랜덤넘버를 다음과 같이 나타낼 때

$$U_1,\ U_2, \ldots,\ U_n$$

지수분포의 누적함수의 역함수인 $F^{-1}(x) = -\frac{1}{\lambda}\log(1 - x)$ 를 통하여 다음과 같이 변환한 랜덤넘버는 지수분포를 따르게 된다.

$$-\frac{1}{\lambda}\log(1 - U_1),\ -\frac{1}{\lambda}\log(1 - U_2), \ldots,\ -\frac{1}{\lambda}\log(1 - U_n)$$

연습문제

1. 검정색 공 3개와 흰색 공 2개가 들어있는 상자에서 임의로 2개의 공을 꺼내어 색깔을 관찰하였다.
   a) 표본공간을 구하여라.
   b) 추출된 검정색 공의 수에서 흰색 공의 수를 뺀 값을 X라 할 때 X의 확률분포를 구하여라.
2. 두 개의 주사위를 던지는 실험에서 두 눈의 수 중의 큰 값을 X라 할 때, X의 확률분포를 구하여라.
3. 상자 A에는 무수히 많은 검정색 공이 들어있고 상자 B에는 하나의 흰색 공이 들어있다. 상자 A에서 검정색 공을 꺼내어 상자 B에 집어넣은 후 상자 B에서 임의로 하나의 공을 선택하는 실험을 하였다. 이때 상자 A에서 꺼내어 상자 B로 옮기는 검정색 공의 수는 포아송분포를 따른다고 한다.
   a) 상자 B로부터 꺼낸 공이 흰색일 확률을 구하여라.
   b) 상자 B로부터 꺼낸 공이 흰색이었을 때 상자 A에서 꺼내어 상자 B로 옮긴 검정색 공의 수에 대한 확률분포를 구하여라.
4. 연속형 확률변수 X의 확률밀도함수가 아래와 같을 때

   $$f(x) \ = \ cx^3, \quad 0 \le x \le 1$$

   다음 물음에 답하여라.
   a) 상수 c의 값을 구하여라.
   b) $Y = -\log X$의 확률밀도함수를 구하여라.
5. 이산형 확률변수 X의 확률질량함수가 다음과 같을 때 $Y = (X-1)^2$의 확률질량함수를 구하여라.

x	0	1	2	3	4
$p_X(x)$	0.3	0.1	0.2	0.1	0.3

6. $X \sim \text{Uniform}(-1, 2)$일 때 $Y = X^2$의 확률밀도함수를 구하여라.
7. 연속형 확률변수 X의 확률밀도함수가 $f(x) = xe^{-x},\ x > 0$일 때 $Y = X^2$의 확률밀도함수를 구하여라.
8. 다음과 같이 정의된 확률변수 Y의 확률밀도함수를 구하여라.

a) X가 표준정규분포를 따르는 확률변수일 때 $Y = |X|$이다.

b) X의 확률밀도함수가 $f_X(x) = \dfrac{3}{8}x^2, \quad 0 < x < 2$일 때 $Y = X^2$이다.

9. 연속형 확률변수 X의 확률밀도함수가 $f(x) = 3x^2,\ 0 < x < 1$이다.

   a) $Y = 3\log X$의 확률밀도함수를 구하여라.

   b) $Y = (X - \frac{1}{2})^2$의 확률밀도함수를 구하여라.

10. $X \sim \text{Gamma}(\alpha, \lambda)$일 때 $Y = cX$의 확률분포를 구하여라. 단, c는 양의 상수이다.

11. 성공의 확률이 p인 독립적인 베르누이 시행에서 2회 연속 성공이나 실패가 나올 때까지 시행 횟수를 X라 할 때, X의 확률질량함수를 구하여라.

12. $F(\cdot)$는 정수 값을 갖는 이산형 확률변수의 누적분포함수이고 $U \sim \text{Uniform}(0, 1)$일 때 다음과 같이 정의된 확률변수 X의 확률질량함수를 구하여라.

$$\text{정수 } k\text{에 대하여 } F(k-1) \le U < F(k)\text{일 때} \quad \Longleftrightarrow \quad X = k$$

13. 확률변수 T는 지수분포를 따르는 확률변수일 때 다음과 같이 정의된 확률변수 X의 확률질량함수를 구하여라.

$$k = 0, 1, 2 \ldots\text{인 정수 } k\text{에 대하여 } k \le T < k+1\text{일 때} \quad \Longleftrightarrow \quad X = k$$

제 3 장

결합분포

같은 표본공간에서 정의된 두 개 이상의 확률변수의 확률분포를 **결합분포**(joint distribution) 또는 **결합확률분포**(joint probability distribution)라 한다. 예를 들어 동전을 3번 던지는 실험에서 X를 첫 번째 동전의 앞면의 수, Y를 전체 동전의 앞면의 수라 하자. 이 경우 같은 표본공간 위에 두 개의 확률변수 X와 Y가 정의되어 있다. 이 실험에서 표본공간은 다음과 같다.

$$\Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$$

표본공간의 각 원소는 X와 Y에 의하여 다음과 같이 실수로 옮겨진다.

$$\begin{aligned}
X(HHH) &= 1, \quad Y(HHH) = 3, \\
X(HHT) &= 1, \quad Y(HHT) = 2, \\
&\vdots \\
X(TTT) &= 0, \quad Y(TTT) = 0
\end{aligned}$$

예를 들어 $(X = 0, Y = 0)$인 사건은 표본공간에서 TTT인 경우에 대응된다. 따라서 $P(X = 0, Y = 0) = P(TTT) = \frac{1}{8}$이다. 이와 같은 방법으로 X와 Y가 가지는 각 값에서 확률을 구하면 표 3.1과 같은 결합분포를 구할 수 있다.

확률분포는 누적분포함수에 의하여 결정되는 것과 같이 두 확률변수 X와 Y의 결합분포도

표 3.1: 동전을 3번 던지는 실험에서 X와 Y의 결합분포

		y			
		0	1	2	3
x	0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	0
	1	0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$

$\{X \le x,\ Y \le y\}$ 의 확률에 의하여 결정이 된다. 이 확률을 다음과 같이 나타낼 수 있다.

$$F(x,y) \;=\; P(X \le x, Y \le y)$$

이를 (X,Y) 의 결합누적분포함수라 한다. 일반적으로 n 개의 확률변수 $(X_1, X_2, \ldots, X_n)$ 에 대한 결합누적분포함수 $F(x_1, x_2, \ldots, x_n)$ 은 다음과 같다.

$$F(x_1, x_2, \ldots, x_n) \;=\; P(X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n)$$

결합분포를 갖는 여러 확률변수 중에서 일부 확률변수의 확률분포를 **주변분포**(marginal distribution)라 한다. 예를 들어 결합분포를 갖는 두 확률변수 (X,Y) 가 있을 때 X 의 확률분포 또는 Y 의 확률분포를 주변분포라 한다. 한편 확률분포에 대한 누적분포함수, 확률질량함수 또는 확률밀도함수를 결합분포에서는 결합누적분포함수, 결합확률질량함수 또는 결합확률밀도함수라 한다. 또한 주변분포에 대해서는 주변누적분포함수, 주변확률질량함수 또는 주변확률밀도함수라는 용어를 사용한다. 결합분포에서는 여러 확률변수를 동시에 다루기 때문에 혼동을 피하기 위하여 특정한 확률변수의 누적분포함수나 확률밀도함수(확률질량함수)를 나타낼 때 첨자를 붙여 나타내기도 한다. 예를 들어 $F_X(x)$ 와 $f_X(x)$ 는 X 의 주변누적분포함수와 주변확률밀도함수를 나타낸다.

(X,Y) 에 대한 결합분포가 주어져 있을 때 X 또는 Y 의 주변분포는 결합누적분포함수로부터 구할 수 있다. X 의 주변누적분포함수는 $F_X(x) = P(X \le x)$ 이고 사건 $\{X \le x\}$ 는 Y 에 대한 어떤 제약도 없으므로 다음과 같이 나타낼 수 있다.

$$\{X \le x\} \;=\; \{X \le x,\ Y < \infty\}$$

따라서 X 의 주변누적분포함수는 다음과 같이 구할 수 있다.

$$\begin{aligned} F_X(x) &= P(X \le x) \;=\; P(X \le x,\ Y < \infty) \\ &= \lim_{y\to\infty} F(x,y) \end{aligned}$$

일반적으로 $\lim_{y\to\infty} F(x,y)$ 를 $F(x,\infty)$ 로 간단히 나타내기도 한다. 마찬가지로 Y 의 주변누적분포함수는 다음과 같이 구할 수 있다.

$$F_Y(y) \;=\; P(X < \infty,\ Y \le y) \;=\; F(\infty, y)$$

예 3.1 (X,Y) 의 결합누적분포함수가 다음과 같을 때

$$F(x,y) \;=\; (1 - e^{-x}) - \frac{1}{1+y}(1 - e^{-(1+y)x}), \quad x > 0, y > 0$$

X 와 Y 의 주변누적분포함수를 구하여라.

[풀이] X의 주변누적분포함수는 다음과 같고

$$F_X(x) \;=\; F(x,\infty) \;=\; 1-e^{-x}, \quad x>0$$

Y의 주변누적분포함수는 다음과 같다.

$$F_Y(y) \;=\; F(\infty,y) \;=\; 1-\frac{1}{(1+y)}, \quad y>0$$

■

두 개의 확률변수로 이루어진 결합분포를 **이변량분포**(bivariate distribution)라 하며 이 장에서는 주로 두 개의 확률변수로 이루어진 결합분포를 다룰 것이다. n개의 확률변수로 이루어진 결합분포의 성질은 이변량분포의 성질을 확장하여 유추할 수 있다.

3.1 이산형 결합분포

결합분포를 가지는 두 확률변수 X와 Y가 가질 수 있는 값의 수가 유한개이거나 셀 수 있을 때 (X, Y)를 이산형 결합확률변수라 한다. (X, Y)가 가질 수 있는 값을 다음과 같이 나타낼 때

$$(x_i,\, y_j), \quad i=1,2,\ldots,\; j=1,2,\ldots$$

각 $(x_i,\, y_j)$에서의 확률을 나타낸 함수를 결합확률질량함수(joint probability mass function)라 하며 $p(x_i, y_j)$ 또는 $f(x_i, y_j)$로 나타낸다. 즉 결합확률질량함수 $p(x_i, y_j)$는 다음과 같다.

$$p(x_i,y_j) \;=\; P(X=x_i,\, Y=y_j)$$

(X, Y)의 결합확률질량함수로부터 X와 Y의 주변확률질량함수는 다음과 같이 구할 수 있다. 먼저 X의 주변확률질량함수 $p_X(x_i)$는 $X=x_i$일 확률이고 사건 $\{X=x_i\}$는 Y가 어떤 값을 가지더라도 상관이 없으므로 $\cup_j\{X=x_i,\; Y=y_j\}$로 나타낼 수 있다. 따라서 X의 주변확률질량함수는 다음과 같다.

$$\begin{aligned} p_X(x_i) \;&=\; P(X=x_i) \;=\; P(\cup_j\{X=x_i,\; Y=y_j\}) \\ &=\; \sum_j P(X=x_i, Y=y_j) \\ &=\; \sum_j p(x_i,y_j) \end{aligned} \tag{3.1}$$

같은 방법으로 Y의 주변확률질량함수를 구하면 $p_Y(y_j)$는 다음과 같다.

$$p_Y(y_j) \;=\; \sum_i p(x_i,y_j) \tag{3.2}$$

표 3.2: 동전을 3번 던지는 실험에서 X 와 Y 의 주변분포

X 의 주변확률질량함수

x	0	1
$p_X(x)$	$\frac{1}{2}$	$\frac{1}{2}$

Y 의 주변확률질량함수

y	0	1	2	3
$p_Y(y)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

동전을 3번 던지는 실험에서 X 는 첫 번째 동전의 앞면의 수, Y 는 전체 동전의 앞면의 수라고 정의한 예에서 주변확률질량함수를 구하여보자. 표 3.1에 주어진 (X, Y) 의 결합확률질량함수로부터 X 와 Y 의 주변확률질량함수는 다음과 같이 구할 수 있다. Y 는 0, 1, 2, 3 중의 하나의 값을 가지므로 X 의 주변확률질량함수는 다음과 같다.

$$\begin{aligned} p_X(0) &= p(0,0) + p(0,1) + p(0,2) + p(0,3) = \frac{1}{2} \\ p_X(1) &= p(1,0) + p(1,1) + p(1,2) + p(1,3) = \frac{1}{2} \end{aligned}$$

즉 X 의 주변확률질량함수는 표 3.1의 각 행에 주어진 결합확률의 합이 된다. 같은 방법으로 Y 의 주변확률질량함수는 각 열에 주어진 결합확률의 합이 된다. 위와 같이 구한 X 와 Y 의 주변확률질량함수는 표 3.2와 같다. 이 예에서 X 와 Y 의 주변확률질량함수는 각 확률변수에 대해 개별적으로 구한 확률질량함수와 같음을 쉽게 확인할 수 있다. 즉 처음부터 X 와 Y 의 결합분포를 생각하지 않고 단순히 X 는 첫 번째 동전의 앞면의 수로 생각하면 X 의 확률분포는 위에서 결합분포로부터 구한 주변확률분포와 같음을 알 수 있다.

두 개의 확률변수에 대한 결합확률질량함수로부터 주변확률질량함수를 구하는 방법은 확률변수가 여러 개인 경우로 쉽게 확장될 수 있다. n 개의 확률변수 $(X_1, X_2, \ldots, X_n)$ 의 결합확률질량함수가 $p(x_1, x_2, \ldots, x_n)$ 일 때 X_1 의 주변확률질량함수는 다음과 같이 구할 수 있다.

$$p_{X_1}(x_1) = \sum_{x_2, x_3, \ldots, x_n} p(x_1, x_2, \ldots, x_n)$$

위 식에서 합은 $X_2, X_3, \ldots, X_n$ 의 모든 가능한 값 $x_2, x_3, \ldots, x_n$ 에 대한 결합확률질량함수의 합을 나타낸다.

3.2 연속형 결합분포

결합분포를 가지는 두 확률변수 X 와 Y 가 어느 구간에 속하는 모든 값을 가질 수 있을 때 (X, Y) 를 연속형 결합확률변수라 한다. (X, Y) 의 결합누적분포함수 $F(x, y)$ 를 x 와 y 로 미분한 함수를 결합확률밀도함수(joint probability density function)라 하며 $f(x, y)$ 로 나타낸다. 즉 결합확률밀도함수 $f(x, y)$ 는 다음과 같다.

$$f(x, y) = \frac{d^2}{dxdy} F(x, y)$$

반대로 결합확률밀도함수 $f(x,y)$로부터 결합누적분포함수를 다음과 같이 구할 수 있다.

$$F(x,y) \;=\; \int_{-\infty}^{x}\int_{-\infty}^{y} f(s,t)\; dtds$$

결합확률밀도함수는 다음과 같은 성질을 가진다.

(i) $f(x,y) \ge 0$

(ii) $\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\; dx\, dy \;=\; 1$

(iii) $\displaystyle P((X,Y)\in A) \;=\; \int\int_{A} f(x,y)\; dx\, dy$

예 3.2 (X,Y)의 결합확률밀도함수가 $f(x,y)=\lambda^2 e^{-\lambda y},\; 0 \le x \le y$라 한다.

(a) $P(X \le 1,\, Y \le 2)$를 구하여라.

(b) $P(X \le 1)$를 구하여라.

[풀이]

(a) $\{X \le 1,\, Y \le 2\}$일 확률은 $\{(x,y):\; x \le 1, y \le 1\}$에서 결합확률밀도함수를 적분하여 구할 수 있다. 그런데 결합확률밀도함수는 $0 \le x \le y$인 부분에서만 0이 아닌 값을 가지므로 적분하는 영역은 그림 3.1과 같이 $\{0 \le x \le 1,\, x \le y \le 2\}$이 된다. 따라서 구하는 확률은 다음과 같다.

$$\begin{aligned} P(X \le 1,\, Y \le 2) \;&=\; \int_0^1\int_x^2 \lambda^2 e^{-\lambda y}\; dy\, dx \;=\; \int_0^1 \lambda(e^{-\lambda x} - e^{-2\lambda})\; dx \\ &=\; 1 - e^{-\lambda} - \lambda e^{-2\lambda} \end{aligned}$$

(b) 결합확률밀도함수는 $0 \le x \le y$인 부분에서만 0이 아닌 값을 가지므로 $\{X \le 1\}$일 확률을 구하기 위해 적분하는 영역은 $\{0 \le x \le 1,\, x \le y < \infty\}$이 된다. 따라서 구하는 확률은 다음과 같다.

$$\begin{aligned} P(X \le 1) \;&=\; \int_0^1\int_x^\infty \lambda^2 e^{-\lambda y}\; dydx \;=\; \int_0^1 \lambda e^{-\lambda x}\; dx \\ &=\; 1 - e^{-\lambda} \end{aligned}$$

■

연속형 결합확률변수 (X, Y)의 주변확률밀도함수는 결합확률밀도함수로부터 구할 수 있다. X의 주변누적분포함수 $F_X(x)$는 다음과 같다.

$$\begin{aligned} F_X(x) &= P(X \le x,\, Y < \infty) \\ &= \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(s, y)\, dy\, ds \end{aligned}$$

따라서 X의 주변확률밀도함수 $f_X(x)$는 다음과 같이 구할 수 있다.

$$\begin{aligned} f_X(x) &= \frac{d}{dx} F_X(x) = \frac{d}{dx} \left(\int_{-\infty}^{x} \int_{-\infty}^{\infty} f(s, y)\, dy\, ds \right) \\ &= \int_{-\infty}^{\infty} f(x, y)\, dy \end{aligned} \tag{3.3}$$

즉 X의 주변확률밀도함수를 구하기 위해서는 다른 확률변수인 Y에 대하여 전체 영역에서 결합확률밀도함수를 적분하면 된다. 마찬가지로 Y의 주변확률밀도함수는 다음과 같이 구할 수 있다.

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx \tag{3.4}$$

주변확률밀도함수를 구하는 위의 식 (3.3)과 (3.4)를 이산형 확률변수에 대하여 주변확률질량함수를 구하는 식 (3.1)과 (3.2)에 비교하면 이산형의 경우는 합을 사용하는 반면 연속형의 경우는 적분을 사용한다는 것 외에는 본질적으로 같은 형태라 할 수 있다.

예 3.3 (X, Y)의 결합확률밀도함수가 $f(x, y) = \lambda^2 e^{-\lambda y},\ 0 < x < y$라 한다.

(a) X의 주변확률밀도함수를 구하여라.

(b) Y의 주변확률밀도함수를 구하여라.

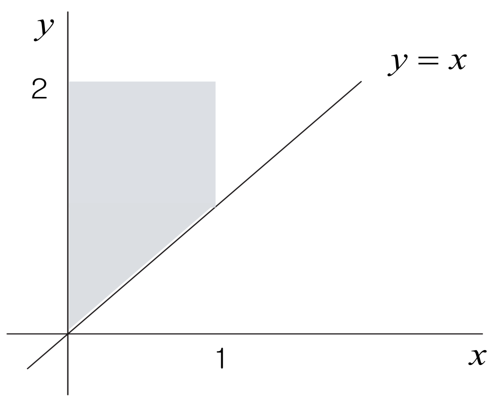

그림 3.1: $\{0 \le x \le 1,\ x \le y \le 2\}$의 영역

[풀이]

(a) 주어진 $X = x$ 에서 Y 가 값을 가지는 범위는 $\{y \,|\, y > x\}$ 이다. 즉 결합확률밀도함수는 $\{y \,|\, y > x\}$ 인 영역에서만 0이 아닌 값을 가지므로 X 의 주변확률밀도함수는 다음과 같다.

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_{x}^{\infty} \lambda^2 e^{-\lambda y}\,dy = \lambda e^{-\lambda x}, \quad x > 0$$

위 확률밀도함수로부터 X 의 주변분포는 지수분포임을 알 수 있다.

(b) 주어진 $Y = y$ 에서 X 가 값을 가지는 범위는 $\{x \,|\, 0 < x < y\}$ 이다. 따라서 Y 의 주변확률밀도함수는 다음과 같다.

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_{0}^{y} \lambda^2 e^{-\lambda y}\,dx = \lambda^2 y e^{-\lambda y}, \quad y > 0$$

위 확률밀도함수로부터 Y 의 주변분포는 $\text{Gamma}(2, \lambda)$ 임을 알 수 있다.

■

3.3 이변량정규분포

정규분포를 두 개의 확률변수에 대한 결합분포로 확장한 것이 **이변량정규분포**(bivariate normal distribution)이다. 이변량정규분포의 결합확률밀도함수는 다음과 같이 주어진다.

$$\begin{aligned} f(x,y) = \; & \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \\ & \cdot \exp\left[-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)\right)\right] \end{aligned}$$

(X, Y) 가 이변량정규분포를 따르는 결합확률변수일 때 기호로 다음과 같이 나타낸다.

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathrm{N}_2\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}\right)$$

위 확률밀도함수에서 μ_X 와 σ_X^2 은 X 의 중심 위치와 퍼진 정도를 나타내는 값이며 μ_Y 와 σ_Y^2 은 Y 의 중심 위치와 퍼진 정도를 나타내는 값이다. 또 ρ 는 X 와 Y 사이의 선형관계의 정도를 상대적으로 나타낸 값이다.[1]

[1] 4장 4.5절의 상관계수에 대한 내용 참조.

(X,Y)가 이변량정규분포를 따를 때 X의 주변확률밀도함수는 다음과 같이 구할 수 있다.

$$\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty} f(x,y)dy \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \\
&\cdot \exp\left[-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)\right)\right] dy \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \\
&\cdot \exp\left[-\frac{1}{2(1-\rho^2)}\left(\left(\frac{y-\mu_Y}{\sigma_Y} - \rho\frac{x-\mu_X}{\sigma_X}\right)^2 + (1-\rho^2)\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right)\right] dy \\
&= \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right) \\
&\cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_Y^2(1-\rho^2)}} \exp\left(-\frac{1}{2\sigma_Y^2(1-\rho^2)}\left(y-\mu_Y-\rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X)\right)^2\right) dy
\end{aligned}$$

위 식에서 적분 안에 있는 함수는 $\mu = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X)$, $\sigma^2 = \sigma_Y^2(1-\rho^2)$인 정규분포의 확률밀도함수이므로 이 함수를 실수 전체에서 적분한 값은 1이다. 따라서 X의 주변확률밀도함수는 다음과 같다.

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right)$$

즉 X의 주변확률분포는 $\mathrm{N}(\mu_X, \sigma_X^2)$이다. 같은 방법으로 Y의 주변확률밀도함수를 구하면 다음과 같다.

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma_Y^2}} \exp\left(-\frac{1}{2}\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right)$$

따라서 Y의 주변분포는 $\mathrm{N}(\mu_Y, \sigma_Y^2)$임을 알 수 있다.

3.4 확률변수의 독립

두 사건 A와 B에 대하여 $P(A\cap B) = P(A)P(B)$일 때 두 사건은 서로 독립이라 한다. 독립의 개념은 확률변수로 확장될 수 있다. 확률변수는 실수에서 값을 가지므로 두 확률변수 X와 Y의 독립은 모든 실수의 부분집합 A와 B에 대하여 다음이 성립하는 경우로 정의하는 것이 타당할 것이다.

$$P(X\in A, Y\in B) = P(X\in A)P(Y\in B) \tag{3.5}$$

2장에서 언급한 것처럼 확률변수의 확률분포는 사건 $\{X \le x\}$의 확률을 나타내는 누적

분포함수에 의하여 결정된다. 마찬가지로 독립의 조건을 나타내는 식 (3.5)는 $A = (-\infty, x]$, $B = (-\infty, y]$인 부분집합에 대해 성립하면 실수의 모든 부분집합에 대하여도 식 (3.5)가 성립함을 보일 수 있다. 따라서 독립의 정의는 다음과 같이 할 수 있다.

정의 3.1 (X, Y)의 결합누적분포함수가 X와 Y의 주변누적분포함수의 곱일 때, 즉 모든 x와 y에 대하여 다음 식이 성립할 때 두 확률변수 X와 Y는 서로 독립이라 한다.

$$F(x, y) = F_X(x)F_Y(y) \tag{3.6}$$

이산형 확률변수인 경우 위의 독립의 정의와 동치인 조건은 모든 x_i와 y_j에 대하여 다음 식이 성립하는 것이다.

$$p(x_i, y_j) = p_X(x_i)\, p_Y(y_j)$$

즉 결합확률질량함수가 X와 Y의 주변확률질량함수의 곱이 될 때 서로 독립이 된다. 연속형 확률변수인 경우는 모든 x와 y에 대하여 다음 식이 성립할 때 서로 독립이 된다.

$$f(x, y) = f_X(x)f_Y(y)$$

즉 연속형의 경우 결합확률밀도함수가 X와 Y의 주변확률밀도함수의 곱이 될 때 서로 독립이 된다. 연속형 확률변수인 경우 독립의 조건 (3.6)을 미분하면 위의 식을 얻을 수 있다.

두 개의 확률변수에 대한 독립의 정의를 확장하여 n개의 확률변수 $X_1, X_2, \ldots, X_n$에 대한 독립의 정의를 다음과 같이 할 수 있다. 결합누적분포함수가 각 주변누적분포함수의 곱이 될 때, 즉 다음이 성립할 때 $X_1, X_2, \ldots, X_n$는 서로 독립이라 한다.

$$F(x_1, x_2, \ldots, x_n) \;=\; F_{X_1}(x_1) \cdot F_{X_2}(x_2) \cdots F_{X_n}(x_n)$$

이산형 확률변수인 경우 위 정의는 결합확률질량함수가 각 주변확률질량함수의 곱이라는 것과 같으며, 연속형 확률변수인 경우는 결합확률밀도함수가 각 주변확률밀도함수의 곱이라는 것과 같다.

예 3.4 (X, Y)의 결합확률밀도함수가 $f(x, y) = \lambda^2 e^{-\lambda y},\ 0 \le x \le y$일 때 X와 Y는 서로 독립인지 확인하여라.

[풀이] 예 3.3에서 구한 것과 같이 X와 Y의 주변확률밀도함수는 각각 다음과 같다.

$$\begin{aligned} f_X(x) &= \lambda e^{-\lambda x}, \quad x > 0 \\ f_Y(y) &= \lambda^2 y e^{-\lambda y}, \quad y > 0 \end{aligned}$$

따라서 $f(x,y) \neq f_X(x)f_Y(y)$ 이므로 X 와 Y 는 서로 독립이 아니다. ■

예 3.5 (X,Y) 가 이변량정규분포를 따르는 확률변수일 때 X 와 Y 가 서로 독립일 조건을 구하여라.

[풀이] X 와 Y 의 주변확률밀도함수는 다음과 같으므로

$$\begin{aligned} f_X(x) &= \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right) \\ f_Y(y) &= \frac{1}{\sqrt{2\pi\sigma_Y^2}} \exp\left(-\frac{1}{2}\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right) \end{aligned}$$

$\rho = 0$ 일 때 $f(x,y) = f_X(x)f_Y(y)$ 가 된다. 따라서 이변량정규분포에 대한 독립의 조건은 $\rho = 0$ 이다. ■

3.5 조건부분포

어느 사건에 대한 정보가 주어졌을 때 다른 사건이 발생할 확률 (조건부확률) 을 구할 수 있는 것과 같이 어떤 확률변수가 주어졌을 때 다른 확률변수의 확률분포를 구할 수 있다. 이러한 확률분포를 **조건부분포**(conditional distribution)라 한다. $Y = y$ 로 주어졌을 때 X 의 조건부분포를 나타낼 때 흔히 $X|Y = y$ 를 이용하여 표현하기도 한다. 예를 들어 $Y = y$ 일 때 X 의 조건부분포가 $\mathrm{N}(y,\sigma^2)$ 이라면 이를 다음과 같이 나타낸다.

$$X|Y = y \ \sim \mathrm{N}(y,\sigma^2)$$

3.5.1 이산형 확률변수의 조건부분포

결합분포를 가지는 이산형 확률변수 $(X,\, Y)$ 가 가질 수 있는 값을 다음과 같이 나타내도록 하자.

$$(x_i,\, y_j), \quad i = 1,2,\ldots,\; j = 1,2,\ldots$$

두 확률변수 중에서 Y 의 값이 y_j 로 주어졌을 때 X 의 조건부분포를 구하여보자. 이때 Y 의 값은 y_j 로 고정 되어 있지만 X 는 $x_1, x_2, \ldots$ 중의 하나의 값을 가진다. 조건부 확률의 정의에 의하여 $Y = y_j$ 일 때 $X = x_i$ 일 확률은 다음과 같다.

$$P(X = x_i|Y = y_j) \ = \ \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)}$$

이 조건부 확률은 X가 가질 수 있는 각 값 $x_1, x_2, \ldots$에서 구할 수 있다. 각 x_i에서 위의 조건부 확률을 나타낸 함수를 조건부확률질량함수라 하며 $p_{X|Y}(x_i|y_j)$로 나타낸다. 즉 $Y = y_j$일 때 X의 조건부확률질량함수는 다음과 같다.

$$p_{X|Y}(x_i|y_j) \;=\; \frac{p(x_i, y_j)}{p_Y(y_j)} \tag{3.7}$$

마찬가지로 $X = x_i$일 때 Y의 조건부확률질량함수는 다음과 같다.

$$p_{Y|X}(y_j|x_i) \;=\; \frac{p(x_i, y_j)}{p_X(x_i)}$$

조건부확률질량함수 $p_{X|Y}(x_i|y_j)$는 $Y = y_j$일 때 X의 조건부분포에 대한 확률을 나타내므로 다음과 같은 성질을 가지고 있다.

(i) $p_{X|Y}(x_i|y_j) \;\geq\; 0$

(ii) $\sum_i p_{X|Y}(x_i|y_j) \;=\; 1$

(iii) $P(a \leq X \leq b|Y = y_j) \;=\; \sum_{i:a \leq x_i \leq b} p_{X|Y}(x_i|y_j)$

이는 일반적인 확률질량함수가 가지는 성질과 같다. 조건부분포는 각 값에서의 확률이 조건부 확률로 주어진다는 점을 제외하고는 다른 확률분포와 차이점이 없다. 기본적으로 조건부분포도 하나의 확률분포이므로 조건부확률질량함수도 일반적인 확률질량함수가 가지는 성질을 그대로 가진다.

예 3.6 동전을 3번 던지는 실험에서 X를 첫 번째 동전의 앞면의 수, Y를 전체 동전의 앞면의 수라 한다.

(a) $Y = 1$일 때 X의 조건부분포를 구하여라.

(b) $X = 0$일 때 Y의 조건부분포를 구하여라.

[풀이]

(a) X와 Y의 결합분포와 주변분포는 표 3.1과 표 3.2에 주어져 있다. 이를 이용하여 $Y = 1$일 때 X의 조건부확률질량함수를 구하면 다음과 같으므로

$$p_{X|Y}(x_i|1) \;=\; \frac{p(x_i, 1)}{p_Y(1)} \;=\; \frac{p(x_i, 1)}{3/8}$$

X의 조건부분포는 다음과 같다.

x	0	1
$p_{X\|Y}(x\|1)$	$\frac{2}{3}$	$\frac{1}{3}$

(b) $X = 0$일 때 Y의 조건부확률질량함수를 구하면 다음과 같으므로

$$p_{Y|X}(y_j|0) \;=\; \frac{p(0,y_j)}{p_X(0)} \;=\; \frac{p(0,y_j)}{1/2}$$

Y의 조건부분포는 다음과 같다.

y	0	1	2	3
$p_{Y\|X}(y\|0)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	0

■

한편 조건부확률질량함수의 식 (3.7)으로부터 결합확률질량함수를 다음과 같이 구할 수 있다.

$$p(x_i, y_j) \;=\; p_{X|Y}(x_i|y_j)p_Y(y_j) \tag{3.8}$$

즉 Y의 확률분포와 Y가 주어졌을 때 X에 대한 조건부분포로부터 X와 Y의 결합분포를 구할 수 있다. 다음 예는 이러한 방법으로 결합분포를 구하는 것을 보여주고 있다.

예 3.7 N은 포아송분포를 따르는 확률변수이고 $N = n$일 때 X의 조건부분포는 $\mathrm{B}(n,p)$라 한다.

(a) X와 N의 결합분포를 구하여라.

(b) X의 주변분포를 구하여라.

[풀이]

(a) 식 (3.8)에 의하여 X와 N의 결합확률질량함수는 다음과 같다.

$$\begin{aligned} P(X=k, N=n) \;&=\; P(X=k|N=n)P(N=n) \\ &=\; \frac{n!}{k!(n-k)!}\, p^k(1-p)^{n-k}\frac{\lambda^n e^{-\lambda}}{n!} \\ &=\; \frac{1}{k!(n-k)!}\,(\lambda p)^k(\lambda(1-p))^{n-k}e^{-\lambda}, \quad n = 0,1,2\ldots,\; k = 0,1,\ldots,n \end{aligned}$$

(b) X의 주변확률질량함수는 (a)에서 구한 결합확률질량함수를 이용하여 구할 수 있다. $X = k$일 때 N의 값의 범위는 $\{n : n \geq k\}$이므로 X의 주변확률질량함수는 다음과

같다.

$$\begin{aligned}
P(X=k) &= \sum_{n=k}^{\infty} P(X=k, N=n) \\
&= \sum_{n=k}^{\infty} \frac{1}{k!(n-k)!} (\lambda p)^k (\lambda(1-p))^{n-k} e^{-\lambda} \\
&= \frac{(\lambda p)^k e^{-\lambda}}{k!} \sum_{n=k}^{\infty} \frac{1}{(n-k)!} (\lambda(1-p))^{n-k} \\
&= \frac{(\lambda p)^k e^{-\lambda}}{k!} e^{\lambda(1-p)} = \frac{(\lambda p)^k e^{-\lambda p}}{k!}, \quad k=0,1,2,\dots
\end{aligned}$$

위에서 $\sum_{i=0}^{\infty} \frac{a^i}{i!} = e^a$ 이므로 네 번째 등식을 얻을 수 있다. 따라서 X 의 주변분포는 Poisson(λp) 이다.

■

3.5.2 연속형 확률변수의 조건부분포

결합분포를 가지는 연속형 확률변수 (X, Y) 에 대하여 $Y = y$ 일 때 X 의 조건부분포를 구하여보자. 연속형 확률변수의 경우 $\{Y = y\}$ 라는 사건의 확률은 0이므로 주어진 조건 $Y = y$ 는 어떤 의미에서 보면 발생할 수 없는 사건에 대한 정보가 주어진 것이다. 연속형 변수의 경우 주어진 조건 $\{Y = y\}$ 는 Y 의 값이 y 근처인 $[y,\, y + \Delta y]$ 에 속한다는 것으로 해석할 수 있다. 이 조건 하에서 X 가 $[x,\, x + \Delta x]$ 에 속할 확률을 구하여보자. 충분히 작은 Δx 와 Δy 에 대하여 다음이 성립하므로

$$P(x \le X \le x + \Delta x,\ y \le Y \le y + \Delta y) \approx f(x,y)\Delta x \Delta y$$
$$P(y \le Y \le y + \Delta y) \approx f_Y(y)\Delta y$$

조건부확률은 다음과 같다.

$$\begin{aligned}
P(x \le X \le x + \Delta x \,|\, y \le Y \le y + \Delta y) &\approx \frac{f(x,y)\Delta x \Delta y}{f_Y(y)\Delta y} \\
&= \frac{f(x,y)}{f_Y(y)}\Delta x
\end{aligned}$$

즉 $Y \in [y, y + \Delta y]$ 일 때 X 의 값이 구간 $[x, x + \Delta x]$ 에 속할 조건부 확률은 $\frac{f(x,y)}{f_Y(y)}\Delta x$ 이다. 이는 주어진 조건 하에서 X 의 값이 x 근처의 작은 구간에 속할 확률은 $\frac{f(x,y)}{f_Y(y)}$ 와 구간의 길이의 곱으로 주어진다는 것을 의미한다. 연속형 확률변수가 작은 구간에 속할 확률은 근사적으로 확률밀도함수와 구간의 길이의 곱으로 주어지는 것과 비교할 때 $\frac{f(x,y)}{f_Y(y)}$ 를 X 의 조건부분포에 대한 확률밀도함수라 할 수 있다. 이를 조건부확률밀도함수라 하며 $f_{X|Y}(x|y)$ 로 나타낸다. 따라서

$Y = y$ 일 때 X 의 조건부확률밀도함수는 다음과 같이 정의할 수 있다.

$$f_{X|Y}(x|y) \;=\; \frac{f(x,y)}{f_Y(y)} \tag{3.9}$$

또 $X = x$ 일 때 Y 의 조건부확률밀도함수는 다음과 같이 정의할 수 있다.

$$f_{Y|X}(y|x) \;=\; \frac{f(x,y)}{f_X(x)}$$

조건부확률밀도함수 $f_{X|Y}(x|y)$ 는 $Y = y$ 일 때 X 의 확률밀도함수를 나타내므로 다음과 같은 성질을 가지고 있다.

(i) $f_{X|Y}(x|y) \geq 0$

(ii) $\displaystyle\int_{-\infty}^{\infty} f_{X|Y}(x|y)\, dx \;=\; 1$

(iii) $\displaystyle P(a < X \leq b|Y = y) \;=\; \int_a^b f_{X|Y}(x|y)\, dx$

한편 조건부확률밀도함수의 식 (3.9) 로부터 결합확률밀도함수를 다음과 같이 구할 수 있다.

$$f(x,y) \;=\; f_{X|Y}(x|y) f_Y(y)$$

이는 Y 의 주변확률밀도함수와 Y 가 주어졌을 때 X 의 조건부확률밀도함수의 곱으로부터 X 와 Y 의 결합확률밀도함수를 구할 수 있음을 나타낸다. 위의 식은 이산형 확률변수의 경우 확률질량함수에 대해 성립하는 사실이 연속형 확률변수인 경우 확률밀도함수에 대하여도 성립한다는 사실을 보여주고 있다.

예 3.8 (X, Y) 의 결합확률밀도함수가 $f(x,y) = \lambda^2 e^{-\lambda y},\ 0 \leq x \leq y$ 라 한다.

(a) $Y = y$ 일 때 X 의 조건부확률밀도함수를 구하여라.

(b) $X = x$ 일 때 Y 의 조건부확률밀도함수를 구하여라.

[풀이]

(a) 예 3.3에서 구한 것과 같이 Y 의 주변확률밀도함수는 $f_Y(y) = \lambda^2 y e^{-\lambda y},\ y > 0$ 이다. 따라서 $Y = y$ 일 때 X 의 조건부 확률밀도함수는 다음과 같다.

$$f_{X|Y}(x|y) \;=\; \frac{f(x,y)}{f_Y(y)} \;=\; \frac{\lambda^2 e^{-\lambda y}}{\lambda^2 y e^{-\lambda y}} \;=\; \frac{1}{y}, \quad 0 \leq x \leq y$$

따라서 $Y = y$ 일 때 X 의 조건부분포는 $\text{Uniform}(0, y)$ 이다.

(b) X의 주변확률밀도함수는 $f_X(x) = \lambda e^{-\lambda x},\ x > 0$이므로 $X = x$일 때 Y의 조건부확률밀도함수는 다음과 같다.

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{\lambda^2 e^{-\lambda y}}{\lambda e^{-\lambda x}} = \lambda e^{-\lambda(y-x)}, \quad y \geq x$$

따라서 $X = x$일 때 Y의 조건부분포는 지수분포를 x만큼 평행 이동한 것이다.

■

예 3.9 (X, Y)는 이변량정규분포를 따르는 확률변수라 한다.

(a) $X = x$일 때 Y의 조건부확률밀도함수를 구하여라.

(b) $Y = y$일 때 X의 조건부확률밀도함수를 구하여라.

[풀이]

(a) 3.3절에서 구한 X의 주변확률밀도함수는 다음과 같으므로

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right)$$

$X = x$일 때 Y의 조건부확률밀도함수는 다음과 같다.

$$\begin{aligned} f_{Y|X}(y|x) &= \frac{f(x,y)}{f_X(x)} \\ &= \frac{\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left(\left(\frac{y-\mu_Y}{\sigma_Y} - \rho\frac{x-\mu_X}{\sigma_X}\right)^2 + (1-\rho^2)\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right)\right]}{\frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right)} \\ &= \frac{1}{\sqrt{2\pi\sigma_Y^2(1-\rho^2)}} \exp\left(-\frac{1}{2\sigma_Y^2(1-\rho^2)}\left(y - \mu_Y - \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X)\right)^2\right) \end{aligned}$$

즉 $X = x$일 때 Y의 조건부분포는 다음과 같다.

$$Y|X = x \ \sim \ \mathrm{N}\left(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X),\ (1-\rho^2)\sigma_Y^2\right)$$

(b) 같은 방법으로 $Y = y$일 때 X의 조건부분포는 다음과 같음을 보일 수 있다.

$$X|Y = y \ \sim \ \mathrm{N}\left(\mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y-\mu_Y),\ (1-\rho^2)\sigma_X^2\right)$$

■

3.6 결합확률변수의 변환

이 절에서는 X와 Y의 결합분포를 이용하여 $Z = g(X, Y)$의 확률분포를 구하는 방법을 알아보고자 한다. 먼저 $Z = X + Y$인 경우부터 다루도록 하자.

3.6.1 합의 분포

X와 Y가 이산형 확률변수일 때 두 확률변수의 합이 z인 사건은 여러 가능한 경우가 있다. 예를 들어 음이 아닌 정수의 값을 가지는 확률변수 X와 Y에 대하여 합이 5인 경우는 다음과 같은 6가지 경우가 있을 수 있다.

$$(X = 0, Y = 5),\ (X = 1, Y = 4),\ (X = 2, Y = 3),$$
$$(X = 3, Y = 2),\ (X = 4, Y = 5),\ (X = 5, Y = 0)$$

이처럼 두 이산형 확률변수의 합이 z인 사건은 여러 가지 가능한 x에 대하여 $X = x$이고 $Y = z - x$일 때이다. 따라서 $X + Y = z$일 확률은 다음과 같이 구할 수 있다.

$$P(Z = z) \ = \ P(X + Y = z) \ = \ \sum_x P(X = x, Y = z - x)$$

이를 확률질량함수로 표현하면 다음과 같다.

$$p_Z(z) \ = \ \sum_x p(x, z - x) \tag{3.10}$$

만약 X와 Y가 서로 독립이면 Z의 확률질량함수는 다음과 같이 표현할 수 있다.

$$p_Z(z) \ = \ \sum_x p_X(x) p_Y(z - x)$$

예 3.10 $X \sim \text{Poisson}(\lambda_1)$, $Y \sim \text{Poisson}(\lambda_2)$이고 서로 독립일 때 $Z = X + Y$의 확률분포를 구하여라.

[풀이] 음이 아닌 정수 값을 가지는 포아송 확률변수 X와 Y의 합이 n이 되는 경우는 $\{X = 0, Y = n\}$, $\{X = 1, Y = n - 1\}$, $\ldots, \{X = n, Y = 0\}$이 있다. 따라서 $Z = X + Y$의

확률질량함수는 다음과 같다.

$$\begin{aligned}
P(Z=n) &= \sum_{k=0}^{n} P(X=k,\, Y=n-k) \;=\; \sum_{k=0}^{n} P(X=k)P(Y=n-k) \\
&= \sum_{k=0}^{n} \frac{\lambda_1^k e^{-\lambda_1}}{k!}\,\frac{\lambda_2^{n-k} e^{-\lambda_2}}{(n-k)!} \;=\; \frac{e^{-(\lambda_1+\lambda_2)}}{n!} \sum_{k=0}^{n} \frac{n!}{k!(n-k)!}\,\lambda_1^k \lambda_2^{n-k} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{n!}(\lambda_1+\lambda_2)^n \;=\; \frac{(\lambda_1+\lambda_2)^n e^{-(\lambda_1+\lambda_2)}}{n!}
\end{aligned}$$

따라서 $Z = X + Y \ \sim \mathrm{Poisson}(\lambda_1+\lambda_2)$ 이다. ■

이번에는 X와 Y가 연속형 확률변수일 때 $Z = X + Y$의 확률밀도함수를 구하여보자. X와 Y의 합이 z보다 작은 영역은 $\{-\infty < X < \infty,\, Y < z - X\}$로 나타낼 수 있으므로 Z의 누적분포함수 $F_Z(z)$는 다음과 같다.

$$\begin{aligned}
F_Z(z) &= P(Z \le z) = P(X+Y \le z) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{z-x} f(x,y)\, dy\, dx \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{z} f(x,v-x)\, dv\, dx \quad (\text{안쪽 적분에서 } v = y + x \text{로 변환하였을 때}) \\
&= \int_{-\infty}^{z}\int_{-\infty}^{\infty} f(x,v-x)\, dx\, dv \quad (\text{적분 순서를 바꾸었을 때})
\end{aligned}$$

이를 z로 미분한 Z의 확률밀도함수 $f_Z(z)$는 다음과 같다.

$$f_Z(z) \;=\; \frac{d}{dz}F_Z(z) \;=\; \int_{-\infty}^{\infty} f(x, z-x)\, dx \tag{3.11}$$

만약 X와 Y가 서로 독립이면 Z의 확률밀도함수는 다음과 같이 표현할 수 있다.

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

연속형 확률변수에 대하여 합의 확률밀도함수를 구하는 식 (3.11)과 이산형 확률변수에 대해 합의 확률질량함수를 구하는 식 (3.10)을 비교하면 합 대신에 적분을 사용하였고 확률질량함수 대신에 확률밀도함수를 사용한다는 점을 제외하고는 근본적으로 같다는 것을 알 수 있다.

예 3.11 X와 Y 모두 지수분포를 따르는 확률변수이며 서로 독립일 때 $Z = X + Y$의 확률분포를 구하여라.

[풀이] 합의 확률밀도함수를 구하는 식 (3.11) 에 의하여 $Z = X + Y$ 의 확률밀도함수는 다음과 같이 구할 수 있다.

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

지수분포의 확률밀도함수는 0보다 큰 영역에서만 양의 값을 가지므로 위의 적분에서 함수 $f_X(x)f_Y(z-x)$ 는 $0 < x < z$ 인 경우만 양의 값을 가진다. 따라서 Z 의 확률밀도함수는 다음과 같다.

$$\begin{aligned} f_Z(z) &= \int_0^z \lambda e^{-\lambda x} \lambda e^{-\lambda(z-x)}\, dx \\ &= \int_0^z \lambda^2 e^{-\lambda z}\, dx = \lambda^2 z e^{-\lambda z}, \quad z > 0 \end{aligned}$$

이는 감마분포의 확률밀도함수이므로 $Z \sim \text{Gamma}(2, \lambda)$ 이다. ■

같은 λ 를 갖는 서로 독립인 지수분포의 합이 감마분포가 된다는 사실은 같은 λ 를 갖는 서로 독립인 감마분포의 합이 다시 감마분포가 된다는 사실로 확장될 수 있다.

예 3.12 $X \sim \text{Gamma}(\alpha, \lambda)$, $Y \sim \text{Gamma}(\beta, \lambda)$ 이고 서로 독립일 때 $Z = X + Y$ 의 확률분포를 구하여라.

[풀이] 지수분포의 합과 마찬가지로 감마분포의 확률밀도함수는 0보다 큰 영역에서만 양의 값을 가지므로 $Z = X + Y$ 의 확률밀도함수를 구하는 식 (3.11) 에서 x 에 대한 적분의 범위는 $0 < x < z$ 로 축소된다. 따라서 Z 의 확률밀도함수는 다음과 같다.

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx \\ &= \int_0^z \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \frac{\lambda^\beta}{\Gamma(\beta)} (z-x)^{\beta-1} e^{-\lambda(z-x)}\, dx \\ &= \int_0^z \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (z-x)^{\beta-1} e^{-\lambda z}\, dx \\ &= \int_0^1 \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} (uz)^{\alpha-1} (z-uz)^{\beta-1} e^{-\lambda z} z\, du \quad (\text{단}, u = \tfrac{x}{z}) \\ &= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha+\beta)} z^{\alpha+\beta-1} e^{-\lambda z} \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} u^{\alpha-1} (1-u)^{\beta-1}\, du \\ &= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha+\beta)} z^{\alpha+\beta-1} e^{-\lambda z}, \quad z > 0 \end{aligned}$$

위에서 $\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}u^{\alpha-1}(1-u)^{\beta-1}$은 $\mathrm{Beta}(\alpha,\beta)$의 확률밀도함수이미로 이 함수를 적분한 값은 1이 되어 마지막 등식을 얻을 수 있다. 따라서 $Z \sim \mathrm{Gamma}(\alpha+\beta,\lambda)$이다. ■

3.6.2 곱의 분포

확률변수 X와 Y의 합의 확률분포를 구하는 유사한 방법으로 X와 Y의 곱의 확률분포도 구할 수 있다. $Z = XY$의 누적분포함수를 구하기 위해 X의 값이 양수인 부분과 음수인 부분으로 나누어 생각하여 보자. 양수인 부분에서는 XY가 z보다 작은 영역은 $\{X > 0, Y < \frac{z}{X}\}$로 나타낼 수 있으며 음수인 부분에서는 $\{X < 0, Y > \frac{z}{X}\}$로 나타낼 수 있다. 따라서 Z의 확률밀도함수 $F_Z(z)$는 다음과 같다.

$$\begin{aligned}
F_Z(z) &= P(XY \le z) = P(Y \le z/X, X>0) + P(Y \ge z/X, X<0) \\
&= \int_0^\infty \int_{-\infty}^{\frac{z}{x}} f(x,y)\,dy\,dx + \int_{-\infty}^0 \int_{\frac{z}{x}}^\infty f(x,y)\,dy\,dx \\
&= \int_0^\infty \int_{-\infty}^{z} f\left(x,\frac{v}{x}\right)\frac{1}{x}\,dv\,dx + \int_{-\infty}^0 \int_{-\infty}^{z} f\left(x,\frac{v}{x}\right)\left(-\frac{1}{x}\right)\,dy\,dx \\
&\qquad\qquad\qquad\qquad\qquad\qquad\text{(안쪽 적분에서 } v = xy\text{라 할 때)} \\
&= \int_{-\infty}^\infty \int_{-\infty}^{z} \left|\frac{1}{x}\right| f\left(x,\frac{v}{x}\right)\,dv\,dx \\
&= \int_{-\infty}^z \int_{-\infty}^{\infty} \left|\frac{1}{x}\right| f\left(x,\frac{v}{x}\right)\,dx\,dv
\end{aligned}$$

이를 z로 미분한 Z의 확률밀도함수 $f_Z(z)$는 다음과 같다.

$$f_Z(z) = \int_{-\infty}^\infty \left|\frac{1}{x}\right| f\left(x,\frac{z}{x}\right)\,dx$$

같은 방법으로 $Z = \frac{Y}{X}$의 확률밀도함수를 구하면 다음과 같음을 확인 할 수 있다.

$$f_Z(z) = \int_{-\infty}^\infty |x| f(x,xz)\,dx \tag{3.12}$$

예 3.13 $X \sim \mathrm{N}(0,1)$, $Y \sim \mathrm{N}(0,1)$이며 서로 독립일 때 $Z = \frac{Y}{X}$의 확률밀도함수를 구하여라.

[풀이] 위의 주어진 식 (3.12)를 이용하면 $Z = \dfrac{Y}{X}$의 확률밀도함수는 다음과 같다.

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} |x| f(x, xz)\, dx \\ &= \int_{-\infty}^{\infty} |x| \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(xz)^2}{2}}\, dx \\ &= \int_{-\infty}^{\infty} |x| \frac{1}{2\pi} e^{-\frac{x^2(1+z^2)}{2}}\, dx \end{aligned}$$

적분 안의 함수는 y축에 대칭인 함수이므로 Z의 확률밀도함수는 다음과 같다.

$$f_Z(z) = \int_0^{\infty} \frac{1}{\pi} x e^{-\frac{x^2(1+z^2)}{2}}\, dx = \frac{1}{\pi(1+z^2)}, \quad -\infty < z < \infty$$

위와 같은 확률밀도함수를 가지는 확률분포를 코시분포라 한다. ■

3.6.3 이변량 확률변수의 변환

지금까지는 두 확률변수의 합과 곱의 확률분포를 구하는 방법을 다루었다. 좀더 일반적인 변수변환에 의하여 만들어진 새로운 확률변수의 확률분포를 구하는 방법을 알아보기로 하자. 함수 $g(\cdot)$의 역함수가 존재하고 미분가능할 때 X의 확률밀도함수로부터 식 (2.15)에 의하여 $Y = g(X)$의 확률밀도함수를 구할 수 있는 것과 같이 이변량 확률변수에 대하여도 변환하는 함수의 역함수가 존재하고 미분가능하면 (X, Y)의 결합확률밀도함수로부터 변환한 확률변수의 결합확률밀도함수를 구할 수 있다.

X와 Y는 결합확률밀도함수 $f_{X,Y}(x, y)$를 갖는 연속형 확률변수이고 새로운 확률변수 U, V는 다음과 같다고 하자.

$$U = g_1(X, Y), \quad V = g_2(X, Y)$$

이때 다음이 성립하는 함수 h_1과 h_2가 존재하며

$$X = h_1(U, V), \quad Y = h_2(U, V)$$

이 함수가 미분가능하면 다음 정리에 의하여 (U, V)의 결합확률밀도함수를 구할 수 있다.

정리 3.1 $u = g_1(x, y)$, $v = g_2(x, y)$에 대하여 $x = h_1(u, v)$, $y = h_2(u, v)$인 함수 $h_1(\cdot, \cdot)$와 $h_2(\cdot, \cdot)$가 존재하고 이 함수가 미분가능할 때, $U = g_1(X, Y), V = g_2(X, Y)$의 결합확률밀도함수 $f_{U,V}(u, v)$는 다음과 같다.

$$f_{U,V}(u, v) = f_{X,Y}(h_1(u, v), h_2(u, v))\, |J(u, v)|$$

단, $J(u,v) = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} \frac{\partial}{\partial u}h_1(u,v) & \frac{\partial}{\partial v}h_1(u,v) \\ \frac{\partial}{\partial u}h_2(u,v) & \frac{\partial}{\partial v}h_2(u,v) \end{vmatrix}$ 이다.

이 정리는 2장에서 변수변환에 관한 정리를 이변량 확률분포로 확장한 것으로 증명은 생략하기로 하겠다. 위 정리를 이용하여 $X \sim \mathrm{N}(0,1)$, $Y \sim \mathrm{N}(0,1)$ 이며 서로 독립일 때 X 와 $X+Y$ 의 결합확률밀도함수를 구하여 보자. X 와 Y 는 독립이므로 결합확률밀도함수는 다음과 같다.

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) = \frac{1}{2\pi}\exp\left(-\frac{x^2}{2} - \frac{y^2}{2}\right)$$

$U = X$, $V = X+Y$ 라 할 때 g_1, g_2 는 다음과 같다.

$$g_1(x,y) = x, \quad g_2(x,y) = x+y$$

이때 $X = U,\ Y = V - U$ 이므로 h_1 과 h_2 는 다음과 같다.

$$h_1(u,v) = u, \quad h_2(u,v) = v-u$$

따라서 J 는 다음과 같고

$$J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix} = 1$$

U, V 의 결합확률밀도함수는 다음과 같다.

$$\begin{aligned} f_{U,V}(u,v) &= f_{X,Y}(u, v-u) \\ &= \frac{1}{2\pi}\exp\left(-\frac{u^2}{2} - \frac{(v-u)^2}{2}\right) \\ &= \frac{1}{2\pi}\exp\left(-u^2 + uv - \frac{v^2}{2}\right) \end{aligned}$$

즉 U 와 V 의 결합분포는 다음과 같다.

$$\begin{pmatrix} U \\ V \end{pmatrix} \sim \mathrm{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}\right)$$

연습문제

1. 세 개의 상자에는 각각 1, 2, 3, 4가 적혀있는 네 장의 카드가 들어있다. 각 상자에서 임의로 카드를 한 장씩 선택하여 X와 Y를 다음과 같이 정의할 때, X와 Y의 결합확률분포를 구하여라.

$$X = \text{세 장의 카드 중 가장 작은 값}, \quad Y = \text{세 장의 카드 중 가장 큰 값}$$

2. X와 Y는 다음과 같은 결합분포를 가지고 있다.

		y		
		10	20	30
	0	$\frac{1}{9}$	$\frac{1}{9}$	0
x	1	$\frac{1}{6}$	0	$\frac{1}{6}$
	2	0	$\frac{2}{9}$	$\frac{2}{9}$

a) X의 주변분포를 구하여라.

b) Y의 주변분포를 구하여라.

c) $Y = 10$일 때 X의 조건부분포를 구하여라.

3. 두 연속형 확률변수 X와 Y의 결합확률밀도함수가 다음과 같다.

$$f(x, y) \ = \ 2, \quad 0 < x < y < 1$$

a) X의 주변확률밀도함수를 구하여라.

b) Y의 주변확률밀도함수를 구하여라.

c) X와 Y는 서로 독립인지 확인하여라.

4. 두 연속형 확률변수 X와 Y의 결합확률밀도함수가 다음과 같다.

$$f(x, y) \ = \ 2xe^{-(x+2xy)} \quad 0 < x < \infty, \ 0 < y < \infty$$

a) X의 주변확률밀도함수를 구하여라.

b) Y의 주변확률밀도함수를 구하여라.

c) X와 Y는 서로 독립인지 확인하여라.

5. $X \sim \text{Exp}(\lambda_1), \ Y \sim \text{Exp}(\lambda_2)$이며 서로 독립일 때 $X + Y$의 확률밀도함수를 구하여라.

6. $X \sim \text{Poisson}(\lambda_1)$, $Y \sim \text{Poisson}(\lambda_2)$ 이며 서로 독립이다. $X + Y = n$ 일 때 X 의 조건부분포를 구하여라.

7. $N \sim \text{Poisson}(\lambda)$ 이고, $N = n$ 일 때 X 의 조건부분포는 $\text{B}(n, p)$ 이다.

   a) $P(X = 0)$ 을 구하여라.

   b) X 의 확률분포를 구하여라.

8. 두 연속형 확률변수 X 와 Y 의 결합확률밀도함수가 다음과 같다.

$$f(x, y) \;=\; \frac{\lambda^2}{2} e^{-\frac{\lambda}{2}(x+y)}, \quad 0 < x < y < \infty$$

   a) X 의 주변확률밀도함수를 구하여라.

   b) $X = x$ 일 때 Y 의 조건부 확률밀도함수를 구하여라.

9. 두 연속형 확률변수 X 와 Y 에 대하여 $Z = X - Y$ 의 확률밀도함수를 구하여라.

10. 두 연속형 확률변수 X 와 Y 에 대하여 $Z = \dfrac{Y}{X}$ 의 확률밀도함수는 식 (3.12) 로 주어짐을 보여라.

11. $N \sim \text{Poisson}(\lambda)$ 이고, $N = n$ 일 때 X 의 조건부분포는 $\text{N}(n, 1)$ 이라고 한다. $P(X < 0)$ 을 표준정규분포의 누적확률 $\Phi(z) = P(\text{N}(0, 1) < z)$ 를 사용하여 나타내어라.

12. X 와 Y 는 지수분포를 따르는 확률변수이고 서로 독립이다. S 와 T 를 아래와 같이 정의할 때

$$S = X + Y, \quad T = \frac{X}{X + Y}$$

    다음 물음에 답하여라.

    a) S 와 T 의 결합확률밀도함수를 구하여라.

    b) T 의 확률분포를 구하여라.

13. $X \sim \text{Gamma}(\alpha, \lambda)$, $Y \sim \text{Gamma}(\beta, \lambda)$ 이고 서로 독립이다. S 와 T 를 아래와 같이 정의할 때

$$S = X + Y, \quad T = \frac{X}{X + Y}$$

    다음 물음에 답하여라.

    a) S 와 T 의 결합확률밀도함수를 구하여라.

    b) T 의 확률분포를 구하여라.

14. X와 Y의 결합확률밀도함수 $f(x, y)$를 다음과 같이 나타낼 수 있을 때 X와 Y는 독립임을 보여라.

$$f(x, y) = g(x)h(y), \quad a \leq x \leq b,\ c \leq y \leq d$$

제 4 장

기댓값

4.1 확률변수의 기댓값

자료 $x_1, x_2, \ldots, x_m$ 를 대표하는 값으로 자료의 평균은 다음과 같다.

$$\overline{x} \;=\; \frac{1}{m}\sum_{i=1}^{m} x_i$$

만약 각 자료 값에서 도수(frequency)가 n_i 라면 자료의 평균은 다음과 같이 나타낼 수 있다.

$$\overline{x} \;=\; \sum_{i=1}^{m}\left(\frac{n_i}{N}\right)x_i, \quad \text{단, } N = \textstyle\sum_{i=1}^{m} n_i$$

이는 각 자료 값에서 도수의 비율을 가중값으로 사용한 가중평균이다. 위의 평균을 구하는 식과 다른 점은 가중값으로 $\frac{1}{m}$ 대신에 $\dfrac{n_i}{N}$ 를 사용한 것이다.

대푯값으로 가중평균이 적절한 한 가지 예를 들어 살펴보자. N장의 카드에서 임의로 카드를 한 장 선택하여 그 카드에 쓰여 있는 금액만큼 상금을 받는다고 할 때 평균 상금을 구해보도록 하자. N장의 카드 중 n_1장에는 상금이 x_1, n_2장에는 상금이 x_2, ..., n_m장에는 상금이 x_m 이라고 쓰여 있다고 한다. 이때 상금의 종류별 단순한 평균인 $\overline{x} \;=\; \frac{1}{m}\sum_{i=1}^{m} x_i$ 를 평균 상금이라고 말할 수는 없다. 이 평균은 상금의 종류 $x_1, x_2, \ldots, x_m$ 에 같은 가중값을 주고 구한 값이기 때문이다. 만약 상금 x_1 이 쓰여 있는 카드가 다른 카드에 비해 많다면 x_1 이 쓰여 있는 카드를 선택할 확률이 다른 상금이 쓰여 있는 카드를 선택할 확률에 비해 클 것이다. 예를 들어 x_1 이 쓰여 있는 카드가 x_2 가 쓰여 있는 카드의 2배라면 상금이 x_1 이 될 확률은 x_2가 될 확률의 2배라고 할 수 있다. 따라서 평균 상금을 구하기 위해서는 x_1 의 가중값은 x_2 의 가중값의 2배일 것이다. 이처럼 각 값을 가질 확률이 다르면 이러한 값의 평균을 구할 때 확률을 가중값으로 사용한 가중평균을 사용하여야 할 것이다.

여러 가지 값의 중심 위치를 나타내는 대푯값으로 가중평균을 이용하는 개념은 확률변수에 대해서도 적용될 수 있다. 확률변수는 여러 가지 값 중에서 하나의 값을 가지며 각 값을 가질

확률은 서로 다를 수 있다. 따라서 확률변수가 가지는 평균적인 값[1]은 각 값의 확률을 가중값으로 한 가중평균이 될 것이다. 이산형 확률변수 X가 가질 수 있는 값이 $x_1, x_2, \ldots$이고 각 값의 확률이 $p(x_i)$일 때 X의 평균적인 값은 다음과 같다.

$$\sum_i x_i\, p(x_i) \tag{4.1}$$

이번에는 연속형 확률변수의 평균적인 값을 구하는 방법을 생각하여 보자. 연속형 확률변수에 대하여는 각 값의 확률이 주어지지 않으므로 이산형 확률변수처럼 식 (4.1)에 주어진 방법으로 평균적인 값을 구할 수는 없다. 구간 $[a,\, b)$에서 값을 가지는 연속형 확률변수 X의 평균적인 값을 근사적으로 구하도록 하자. 구간 $[a,\, b)$를 다음과 같은 n개의 소 구간으로 나누고

$$[x_1, x_2),\ [x_2, x_3), \ldots, [x_n, x_{n+1}), \quad \text{단, } a = x_1 < x_2 < x_3 < \cdots < x_{n+1} = b$$

확률변수 Y를 다음과 같이 정의하자.

$$X \in [x_i,\, x_{i+1})\text{ 일 때} \quad \Longleftrightarrow \quad Y = x_i$$

확률변수 Y는 X의 값이 소 구간 $[x_i,\, x_{i+1})$에 속할 때 이 구간의 좌측 끝 값인 x_i로 정의했으므로 X와 Y의 값의 차이는 소 구간의 폭보다 작게 된다. 따라서 소 구간이 폭이 작을수록 X와 Y는 가까워진다고 할 수 있다.

한편 Y가 가질 수 있는 값은 $x_1, x_2, \ldots, x_n$ 중의 하나이므로 이산형 확률변수이고 각 값에서의 확률은 다음과 같다.

$$P(Y = x_i) \;=\; P(x_i \le X < x_{i+1})$$

따라서 Y의 평균적인 값은 식 (4.1)에 의하여 다음과 같이 구할 수 있다.

$$\sum_{i=1}^{n} x_i P(x_i \le X < x_{i+1})$$

소 구간의 길이가 짧을 때 $P(x_i \le X < x_{i+1}) \approx f(x_i)(x_{i+1} - x_i)$이므로 위의 평균은 근사적으로 다음과 같이 나타낼 수 있다.

$$\sum_{i=1}^{n} x_i f(x_i)(x_{i+1} - x_i)$$

[1]보통 평균이라는 용어는 여러 가지 값이 있을 때 이 값의 중심을 나타내는 대푯값을 나타낼 때 사용한다. 확률변수는 여러 가지 가능한 값 중 하나의 값만 가지므로 여기서는 평균적인 값이라는 모호한 표현을 사용하였다. 한편 확률변수의 평균을 같은 확률분포를 가지는 확률변수의 값을 무한히 반복하여 얻은 값의 평균으로 해석할 수도 있을 것이다.

각 소 구간의 길이가 0에 가까이 갈 때 이 값은 다음 값으로 수렴한다.

$$\int_a^b xf(x)\,dx \tag{4.2}$$

즉 소 구간의 폭이 작아질 때 Y는 X에 근접하게 되고 Y의 평균은 식 (4.2)에 수렴하므로 연속형 확률변수 X의 평균적인 값은 (4.2)라 할 수 있다. 마찬가지로 확률변수 X의 값의 범위가 $(-\infty, \infty)$일 때 X의 평균적인 값은 다음과 같다고 할 수 있다.

$$\int_{-\infty}^{\infty} xf(x)\,dx \tag{4.3}$$

이산형 확률변수의 평균적인 값 (4.1)과 비교하면 확률질량함수와 합 대신에 확률밀도함수와 적분을 이용한 점 외에는 근본적으로 같은 식이다.

식 (4.1) 또는 (4.3)의 값이 존재할 때, 이 값을 확률변수의 **기댓값**(expected value)이라 하고 $E(X)$ 또는 μ로 나타낸다.

정의 4.1 이산형 확률변수 X의 기댓값은 다음과 같고

$$E(X) = \sum_i x_i p(x_i)$$

연속형 확률변수 X의 기댓값은 다음과 같다.

$$E(X) = \int_{-\infty}^{\infty} xf(x)\,dx$$

확률변수의 기댓값은 확률변수를 하나의 실수 값으로 표현한 대푯값이다. 기댓값은 확률을 가중값으로 사용한 확률분포의 중심위치를 나타낸 값이라 할 수 있다. 이런 관점에서 확률변수의 기댓값을 모집단(확률분포)의 중심위치를 나타낸 모평균(population mean)이라 한다.

예 4.1 이산형 확률변수 X의 확률질량함수가 다음과 같을 때 X의 기댓값을 구하여라.

x	-1	0	1	2
$p_X(x)$	0.1	0.2	0.3	0.4

[풀이] 이산형 확률변수 X의 기댓값은 다음과 같다.

$$E(X) = (-1)\times 0.1 + 0\times 0.2 + 1\times 0.3 + 2\times 0.4 = 1.0$$

■

2장에서 다룬 여러 확률분포의 기댓값을 구하여 보자. 기댓값을 구하는데 주로 이용되는 사실은 전체 영역에서 확률질량함수의 합과 확률밀도함수의 적분은 1이 된다는 사실이다. 여기에서 여러 확률분포의 기댓값을 구하는 방법을 자세히 설명하는 이유는 다른 계산에서도 활용될 수 있기 때문이다.

먼저 이항분포의 기댓값을 구하여 보자. X가 시행횟수는 n이고 성공의 확률이 p인 이항분포를 따르는 확률변수일 때 X의 기댓값은 다음과 같다.

$$\begin{aligned}
E(X) &= \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!}\, p^k (1-p)^{n-k} \\
&= \sum_{l=0}^{n-1} \frac{n!}{l!(n-1-l)!}\, p^{l+1} (1-p)^{n-1-l} \quad (\text{단, } l = k-1) \\
&= np \sum_{l=0}^{n-1} \frac{(n-1)!}{l!(n-1-l)!}\, p^l (1-p)^{n-1-l}
\end{aligned}$$

위 식에서 $\sum_{l=0}^{n-1} \dfrac{(n-1)!}{l!(n-1-l)!}\, p^l (1-p)^{n-1-l}$ 은 시행횟수 $n-1$, 성공의 확률 p인 이항분포의 모든 확률의 합이므로 1이 된다. 따라서 이항분포의 기댓값은 다음과 같다.

$$E(X) \;=\; np$$

이항분포에서 $n = 1$인 경우인 베르누이분포의 기댓값은 성공의 확률 p가 된다.

유사한 방법으로 포아송분포의 기댓값을 구할 수 있다. 포아송분포를 따르는 확률변수 X의 기댓값은 다음과 같이 나타낼 수 있다.

$$\begin{aligned}
E(X) \;&=\; \sum_{k=0}^{\infty} k\, \frac{\lambda^k}{k!}\, e^{-\lambda} \;=\; \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!}\, e^{-\lambda} \\
&=\; \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}\, e^{-\lambda} \;=\; \lambda \sum_{l=0}^{\infty} \frac{\lambda^l}{l!}\, e^{-\lambda}
\end{aligned}$$

위 식에서 합 $\sum_{l=0}^{\infty} \frac{\lambda^l}{l!}\, e^{-\lambda}$은 포아송분포의 모든 확률의 합이므로 1이 된다. 따라서 포아송분포의 기댓값은 λ이다.

기하분포의 기댓값을 구하는 방법은 두 가지가 있다. 첫 번째는 급수의 합을 이용하는 방법이다. 기하분포를 따르는 확률변수 X의 기댓값은 다음과 같이 나타낼 수 있다.

$$\begin{aligned}
E(X) \;&= \sum_{k=1}^{\infty} k\,(1-p)^{k-1} p \;=\; p \sum_{k=1}^{\infty} k (1-p)^{k-1} \\
&=\; p\Big(1 + 2(1-p) + 3(1-p)^2 + \cdots\Big) \qquad (4.4)
\end{aligned}$$

양변에 $(1-p)$를 곱하면 위의 식은 다음과 같이 나타낼 수 있다.

$$(1-p)E(X) \;=\; p\Big((1-p)+2(1-p)^2+3(1-p)^3+\cdots\Big) \qquad (4.5)$$

식 (4.4)와 식 (4.5)의 차를 구하면 다음과 같으므로

$$\begin{aligned} pE(X) \;&=\; p\Big(1+(1-p)+(1-p)^2+\cdots\Big) \\ &=\; p\left(\frac{1}{1-(1-p)}\right) \;=\; 1 \end{aligned}$$

기하분포의 기댓값은 $E(X)=\frac{1}{p}$이다.

기하분포의 기댓값을 구하는 두 번째 방법은 미분을 이용하는 것이다. 기하분포의 기댓값은 다음과 같이 나타낼 수 있다.

$$E(X) \;=\; \sum_{k=1}^{\infty} k(1-p)^{k-1}p \;=\; p\sum_{k=1}^{\infty} k(1-p)^{k-1}$$

위 식에서 $k(1-p)^{k-1}$는 $-(1-p)^k$를 p에 대하여 미분한 것으로 볼 수 있다. 합과 미분의 순서를 바꿀 수 있다면 기댓값은 다음과 같이 구할 수 있다.

$$\begin{aligned} E(X) \;&= p\sum_{k=1}^{\infty}\frac{d}{dp}\left(-(1-p)^k\right) \;=\; p\,\frac{d}{dp}\left(-\sum_{k=1}^{\infty}(1-p)^k\right) \\ &= p\,\frac{d}{dp}\left(-\frac{1-p}{p}\right) \;=\; \frac{1}{p} \end{aligned}$$

합과 미분의 순서를 항상 바꿀 수 있는 것은 아니지만 위 식에서는 합과 미분의 순서를 바꾸어도 상관이 없다.

이번에는 연속형 확률분포의 기댓값을 구하여 보자. 확률밀도함수가 (2.8)로 주어지는 감마분포를 따르는 확률변수 X의 기댓값은 다음과 같이 나타낼 수 있다.

$$\begin{aligned} E(X) \;&= \int_0^{\infty} x\,\frac{\lambda^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}\,dx \;=\; \int_0^{\infty}\frac{\lambda^{\alpha}}{\Gamma(\alpha)}x^{\alpha}e^{-\lambda x}\,dx \\ &= \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)\lambda}\int_0^{\infty}\frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)}x^{(\alpha+1)-1}e^{-\lambda x}\,dx \end{aligned}$$

이항분포나 포아송분포의 기댓값을 구할 때 확률질량함수의 합은 1이라는 사실을 이용한 것처럼 감마분포의 기댓값을 구할 때에도 확률밀도함수의 적분은 1이라는 사실을 이용할 수 있다. 위 식에서 적분 안의 함수는 $\mathrm{Gamma}(\alpha+1,\lambda)$의 확률밀도함수이므로 적분은 1이 된다. 따라서 기댓값은 다음과 같다.

$$E(X) \;=\; \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)\lambda} \;=\; \frac{\alpha\Gamma(\alpha)}{\Gamma(\alpha)\lambda} \;=\; \frac{\alpha}{\lambda}$$

지수분포나 카이제곱분포는 감마분포의 특별한 경우이므로 감마분포의 기댓값으로부터 지수분포의 기댓값을 구할 수 있다. 지수분포는 감마분포에서 $\alpha = 1$인 경우이므로 $X \sim \text{Exp}(\lambda)$일 때 $E(X) = \frac{1}{\lambda}$이다. 또한 자유도가 n인 카이제곱분포는 $\alpha = \frac{n}{2}$이고 $\lambda = \frac{1}{2}$인 감마분포이므로 $X \sim \chi^2_n$일 때 $E(X) = n$이다

정규분포의 기댓값은 적분을 통해 직접 구하는 것은 쉽지 않다. 그러나 정규분포의 확률밀도함수가 대칭이라는 사실을 이용하여 기댓값을 유추할 수 있다. $\text{N}(\mu, \sigma^2)$를 따르는 확률변수 X의 기댓값은 다음과 같이 나타낼 수 있다.

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx \\ &= \mu + \int_{-\infty}^{\infty} (x-\mu) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx \\ &= \mu + \int_{-\infty}^{\infty} t \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{t^2}{2\sigma^2}}\, dt \quad (\text{단}, t = x - \mu) \end{aligned}$$

위 식에서 적분 안의 함수는 원점에 대칭인 함수이다. 따라서 적분 값이 존재한다면 그 값은 0이 될 수밖에 없다. 여기서 설명하지는 않겠지만 이 적분 값이 존재한다는 사실을 확인할 수 있다. 따라서 정규분포 $\text{N}(\mu, \sigma^2)$의 기댓값은 μ이다.

참고로 모든 확률분포의 기댓값이 항상 존재하는 것은 아니다. 예를 들어 코시분포의 기댓값은 존재하지 않는다. 코시분포의 확률밀도함수는 다음과 같다.

$$f(x) = \frac{1}{\pi(1+x^2)}, \; -\infty < x < \infty$$

따라서 기댓값은 다음과 같다.

$$\int_{-\infty}^{\infty} \frac{x}{\pi(1+x^2)}\, dx$$

적분 안의 함수는 원점에 대칭인 함수이므로 적분 값이 존재한다면 0이 될 것이다. 문제는 위의 적분 값이 존재하지 않는다는 사실이다. 이는 적분의 범위를 0보다 큰 범위에서 적분 값은 다음과 같이 무한대이다.

$$\int_0^{\infty} \frac{x}{(1+x^2)}\, dx \geq \int_1^{\infty} \frac{x}{(1+x^2)}\, dx \int_1^{\infty} \frac{1}{2x}\, dx \; = \infty$$

마찬가지로 0보다 작은 범위에서도 적분 값은 음의 무한대이다. 따라서 전체 범위에서 적분 값은 $\infty - \infty$ 형태가 되어 적분 값이 존재하지 않는다. 즉 코시분포는 확률밀도함수가 $x = 0$에 대칭이므로 중심은 0이지만 기댓값은 존재하지 않는다.

4.2 확률변수의 함수의 기댓값

앞 절에서는 확률변수의 기댓값을 구하는 방법을 다루었다. 이 절에서는 확률변수 X의 함수인 $Y = g(X)$의 기댓값을 구하는 방법을 살펴보도록 하자. X의 확률분포로부터 X의 기댓값을 구할 수 있는 것처럼 $Y = g(X)$의 기댓값을 구하기 위해서는 Y의 확률분포를 알아야 한다. 따라서 Y의 기댓값을 구하기 위해서는 먼저 X의 확률분포(확률질량함수 또는 확률밀도함수)로부터 $Y = g(X)$의 확률분포(확률질량함수 또는 확률밀도함수)를 구해야 한다.

Y의 확률질량함수 또는 확률밀도함수를 $p_Y(y)$ 또는 $f_Y(y)$라 할 때 $Y = g(X)$의 기댓값은 다음과 같다.

$$E[g(X)] \;=\; E(Y) \;=\; \begin{cases} \sum_i y_i p_Y(y_i) & (\text{이산형 확률변수의 경우}) \\ \int_{-\infty}^{\infty} y f_Y(y)\,dy & (\text{연속형 확률변수의 경우}) \end{cases}$$

예를 들어 X가 지수분포를 따르는 확률변수일 때 $Y = 2X$의 기댓값을 구하여 보자. X의 확률밀도함수는 $f_X(x) = \lambda e^{-\lambda x}$, $x \geq 0$이므로 정리 2.1에 의하여 $Y = 2X$의 확률밀도함수는 다음과 같다.

$$\begin{aligned} f_Y(y) &= \frac{1}{2} f_X\left(\frac{y}{2}\right) \\ &= \frac{\lambda}{2} e^{-\frac{\lambda}{2}y}, \quad y \geq 0 \end{aligned}$$

따라서 Y의 기댓값은 다음과 같다.

$$\begin{aligned} E(Y) &= \int_0^{\infty} y f_Y(y)\,dy \\ &= \int_0^{\infty} y\,\frac{\lambda}{2} e^{-\frac{\lambda}{2}y} \;=\; \frac{2}{\lambda} \end{aligned}$$

위 방법을 이용하여 $Y = g(X)$의 기댓값을 구하기 위해서는 먼저 Y의 확률분포를 구해야 하는 번거로움이 있다. 그러나 다음 정리에 의하면 Y의 확률분포를 구하지 않고 X의 확률분포를 이용하여 Y의 기댓값을 바로 구할 수 있음을 알 수 있다.

정리 4.1 이산형 확률변수 X의 확률질량함수를 $p_X(x)$라 할 때 $g(X)$의 기댓값은 다음과 같다.

$$E[g(X)] \;=\; \sum_i g(x_i) p_X(x_i)$$

또 연속형 확률변수 X의 확률밀도함수를 $f_X(x)$라 할 때 $g(X)$의 기댓값은 다음과 같다.

$$E[g(X)] \;=\; \int_{-\infty}^{\infty} g(x) f_X(x)\,dx \tag{4.6}$$

[증명] 이 정리의 증명은 X가 연속형 확률변수이고 $g(\cdot)$는 역함수가 존재하고 미분가능한 경우만 다루기로 하겠다. 다른 경우도 근본적으로 증명방법은 같다고 할 수 있다. $g(\cdot)$의 역함수가 존재하고 미분가능할 때 정리 2.1에 의하여 $Y = g(X)$의 확률밀도함수 $f_Y(y)$는 다음과 같다.

$$f_Y(y) \ = \ f_X(g^{-1}(y)) \cdot \left|\frac{d}{dy}g^{-1}(y)\right|$$

따라서 Y의 기댓값은 다음과 같이 나타낼 수 있다.

$$\begin{aligned} E(Y) &= \int_{-\infty}^{\infty} y f_Y(y)\, dy \\ &= \int_{-\infty}^{\infty} y f_X(g^{-1}(y)) \cdot \left|\frac{d}{dy}g^{-1}(y)\right|\, dy \end{aligned}$$

$g^{-1}(y) = t$라 할 때 위의 적분 값은 다음과 같이 표현할 수 있다.

$$\begin{aligned} &\int_{-\infty}^{\infty} y f_X(g^{-1}(y)) \cdot \left|\frac{d}{dy}g^{-1}(y)\right|\, dy \\ &\quad = \int_{-\infty}^{\infty} g(t) f_X(t) \cdot \left|\frac{d}{dy}g^{-1}(y)\right| \cdot \left|\frac{dy}{dt}\right|\, dt \end{aligned}$$

한편 $t = g^{-1}(y)$일 때 다음을 얻을 수 있으므로

$$1 \ = \ \frac{dy}{dy} \ = \ \frac{dy}{dt} \cdot \frac{dt}{dy} \ = \ \frac{dy}{dt} \cdot \frac{dg^{-1}(y)}{dy}$$

Y의 기댓값은 다음과 같이 나타낼 수 있다.

$$\int_{-\infty}^{\infty} g(t) f_X(t)\, dt$$

■

예 4.2 이산형 확률변수 X의 확률질량함수가 다음과 같다.

x	-1	0	1	2
$p_X(x)$	0.1	0.2	0.3	0.4

(a) X^2의 확률분포를 구하여라.

(b) X^2의 확률분포를 이용하여 X^2의 기댓값을 구하여라.

(c) 정리 4.1을 이용하여 X^2의 기댓값을 구하여라.

[풀이]

(a) X의 가능한 값 $-1, 0, 1, 2$를 제곱하면 $1, 0, 1, 4$가 되므로 $Y = X^2$의 확률질량함수를 구하면 다음과 같다.

y	0	1	4
$p_Y(y)$	0.2	0.4	0.4

(b) (a)에서 구한 확률질량함수를 이용하여 Y의 기댓값을 구하면 다음과 같다.

$$E(Y) \;=\; \sum_i y_i\, p_Y(y_i) \;=\; 0 \times 0.2 + 1 \times 0.4 + 4 \times 0.4 \;=\; 2$$

(c) 정리 4.1을 이용하여 X^2의 기댓값을 구하면 다음과 같다.

$$E(X^2) \;=\; \sum_i x_i^2\, p_X(x_i) \;=\; (-1)^2 \times 0.1 + 0^2 \times 0.2 + 1^2 \times 0.3 + 2^2 \times 0.4 \;=\; 2$$

이 값은 (b)에서 구한 값과 같음을 확인할 수 있다.

■

$E[g(X)]$의 기댓값을 구할 때 한 가지 주의할 사항은 $E[g(X)] \neq g[E(X)]$라는 사실이다. 예를 들어 $E\left(\frac{1}{X}\right)$의 값을 구할 때 $E(X)$의 값을 구한 후 역수를 취한 $\frac{1}{E(X)}$로 계산하는 것은 올바른 방법이 아니다. 예를 통해 이를 확인하여 보자. X의 확률분포가 다음과 같을 때

$$X = \begin{cases} 1, & 0.5\text{의 확률로} \\ 2, & 0.5\text{의 확률로} \end{cases}$$

$\frac{1}{X}$의 확률분포는 다음과 같다.

$$\frac{1}{X} = \begin{cases} 1, & 0.5\text{의 확률로} \\ \frac{1}{2}, & 0.5\text{의 확률로} \end{cases}$$

이때 $E(X) = 1.5$이고 $E\left(\frac{1}{X}\right) = 0.75$이다. 따라서 $\frac{1}{E(X)} \neq E\left(\frac{1}{X}\right)$임을 알 수 있다.

X의 확률분포를 이용하여 $g(X)$의 기댓값을 구하는 정리 4.1은 결합확률변수로도 확장될 수 있다.

정리 4.2 이산형 확률변수 $(X_1, X_2, \dots, X_n)$의 결합확률질량함수를 $p(x_1, x_2, \dots, x_n)$이라 할 때 $g(X_1, X_2, \dots, X_n)$의 기댓값은 다음과 같다.

$$E[g(X_1, X_2, \dots, X_n)] = \sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} g(x_1, x_2, \dots, x_n) p(x_1, x_2, \dots, x_n)$$

또 연속형 확률변수 $(X_1, X_2, \dots, X_n)$의 결합확률밀도함수를 $f(x_1, x_2, \dots, x_n)$이라 할 때 $g(X_1, X_2, \dots, X_n)$의 기댓값은 다음과 같다.

$$\begin{aligned} &E[g(X_1, X_2, \dots, X_n)] \\ &\quad = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, x_2, \dots, x_n) f(x_1, x_2, \dots, x_n)\ dx_1\, dx_2 \cdots dx_n \end{aligned}$$

위의 정리를 이용하면 다음과 같은 결과를 얻을 수 있다.

정리 4.3 X와 Y가 서로 독립일 때 다음이 성립한다.

$$E[g(X)h(Y)] = E[g(X)]\, E[h(Y)]$$

[증명] 증명은 X와 Y 모두 연속형 확률변수인 경우만 다루기로 하겠다. X와 Y가 독립이므로 결합확률밀도함수는 $f(x, y) = f_X(x) f_Y(y)$이 된다. 따라서 다음 결과를 얻을 수 있다.

$$\begin{aligned} E[g(X)h(Y)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x)h(y)\ f(x, y)\ dx\, dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x)h(y)\ f_X(x) f_Y(y)\ dx\, dy \\ &= \left(\int_{-\infty}^{\infty} g(x) f_X(x)\ dx \right) \left(\int_{-\infty}^{\infty} h(y) f_Y(y)\ dy \right) \\ &= E[g(X)]E[h(Y)] \end{aligned}$$

■

위 정리에서 $g(x) = x$ 이고 $h(y) = y$ 이면 서로 독립인 X 와 Y 에 대하여 다음이 성립한다.

$$E(XY) = E(X)\,E(Y)$$

즉 서로 독립인 확률변수에 대하여 곱의 기댓값은 각각 기댓값의 곱이 된다. 이러한 관계는 서로 독립인 확률변수에 대해서만 성립하며 일반적인 확률변수에 대해서는 성립하지 않는다.

예 4.3 $X \sim \mathrm{Exp}(\lambda_1),\ Y \sim \mathrm{Exp}(\lambda_2)$ 이고 서로 독립일 때 XY 의 기댓값을 구하여라.

[풀이] X 와 Y 가 서로 독립이므로 정리 4.3에 의하여 $E(XY) = E(X)E(Y)$ 이다. X 와 Y 의 기댓값은 각각 $\dfrac{1}{\lambda_1}$ 과 $\dfrac{1}{\lambda_2}$ 이므로 $E(XY) = \dfrac{1}{\lambda_1\lambda_2}$ 이다. ■

정리 4.4 상수 $a,\ b_1, b_2, \ldots, b_n$ 에 대하여 다음이 성립한다.

$$E\left(a + \sum_{i=1}^{n} b_i X_i\right) = a + \sum_{i=1}^{n} b_i E(X_i)$$

[증명] 이 정리의 증명도 연속형 확률변수의 경우만 다루도록 하겠다. $X_1, X_2, \ldots,\ X_n$ 의 결합확률밀도함수를 $f(x_1, x_2, \ldots, x_n)$ 라 할 때 다음 결과를 얻을 수 있다.

$$\begin{aligned}
&E\left(a + \sum_{i=1}^{n} b_i X_i\right) \\
&\quad = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}(a + b_1x_1 + b_2x_2 + \cdots + b_nx_n)f(x_1, x_2, \ldots, x_n)\,dx_1\,dx_2\cdots dx_n \\
&\quad = a + b_1\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} x_1 f(x_1, x_2, \ldots, x_n)\,dx_1\,dx_2\cdots dx_n \\
&\qquad\qquad + \cdots + b_n\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} x_n f(x_1, x_2, \ldots, x_n)\,dx_1\,dx_2\cdots dx_n \\
&\quad = a + b_1E(X_1) + \cdots + b_nE(X_n)
\end{aligned}$$

■

위 정리는 확률변수의 선형결합[2]의 기댓값은 각 확률변수의 기댓값의 선형결합과 같다는 사실을 말해주고 있다. 즉 확률변수에 대해 합을 하는 것과 기댓값을 구하는 것은 순서를 바꾸어

[2]확률변수의 선형결합이란 각 확률변수에 상수를 곱하여 더한 형태를 말한다

계산하여도 된다는 것을 말한다. 이 정리는 확률변수의 독립성과는 무관하게 항상 성립한다. 반면 정리 4.3은 확률변수가 독립일 경우에만 성립한다. 즉 곱과 기댓값의 순서를 바꾸기 위해서는 확률변수들이 서로 독립이어야 한다.

정리 4.4를 이용하면 베르누이분포의 기댓값으로부터 이항분포의 기댓값을 쉽게 구할 수 있다. 시행횟수가 n 인 이항분포는 베르누이 시행을 n 번 반복하여 성공한 횟수에 대한 확률분포이므로 이항분포를 따르는 확률변수 X 는 베르누이분포를 따르는 확률변수 n 개의 합이라 할 수 있다. 즉 $X \sim \mathrm{B}(n, p)$ 이고 $Y_i \sim \mathrm{Bernoulli}(p)$ 일 때, X 는 다음과 같이 나타낼 수 있다.[3]

$$X = \sum_{i=1}^{n} Y_i$$

베르누이분포를 따르는 확률변수 Y_i 의 기댓값은 p 이므로 $E(X) = \sum_{i=1}^{n} E(Y_i) = np$ 임을 알 수 있다.

4.3 분산과 표준편차

모평균 (기댓값) 이 모집단 (확률분포) 의 중심위치를 나타낸 값인 반면 중심위치로부터 얼마만큼 퍼져있는지를 나타낸 값이 **분산**(variance)이다. 확률변수의 모평균을 μ 라 할 때 분산은 $(X - \mu)^2$ 의 기댓값으로 정의된다. 확률변수 X 의 분산을 $Var(X)$ 또는 σ^2 로 나타낸다. 또 분산의 양의 제곱근을 **표준편차**(standard deviation)라 하며 σ 로 나타낸다.

정의 4.2 확률변수 X 의 분산은 다음과 같다.

$$Var(X) = E(X - \mu)^2 \tag{4.7}$$

분산을 구하는 위 식은 다음과 같이 나타낼 수 있다.

$$\begin{aligned} Var(X) &= E(X - \mu)^2 = E(X^2 - 2\mu X + \mu^2) \\ &= E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - \mu^2 \\ &= E(X^2) - [E(X)]^2 \end{aligned} \tag{4.8}$$

즉 분산은 확률변수의 제곱의 기댓값에서 기댓값의 제곱을 뺀 것이다. 식 (4.7) 을 사용하여 분산을 구하는 것보다 식 (4.8) 을 이용한 계산이 간편할 수도 있다.

예 4.4 이산형 확률변수 X 의 확률질량함수가 다음과 같을 때 X 의 분산을 구하여라.

[3] 위 식에서 좌변과 우변이 같다는 것은 좌변의 확률분포와 우변의 확률분포가 같다는 것을 의미한다. 두 확률변수의 확률분포가 같을 때 두 확률변수가 같다고 표현하기도 한다.

x	-1	0	1	2
$p_X(x)$	0.1	0.2	0.3	0.4

[풀이] X의 기댓값은 1이므로 X의 분산은 다음과 같다.

$$\begin{aligned} Var(X) &= E(X-1)^2 \\ &= (-1-1)^2 \times 0.1 + (0-1)^2 \times 0.2 + (1-1)^2 \times 0.3 + (2-1)^2 \times 0.4 \\ &= 1.0 \end{aligned}$$

한편 예 4.2에서 구한 X^2의 기댓값은 $E(X^2) = 2$이므로 다음과 같이 분산을 구할 수도 있다.

$$Var(X) = E(X^2) - [E(X)]^2 = 2 - 1 = 1$$

■

2장에서 다룬 몇 가지 확률분포의 분산을 구하여 보자. X가 성공의 확률이 p인 베르누이분포를 따르는 확률변수일 때 X^2의 기댓값은 다음과 같다.

$$E(X^2) = 0^2 \times (1-p) + 1^2 \times p = p$$

따라서 X의 분산은 다음과 같다.

$$Var(X) = E(X^2) - E(X)^2 = p - p^2 = p(1-p)$$

이항분포를 따르는 확률변수 X에 대하여 X^2의 기댓값을 직접 구하는 것은 쉽지 않다. 한편 $X^2 = X(X-1) + X$이므로 X^2의 기댓값은 다음과 같이 구할 수 있다.

$$E(X^2) = E[X(X-1)] + E(X)$$

한편 $X(X-1)$ 의 기댓값은 다음과 같다.

$$\begin{aligned}
E[X(X-1)] &= \sum_{k=0}^{n} k(k-1)\frac{n!}{k!(n-k)!}\, p^k(1-p)^{n-k} \\
&= \sum_{k=2}^{n} \frac{n!}{(k-2)!(n-k)!}\, p^k(1-p)^{n-k} \\
&= n(n-1)p^2 \sum_{k=2}^{n} \frac{(n-2)!}{(k-2)!(n-k)!}\, p^{k-2}(1-p)^{n-k} \\
&= n(n-1)p^2 \sum_{l=0}^{n-2} \frac{(n-2)!}{l!(n-2-l)!}\, p^{l}(1-p)^{n-2-l} \quad (\text{단}, l=k-2)
\end{aligned}$$

위 식에서 $\sum_{l=0}^{n-2} \frac{(n-2)!}{l!(n-2-l)!}\, p^l(1-p)^{n-2-l}$ 은 시행횟수는 $n-2$ 이고 성공의 확률이 p 인 이항분포의 모든 확률의 합이므로 1이 된다. 따라서 $E[X(X-1)] = n(n-1)p^2$ 이며 X^2 의 기댓값은 다음과 같다.

$$E(X^2) \;=\; E[X(X-1)] + E(X) \;=\; n(n-1)p^2 + np$$

따라서 X 의 분산은 다음과 같다.

$$\begin{aligned}
Var(X) \;&=\; E(X^2) - [E(X)]^2 \;=\; n(n-1)p^2 + np - (np)^2 \\
&=\; np(1-p)
\end{aligned}$$

포아송분포의 경우도 이와 같은 방법으로 계산하면 분산이 λ 임을 알 수 있다.

연속형 확률분포의 예로 감마분포의 분산을 구하여 보자. 확률밀도함수가 (2.8) 로 주어지는 감마분포를 따르는 확률변수 X 에 대하여 X^2 의 기댓값은 다음과 같이 나타낼 수 있다.

$$\begin{aligned}
E(X^2) \;&= \int_0^\infty x^2 \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}\, dx = \int_0^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha+1} e^{-\lambda x}\, dx \\
&= \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)\lambda^2} \int_0^\infty \frac{\lambda^{\alpha+2}}{\Gamma(\alpha+2)} x^{(\alpha+2)-1} e^{-\lambda x}\, dx
\end{aligned}$$

위 식에서 적분 안의 함수는 Gamma$(\alpha+2,\lambda)$ 의 확률밀도함수이므로 적분은 1 이 된다. 따라서 X^2 의 기댓값은 다음과 같다.

$$E(X^2) \;== \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)\lambda^2} = \frac{\alpha(\alpha+1)}{\lambda^2}$$

한편 $E(X) = \frac{\alpha}{\lambda}$ 이므로 X 의 분산은 $Var(X) = \frac{\alpha}{\lambda^2}$ 이다. 이로부터 지수분포의 분산은 $\frac{1}{\lambda^2}$ 이고, 자유도가 n 인 카이제곱분포의 분산은 $2n$ 임을 알 수도 있다.

한편 $\mathrm{N}(\mu, \sigma^2)$ 를 따르는 확률변수 X 의 분산은 다음과 같이 구할 수 있다.

$$\begin{aligned}
Var(X) &= E(X-\mu)^2 = \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx \\
&= \sigma^2 \int_{-\infty}^{\infty} \frac{t^2}{\sqrt{2\pi}}\, e^{-\frac{1}{2}t^2}\, dt \quad (\text{단},\ t = \frac{x-\mu}{\sigma}\ ) \\
&= 2\sigma^2 \int_0^{\infty} t^2 \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}t^2}\, dt \quad (\text{적분 안의 함수가 } y \text{축에 대칭이므로}) \\
&= \sigma^2 \int_0^{\infty} u \frac{1}{\sqrt{2\pi}}\, e^{-\frac{u}{2}} \frac{1}{\sqrt{u}}\, du \quad (\text{단},\ u = t^2\ ) \\
&= \sigma^2 \int_0^{\infty} \frac{1}{\frac{1}{2}\sqrt{\pi}} \left(\frac{1}{2}\right)^{\frac{3}{2}} u^{\frac{3}{2}-1}\, e^{-\frac{u}{2}}\, du
\end{aligned}$$

위 식에서 적분 안의 함수는 $\mathrm{Gamma}\left(\frac{3}{2}, \frac{1}{2}\right)$의 확률밀도함수이다. 참고로 $\Gamma\left(\frac{3}{2}\right) = \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \left(\frac{1}{2}\right)\sqrt{\pi}$ 이다. 따라서 적분 값은 1 이며 $\mathrm{N}(\mu, \sigma^2)$ 의 분산은 $Var(X) = \sigma^2$ 이다.

다음 정리는 확률변수 X 에 상수를 더하거나 곱했을 때 분산이 어떻게 변하는지를 보여주고 있다.

정리 4.5 상수 a 와 b 에 대하여 다음이 성립한다.

(i) $Var(a + X) = Var(X)$

(ii) $Var(bX) = b^2 Var(X)$

(iii) $Var(a + bX) = b^2 Var(X)$

[증명] $E(a + X) = a + E(X)$ 이므로 $a + X$ 의 분산은 다음과 같다.

$$Var(a+X) = E[a + X - (a + E(X))]^2 = E[X - E(X)]^2 = Var(X)$$

또 $E(bX) = bE(X)$ 이므로 bX 의 분산은 다음과 같다.

$$Var(bX) = E[bX - bE(X)]^2 = b^2 E[X - E(X)]^2 = b^2 Var(X)$$

앞의 두 성질에 의하여 $Var(a + bX) = Var(bX) = b^2 Var(X)$ 임을 확인할 수 있다. ■

위 정리는 확률변수를 a 만큼 평행이동 하여도 퍼진 정도를 측정하는 분산은 변하지 않는다는 사실과 분산은 제곱의 기댓값 형태이기 때문에 확률변수에 b 를 곱할 때 분산은 b^2 배가 된다는 사실을 말해주고 있다.

다음은 확률변수가 기댓값으로부터 일정한 거리 이상 떨어질 확률의 부등식에 대한 정리이다.

정리 4.6 [체비세프 부등식, Tchebychev inequality] X의 기댓값이 μ이고 분산이 σ^2일 때 임의의 실수 $t > 0$에 대하여 다음 부등식이 성립한다.

$$P(|X - \mu| \geq t) \;\leq\; \frac{\sigma^2}{t^2}$$

[**증명**] 증명은 X가 연속형 확률변수인 경우만 다루기로 하겠다. 확률변수 X가 기댓값 μ로부터 t 이상 떨어질 확률은 다음과 같이 나타낼 수 있다.

$$\begin{aligned} P(|X - \mu| \geq \; t) \;&=\; P(X \leq \mu - t \text{ 또는 } X \geq \mu + t) \\ &=\; \int_{-\infty}^{\mu-t} f(x)\,dx \;+\; \int_{\mu+t}^{\infty} f(x)\,dx \end{aligned}$$

$x \leq \mu - t$이거나 $x \geq \mu + t$일 때 $\left(\dfrac{x-\mu}{t}\right)^2 \;\geq\; 1$이므로 다음 부등식이 성립한다.

$$\begin{aligned} &\int_{-\infty}^{\mu-t} f(x)\,dx \;+\; \int_{\mu+t}^{\infty} f(x)\,dx \\ &\quad\leq\; \int_{-\infty}^{\mu-t} \left(\frac{x-\mu}{t}\right)^2 f(x)\,dx \;+\; \int_{\mu+t}^{\infty} \left(\frac{x-\mu}{t}\right)^2 f(x)\,dx \\ &\quad\leq\; \int_{-\infty}^{\infty} \left(\frac{x-\mu}{t}\right)^2 f(x)\,dx = \frac{\sigma^2}{t^2} \end{aligned}$$

■

체비세프 부등식에 의하여 $t = k\sigma$일 때 다음 부등식을 얻을 수 있다.

$$P(|X - \mu| \geq k\sigma) \;\leq\; \frac{1}{k^2}$$

즉 확률변수가 기댓값으로부터 표준편차의 k배 이상 떨어진 값을 가질 확률은 $\frac{1}{k^2}$보다 작다는 것을 의미한다. 예를 들어 기댓값으로부터 표준편차의 2배 이상 떨어진 값을 가질 확률은 $\frac{1}{4}$보다 작으며, 3배 이상 떨어진 값을 가질 확률은 $\frac{1}{9}$보다 작다. 이 부등식은 X의 확률분포에 관계없이 항상 성립하는 부등식이다.

4.4 델타 방법

X의 확률분포를 알고 있다고 하더라도 $g(\cdot)$가 복잡한 형태의 함수라면 $Y = g(X)$의 기댓값과 분산을 구하는 것이 쉽지 않을 수 있다. 이 절에서는 X의 기댓값과 분산을 이용하여 $Y = g(X)$의 근사적인 기댓값과 분산을 구하는 방법을 알아보도록 하자.

확률변수 X의 기댓값과 분산을 각각 μ_X와 σ_X^2라 하자. 함수 $Y = g(X)$에 대하여 μ_X를 중심으로 테일러 전개하면 다음을 얻을 수 있다.

$$g(X) = g(\mu_X) + g'(\mu_X)(X - \mu_X) + \frac{g''(\mu_X)}{2}(X - \mu_X)^2 + \cdots \tag{4.9}$$

한편 X의 기댓값이 μ_X이므로 X는 μ_X에 가까울 것으로 생각할 수 있다. 즉 체비세프 부등식이 의미하는 것과 같이 X가 μ_X에서 멀리 떨어진 값을 가질 확률은 작다고 할 수 있다. $|X - \mu_X|$의 값이 작으면 k가 클수록 $(X - \mu_X)^k$는 0에 가까운 값을 가진다. 따라서 식 (4.9)의 우측에 주어진 값은 처음 몇 항의 값에 크게 의존하게 될 것이다.

테일러 전개에서 처음 세 항을 이용하여 $Y = g(X)$의 근사적인 기댓값을 구하면 다음과 같다.

$$\begin{aligned} E(g(X)) &\approx g(\mu_X) + g'(\mu_X)E(X - \mu_X) + \frac{g''(\mu_X)}{2}E(X - \mu_X)^2 \\ &= g(\mu_X) + \frac{g''(\mu_X)}{2}\sigma_X^2 \end{aligned}$$

또 테일러 전개에서 처음 두 항을 이용하여 $Y = g(X)$의 근사적 분산을 구하면 다음과 같다.

$$\begin{aligned} Var(g(X)) &\approx Var\Big(g(\mu_X) + g'(\mu_X)(X - \mu_X)\Big) \\ &= (g'(\mu_X))^2\sigma_X^2 \end{aligned}$$

한편 $Y = g(X)$의 근사적 기댓값과 분산을 구할 때 테일러 전개에서 사용하는 항의 수를 늘리면 좀 더 정확한 값을 알 수 있겠지만 계산하는 과정이 복잡해질 것이다.

예 4.5 $X \sim \text{Uniform}(0,1)$일 때 다음 물음에 답하여라.

(a) $Y = \sqrt{X}$의 기댓값과 분산을 구하여라.

(b) 델타 방법을 사용하여 $Y = \sqrt{X}$의 근사적인 기댓값과 분산을 구하여라.

[풀이]

(a) $Y = \sqrt{X}$의 기댓값과 분산은 다음과 같다.

$$\begin{aligned} E(\sqrt{X}) &= \int_0^1 \sqrt{x}\, dx = \frac{2}{3} \\ Var(\sqrt{X}) &= E(X) - [E(\sqrt{X})]^2 = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18} \end{aligned}$$

(b) $g(x) = \sqrt{x}$ 일 때 $g'(x) = \frac{1}{2\sqrt{x}}$ 이고 $g''(x) = -\frac{1}{4}x^{-3/2}$ 이다. 한편 $\mu_X = \frac{1}{2}$ 이고 $\sigma_X^2 = \frac{1}{12}$ 이므로 $\sqrt{X}$ 의 근사적인 기댓값과 분산은 다음과 같다.

$$\begin{aligned}
E(\sqrt{X}) &\approx g(\mu_X) + \frac{g''(\mu_X)}{2}\sigma_X^2 \\
&= \sqrt{\frac{1}{2}} + \frac{1}{2}\left(-\frac{1}{4}2\sqrt{2}\right)\frac{1}{12} = 0.678 \\
Var(\sqrt{X}) &\approx \left(g'(\mu_X)\right)^2 \sigma_X^2 \\
&= \left(\frac{\sqrt{2}}{2}\right)^2 \frac{1}{12} = \frac{1}{24}
\end{aligned}$$

■

4.5 공분산

공분산(covariance)은 두 확률변수의 선형관계의 정도를 측정한 값이다. 두 확률변수 X와 Y의 기댓값을 각각 μ_X와 μ_Y라 할 때 $(X-\mu_X)(Y-\mu_Y)$의 기댓값을 공분산이라 하며 $Cov(X,Y)$로 나타낸다.

정의 4.3 확률변수 X와 Y의 공분산은 다음과 같다.

$$Cov(X,Y) \ = \ E[(X-\mu_X)(Y-\mu_Y)]$$

공분산을 구하는 식은 다음과 같이 나타낼 수 있다.

$$\begin{aligned} Cov(X,Y) &= E[(X-\mu_X)(Y-\mu_Y)] \ = \ E(XY-\mu_X Y-\mu_Y X+\mu_X\mu_Y) \\ &= E(XY)-\mu_X\mu_Y-\mu_Y\mu_X+\mu_X\mu_Y \ = \ E(XY)-\mu_X\mu_Y \\ &= E(XY)-E(X)E(Y) \end{aligned}$$

즉 공분산은 두 확률변수의 곱의 기댓값에서 기댓값의 곱을 뺀 값이다. 한편 $Cov(X,X)=Var(X)$이다. 분산은 하나의 확률변수에 대하여 퍼진 정도를 측정하는 값인 반면 공분산은 이차원에서 분포 형태(퍼진 정도)를 이용하여 두 확률변수 사이의 선형관계를 측정한 값이다. X가 증가할 때 Y가 증가하면 $(X-\mu_X)(Y-\mu_Y)$는 양의 값을 가지며 X가 증가할 때 Y가 감소하면 $(X-\mu_X)(Y-\mu_Y)$는 음의 값을 가진다. 따라서 X와 Y의 결합분포가 양의 기울기를 가지는 직선에 집중되어 있을수록 공분산은 크기가 큰 양의 값을 가지며 음의 기울기를 가지는 직선에 집중되어 있을수록 공분산은 크기가 큰 음의 값을 가진다. 한편 공분산의 값이 0이라는 것은 두 변수 사이에 선형관계가 없음을 나타낸다.

예 4.6 다음과 같은 결합확률질량함수를 가지는 확률변수 X와 Y의 공분산을 구하여라.

X와 Y의 결합분포

		y				
		-1	$-\sqrt{\frac{1}{2}}$	0	$\sqrt{\frac{1}{2}}$	1
	-1	0	0	$\frac{1}{8}$	0	0
	$-\sqrt{\frac{1}{2}}$	0	$\frac{1}{8}$	0	$\frac{1}{8}$	0
x	0	$\frac{1}{8}$	0	0	0	$\frac{1}{8}$
	$\sqrt{\frac{1}{2}}$	0	$\frac{1}{8}$	0	$\frac{1}{8}$	0
	1	0	0	$\frac{1}{8}$	0	0

[풀이] 주어진 결합분포로부터 X의 주변확률분포를 구하면 다음과 같다.

x	-1	$-\sqrt{\frac{1}{2}}$	0	$\sqrt{\frac{1}{2}}$	1
$p_X(x)$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{8}$

이로부터 X의 기댓값은 0임을 알 수 있다. 결합분포가 X와 Y에 대하여 대칭이므로 Y의 주변확률분포도 X의 주변확률분포와 같고 Y의 기댓값도 0이다. 또한 $E(XY) = 0$임을 쉽게 확인할 수 있다. 따라서 $Cov(X,Y) = 0$이다. ■

예 4.6에서 주어진 결합분포는 원점을 중심으로 반지름이 1인 원 위의 대칭적인 8개의 점에 균일한 확률을 가지는 이산형 분포이다. 따라서 X와 Y의 결합분포는 원 위에서 대칭적인 확률을 가지므로 선형관계는 없다고 할 수 있다.

예 4.7 다음과 같은 결합확률밀도함수를 가지는 확률변수 X와 Y의 공분산을 구하여라.

$$f(x,y) \;=\; x+y, \quad 0 \le x \le 1,\ 0 \le y \le 1$$

[풀이] 결합확률밀도함수를 이용하여 X와 Y의 기댓값을 구하면 다음과 같다.

$$\begin{aligned} E(X) &= \int_0^1 \int_0^1 x(x+y)\,dy\,dx = \int_0^1 \left(x^2 + \frac{x}{2}\right)\,dx = \frac{7}{12} \\ E(Y) &= \int_0^1 \int_0^1 y(x+y)\,dx\,dy = \int_0^1 \left(\frac{y}{2} + y^2\right)\,dy = \frac{7}{12} \end{aligned}$$

또 XY의 기댓값은 다음과 같으므로

$$E(XY) \;=\; \int_0^1 \int_0^1 xy(x+y)\,dy\,dx \;=\; \int_0^1 \left(\frac{x^2}{2} + \frac{x}{3}\right)\,dx = \frac{1}{3}$$

공분산은 $Cov(X,Y) = \frac{1}{3} - \left(\frac{7}{12}\right)^2 = -\frac{1}{144}$ 이다. ■

X와 Y가 서로 독립이면 두 확률변수 사이에는 어떤 연관성도 없다고 할 수 있으므로 X와 Y의 선형관계를 측정하는 공분산의 값은 0이어야 할 것이다. 이는 두 확률변수가 독립일 때 $E(XY) = E(X)E(Y)$이므로 X와 Y의 공분산이 0임을 쉽게 확인할 수 있다. 한편 두 확률변수의

공분산이 0이라는 사실이 서로 독립임을 의미하지는 않는다. 이는 앞에서 다룬 원형관계를 가지는 예 4.6을 통하여 확인할 수 있다. 원형관계를 가지는 두 확률변수의 공분산은 0이지만 두 확률변수 X와 Y는 서로 독립이 아니다. 이는 $P(X=1)=P(Y=1)=\frac{1}{8}$ 이지만 $P(X=1, Y=1)=0$ 라는 것으로부터 X와 Y가 서로 독립이 아님을 확인할 수 있다. 즉 공분산이 0이라는 것은 두 확률변수 사이에 선형관계가 없다는 것이지 서로 독립이라는 것을 의미하지는 않는다.

확률변수에 상수를 더하거나 곱했을 때 분산과 유사하게 공분산도 다음과 같은 성질을 가지고 있다.

정리 4.7 상수 a와 확률변수 X, X_1, X_2, Y에 대하여 다음이 성립한다.

(i) $Cov(X+a, Y) = Cov(X, Y)$

(ii) $Cov(aX, Y) = a\,Cov(X, Y)$

(iii) $Cov(X_1+X_2, Y) = Cov(X_1, Y) + Cov(X_2, Y)$

[증명]

(i) 상수 a에 대하여 $E(X+a) = E(X)+a$, $E(aX) = aE(X)$ 이므로 $X+a$와 Y의 공분산은 다음과 같다.

$$\begin{aligned} Cov(X+a, Y) &= E[(X+a-E(X+a))(Y-E(Y))] = E[(X-E(X))(Y-E(Y)] \\ &= Cov(X, Y) \end{aligned}$$

(ii) aX와 Y의 공분산은 다음과 같다.

$$\begin{aligned} Cov(aX, Y) &= E[(aX-E(aX))(Y-E(Y))] = a\,E[(X-E(X))(Y-E(Y))] \\ &= a\,Cov(X, Y) \end{aligned}$$

(iii) X_1+X_2와 Y의 공분산은 다음과 같이 구할 수 있다.

$$\begin{aligned} Cov(X_1+X_2, Y) &= E[(X_1+X_2-E(X_1+X_2))(Y-E(Y)] \\ &= E[(X_1-E(X_1))(Y-E(Y))] + E[(X_2-E(X_2))(Y-E(Y))] \\ &= Cov(X_1, Y) + Cov(X_2, Y) \end{aligned}$$

■

정리 4.7의 세 번째 성질을 확장하면 $\sum_{i=1}^{m} a_i X_i$와 $\sum_{j=1}^{n} b_j Y_j$의 공분산을 구하는 식은 다음과 같이 나타낼 수 있다.

$$
\begin{aligned}
& Cov\Big(\sum_{i=1}^{m} a_i X_i,\ \sum_{j=1}^{n} b_j Y_j\Big) \\
&= \sum_{i=1}^{m} Cov\Big(a_i X_i,\ \sum_{j=1}^{n} b_j Y_j\Big) \ = \ \sum_{i=1}^{m}\sum_{j=1}^{n} Cov(a_i X_i, b_j Y_j) \\
&= \ \sum_{i=1}^{m}\sum_{j=1}^{n} a_i b_j Cov(X_i, Y_j)
\end{aligned}
$$

한편 $Var(X) = Cov(X, X)$이므로 위 식으로부터 두 확률변수의 합의 분산은 다음과 같이 나타낼 수 있다.

$$
\begin{aligned}
Var(X+Y) \ &= \ Cov(X+Y, X+Y) \\
&= \ Cov(X,X) + 2Cov(X,Y) + Cov(Y,Y) \ = \ Var(X) + 2Cov(X,Y) + Var(Y)
\end{aligned}
$$

같은 방법으로 n개의 확률변수 $X_1, X_2, \ldots, X_n$의 합의 분산은 다음과 같이 나타낼 수 있다.

$$
Var\Big(\sum_{i=1}^{n} X_i\Big) \ = \ Cov\Big(\sum_{i=1}^{n} X_i,\ \sum_{j=1}^{n} X_j\Big) \ = \ \sum_{i=1}^{n}\sum_{j=1}^{n} Cov(X_i, X_j)
$$

따라서 서로 독립인 확률변수 $X_1, X_2, \ldots, X_n$의 합의 분산은 다음과 같다.

$$
Var\Big(\sum_{i=1}^{n} X_i\Big) \ = \sum_{i=1}^{n} Var(X_i) \tag{4.10}
$$

서로 독립인 확률변수의 합의 분산을 구하는 식 (4.10)을 이용하면 베르누이분포의 분산으로부터 이항분포의 분산을 구할 수 있다. 시행횟수가 n인 이항분포를 따르는 확률변수는 베르누이분포를 따르는 서로 독립인 확률변수 n개의 합이라 할 수 있다. 베르누이분포의 분산은 $p(1-p)$이므로 이항분포의 분산은 $np(1-p)$임을 알 수 있다.

공분산은 두 확률변수 사이의 선형관계를 측정하는 값이라 하였다. 그런데 공분산은 상대적인 값이 아니어서 이 값을 이용하여 선형관계의 정도를 측정하는 것은 쉽지 않다. 예를 들어 $2X$와 $2Y$의 선형관계는 X와 Y의 선형관계와 같다고 할 수 있지만 $Cov(2X, 2Y) = 4Cov(X, Y)$로 공분산이 4배가 된다. 구체적인 예로 학생 30명 중 임의로 선택된 학생의 키와 체중을 나타내는 확률변수를 X와 Y라 하자. 일반적으로 키가 클수록 체중도 증가하는 경향이 있으므로 키와 체중 사이에는 양의 선형관계가 있을 것이다. 이때 키를 나타내는 단위로 m를 사용할 수도 있고 cm를 사용할 수도 있다. 그런데 키의 단위로 cm를 사용했을 때의 공분산은 m를 사용했을 때의 100배가 된다. 이처럼 같은 확률변수에 대하여도 사용하는 단위에 따라 공분산은 다른 값을 가지게 되므로 공분산의 값을 이용하여 선형관계의 정도를 판단하는 것은 쉽지 않다.

상관계수(correlation coefficient)는 두 확률변수 X와 Y 사이의 선형관계를 상대적인 값으로

나타낸 것으로 $Corr(X,Y)$ 또는 ρ로 나타낸다.

정의 4.4 확률변수 X와 Y의 상관계수는 다음과 같다.

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$$

예 4.8 다음과 같은 결합확률밀도함수를 가지는 확률변수 X와 Y의 상관계수를 구하여라.

$$f(x,y) = x+y, \quad 0 \le x \le 1,\ 0 \le y \le 1.$$

[**풀이**] 예 4.7에서 구한 것처럼 X와 Y의 기댓값은 모두 $\frac{7}{12}$이며 공분산은 $-\frac{1}{144}$이다. 한편 X^2의 기댓값은 다음과 같으므로

$$E(X^2) = \int_0^1\int_0^1 x^2(x+y)\,dydx = \int_0^1 \left(x^3 + \frac{x^2}{2}\right)\,dx = \frac{5}{12}$$

X의 분산은 $Var(X) = E(X^2) - [E(X)]^2 = \frac{5}{12} - \left(\frac{7}{12}\right)^2 = \frac{11}{144}$ 이다. 같은 방법으로 Y의 분산도 $Var(Y) = \frac{11}{144}$이 된다. 따라서 X와 Y의 상관계수는 다음과 같다.

$$\rho = \frac{-1/144}{\sqrt{11/144}\,\sqrt{11/144}} = -\frac{1}{11}$$

■

상관계수의 정의로부터 상수 a, b, c, d에 대하여 다음이 성립한다.

$$Corr(aX+b, cY+d) = \begin{cases} Corr(X,Y), & ac > 0\text{일 때} \\ -Corr(X,Y), & ac < 0\text{일 때} \end{cases}$$

즉 상관계수는 선형변환에 관계없이 일정한 크기의 값을 가진다. 또한 상관계수는 다음과 같은 성질을 가지고 있음을 보일 수 있다.

정리 4.8 두 확률변수 X와 Y의 상관계수 ρ는 다음과 같은 성질을 갖는다.

(i) $-1 \le \rho \le 1$

(ii) $\rho = 1$일 필요충분 조건은 상수 a와 양의 상수 b에 대하여 $P(Y = a + bX) = 1$이다.

(iii) $\rho = -1$일 필요충분 조건은 상수 a와 음의 상수 b에 대하여 $P(Y = a + bX) = 1$이다.

여기서 $P(Y = a + bX) = 1$이라는 것은 (X, Y)는 직선 $y = a + bx$ 위에서만 값을 가질 확률이 1임을 의미한다.

[증명] X와 Y의 표준편차를 σ_X와 σ_Y라 할 때 $\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}$의 분산은 다음과 같다.

$$\begin{aligned} Var\left(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right) &= Var\left(\frac{X}{\sigma_X}\right) + Var\left(\frac{Y}{\sigma_Y}\right) + 2Cov\left(\frac{X}{\sigma_X}, \frac{Y}{\sigma_Y}\right) \\ &= \frac{Var(X)}{\sigma_X^2} + \frac{Var(Y)}{\sigma_Y^2} + 2\frac{Cov(X,Y)}{\sigma_X \sigma_Y} \; = \; 2(1+\rho) \end{aligned}$$

분산은 항상 음이 아닌 값을 가지므로 위 식에서 $\rho \geq -1$이어야 한다. 또한 위 식으로부터 $\rho = -1$인 경우는 $Var\left(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right) = 0$일 때이다. 한편 확률변수의 분산이 0이라는 것은 확률변수가 오직 하나의 값만 가진다는 것을 의미한다. 즉 $\rho = -1$인 경우는 $\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}$이 하나의 값만 가질 때이며, 이 값을 상수 c라 할 때 다음이 성립한다.

$$\begin{aligned} \rho = -1 \quad &\iff \quad \frac{X}{\sigma_X} + \frac{Y}{\sigma_Y} = c \\ &\iff \quad Y \; = \; c\sigma_Y - \frac{\sigma_Y}{\sigma_X}X \end{aligned}$$

또한 $\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}$의 분산은 다음과 같으므로

$$\begin{aligned} Var\left(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right) &= Var\left(\frac{X}{\sigma_X}\right) + Var\left(\frac{Y}{\sigma_Y}\right) - 2Cov\left(\frac{X}{\sigma_X}, \frac{Y}{\sigma_Y}\right) \\ &= \frac{Var(X)}{\sigma_X^2} + \frac{Var(Y)}{\sigma_Y^2} - 2\frac{Cov(X,Y)}{\sigma_X \sigma_Y} \; = \; 2(1-\rho) \end{aligned}$$

$\rho \leq 1$이다. 위 식으로부터 다음이 성립한다.

$$\begin{aligned} \rho = 1 \quad &\iff \quad Var\left(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right) = 0 \\ &\iff \quad \frac{X}{\sigma_X} - \frac{Y}{\sigma_Y} = c \text{ (상수)} \\ &\iff \quad Y \; = \; -c\sigma_Y + \frac{\sigma_Y}{\sigma_X}X \end{aligned}$$

■

상관계수는 두 확률변수의 선형관계를 상대적인 값으로 나타낸 것이며, 이 값의 절대값이 클수록 두 확률변수의 선형관계는 강하다고 할 수 있다. 정리 4.8에 의하여 상관계수는 -1과 1

사이의 값을 가지므로 선형관계에 가까울수록 상관계수의 절대값은 1에 가까워진다. X와 Y 사이에 완벽한 양의 선형관계가 있다면 상관계수는 1의 값을 가지며 반대로 완벽한 음의 선형관계가 있다면 상관계수는 -1이 된다. 그림 4.1은 네 가지 확률분포의 상관계수의 값을 보여주고 있다. 확률분포는 각 점에서 균등한 확률을 가지는 경우이다. 그림에서 볼 수 있는 것처럼 상관계수의 절대값이 클수록 강한 선형관계가 있음을 알 수 있다.

4.6 조건부분포의 기댓값과 분산

확률분포의 기댓값과 분산을 구하는 것과 같이 조건부분포에 대해서도 기댓값과 분산을 구할 수 있다. 조건부분포를 이용한 기댓값과 분산을 **조건부기댓값**(conditional expectation)과 **조건부분산**(conditional variance)이라 한다. $Y = y$로 주어졌을 때 X의 조건부기댓값을 $E(X|Y = y)$, 조건부분산을 $Var(X|Y = y)$로 나타낸다.

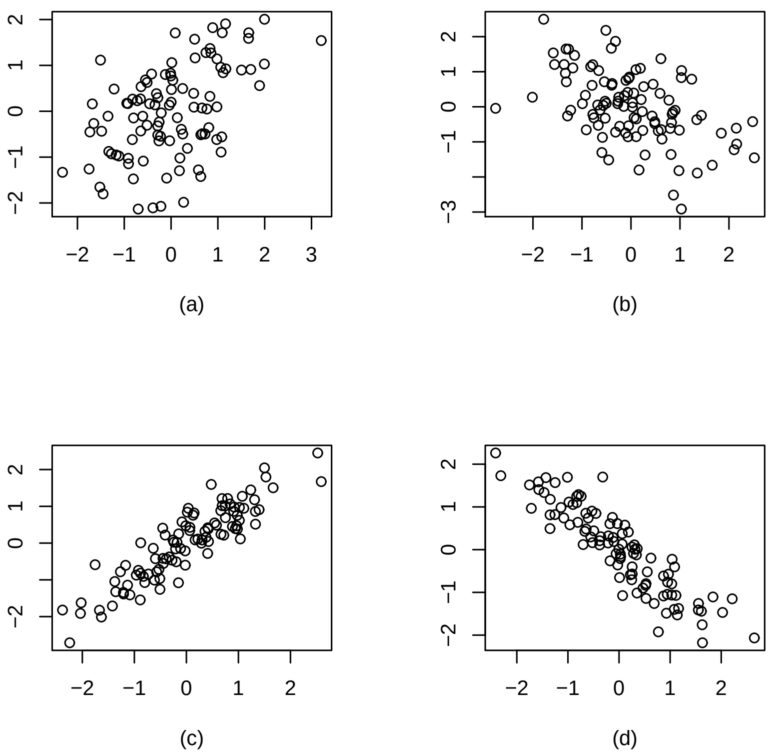

그림 4.1: 네 가지 확률분포의 상관계수: (a) $\rho = 0.5$, (b) $\rho = -0.5$, (c) $\rho = 0.9$, (d) $\rho = -0.9$

정의 4.5 $Y=y$로 주어졌을 때 이산형 확률변수 X의 조건부기댓값은 다음과 같고

$$E(X|Y=y) \ = \ \sum_i x_i p_{X|Y}(x_i|y)$$

연속형 확률변수 X의 조건부기댓값은 다음과 같다.

$$E(X|Y=y) \ = \ \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\, dx$$

마찬가지로 $Y=y$로 주어졌을 때 $g(X)$의 조건부기댓값은 다음과 같다.

$$E[g(X)|Y=y] \ = \ \begin{cases} \sum_i g(x_i)p_{X|Y}(x_i|y) & \text{(이산형 확률변수의 경우)} \\ \int_{-\infty}^{\infty} g(x) f_{X|Y}(x|y)\, dx & \text{(연속형 확률변수의 경우)} \end{cases}$$

확률분포의 기댓값과 분산은 확률밀도함수나 확률질량함수를 이용하여 구하는 것과 같이 조건부기댓값과 조건부분산은 조건부확률밀도함수나 조건부확률질량함수로부터 구할 수 있다. X의 분산을 $Var(X)=E(X^2)-(E(X))^2$로 구하는 것처럼 $Y=y$로 주어졌을 때 X의 조건부분산 $Var(X|Y=y)$은 다음과 같이 구할 수 있다.

$$Var(X|Y=y) = E(X^2|Y=y) - (E(X|Y=y))^2$$

예 4.9 X와 Y의 결합확률질량함수가 다음과 같다.

		y		
		1	2	3
x	0	$\frac{1}{12}$	$\frac{3}{12}$	$\frac{1}{12}$
	1	$\frac{2}{12}$	$\frac{2}{12}$	$\frac{3}{12}$

(a) $E(Y|X=0)$과 $Var(Y|X=0)$을 구하여라.

(b) $E(Y|X=1)$과 $Var(Y|X=1)$을 구하여라.

[풀이]

(a) $X=0$일 때 Y의 조건부분포는 다음과 같다.

y	1	2	3
$P(Y=y\|X=0)$	$\frac{1}{5}$	$\frac{3}{5}$	$\frac{1}{5}$

따라서 조건부기댓값 $E(Y|X=0)$은 다음과 같다.

$$E(Y|X=0) \;=\; 1\times\frac{1}{5}+2\times\frac{3}{5}+3\times\frac{1}{5} \;=\; 2$$

한편 $E(Y^2|X=0)$은 다음과 같으므로

$$E(Y^2|X=0) \;=\; 1^2\times\frac{1}{5}+2^2\times\frac{3}{5}+3^2\times\frac{1}{5} \;=\; \frac{22}{5}$$

조건부분산 $Var(Y|X=0)$은 다음과 같다.

$$Var(Y|X=0) \;=\; E(Y^2|X=0)-[E(Y|X=0)]^2 \;=\; \frac{2}{5}$$

(b) $X=1$일 때 Y의 조건부분포는 다음과 같으므로

y	1	2	3
$P(Y=y\|X=0)$	$\frac{2}{7}$	$\frac{2}{7}$	$\frac{3}{7}$

조건부기댓값과 조건부분산은 다음과 같다.

$$E(Y|X=1)=\frac{15}{7}, \quad Var(Y|X=1)=\frac{34}{49}$$

■

예 4.10 (X,Y)는 이변량정규분포를 따르는 확률변수라 한다. $X=x$일 때 Y의 조건부기댓값을 구하여라.

[풀이] 예 3.9에서 구한 것처럼 $X=x$일 때 Y의 조건부분포는 다음과 같다.

$$Y|X=x \;\sim\; \mathrm{N}\left(\mu_Y+\rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X),\; (1-\rho^2)\sigma_Y^2\right)$$

따라서 $X=x$일 때 Y의 조건부기댓값은 다음과 같다.

$$E(Y|X=x) \;=\; \mu_Y+\rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X)$$

참고로 이 조건부기댓값을 다르게 표현하면 다음과 같이 나타낼 수 있다.

$$\mu_Y + \frac{Cov(X,Y)}{Var(X)}(x - \mu_X)$$

자료 $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ 에 대하여 위의 조건부기댓값에 대응되는 값은 다음과 같다.

$$\overline{Y} + \frac{\sum_{i=1}^n (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^n (X_i - \overline{X})^2}(x - \overline{X})$$

이는 단순회귀분석에서 회귀 추정식과 일치함을 쉽게 확인할 수 있다. ■

예 4.11 X 와 Y 의 결합확률밀도함수가 $f(x,y) = \lambda^2 e^{-\lambda y},\ 0 \le x \le y$ 라 한다.

(a) $Y = y$ 일 때 X 의 조건부기댓값과 조건부분산을 구하여라.

(b) $X = x$ 일 때 Y 의 조건부기댓값과 조건부분산을 구하여라.

[풀이]

(a) 예 3.8에서 구한 것과 같이 $Y = y$ 일 때 X 의 조건부확률밀도함수는 $f_{X|Y}(x|y) = \frac{1}{y},\ 0 \le x \le y$ 이므로 X 와 X^2 의 조건부기댓값은 다음과 같다.

$$\begin{aligned} E(X|Y=y) &= \int_0^y x\,\frac{1}{y}\,dx = \frac{y}{2} \\ E(X^2|Y=y) &= \int_0^y x^2\,\frac{1}{y}\,dx = \frac{y^2}{3} \end{aligned}$$

따라서 X 의 조건부기댓값은 $\frac{y}{2}$ 이며 조건부분산은 $\frac{y^2}{12}$ 이다.

(b) $X = x$ 일 때 Y 의 조건부확률밀도함수는 $f_{Y|X}(y|x) = \lambda e^{-\lambda(y-x)},\ y \ge x$ 이므로 Y 와 Y^2 의 조건부기댓값은 다음과 같다.

$$\begin{aligned} E(Y|X=x) &= \int_x^\infty y\lambda e^{-\lambda(y-x)}\,dy = x + \frac{1}{\lambda} \\ E(Y^2|X=x) &= \int_x^\infty y^2\lambda e^{-\lambda(y-x)}\,dy = x^2 + \frac{2x}{\lambda} + \frac{2}{\lambda^2} \end{aligned}$$

따라서 Y 의 조건부기댓값은 $x + \frac{1}{\lambda}$ 이며 조건부분산은 $\frac{1}{\lambda^2}$ 이다.

■

$Y = y$일 때 X의 조건부기댓값 $E(X|Y = y)$은 주어진 y의 값이 변함에 따라 달라진다. 즉 조건부기댓값은 주어진 Y의 값에 의해 결정되므로 확률변수 Y의 함수라 할 수 있다. 이를 나타내기 위해 조건부기댓값을 $E(X|Y)$로 표현하기도 한다. $E(X|Y)$는 확률변수 Y의 함수이므로 조건부기댓값에 대하여 Y의 확률분포를 이용하여 기댓값을 구할 수 있다. 다음 정리는 이렇게 구한 값이 X의 기댓값과 같다는 사실을 보여주고 있다.

정리 4.9 확률변수 X와 Y에 대하여 다음이 성립한다.

$$E[E(X|Y)] = E(X) \tag{4.11}$$

[증명] 여기에서도 증명은 연속형 확률변수인 경우만 다루도록 하겠다. 주어진 Y의 값에 따라 달라지는 조건부기댓값 $E[E(X|Y)]$에 대하여 다시 기댓값을 구하면 다음과 같이 나타낼 수 있다.

$$E[E(X|Y)] = \int_{-\infty}^{\infty} E(X|Y = y)\, f_Y(y)\, dy$$

위 식은 $g(Y)$의 기댓값을 구하는 식 $E[g(Y)] = \int_{-\infty}^{\infty} g(y)\, f_Y(y)\, dy$에서 $g(y) = E(X|Y = y)$로 생각하면 이해가 빠를 것이다. $Y = y$일 때 X의 조건부기댓값 $E(X|Y = y)$는 다음과 같으므로

$$E(X|Y = y) = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\, dx$$

$E[E(X|Y)]$는 다음과 같다.

$$E[E(X|Y)] = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x f_{X|Y}(x|y)\; dx \right) f_Y(y)\, dy$$

한편 $f_{X|Y}(x|y) f_Y(y) = f(x, y)$이므로 다음이 성립한다.

$$E[E(X|Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x, y)\, dx\, dy = E(X)$$

■

위 정리의 의미를 알아보기 위해 다음과 같은 예를 생각하여 보자. 어느 고등학교 1학년에는 1반 40명, 2반 30명, 3반 30명의 총 100명의 학생이 있다고 한다. X를 이 학교 1학년 학생 중에서

임의로 선택된 학생의 키라 하고, Y를 이 학생이 소속된 반이라 하자. 이때 Y의 확률분포는 다음과 같다.

$$Y = \begin{cases} 1, & \frac{40}{100} \text{의 확률로} \\ 2, & \frac{30}{100} \text{의 확률로} \\ 3, & \frac{30}{100} \text{의 확률로} \end{cases}$$

$E(X|Y=1)$은 선택된 학생이 1반에 속한다는 조건이 주어졌으므로 1반 학생의 평균 키가 된다. 마찬가지로 $E(X|Y=2)$는 2반 학생의 평균 키이며, $E(X|Y=3)$은 3반 학생의 평균 키이다. 따라서 $E(X|Y)$의 기댓값은 다음과 같다.

$$\begin{aligned} E[E(X|Y)] &= E(X|Y=1)P(Y=1) + E(X|Y=2)P(Y=2) + E(X|Y=3)P(Y=3) \\ &= (1\text{반 평균 키}) \times \frac{40}{100} + (2\text{반 평균 키}) \times \frac{30}{100} + (3\text{반 평균 키}) \times \frac{30}{100} \end{aligned}$$

이 값은 전체 1학년 학생 100명의 평균 키임을 쉽게 확인 할 수 있다. $E(X)$ 역시 전체 1학년 학생의 평균 키이므로 $E[E(X|Y)]$와 $E(X)$는 같음을 알 수 있다. 일반적으로 전체 모집단을 Y의 값에 따라 나눌 때(예를 들어 반으로 나누는 경우) $E(X|Y)$는 각 부분에서 X의 기댓값 또는 평균(예를 들어 반 평균)이 된다. $E[E(X|Y)]$는 각 부분의 평균 값에 대하여 다시 평균(각 반 평균의 평균)을 구한 것이며, 이 값은 전체 모집단을 나누지 않고 구한 전체 평균과 같다. 한편 정리 4.9는 모든 확률변수에 대해 성립하므로 X와 Y의 역할을 바꾸어도 성립한다. 즉 다음 식이 성립한다.

$$E[E(Y|X)] = E(Y) \tag{4.12}$$

예 4.12 예 4.9에서 주어진 확률분포에 대하여 $E[E(Y|X)]$를 구하고, 이 값이 Y의 주변분포를 이용하여 구한 기댓값 $E(Y)$와 같음을 보여라.

[풀이] 이산형 확률변수 X는 0 또는 1의 값을 가지며 $P(X=0) = \frac{5}{12}$, $P(X=1) = \frac{7}{12}$이다. 예 4.9에서 구한 조건부기댓값으로부터 $E(Y|X)$의 기댓값은 다음과 같다.

$$\begin{aligned} E[E(Y|X)] &= E(Y|X=0) \times P(X=0) + E(Y|X=1) \times P(X=1) \\ &= 2 \times \frac{5}{12} + \frac{15}{7} \times \frac{7}{12} = \frac{25}{12} \end{aligned}$$

한편 Y의 주변분포는 다음과 같으므로

$$P(Y=1) = \frac{3}{12},\ P(Y=2) = \frac{5}{12},\ P(Y=3) = \frac{4}{12}$$

Y 의 기댓값은 다음과 같다.

$$E(Y) \;=\; 1 \times P(Y=1) + 2 \times P(Y=2) + 3 \times P(Y=3) = \frac{25}{12}$$

이 값은 앞에서 구한 $E[E(Y|X)]$ 와 같음을 알 수 있다. ■

예 4.13 X 와 Y 의 결합확률밀도함수가 $f(x,y) = \lambda^2 e^{-\lambda y},\ 0 \le x \le y$ 라 한다.

(a) $E[E(X|Y)]$ 를 구하고 이 값이 $E(X)$ 와 같음을 확인하여라.

(b) $E[E(Y|X)]$ 를 구하고 이 값이 $E(Y)$ 와 같음을 확인하여라.

[풀이]

(a) 앞의 예 4.11에서 구한 것처럼 $E(X|Y=y) = \frac{y}{2}$ 이고 Y 의 확률밀도함수는 $f_Y(y) = \lambda^2 y e^{-\lambda y},\ y > 0$ 이므로 $E(X|Y)$ 의 기댓값은 다음과 같다.

$$E[E(X|Y)] \;=\; E\left(\frac{Y}{2}\right) \;=\; \int_0^\infty \frac{y}{2} f_Y(y)\, dy \;=\; \int_0^\infty \frac{1}{2} \lambda^2 y^2 e^{-\lambda y}\, dy \;=\; \frac{1}{\lambda}$$

한편 X 의 확률밀도함수는 $f_X(x) = \lambda e^{-\lambda x},\ x > 0$ 이므로 X 의 기댓값은 다음과 같다.

$$E(X) \;=\; \int_0^\infty x\, f_X(x)\, dx \;=\; \int_0^\infty x\, \lambda e^{-\lambda x}\, dx \;=\; \frac{1}{\lambda}$$

위의 두 값으로부터 $E[E(X|Y)] = E(X)$ 임을 확인할 수 있다.

(b) $E(Y|X=x) = x + \frac{1}{\lambda}$ 이고 X 의 확률밀도함수는 $f_X(x) = \lambda e^{-\lambda x},\ x > 0$ 이므로 $E(Y|X)$ 의 기댓값은 다음과 같다.

$$E[E(Y|X)] \;=\; E\left(X + \frac{1}{\lambda}\right) \;=\; \int_0^\infty \left(x + \frac{1}{\lambda}\right) \lambda e^{-\lambda x}\, dx \;=\; \frac{2}{\lambda}$$

한편 Y 의 확률밀도함수는 $f_Y(x) = \lambda^2 y e^{-\lambda y},\ x > 0$ 이므로 Y 의 기댓값은 다음과 같다.

$$E(Y) \;=\; \int_0^\infty \lambda^2 y^2 e^{-\lambda y}\, dy \;=\; \frac{2}{\lambda}$$

위의 두 값으로부터 $E[E(Y|X)] = E(Y)$ 임을 확인할 수 있다.

■

주어진 Y의 값에 따라 달라지는 $E(X|Y)$의 기댓값을 구할 수 있는 것처럼 $E(X|Y)$의 분산도 구할 수 있다. 또한 Y의 확률분포를 이용하여 조건부분산 $Var(X|Y)$의 기댓값도 구할 수 있다. 다음 정리는 X의 분산이 조건부기댓값의 분산과 조건부분산의 기댓값의 합이라는 사실을 보여주고 있다.

정리 4.10 확률변수 X와 Y에 대하여 다음이 성립한다.

$$Var(X) = Var[E(X|Y)] + E[Var(X|Y)]$$

[증명] $E(X|Y)$의 분산은 다음과 같이 구할 수 있다.

$$\begin{aligned} Var[E(X|Y)] &= E[E(X|Y)]^2 - [E(E(X|Y))]^2 \\ &= E[E(X|Y)]^2 - [E(X)]^2 \end{aligned} \tag{4.13}$$

위 식은 $g(Y)$의 분산을 구하는 식 $Var(g(Y)) = E[g(Y)]^2 - [E(g(Y))]^2$에서 $g(Y) = E(X|Y)$로 생각하면 된다. 한편 조건부분산은 $Var(X|Y) = E(X^2|Y) - (E(X|Y))^2$이므로 조건부분산의 기댓값은 다음과 같다.

$$\begin{aligned} E[Var(X|Y)] &= E[E(X^2|Y) - (E(X|Y))^2] \\ &= E(X^2) - E[E(X|Y)]^2 \end{aligned} \tag{4.14}$$

식 (4.13)과 (4.14)에 의하여 다음이 성립한다.

$$Var(E(X|Y)) + E(Var(X|Y)) = E(X^2) - [E(X)]^2 = Var(X)$$

■

위 정리의 의미를 알아보기 위하여 다시 n_1명, n_2명, n_3명으로 이루어진 세 반에서 임의로 선택된 학생의 키에 대한 예를 살펴보도록 하자. $E(X|Y)$는 각 반의 평균이므로 $Var[E(X|Y)]$는 반 평균 간의 분산을 나타낸다. 또 $Var(X|Y)$는 각 반에서 학생 키의 분산을 나타낸다. 따라서 $E[Var(X|Y)]$는 각 반의 분산을 평균한 값이다. 이를 좀더 자세히 알아보기 위해 i반에 소속된 j 번째 학생의 키를 a_{ij}라 하자. 전체 학생의 수를 n이라 하고 i반의 학생의 수를 n_i라 할 때 각

반의 평균과 분산은 다음과 같다.

$$
\begin{aligned}
E(X|Y=i) &= \frac{1}{n_i}\sum_{j=1}^{n_i} a_{ij} \\
Var(X|Y=i) &= \frac{1}{n_i}\sum_{j=1}^{n_i}(a_{ij}-\overline{a}_i)^2
\end{aligned}
$$

위에서 $\overline{a}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} a_{ij}$ 이다. 전체 학생 키의 평균을 $\overline{a} = \frac{1}{n}\sum_{i=1}^{3}\sum_{j=1}^{n_i} a_{ij}$ 라 할 때 반 평균 간의 분산 $Var[E(X|Y)]$ 은 다음과 같다.

$$
Var[E(X|Y)] = \frac{1}{n}\sum_{i=1}^{3} n_i(\overline{a}_i - \overline{a})^2
$$

각 반의 인원이 n_i 이므로 반 평균의 분산을 구할 때 n_i 를 가중값으로 하여 분산을 구하여야 한다. 또 각 반의 분산의 평균은 다음과 같다.

$$
\begin{aligned}
E[Var(X|Y)] &= \sum_{i=1}^{3} Var(X|Y=i)P(Y=i) \\
&= \sum_{i=1}^{3}\left(\frac{1}{n_i}\sum_{j=1}^{n_i}(a_{ij}-\overline{a}_i)^2\right)\frac{n_i}{n} = \frac{1}{n}\sum_{i=1}^{3}\sum_{j=1}^{n_i}(a_{ij}-\overline{a}_i)^2
\end{aligned}
$$

따라서 다음 결과를 얻을 수 있다.

$$
Var[E(X|Y)] + E[Var(X|Y)] = \frac{1}{n}\sum_{i=1}^{3} n_i(\overline{a}_i - \overline{a})^2 + \frac{1}{n}\sum_{i=1}^{3}\sum_{j=1}^{n_i}(a_{ij}-\overline{a}_i)^2
$$

한편 임의로 선택된 학생의 분산은 다음과 같고

$$
Var(X) = \frac{1}{n}\sum_{i=1}^{3}\sum_{j=1}^{n_i}(a_{ij}-\overline{a})^2
$$

위의 식은 다음과 같이 나타낼 수 있다.

$$
\frac{1}{n}\sum_{i=1}^{3}\sum_{j=1}^{n_i}(a_{ij}-\overline{a})^2 = \frac{1}{n}\sum_{i=1}^{3} n_i(\overline{a}_i - \overline{a})^2 + \frac{1}{n}\sum_{i=1}^{3}\sum_{j=1}^{n_i}(a_{ij}-\overline{a}_i)^2 \tag{4.15}
$$

따라서 다음 식이 성립함을 확인할 수 있다.

$$
Var(X) = Var[E(X|Y)] + E[Var(X|Y)]
$$

식 (4.15) 는 일원배치 실험계획법에서 다음을 나타내는 식이다.

$$\begin{aligned}&\text{총변동 (total variation)}\\&\quad = \text{계급간변동 (between class variation)} + \text{계급내변동 (within class variation)}\end{aligned}$$

예 4.14 예 4.9에서 주어진 확률분포에 대하여 다음 물음에 답하여라.

(a) $Var[E(Y|X)]$ 를 구하여라.

(b) $E[Var(Y|X)]$ 를 구하여라.

(c) 위에서 구한 두 값의 합이 $Var(Y)$ 와 같음을 확인하여라.

[풀이]

(a) 예 4.9에서 구한 $E(Y|X)$ 의 값을 이용하여 $(E(Y|X))^2$ 의 기댓값을 구하면 다음과 같고

$$\begin{aligned}E[E(Y|X)]^2 &= [E(Y|X=0)]^2 \times P(X=0) + [E(Y|X=1)]^2 \times P(X=1)\\&= 2^2 \times \frac{5}{12} + \left(\frac{15}{7}\right)^2 \times \frac{7}{12} \;=\; \frac{365}{84}\end{aligned}$$

$E(Y|X)$ 의 기댓값은 다음과 같다.

$$E[E(Y|X)] \;=\; E(Y) \;=\; \frac{25}{12}$$

따라서 조건부기댓값의 분산은 다음과 같다.

$$Var[E(Y|X)] \;=\; E[E(Y|X)]^2 - [E(E(Y|X))]^2 = \frac{5}{1008}$$

(b) 조건부분산 $Var(Y|X)$ 의 기댓값은 다음과 같다.

$$\begin{aligned}E[Var(Y|X)] &= Var(Y|X=0) \times P(X=0) + Var(Y|X=1) \times P(X=1)\\&= \frac{2}{5} \times \frac{5}{12} + \frac{34}{49} \times \frac{7}{12} \;=\; \frac{4}{7}\end{aligned}$$

(c) Y 의 주변분포를 이용하면 $Var(Y) = \frac{83}{144}$ 임을 알 수 있다. 이 값은 (a) 와 (b) 에서 구한 값의 합임을 확인할 수 있다.

■

앞의 정리 4.9와 4.10를 활용하여 랜덤합(random sum)의 기댓값과 분산을 구할 수 있다. 더하는 항의 수가 랜덤인 경우를 랜덤합이라 한다. 랜덤합의 한 가지 예로 다음 경우를 생각할 수 있다. 수리공이 i 번째 고장이 난 부품을 수리하는데 걸리는 시간을 X_i 라 하고 고장 난 부품의 수를 N 이라 하자. 이 경우 X_i 와 N 모두 랜덤이다. 이 때 고장 난 부품을 수리하는데 걸리는 총 시간을 T 라 하면 $T = \sum_{i=1}^{N} X_i$ 로 나타낼 수 있다. X_i 는 모두 독립이고 같은 분포를 가지고 있으며 N 과 독립일 때 T 의 기댓값과 분산을 구해보도록 하자.

$N = n$ 으로 주어지면 $T = \sum_{i=1}^{n} X_i$ 이 되어 고정된 항의 합이 된다. 따라서 $N = n$ 일 때 T 의 조건부기댓값은 다음과 같다.

$$E(T|N=n) \;=\; E\Big(\sum_{i=1}^{N} X_i \Big| N=n\Big) \;=\; E\Big(\sum_{i=1}^{n} X_i\Big) \;=\; nE(X)$$

위에서 두 번째 등식은 X_i 와 N 이 독립이기 때문에 성립한다. 위 식으로부터 $E(T|N) = NE(X)$ 임을 알 수 있다. 정리 4.9에 의하여 T 의 기댓값은 다음과 같다.

$$E(T) \;=\; E[E(T|N)] \;=\; E[NE(X)] = E(X)E(N)$$

한편 $N = n$ 일 때 T 의 조건부분산은 다음과 같으므로

$$Var(T|N=n) \;=\; Var\Big(\sum_{i=1}^{N} X_i \Big| N=n\Big) \;=\; Var\Big(\sum_{i=1}^{n} X_i\Big) \;=\; nVar(X)$$

T 의 분산은 다음과 같다.

$$\begin{aligned} Var(T) &= Var[E(T|N)] + E[Var(T|N)] \\ &= Var[NE(X)] + E[NVar(X)] = [E(X)]^2 Var(N) + Var(X)E(N) \end{aligned}$$

예 4.15 어느 상점에 하루 동안 방문하는 손님의 수는 Poisson(5) 를 따른다고 한다. 또 손님 한 명이 구매하는 금액의 기댓값과 표준편차는 각각 10만원과 3만원이라고 한다. 손님의 수와 손님이 구매하는 금액은 서로 독립일 때, 이 상점 하루 판매액의 기댓값과 분산을 구하여라.

[풀이] 하루 동안 방문하는 손님의 수를 N, i 번째 손님이 구매하는 금액을 X_i 라 할 때 이 상점의 하루 판매액은 $T = \sum_{i=1}^{N} X_i$ 로 나타낼 수 있다. 따라서 하루 판매액의 기댓값과 분산은 다음과 같다.

$$\begin{aligned} E(T) &= E(X)E(N) \;=\; 10 \cdot 5 \;=\; 50 \\ Var(T) &= [E(X)]^2 Var(N) + Var(X)E(N) \;=\; 10^2 \cdot 5 + 9 \cdot 5 \;=\; 545 \end{aligned}$$

■

4.7 적률생성함수

확률변수 X에 대하여 e^{tX}의 기댓값을 **적률생성함수**(moment generating function)라 한다. 적률생성함수는 t의 값에 따라 달라지므로 t의 함수가 된다.

정의 4.6 확률변수 X에 대하여 e^{tX}의 기댓값을 적률생성함수라 하며 $M(t)$로 나타낸다. 즉 적률생성함수는 다음과 같다.

$$M(t) = E(e^{tX})$$

적률생성함수가 가지는 한 가지 특징은 확률분포를 결정지어준다는 것이다. 즉 서로 다른 확률분포는 같은 적률생성함수를 가질 수 없다. 따라서 적률생성함수를 보면 어떤 확률분포인지 알 수 있게 된다. 적률생성함수의 이러한 특징을 이용하여 확률분포를 구하는데 이용되기도 한다.

일반적으로 모든 확률분포에 대하여 적률생성함수가 존재하는 것은 아니다. 그러나 이 책에서 다루는 대부분의 확률분포에 대해서는 적률생성함수가 존재하며 다음과 같은 성질을 가지고 있다.

(i) $t = 0$일 때 $e^{tX} = 1$이므로 $M(0) = 1$이다.

(ii) 적률생성함수를 미분하면 다음과 같으므로

$$\begin{aligned} M'(t) &= \frac{d}{dt}\int_{-\infty}^{\infty} e^{tx}f(x)dx = \int_{-\infty}^{\infty} \frac{d}{dt}(e^{tx}f(x))dx \\ &= \int_{-\infty}^{\infty} xe^{tx}f(x)dx \end{aligned}$$

$M'(0) = E(X)$이다.

(iii) 적률생성함수를 k 번 미분하면 다음과 같으므로

$$\begin{aligned} M^{(k)}(t) &= \frac{d^k}{dt^k}\int_{-\infty}^{\infty} e^{tx}f(x)dx = \int_{-\infty}^{\infty} \frac{d^k}{dt^k}(e^{tx}f(x))dx \\ &= \int_{-\infty}^{\infty} x^k e^{tx}f(x)dx \end{aligned}$$

$M^{(k)}(0) = E(X^k)$이다.

$E(X^k)$를 확률변수 X의 k차 적률(moment)이라 한다. 위의 성질에 의하여 X의 모든 적률은 $M(t)$로부터 구할 수 있다. 이러한 이유로 $M(t)$를 적률생성함수라 한다. 이처럼 적률생성함수는 확률분포의 모든 적률을 결정하고 나아가서 확률분포까지 결정하게 된다. 즉 적률생성함수를 알면 확률분포를 알 수 있게 된다.

여러 확률분포의 적률생성함수를 구하여 보자. X가 성공의 확률이 p인 베르누이분포를 따르는 확률변수일 때 X의 적률생성함수는 다음과 같다.

$$M(t) \;=\; E(e^{tX}) \;=\; (1-p)+pe^t$$

이항분포를 따르는 확률변수 X의 적률생성함수는 다음과 같이 나타낼 수 있다.

$$\begin{aligned} M(t) \;&=\; E(e^{tX}) \;=\; \sum_{k=0}^{n} e^{tk}\frac{n!}{k!(n-k)!}\,p^k(1-p)^{n-k} \\ &=\; \sum_{k=0}^{n}\frac{n!}{k!(n-k)!}\,(pe^t)^k(1-p)^{n-k} \end{aligned}$$

이 합은 이항정리에 의하여 $(1-p+pe^t)^n$이다. 즉 이항분포의 적률생성함수는 다음과 같다.

$$M(t) \;=\; (1-p+pe^t)^n$$

이 적률생성함수의 1차 미분과 2차 미분을 구하면 다음과 같으므로

$$\begin{aligned} M'(t) \;&=\; n(1-p+pe^t)^{n-1}pe^t \\ M''(t) \;&=\; n(n-1)(1-p+pe^t)^{n-2}(pe^t)^2+n(1-p+pe^t)^{n-1}pe^t \end{aligned}$$

X와 X^2의 기댓값은 다음과 같음을 알 수 있다.

$$\begin{aligned} E(X) \;&=\; M'(0) \;=\; np \\ E(X^2) \;&=\; M''(0) \;=\; n(n-1)p^2+np \end{aligned}$$

포아송분포의 적률생성함수는 다음과 같이 나타낼 수 있다.

$$M(t) \;=\; \sum_{k=0}^{\infty} e^{tk}\frac{\lambda^k}{k!}e^{-\lambda} \;=\; e^{-\lambda}\sum_{k=0}^{\infty}\frac{(\lambda e^t)^k}{k!}$$

실수 a에 대하여 $\sum_{k=0}^{\infty}\frac{a^k}{k!} \;=\; e^a$이므로 포아송분포의 적률생성함수는 다음과 같다.

$$M(t) \;=\; e^{-\lambda}e^{\lambda e^t} \;=\; e^{\lambda(e^t-1)} \tag{4.16}$$

연속형 확률분포인 감마분포의 적률생성함수는 다음과 같이 나타낼 수 있다.

$$\begin{aligned} M(t) \;&=\int_0^\infty e^{tx}\frac{\lambda^\alpha}{\Gamma(\alpha)}\,x^{\alpha-1}e^{-\lambda x}\,dx \;=\int_0^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)}\,x^{\alpha-1}e^{-(\lambda-t)x}\,dx \\ &=\left(\frac{\lambda}{\lambda-t}\right)^\alpha\int_0^\infty\frac{(\lambda-t)^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-(\lambda-t)x}\,dx \end{aligned}$$

$t < \lambda$ 일 때 적분 안의 함수는 Gamma$(\alpha, \lambda - t)$의 확률밀도함수이므로 적분은 1이 된다. 따라서 감마분포의 적률생성함수는 다음과 같다.

$$M(t) \;=\; \left(\frac{\lambda}{\lambda - t}\right)^{\alpha}, \quad \text{단}, \; t < \lambda \tag{4.17}$$

표준정규분포의 적률생성함수는 다음과 같이 나타낼 수 있다.

$$\begin{aligned} M(t) \;&=\; \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx \;=\; \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-t)^2 + \frac{1}{2}t^2}\, dx \\ &=\; e^{\frac{1}{2}t^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-t)^2}\, dx \end{aligned}$$

위 식에서 적분 안의 함수는 N$(t,\, 1)$의 확률밀도함수이므로 적분은 1이 된다. 따라서 정규분포의 적률생성함수는 다음과 같다.

$$M(t) \;=\; e^{\frac{1}{2}t^2}$$

적률생성함수는 다음과 같은 성질을 가지고 있음을 보일 수 있다.

정리 4.11 (i) X의 적률생성함수가 $M_X(t)$ 일 때 $Y = a + bX$의 적률생성함수는 $M_Y(t) = e^{at}M_X(bt)$이다.

(ii) 확률변수 X와 Y가 서로 독립이고 적률생성함수가 각각 $M_X(t)$와 $M_Y(t)$일 때 $Z = X + Y$의 적률생성함수는 $M_Z(t) = M_X(t)M_Y(t)$이다.

[증명]

(i) $Y = a + bX$의 적률생성함수는 다음과 같이 구할 수 있다.

$$M_Y(t) \;=\; E(e^{tY}) \;=\; E(e^{at+btX}) \;=\; e^{at}E(e^{btX}) \;=\; e^{at}M_X(bt)$$

(ii) X와 Y의 적률생성함수가 $M_X(t)$와 $M_Y(t)$일 때 $Z = X + Y$의 적률생성함수는 다음과 같이 나타낼 수 있다.

$$M_Z(t) \;=\; E(e^{tZ}) \;=\; E(e^{tX+tY})$$

X와 Y는 서로 독립이므로 정리 4.3에 의하여 $E(e^{tX}e^{tY}) = E(e^{tX})E(e^{tY})$ 이다. 따라서 Z의 적률생성함수는 다음과 같다.

$$M_Z(t) \;=\; E(e^{tX})\, E(e^{tY}) \;=\; M_X(t)M_Y(t)$$

■

예 4.16 $\mathrm{N}(\mu, \sigma^2)$를 따르는 확률변수 X의 적률생성함수를 구하여라.

[풀이] 표준정규분포를 따르는 확률변수 Z라 할 때 $X = \mu + \sigma Z$로 나타낼 수 있다. Z의 적률생성함수는 $M_Z(t) = e^{\frac{t^2}{2}}$이므로 X의 적률생성함수는 다음과 같다.

$$M_X(t) \;=\; e^{\mu t} M_Z(\sigma t) \;=\; e^{\mu t + \frac{1}{2}\sigma^2 t^2} \tag{4.18}$$

■

적률생성함수를 이용하면 다음 예와 같이 서로 독립인 확률변수 X와 Y의 합의 확률분포를 구할 수 있다.

예 4.17 서로 독립인 두 확률변수 X와 Y에 대하여 적률생성함수를 이용하여 $Z = X + Y$의 확률분포를 구하여라.

(a) $X \sim \mathrm{Poisson}(\lambda_1)$, $Y \sim \mathrm{Poisson}(\lambda_2)$일 때 $Z = X + Y$의 확률분포를 구하여라.

(b) $X \sim \mathrm{Gamma}(\alpha_1, \lambda)$, $Y \sim \mathrm{Gamma}(\alpha_2, \lambda)$일 때 $Z = X + Y$의 확률분포를 구하여라.

(c) $X \sim \mathrm{N}(\mu_1, \sigma_1^2)$, $Y \sim \mathrm{N}(\mu_2, \sigma_2^2)$일 때 $Z = X + Y$의 확률분포를 구하여라.

[풀이]

(a) 기댓값이 λ인 포아송분포의 적률생성함수는 $M(t) = e^{\lambda(e^t-1)}$이므로 X와 Y의 적률생성함수는 $M_X(t) = e^{\lambda_1(e^t-1)}$와 $M_Y(t) = e^{\lambda_2(e^t-1)}$이다. 정리 4.11에 의하여 $Z = X+Y$의 적률생성함수는 다음과 같다.

$$\begin{aligned} M_Z(t) &= M_X(t) M_Y(t) \;=\; e^{\lambda_1(e^t-1)} e^{\lambda_2(e^t-1)} \\ &= e^{(\lambda_1+\lambda_2)(e^t-1)} \end{aligned}$$

이를 포아송분포의 적률생성함수를 나타낸 식 (4.16)과 비교할 때 λ 대신에 $\lambda_1 + \lambda_2$를 사용한 것 외에는 같은 형태임을 알 수 있다. 따라서 $M_Z(t)$는 기댓값이 $\lambda_1 + \lambda_2$인 포아송분포의 적률생성함수이므로 $X + Y \sim \mathrm{Poisson}(\lambda_1 + \lambda_2)$이다.

(b) Gamma(α, λ)의 적률생성함수는 $\left(\frac{\lambda}{\lambda-t}\right)^{\alpha}$ 이므로 $Z = X + Y$의 적률생성함수는 다음과 같다.

$$\begin{aligned} M_Z(t) &= M_X(t)M_Y(t) = \left(\frac{\lambda}{\lambda-t}\right)^{\alpha_1}\left(\frac{\lambda}{\lambda-t}\right)^{\alpha_2} \\ &= \left(\frac{\lambda}{\lambda-t}\right)^{\alpha_1+\alpha_2} \end{aligned}$$

이를 감마분포의 적률생성함수를 나타낸 식 (4.17)과 비교할 때 $M_Z(t)$는 $\alpha = \alpha_1 + \alpha_2$인 감마분포의 적률생성함수임을 알 수 있다. 따라서 $X + Y \sim$ Gamma($\alpha_1 + \alpha_2$, λ)이다.

(c) N(μ, σ^2)의 적률생성함수는 $e^{\mu t+\frac{1}{2}\sigma^2 t^2}$ 이므로 $Z = X + Y$의 적률생성함수는 다음과 같다.

$$\begin{aligned} M_Z(t) &= M_X(t)M_Y(t) = e^{\mu_1 t+\frac{1}{2}\sigma_1^2 t^2} e^{\mu_2 t+\frac{1}{2}\sigma_2^2 t^2} \\ &= e^{(\mu_1+\mu_2)t+\frac{1}{2}(\sigma_1^2+\sigma_2^2)t^2} \end{aligned}$$

이를 정규분포의 적률생성함수를 나타낸 식 (4.18)과 비교할 때 $M_Z(t)$는 $\mu = \mu_1 + \mu_2$, $\sigma^2 = \sigma_1^2 + \sigma_2^2$인 정규분포의 적률생성함수임을 알 수 있다. 따라서 $X + Y \sim$ N($\mu_1 + \mu_2$, $\sigma_1^2 + \sigma_2^2$)이다.

■

이 장에서 다룬 여러 확률분포의 기댓값과 분산, 적률생성함수를 정리하면 표 4.1와 같다.

표 4.1: 여러 확률분포의 기댓값, 분산과 적률생성함수

확률분포	기댓값	분산	적률생성함수
Bernoullli(p)	p	$p(1-p)$	$(1-p+pe^t)$
B(n,p)	np	$np(1-p)$	$(1-p+pe^t)^n$
Poisson(λ)	λ	λ	$e^{\lambda(e^t-1)}$
Uniform(a,b)	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{1}{(b-a)t}(e^{bt}-e^{at})$
Exp(λ)	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{\lambda}{\lambda-t}$
Gamma(α,λ)	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$	$\left(\frac{\lambda}{\lambda-t}\right)^{\alpha}$
$N(\mu,\sigma^2)$	μ	σ^2	$e^{\mu t+\frac{1}{2}\sigma^2 t^2}$

연습문제

1. $N \sim \text{Poisson}(\lambda)$ 일 때 $E\left(\frac{1}{(N+1)(N+2)}\right)$ 을 구하여라.

2. $X \sim \text{Gamma}(\alpha,\, \lambda)$ 일 때 $E(X^3)$ 을 구하여라.

3. $X \sim \text{Beta}(\alpha,\, \beta)$ 일 때 $E[X^3(1-X)^2]$ 을 구하여라.

4. X 와 Y 는 다음과 같은 결합분포를 가지고 있다.

		y		
		10	20	30
	0	$\frac{1}{9}$	$\frac{1}{9}$	0
x	1	$\frac{1}{6}$	0	$\frac{1}{6}$
	2	0	$\frac{2}{9}$	$\frac{2}{9}$

a) X 의 주변분포를 이용하여 $E(X)$ 를 구하여라.
b) $Y = 10, 20, 30$ 일 때 각각 X 의 조건부기댓값을 구하여라.
c) 위 결과를 이용하여 $E(E(X|Y))$ 를 구하고 (a) 에서 구한 값과 같은지 확인하여라.
d) $Y = 10$ 일 때 X 의 조건부분산을 구하여라.

5. X 와 Y 는 다음과 같은 결합분포를 가지고 있다.

		y			
		0	1	2	3
x	0	0.1	0.2	0.3	0.0
	1	0.0	0.2	0.1	0.1

a) 각 X 의 값에서 $E(Y|X)$ 와 $Var(Y|X)$ 를 구하여라.
b) 위 결과를 이용하여 $E(E(Y|X))$ 를 구하여라.
c) $Var(E(Y|X))$ 를 구하여라.
d) $E(Var(Y|X))$ 를 구하여라.

6. 두 연속형 확률변수 X 와 Y 의 결합확률밀도함수가 다음과 같다고 한다.

$$f(x,y) \;=\; \frac{\lambda^2}{2}e^{-\frac{\lambda}{2}(x+y)}, \quad 0 < x < y < \infty$$

a) $E(Y)$를 구하여라.

b) $X = x$일 때 Y의 조건부기댓값을 구하여라.

c) 위의 결과를 이용하여 $E(E(Y|X))$를 구하여라.

d) $X = x$일 때 Y의 조건부분산을 구하여라.

7. 두 연속형 확률변수 X와 Y의 결합확률밀도함수가 다음과 같다고 한다.

$$f(x, y) \ = e^{-y}, \quad 0 < x < y < \infty$$

a) $E(X|Y)$와 $Var(X|Y)$를 구하여라.

b) 위 결과를 이용하여 $E(E(X|Y))$를 구하여라.

c) $Var(E(X|Y))$를 구하여라.

d) $E(Var(X|Y))$를 구하여라.

8. X와 Y가 서로 독립일 때 $E(X|Y) = E(X)$임을 보여라.

9. $X \sim \text{Exp}(\lambda)$이며 $X = x$일 때 U의 조건부분포는 Uniform(0, x)이다. 이때 U의 기댓값과 분산을 구하여라.

10. $N \sim \text{Poisson}(\lambda)$이며 $N = n$일 때 X의 조건부분포는 Binomial(n, p)이다. 이때 X의 기댓값과 분산을 구하여라.

11. $X \sim \text{Uniform}(0, 1)$이며 $X = x$일 때 Y의 조건부분포는 $\text{N}(x, x^2)$이다. 이때 $Cov(X, Y)$의 값을 구하여라.

12. $X \sim \text{N}(5,\ 2^2)$일 때 델타 방법을 사용하여 $Y = X^2$의 근사적 기댓값과 분산을 구하여라.

13. $X \sim \text{Exp}(2)$일 때 델타 방법을 사용하여 $Y = \frac{1}{X}$의 근사적 기댓값과 분산을 구하여라.

14. 확률변수 X와 Y는 서로 독립이며 각각의 기댓값이 μ_X, μ_Y이고 분산이 σ_X^2, σ_Y^2일 때 XY의 분산을 구하여라.

15. 확률변수 X, Y, Z는 서로 독립이며 각각의 기댓값이 μ_X, μ_Y, μ_Z이고 분산이 σ_X^2, σ_Y^2, σ_Z^2일 때 XYZ의 분산을 구하여라.

16. $N \sim \text{Poisson}(\lambda)$일 때 $Y = X_1 \cdot X_2 \cdots X_N$의 기댓값을 구하여라. 단, $N = 0$인 경우에 $Y = 1$로 정의하며 X_i의 기댓값은 μ이고 서로 독립이며 또한 N과 서로 독립이다.

17. 확률변수 X의 적률생성함수가 $M(t) = \frac{1}{4}e^t + \frac{3}{4}e^{2t}$일 때 X의 기댓값과 분산을 구하여라.

18. $X \sim \text{Gamma}(\alpha, \lambda)$일 때 $Y = cX$의 적률생성함수를 구하고 이를 이용하여 Y의 확률분포를 밝혀라. 단, c는 양의 상수이다.

19. X_1과 X_2는 서로 독립이며 다음과 같은 분포를 가지는 확률변수이다.

$$X_i = \begin{cases} 0, & p_0 \text{의 확률로} \\ 1, & p_1 \text{의 확률로} \\ 2, & p_2 \text{의 확률로} \end{cases} \qquad (\text{단}, \ p_0 + p_1 + p_2 = 1)$$

a) X_1의 적률생성함수를 구하여라.

b) $Y = X_1 + X_2$의 적률생성함수를 구하여라.

c) $Y = X_1 + X_2$의 적률생성함수를 이용하여 $E(Y^2)$을 구하여라.

제 5 장

랜덤샘플

확률변수 $X_1, X_2, \ldots, X_n$ 이 서로 독립이며 같은 분포를 가질 때 이를 **랜덤샘플**(random sample)이라 한다. 이 장에서는 랜덤샘플의 여러 성질에 대해 알아보고자 한다.

5.1 표본평균과 표본분산

통계학의 주요한 목표 중의 하나는 자료인 랜덤샘플을 이용하여 모집단을 규명하고자 하는 것이다. 모집단의 특성을 나타내는 대표적인 값으로 중심 위치와 퍼진 정도를 나타내는 모평균과 모분산이 있다. 랜덤샘플에 대해서 이 두 값에 대응되는 것이 **표본평균**(sample mean)과 **표본분산**(sample variance)이다. 랜덤샘플 $X_1, X_2, \ldots, X_n$ 에 대하여 표본평균 $\overline{X}$ 와 표본분산 S^2 은 다음과 같다.

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\,X_i - \overline{X}\,)^2$$

다음 정리는 모집단의 확률분포와 상관없이 표본평균과 표본분산의 기댓값은 모평균과 모분산임을 보여주고 있다.

정리 5.1 $X_1, X_2, \ldots, X_n$ 이 $E(X_i) = \mu$, $Var(X_i) = \sigma^2$ 인 랜덤샘플일 때 다음이 성립한다.

(i) $E(\overline{X}\,) = \mu$, $Var(\overline{X}) = \frac{\sigma^2}{n}$

(ii) $E(S^2) = \sigma^2$

[증명]

(i) 기댓값의 성질에 의하여 표본평균의 기댓값은 다음과 같다.

$$E(\,\overline{X}\,) \;=\; E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \;=\; \frac{1}{n}\sum_{i=1}^{n} E(X_i) \;=\; \mu$$

한편 $X_1, X_2, \ldots, X_n$ 은 서로 독립이므로 표본평균의 분산은 다음과 같다.

$$Var(\ \overline{X}\ ) \ = \frac{1}{n^2} Var\left(\sum_{i=1}^{n} X_i\right) \ = \ \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{\sigma^2}{n}$$

(ii) 표본분산은 다음과 같이 나타낼 수 있으므로

$$S^2 \ = \ \frac{1}{n-1}\sum_{i=1}^{n}(\ X_i - \overline{X}\ )^2 \ = \ \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right)$$

S^2 의 기댓값은 다음과 같이 나타낼 수 있다.

$$E(S^2) \ = \ \frac{1}{n-1}\left(\sum_{i=1}^{n} E(X_i^2) - nE(\overline{X}^2)\right)$$

한편 $E(X_i^2) = Var(X_i) + (E(X_i))^2 = \sigma^2 + \mu^2$ 이고 $E(\overline{X}^2) = Var(\overline{X}) + (E(\overline{X}))^2 = \frac{\sigma^2}{n} + \mu^2$ 이므로 표본분산의 기댓값은 다음과 같다.

$$\begin{aligned} E(S^2) \ &= \ \frac{1}{n-1}\left(\sum_{i=1}^{n}(\mu^2+\sigma^2) - n\left(\mu^2 + \frac{\sigma^2}{n}\right)\right) \\ &= \ \sigma^2 \end{aligned}$$

■

5.2 카이제곱분포, t-분포, F-분포

정규분포를 따르는 랜덤샘플의 성질을 규명하기 위해 정규분포로부터 유도되는 몇 가지 확률분포를 알아보도록 하자.

5.2.1 카이제곱분포

카이제곱분포는 감마분포의 특별한 경우이다. 확률밀도함수가 (2.8) 로 주어지는 감마분포에서 $\alpha = \frac{n}{2}$(n 은 양의 정수) 이고 $\lambda = \frac{1}{2}$ 인 확률분포를 자유도 n 인 카이제곱분포라 한다.

정의 5.1 $\alpha = \frac{n}{2}$, $\lambda = \frac{1}{2}$ 인 감마분포를 자유도(degree of freedom) n 인 카이제곱분포라 하며 χ_n^2 로 나타낸다.

자유도 n 인 카이제곱분포는 Gamma$\left(\frac{n}{2}, \frac{1}{2}\right)$ 이므로 $X \sim \chi_n^2$ 일 때 X 의 확률밀도함수는 다음과 같다.

$$f(x) = \frac{1}{\Gamma\left(\frac{n}{2}\right)}\left(\frac{1}{2}\right)^{\frac{n}{2}} x^{\frac{n}{2}-1}e^{-\frac{x}{2}}, \quad x \geq 0$$

감마분포의 기댓값과 분산으로부터 자유도 n 인 카이제곱분포의 기댓값과 분산은 n 과 $2n$ 임을 알 수 있다. 또한 카이제곱분포의 적률생성함수는 $M(t) = (1-2t)^{-\frac{n}{2}}$ 이다.

정규분포와 카이제곱분포의 연관성은 예 2.16를 통하여 알 수 있다. $Z \sim \mathrm{N}(0,1)$ 일 때 $X = Z^2$ 의 확률밀도함수는 다음과 같음을 보였다.

$$f_X(x) = \frac{1}{\sqrt{\pi}}\left(\frac{1}{2}\right)^{\frac{1}{2}} x^{\frac{1}{2}-1}e^{-\frac{1}{2}x}, \quad y \geq 0$$

이는 자유도가 1 인 카이제곱분포의 확률밀도함수이다. 즉 $Z \sim \mathrm{N}(0,1)$ 일 때 $Z^2 \sim \chi_1^2$ 이다.

카이제곱분포가 가지는 한 가지 특징은 서로 독립인 카이제곱분포의 합은 다시 카이제곱분포가 된다는 사실이다. 예 4.17에서 $X_1 \sim \mathrm{Gamma}(\alpha, \lambda)$, $X_2 \sim \mathrm{Gamma}(\beta, \lambda)$ 이고 서로 독립일 때 $X_1 + X_2 \sim \mathrm{Gamma}(\alpha+\beta, \lambda)$ 이므로 다음 정리를 얻을 수 있다.

정리 5.2 $X_1 \sim \chi_m^2$, $X_2 \sim \chi_n^2$ 이고 서로 독립일 때 $X_1 + X_2 \sim \chi_{m+n}^2$ 이다.

다음 정리는 카이제곱분포를 따르는 확률변수의 차도 적절한 조건 하에서 카이제곱분포를 따른다는 것을 보여주고 있다.

정리 5.3 서로 독립인 확률변수 X_1 과 X_2 에 대하여 $Y = X_1 + X_2$ 라 하자. $m > n$ 인 자연수 m 과 n 에 대하여 $Y \sim \chi_m^2$, $X_1 \sim \chi_n^2$ 이면 $X_2 \sim \chi_{m-n}^2$ 이다.

[증명] X_1, X_2, Y 의 적률생성함수를 각각 $M_{X_1}(t)$, $M_{X_2}(t)$, $M_Y(t)$ 라 하자. X_1 과 X_2 는 독립이므로 Y 의 적률생성함수는 $M_Y(t) = M_{X_1}(t)M_{X_2}(t)$ 이다. Y 와 X_1 은 자유도가 m 과 n 인 카이제곱분포를 따르므로 $M_Y(t) = (1-2t)^{-\frac{m}{2}}$, $M_{X_1}(t) = (1-2t)^{-\frac{n}{2}}$ 이다. 따라서 X_2 의 적률생성함수는 $M_{X_2}(t) = (1-2t)^{-\frac{m-n}{2}}$ 이다. 이는 자유도가 $m-n$ 인 카이제곱분포의 적률생성함수이므로 $X_2 \sim \chi_{m-n}^2$ 임을 알 수 있다. ■

5.2.2 t-분포

t-분포는 표준정규분포와 유사한 분포를 가지는 확률분포이다. 표준정규분포와 같이 0을 중심으로 대칭적인 확률분포를 가지고 있다. 한 가지 차이점은 표준정규분포에 비하여 꼬리가 두껍다는 것이다.

정의 5.2 $Z \sim \mathrm{N}(0,1)$, $U \sim \chi^2_n$ 이며 서로 독립일 때

$$\frac{Z}{\sqrt{U/n}}$$

의 확률분포를 자유도 n 인 t-분포라 하며 t_n 로 나타낸다.

$X \sim t_n$ 일 때 X 의 확률밀도함수는 다음과 같다.

$$f(x) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad -\infty < x < \infty$$

이 확률밀도함수는 $x = 0$ 에 대칭이다. 그림 5.1은 표준정규분포와 자유도 2, 5, 20인 t-분포의 확률밀도함수를 나타낸 것이다. 자유도가 증가할수록 t-분포는 표준정규분포에 가까이 가는 것을 알 수 있다.

5.2.3 F-분포

정의 5.3 $U \sim \chi^2_m$, $V \sim \chi^2_n$ 이며 서로 독립일 때

$$\frac{U/m}{V/n}$$

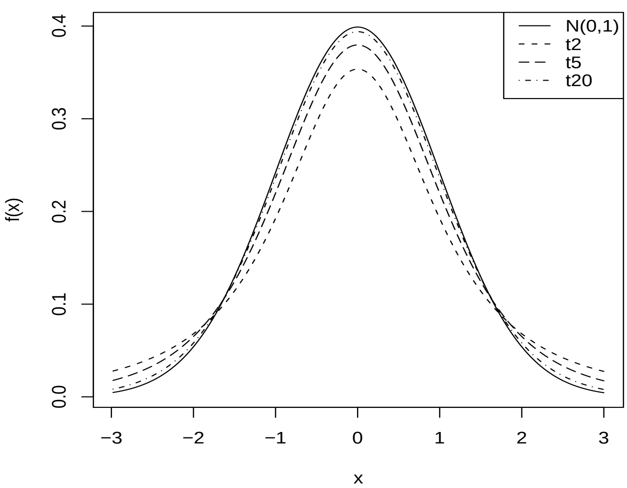

그림 5.1: t-분포의 확률밀도함수

의 확률분포를 자유도 $m,\,n$ 인 F-분포라 하며 $F_{m,\,n}$ 로 나타낸다.

$X \;\sim\; F_{m,\,n}$ 일 때 X 의 확률밀도함수는 다음과 같다.

$$f(x) = \frac{\Gamma((m+n)/2)}{\Gamma(m/2)\Gamma(n/2)} \left(\frac{m}{n}\right)^{m/2} x^{m/2-1} \left(1+\frac{m}{n}x\right)^{-(m+n)/2}, \quad x \geq 0$$

F-분포는 다음과 같은 성질을 가지고 있다.

- $X \;\sim\; t_n$ 일 때 $X = \frac{\mathrm{N}(0,1)}{\sqrt{\chi^2_n/n}}$ 로 나타낼 수 있으므로 X^2 의 확률분포는 다음과 같다.

$$X^2 \;=\; \frac{N^2(0,1)}{\chi^2_n/n} \;=\; \frac{\chi^2_1/1}{\chi^2_n/n} \;\sim\; F_{1,\,n}$$

  즉 자유도 n 인 t-분포를 따르는 확률변수를 제곱하면 자유도 $1,\,n$ 인 F-분포를 따른다.

- $X \sim\; F_{m,\,n}$ 일 때 $X = \frac{\chi^2_m/m}{\chi^2_n/n}$ 로 나타낼 수 있으므로 X^{-1} 의 확률분포는 다음과 같다.

$$X^{-1} \;=\; \frac{\chi^2_n/n}{\chi^2_m/m} \;\sim\; F_{n,\,m}$$

  즉 자유도 $m,\,n$ 인 F-분포를 따르는 확률변수의 역수는 자유도 $n,\,m$ 인 F-분포를 따른다.

5.3 정규 랜덤샘플

이 절에서는 정규분포에서 나온 랜덤샘플의 성질을 살펴보도록 하자.

다음 정리는 서로 독립이며 정규분포를 따르는 확률변수의 선형결합은 다시 정규분포를 따른다는 것을 보여주고 있다.

정리 5.4 $X_i \sim \mathrm{N}(\mu_i,\,\sigma_i^2)$ 이며 서로 독립일 때 $\sum_{i=1}^n a_iX_i$ 의 확률분포는 다음과 같다.

$$\sum_{i=1}^n a_iX_i \;\sim\; \mathrm{N}\left(\sum_{i=1}^n a_i\mu_i,\, \sum_{i=1}^n a_i^2\sigma_i^2\right)$$

[증명] 정규분포 $\mathrm{N}(\mu_i,\,\sigma_i^2)$ 을 따르는 확률변수 X_i 의 적률생성함수는 $e^{\mu_i t+\frac{1}{2}\sigma_i^2t^2}$ 이므로 $Y\;=$

$\sum_{i=1}^n a_iX_i$ 의 적률생성함수 $M_Y(t)$ 는 다음과 같다.

$$\begin{aligned} M_Y(t) &= E\Big(e^{t\sum_{i=1}^n a_iX_i}\Big) = E(e^{ta_1X_1})\,E(e^{ta_2X_2})\,\cdots\,E(e^{ta_nX_n}) \\ &= e^{a_1\mu_1t+\frac{1}{2}a_1^2\sigma_1^2t^2}\,e^{a_2\mu_2t+\frac{1}{2}a_2^2\sigma_2^2t^2}\,\cdots\,e^{a_n\mu_nt+\frac{1}{2}a_n^2\sigma_n^2t^2} \\ &= e^{(\sum_{i=1}^n a_i\mu_i)t+\frac{1}{2}(\sum_{i=1}^n a_i^2\sigma_i^2)t^2} \end{aligned}$$

이는 $\mu = \sum_{i=1}^n a_i\mu_i$ 이고 $\sigma^2 = \sum_{i=1}^n a_i^2\sigma_i^2$ 인 정규분포의 적률생성함수이므로 Y 의 확률분포는 다음과 같다.

$$Y \sim \mathrm{N}\Big(\sum_{i=1}^n a_i\mu_1,\ \sum_{i=1}^n a_i^2\sigma_i^2\Big)$$

■

$\overline{X}$ 는 $X_1, X_2, \ldots, X_n$ 의 선형결합이므로 정리 5.4에 의하여 표본평균의 확률분포는 정규분포임을 알 수 있다. 한편 $E(\overline{X}) = \mu$, $Var(\overline{X}) = \dfrac{\sigma^2}{n}$ 이므로 다음 정리를 얻을 수 있다.

정리 5.5 $X_1, X_2, \ldots, X_n$ 이 $\mathrm{N}(\mu, \sigma^2)$ 로부터의 랜덤샘플일 때 표본평균 $\overline{X}$ 는 기댓값 μ, 분산 $\dfrac{\sigma^2}{n}$ 인 정규분포를 따른다. 즉 표본평균의 확률분포는 다음과 같다.

$$\overline{X} \sim \mathrm{N}\left(\mu,\ \frac{\sigma^2}{n}\right)$$

위 정리는 표본평균의 확률분포를 말하여주고 있다. 다음으로 표본분산의 확률분포를 구해보도록 하자. 이를 위해 먼저 표본평균과 표본분산은 독립임을 보이도록 하자. 이변량정규분포를 따르는 두 확률변수에 대해서는 공분산이 0일 때 서로 독립인 것과 같이 정규분포를 따르는 여러 확률변수 사이에도 공분산이 0이면 서로 독립이다. $\overline{X}$ 와 $(X_k - \overline{X})$ 모두 $X_1, X_2, \ldots, X_n$ 의 선형결합이므로 정규분포를 따르며, 공분산을 구하면 다음과 같다.

$$\begin{aligned} Cov(\overline{X},\ X_k - \overline{X}) &= Cov\left(\frac{1}{n}\sum_{i=1}^n X_i,\ X_k - \frac{1}{n}\sum_{j=1}^n X_j\right) \\ &= \frac{1}{n}\sum_{i=1}^n Cov(X_i,\ X_k) - \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n Cov(X_i,\ X_j) \\ &= \frac{\sigma^2}{n} - \frac{n\sigma^2}{n^2} \ = \ 0 \end{aligned}$$

따라서 $\overline{X}$ 와 $(X_k - \overline{X})$ 는 서로 독립이다. 표본분산은 다음 통계량의 함수이므로

$$(X_1 - \overline{X}, X_2 - \overline{X}, \ldots, X_n - \overline{X})$$

$\overline{X}$ 와 S^2 도 서로 독립이다. 한편 $\sum_{i=1}^{n}(X_i-\mu)^2$ 을 다음과 같이 나타낼 수 있고

$$\sum_{i=1}^{n}(X_i-\mu)^2 \;=\; \sum_{i=1}^{n}(X_i-\overline{X})^2+n(\overline{X}-\mu)^2$$

다음이 성립하므로

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi_n^2, \quad \frac{n(\overline{X}-\mu)^2}{\sigma^2} \sim \chi_1^2$$

정리 5.3에 의하여 다음 결과를 얻을 수 있다.

정리 5.6 $X_1, X_2, \ldots, X_n$ 이 $\mathrm{N}(\mu,\sigma^2)$ 로부터의 랜덤샘플일 때 $\frac{(n-1)S^2}{\sigma^2}$ 의 확률분포는 다음과 같다.

$$\frac{(n-1)S^2}{\sigma^2} \;\sim\; \chi_{n-1}^2$$

표본평균의 확률분포는 $\mathrm{N}\left(\mu,\, \frac{\sigma^2}{n}\right)$ 이므로 다음 결과를 얻을 수 있다.

$$\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma} \;\sim\; \mathrm{N}(0,1)$$

위 식에서 모표준편차 σ 대신 표본표준편차 S 를 사용하면 t-분포임을 보일 수 있다.

정리 5.7 $X_1, X_2, \ldots, X_n$ 이 $\mathrm{N}(\mu,\sigma^2)$ 로부터의 랜덤샘플일 때 $\frac{\sqrt{n}(\overline{X}-\mu)}{S}$ 의 확률분포는 다음과 같다.

$$\frac{\sqrt{n}(\overline{X}-\mu)}{S} \;\sim\; t_{n-1}$$

[증명] $\dfrac{\sqrt{n}(\overline{X}-\mu)}{S}$ 의 분모와 분자를 모두 σ 로 나누면 다음과 같이 나타낼 수 있다.

$$\frac{\sqrt{n}(\overline{X}-\mu)}{S} \;=\; \frac{\sqrt{n}(\overline{X}-\mu)/\sigma}{\sqrt{(n-1)S^2/\sigma^2(n-1)}}$$

이때 $\dfrac{\sqrt{n}(\overline{X}-\mu)}{\sigma} \sim \mathrm{N}(0,1)$ 이며 $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$ 이므로 $\frac{\sqrt{n}(\overline{X}-\mu)}{S}$ 는 다음과 같이 나타낼

수 있다.

$$\frac{\sqrt{n}(\overline{X}-\mu)}{S} = \frac{\mathrm{N}(0,1)}{\sqrt{\chi^2_{n-1}/(n-1)}}$$

한편 $\overline{X}$와 S^2은 서로 독립이므로 $\dfrac{\sqrt{n}(\overline{X}-\mu)}{S} \sim t_{n-1}$이다. ■

5.4 순서통계량

$X_1, X_2, \ldots, X_n$이 랜덤샘플일 때 이들을 작은 값부터 차례대로 늘어놓은 통계량을 **순서통계량**(order statistic)이라 한다. k번째로 작은 통계량을 $X_{(k)}$라 할 때 순서통계량은 다음을 만족한다.

$$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$$

순서통계량의 확률분포를 구하여 보자. 비록 $X_{(k)}$는 $X_1, X_2, \ldots, X_n$ 중의 하나의 값을 가지지만 $X_{(k)}$의 확률분포가 X_i의 확률분포와 같지는 않다. $X_{(k)}$의 확률분포는 $X_1, X_2, \ldots,$ X_n 모두의 영향을 받기 때문이다. 예를 들어 X_1, X_2, X_3이 1부터 5까지의 정수 값을 가지는 이산형 확률분포의 랜덤샘플일 때 $X_{(1)} = 2$, $X_{(2)} = 3$, $X_{(3)} = 5$일 확률을 구하여 보자. $X_{(1)} = 2$, $X_{(2)} = 3$, $X_{(3)} = 5$인 사건은 X_1, X_2, X_3 중에서 가장 작은 값은 2, 가장 큰 값은 5, 나머지 하나는 3이라는 것을 나타내며 구체적으로 X_1, X_2, X_3가 2, 3, 5 중에서 어떤 값을 가지는지 나타내지는 않는다. $X_1 = 2$, $X_2 = 3$, $X_3 = 5$일 수도 있으며 $X_1 = 3$, $X_2 = 2$, $X_3 = 5$일 수도 있다. X_1, X_2, X_3이 2, 3, 5 중에서 하나씩 값을 가질 경우의 수는 $3!$이며 X_1, X_2, X_3이 랜덤샘플이므로 각 경우의 확률은 같다. 따라서 $X_{(1)} = 2$, $X_{(2)} = 3$, $X_{(3)} = 5$일 확률은 다음과 같다.

$$P(X_{(1)} = 2, X_{(2)} = 3, X_{(3)} = 5) = 3!P(X_1 = 2, X_2 = 3, X_3 = 5)$$

또 $X_{(3)} \le 3$일 확률은 X_1, X_2, X_3 모두 3 이하라는 사건과 같으므로 다음과 같이 나타낼 수 있다.

$$P(X_{(3)} \le 3) = P(X_1 \le 3,\ X_2 \le 3,\ X_3 \le 3) = \Big(P(X_1 \le 3)\Big)^3$$

이처럼 순서통계량의 확률분포는 $X_1, X_2, \ldots, X_n$ 모두의 영향을 받기 때문에 X_i의 확률분포와 같지는 않다.

$X_1, X_2, \ldots, X_n$이 확률밀도함수 $f(x)$, 누적분포함수 $F(x)$를 가지는 확률분포로부터의 랜덤샘플일 때, 순서통계량 $X_{(k)}$의 확률분포를 구하여 보자. 간단히 구할 수 있는 경우가 최댓값과 최솟값이다. $X_{(n)}$은 $X_1, X_2, \ldots, X_n$ 중에서 최댓값이다. 사건 $\{X_{(n)} \le x\}$는 $X_1, X_2, \ldots, X_n$

모두 x 이하라는 것을 의미한다. 따라서 $X_{(n)}$ 의 누적분포함수 $F_{X_{(n)}}(x)$ 는 다음과 같다.

$$\begin{aligned} F_{X_{(n)}}(x) &= P(X_{(n)} \le x) \;=\; P(X_1 \le x, X_2 \le x, \cdots, X_n \le x) \\ &= P(X_1 \le x)\, P(X_2 \le x) \cdots P(X_n \le x) \\ &= (F(x))^n \end{aligned}$$

이를 미분한 $X_{(n)}$ 의 확률밀도함수 $f_{X_{(n)}}(x)$ 는 다음과 같다.

$$f_{X_{(n)}}(x) = n(F(x))^{n-1} f(x)$$

예 5.1 $X_1, X_2, \ldots, X_n$ 이 Uniform$(0,1)$ 로부터의 랜덤샘플일 때 $X_{(n)}$ 의 확률분포를 구하여라.

[풀이] Uniform$(0,1)$ 의 확률밀도함수와 누적분포함수는 각각 $f(x) = 1,\ F(x) = x,\ 0 < x < 1$ 이므로 $X_{(n)}$ 의 확률밀도함수는 다음과 같다.

$$f_{X_{(n)}}(x) \;=\; n(F(x))^{n-1} f(x) \;=\; n x^{n-1}, \quad 0 < x < 1$$

이는 베타분포의 확률밀도함수로 $X_{(n)} \sim$ Beta$(n,\ 1)$ 이다. ■

이번에는 최솟값의 확률분포를 구하도록 하자. $X_{(1)}$ 은 $X_1, X_2, \ldots, X_n$ 중에서 최솟값이다. 사건 $\{X_{(1)} > x\}$ 는 $X_1, X_2, \ldots, X_n$ 모두 x 보다 크다는 것을 의미한다. 따라서 $X_{(1)}$ 의 누적분포함수 $F_{X_{(1)}}(x)$ 는 다음과 같다.

$$\begin{aligned} F_{X_{(1)}}(x) &= P(X_{(1)} \le x) \;=\; 1 - P(X_{(1)} > x) \\ &= 1 - P(X_1 > x, X_2 > x, \cdots, X_n > x) \\ &= 1 - P(X_1 > x)\, P(X_2 > x) \cdots P(X_n > x) \\ &= 1 - (1 - F(x))^n \end{aligned}$$

따라서 $X_{(1)}$ 의 확률밀도함수 $f_{X_{(1)}}(x)$ 는 다음과 같다.

$$f_{X_{(1)}}(x) = n(1 - F(x))^{n-1} f(x)$$

예 5.2 $X_1, X_2, \ldots, X_n$ 이 Uniform$(0,1)$ 로부터의 랜덤샘플일 때 $X_{(1)}$ 의 확률분포를 구하여라.

[풀이] Uniform$(0,1)$의 확률밀도함수와 누적분포함수는 각각 $f(x)=1,\ F(x)=x,\ 0<x<1$이므로 $X_{(1)}$의 확률밀도함수는 다음과 같다.

$$f_{X_{(1)}}(x) \ = \ n(1-F(x))^{n-1}f(x) \ = \ n(1-x)^{n-1}, \quad 0<x<1$$

따라서 $X_{(1)} \sim$ Beta$(1,\,n)$이다. ■

예 5.3 각 부품의 수명이 지수분포를 따를 때 독립인 n개의 부품이 직렬로 연결된 시스템의 수명에 대한 확률분포를 구하여라.

[풀이] 각 부품의 수명을 X_i라 할 때 전체 시스템은 직렬로 연결되어 있기 때문에 시스템의 수명은 n개 중에서 가장 먼저 고장이 나는 시간인 $\min(X_1, X_2, \ldots, X_n) = X_{(1)}$이다. 따라서 시스템의 수명에 대한 확률밀도함수는 다음과 같다.

$$\begin{aligned} f_{X_{(1)}}(x) &= n(1-F(x))^{n-1}f(x) = n(e^{-\lambda x})^{n-1}\lambda e^{-\lambda x} \\ &= n\lambda e^{-n\lambda x}, \quad x>0 \end{aligned}$$

따라서 시스템의 수명의 확률분포는 Exp$(n\lambda)$이다. 지수분포에서 $\frac{1}{\lambda}$은 평균적인 대기시간(수명)이라 하였다. 직렬 연결된 시스템은 Exp$(n\lambda)$이므로 전체 시스템의 평균 수명은 $\frac{1}{n\lambda}$이라는 것을 알 수 있다. ■

$X_1, X_2, \ldots, X_n$ 중에서 k번째로 작은 통계량인 $X_{(k)}$의 확률분포를 구하여 보자. 그림 5.2와 같이 $X_{(k)}$가 x보다 작은 경우는 $X_1, X_2, \ldots, X_n$ 중에서 x보다 작은 것의 수가 k개 이상일 때이다. 따라서 $X_{(k)} \le x$일 확률은 $X_1, X_2, \ldots, X_n$ 중에서 x보다 작은 것의 수가 k개 이상일 확률과 같다. 이 확률은 이항분포의 확률을 이용하여 구할 수 있다.

확률변수 X_i에 대하여 확률변수 I_i를 다음과 같이 정의하자.

$$I_i \ = \ \begin{cases} 1, & X_i \le x \text{일 때} \\ 0, & X_i > x \text{일 때} \end{cases}$$

I_i는 0 또는 1의 값을 가지는 베르누이 확률변수이며 1의 값을 가질 확률은 다음과 같다.

$$P(I_i=1) \ = \ P(X_i \le x) \ = \ F(x)$$

따라서 $X_1, X_2, \ldots, X_n$ 중에서 x보다 작은 것의 수인 $\sum_{i=1}^{n} I_i$는 시행횟수는 n이고 성공의 확률이 $F(x)$인 이항분포를 따르게 된다. 즉 다음이 성립한다.

$$\sum_{i=1}^{n} I_i \ \sim \ \mathrm{B}(n,\,F(x))$$

$X_{(k)}$ 의 누적분포함수 $F_{X_{(k)}}(x)$ 는 다음과 같이 나타낼 수 있다.

$$
\begin{aligned}
F_{X_{(k)}}(x) &= P(X_{(k)} \le x) = P(\, X_1, X_2, \ldots, X_n \text{ 중 } x \text{보다 작은 것의 수 } \ge\ k\,) \\
&= P\left(\sum_{i=1}^{n} I_i\ \ge\ k\right)\ =\ \sum_{l=k}^{n} \frac{n!}{l!(n-l)!}\,[F(x)]^{l}[1-F(x)]^{n-l}
\end{aligned}
$$

이를 미분하여 구한 $X_{(k)}$ 의 확률밀도함수 $f_{X_{(k)}}(x)$ 는 다음과 같다.

$$
\begin{aligned}
f_{X_{(k)}}(x) &= \frac{d}{dx}\left(\sum_{l=k}^{n} \frac{n!}{l!(n-l)!}\,[F(x)]^{l}[1-F(x)]^{n-l}\right) \\
&= \sum_{l=k}^{n} \frac{n!}{l!(n-l)!}\ \big(l[F(x)]^{l-1}[1-F(x)]^{n-1}f(x) - (n-l)[F(x)]^{l}[1-F(x)]^{n-l-1}f(x)\big) \\
&= \sum_{l=k}^{n} \frac{n!}{(l-1)!(n-l)!}\,[F(x)]^{l-1}[1-F(x)]^{n-l}f(x) \\
&\qquad - \sum_{l=k}^{n-1} \frac{n!}{l!(n-l-1)!}\,[F(x)]^{l}[1-F(x)]^{n-l-1}f(x) \\
&= \sum_{l=k}^{n} \frac{n!}{(l-1)!(n-l)!}\,[F(x)]^{l-1}[1-F(x)]^{n-l}f(x) \\
&\quad - \sum_{m=k+1}^{n} \frac{n!}{(m-1)!(n-m)!}\,[F(x)]^{m-1}[1-F(x)]^{n-m}f(x) \quad (\text{단, } m = l+1) \\
&= \frac{n!}{(k-1)!(n-k)!}\,[F(x)]^{k-1}[1-F(x)]^{n-k}f(x)
\end{aligned}
$$

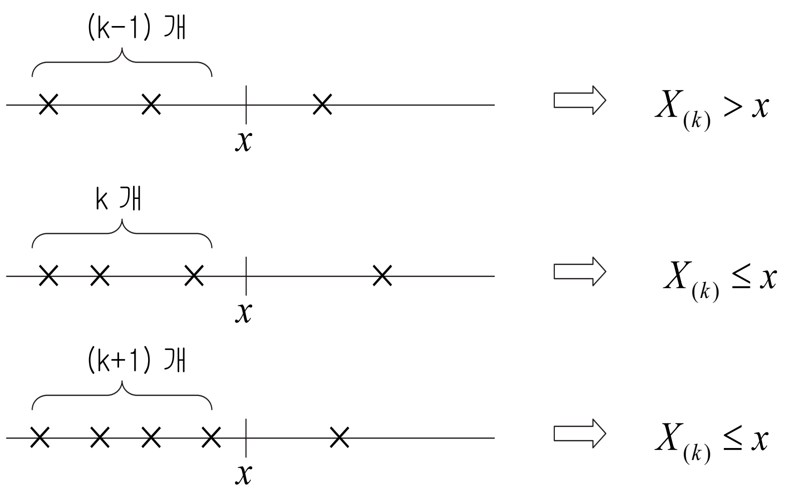

그림 5.2: $\{X_{(k)} \le x\}$ 인 사건

이를 요약하면 다음과 같은 정리를 얻을 수 있다.

정리 5.8 $X_1, X_2, \ldots, X_n$ 이 확률밀도함수 $f(x)$, 누적분포함수 $F(x)$ 를 가지는 확률분포로부터의 랜덤샘플일 때 $X_{(k)}$ 의 확률밀도함수는 다음과 같다.

$$f_{X_{(k)}}(x) \;=\; \frac{n!}{(k-1)!(n-k)!}[F(x)]^{k-1}[1-F(x)]^{n-k}f(x)$$

k 번째 순서통계량 $X_{(k)}$ 가 x 에 위치하기 위해서는 $X_1, X_2, \ldots, X_n$ 중에서 하나는 x 의 값을 가져야 하며 $(k-1)$ 개는 x 보다 작은 값, $(n-k)$ 개는 x 보다 큰 값을 가져야 한다. 이를 바탕으로 앞의 정리는 다음과 같이 해석할 수 있다.

- $\frac{n!}{(k-1)!(n-k)!}$ 은 그림 5.3과 같이 n 개의 자료를 x 보다 작은 첫 번째 그룹, x 값을 가지는 두 번째 그룹, x 보다 큰 세 번째 그룹으로 나눌 때 첫 번째 그룹에 $(k-1)$ 개, 두 번째 그룹에 1개, 세 번째 그룹에 $(n-k)$ 개가 오도록 배정하는 경우의 수이다.
- $[F(x)]^{k-1}$ 은 정해진 $(k-1)$ 개의 자료가 x 보다 작을 확률이다.
- $[1-F(x)]^{n-k}$ 는 정해진 $(n-k)$ 개의 자료가 x 보다 클 확률이다.
- $f(x)$ 는 정해진 하나의 자료가 x 에 위치할 확률밀도함수이다.

예 5.4 $X_1, X_2, \ldots, X_n$ 이 Uniform$(0,1)$ 로부터의 랜덤샘플일 때 $X_{(k)}$ 의 확률분포를 구하여라.

[풀이] Uniform$(0,1)$ 의 확률밀도함수와 누적분포함수는 각각 $f(x)=1,\ F(x)=x,\ 0<x<1$ 이므로 $X_{(k)}$ 의 확률밀도함수는 다음과 같다.

$$f_{X_{(k)}}(x) = \frac{n!}{(k-1)!(n-k)!}x^{k-1}(1-x)^{n-k}, \quad 0<x<1$$

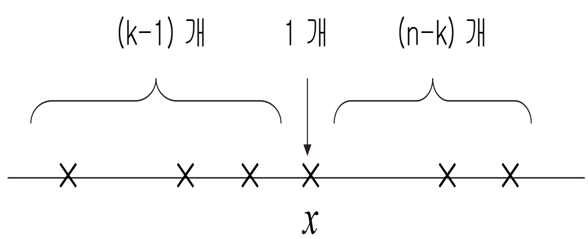

그림 5.3: n 개의 자료를 $k-1$ 개, 1개, $n-k$ 개로 나누는 경우

따라서 $X_{(k)} \sim$ Beta$(k,\, n-k+1)$ 이다. ■

$X_{(k)}$ 의 확률밀도함수에 대한 앞의 설명을 확장하면 $X_{(i)}$ 와 $X_{(j)}$ 의 결합분포를 구할 수 있다.

정리 5.9 $X_1, X_2, \ldots, X_n$ 이 확률밀도함수 $f(x)$, 누적분포함수 $F(x)$ 를 가지는 확률분포로부터의 랜덤샘플일 때 $X_{(i)}$ 와 $X_{(j)}$ (단, $i < j$) 의 결합확률밀도함수는 다음과 같다.

$$\begin{aligned} & f_{X_{(i)},\, X_{(j)}}(u,\, v) \\ & \quad = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} [F(u)]^{i-1}[F(v)-F(u)]^{j-i-1}[1-F(v)]^{n-j} f(u)f(v) \end{aligned}$$

단, 위의 결합확률밀도함수에서 $u < v$ 이다.

위의 결합확률밀도함수는 다음과 같이 해석할 수 있다.

- $\dfrac{n!}{(i-1)!(j-i-1)!(n-j)!}$ 은 n 개의 자료를 u 보다 작은 첫 번째 그룹, u 값을 가지는 두 번째 그룹, u 와 v 사이의 세 번째 그룹, v 값을 가지는 네 번째 그룹, v 보다 큰 다섯 번째 그룹으로 나눌 때 첫 번째 그룹에 $(i-1)$ 개, 두 번째 그룹에 1개, 세 번째 그룹에 $(j-i-1)$ 개, 네 번째 그룹에 1개, 다섯 번째 그룹에 $(n-j)$ 개가 오도록 배정하는 경우의 수이다.
- $[F(u)]^{i-1}$ 은 정해진 $(i-1)$ 개의 자료가 u 보다 작을 확률이다.
- $[F(v)-F(u)]^{j-i-1}$ 은 정해진 $(j-i-1)$ 개의 자료가 u 와 v 사이에 올 확률이다.
- $[1-F(v)]^{n-j}$ 는 정해진 $(n-j)$ 개의 자료가 v 보다 클 확률이다.
- $f(u)$ 와 $f(v)$ 는 정해진 두 개의 자료가 u 와 v 에 위치할 확률밀도함수이다.

예 5.5 $X_1, X_2, \ldots, X_n$ 이 Uniform$(0,1)$ 로부터의 랜덤샘플이라 한다.

(a) $X_{(1)}$ 과 $X_{(n)}$ 의 결합확률밀도함수를 구하여라.

(b) $X_{(n)} - X_{(1)}$ 의 확률분포를 구하여라.

[풀이]

(a) Uniform$(0,1)$ 의 확률밀도함수와 누적분포함수는 각각 $f(x) = 1,\ F(x) = x,\ 0 < x < 1$ 이므로 $U = X_{(1)}$ 와 $V = X_{(n)}$ 의 결합 확률밀도함수는 다음과 같다.

$$\begin{aligned} f_{U,V}(u,v) &= \frac{n!}{(n-2)!}[F(v)-F(u))]^{n-2} f(u)f(v) \\ &= n(n-1)(v-u)^{n-2}, \quad 0 < u < v < 1 \end{aligned}$$

(b) $X_{(n)} - X_{(1)}$의 분포를 구하기 위해 $R = V - U,\ S = U$라 하자. 이때 $U = S,\ V = R + S$이고 U와 V의 범위가 모두 0과 1 사이이므로 R과 S의 범위는 $0 < R < 1,\ 0 < S < 1,\ 0 < R + S < 1$을 만족하는 영역이다. 한편 $J = \begin{vmatrix} \frac{\partial u}{\partial r} & \frac{\partial u}{\partial s} \\ \frac{\partial v}{\partial r} & \frac{\partial v}{\partial s} \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ 1 & 1 \end{vmatrix} = -1$이므로 정리 3.1에 의하여 R과 S의 결합확률밀도함수는 다음과 같다.

$$f_{R,S}(r,s) = n(n-1)r^{n-2}, \quad 0 < r < 1 - s,\ 0 < s < 1$$

R의 확률밀도함수는 R과 S의 결합확률밀도함수를 S에 대하여 적분하여 구할 수 있다. 즉 R의 확률밀도함수는 다음과 같다.

$$\begin{aligned} f_R(r) &= \int_{-\infty}^{\infty} f_{R,S}(r,s)\ ds \\ &= \int_0^{1-r} n(n-1)r^{n-2}\ ds \ = \ n(n-1)r^{n-2}(1-r), \quad 0 < r < 1 \end{aligned}$$

이는 베타분포의 확률밀도함수이므로 다음을 얻을 수 있다.

$$X_{(n)} - X_{(1)} \ \sim \ \text{Beta}(n-1,\ 2)$$

■

연습문제

1. $X \sim \mathrm{N}(1,1),\ Y \sim \mathrm{N}(2,2^2),\ Z \sim \mathrm{N}(3,3^2)$ 이며 서로 독립일 때 다음 통계량의 확률분포를 구하여라.

   a) $(X-1)^2 + \dfrac{(Y-2)^2}{4} + \dfrac{(Z-3)^2}{9}$

   b) $\dfrac{\sqrt{2}(X-1)}{\sqrt{\dfrac{(Y-2)^2}{4} + \dfrac{(Z-3)^2}{9}}}$

   c) $\dfrac{(X-1)^2 + \dfrac{(Y-2)^2}{4}}{\dfrac{2(Z-3)^2}{9}}$

2. $X \sim t_{10},\, Y \sim F_{1,\,10},\, Z \sim F_{10,\,1}$ 이며 α 는 $P(X \geq \alpha) = 0.05$ 를 만족하는 상수이다.

   a) $P(Y \geq \beta) = 0.1$ 일 때 α 와 β 의 관계식을 구하여라.

   b) $P(Z \leq \gamma) = 0.1$ 일 때 α 와 γ 의 관계식을 구하여라.

3. $X,\, Y,\, Z \sim \mathrm{Exp}(\lambda)$ 이며 서로 독립일 때 $\dfrac{2X}{Y+Z}$ 의 확률분포를 구하여라.

4. $X_{ij} \sim \mathrm{N}(\mu_i, \sigma^2), \quad i = 1, 2, \ldots, m,\ j = 1, 2, \ldots, n$ 이며 서로 독립이다.

   a) $\dfrac{1}{\sigma^2} \sum_{i=1}^{m} \sum_{j=1}^{n} (X_{ij} - \overline{X_i})^2$ 의 확률분포를 구하여라. 단, $\overline{X_i} = \dfrac{1}{n} \sum_{j=1}^{n} X_{ij}$ 이다.

   b) $\mu_1 = \mu_2 = \cdots = \mu_m$ 일 때 $\dfrac{n}{\sigma^2} \sum_{i=1}^{m} (\overline{X_i} - \overline{X})^2$ 의 확률분포를 구하여라. 단, $\overline{X} = \dfrac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} X_{ij}$ 이다.

5. $X_1, X_2, \ldots, X_n$ 이 $\mathrm{Gamma}(\alpha, \lambda)$ 로부터의 랜덤샘플이다.

   a) $\overline{X}$ 의 확률분포를 구하여라.

   b) $\alpha = 1$ 일 때 $c\overline{X} \sim \chi^2_{2n}$ 인 상수 c 를 구하여라.

6. $X_1, X_2, \ldots, X_n$ 이 $\mathrm{Uniform}(0,1)$ 로부터의 랜덤샘플이다.

   a) $X_{(j)} = y$ 일 때 $X_{(i)}$ 의 조건부분포를 구하여라. 단, $i < j$ 이다.

   b) $X_{(i)} - X_{(i-2)}$ 의 확률밀도함수를 구하여라. 단, $3 \leq i \leq n$ 이다.

7. $X_1, X_2, \ldots, X_m$ 이 $\mathrm{Exp}(\lambda)$ 로부터의 랜덤샘플일 때 $X_{(i)}$ 와 $X_{(j)}$ 의 결합확률밀도함수를 구하여라.

8. $X_1, X_2, \ldots, X_n$ 이 확률질량함수 $p(k) = (1-p)^{k-1}p,\ \ k = 1, 2, \ldots$ 를 가지는 기하분포로부터의 랜덤샘플이다.

   a) $P(X_{(i)} \leq k)$ 를 구하여라.

   b) $P(X_{(i)} = k)$ 를 구하여라.

9. $X_1, X_2, \ldots, X_m$ 은 $\mathrm{N}(\mu_X, \sigma^2)$ 로부터의 랜덤샘플이고 $Y_1, Y_2, \ldots, Y_n$ 은 $\mathrm{N}(\mu_Y, \sigma^2)$ 로부터의 랜덤샘플이며 서로 독립이다. 이때

$$t \;=\; \frac{(\overline{X}-\overline{Y})-(\mu_X-\mu_Y)}{S_p\sqrt{\frac{1}{m}+\frac{1}{n}}}$$

의 확률분포가 t_{m+n-2} 임을 보여라. 단,

$$S_p^2 = \frac{1}{m+n-2}\left(\sum_{i=1}^{m}(X_i-\overline{X})^2+\sum_{j=1}^{n}(Y_j-\overline{Y})^2\right)$$

이다.

10. $X_k \sim \mathrm{N}(ck, \sigma^2), \quad k=1,2,\ldots,n$ 이며 서로 독립이다. 단, c 는 상수이다.

a) $\frac{\sum_{k=1}^n kX_k}{\sum_{k=1}^n k^2}$ 의 확률분포를 구하여라.

b) $c=0$ 일 때, $\frac{(\sum_{k=1}^n kX_k)^2}{\sigma^2\sum_{k=1}^n k^2}$ 과 $\frac{\sum_{k=1}^n X_k^2}{\sigma^2}$ 의 확률분포를 구하여라.

$\sum_{k=1}^n X_k^2 \;=\; \sum_{k=1}^n \left(X_k - k\frac{\sum_{i=1}^n iX_i}{\sum_{i=1}^n i^2}\right)^2 + \frac{(\sum_{k=1}^n kX_k)^2}{\sum_{k=1}^n k^2}$ 임을 이용하여 다음 물음에 답하여라.

c) 각 k 에 대하여 $\left(X_k - k\frac{\sum_{i=1}^n iX_i}{\sum_{i=1}^n i^2}\right)$ 와 $\sum_{i=1}^n iX_i$ 는 서로 독립임을 설명하여라.

d) $c=0$ 일 때 $\frac{1}{\sigma^2}\sum_{k=1}^n \left(X_k - k\frac{\sum_{i=1}^n iX_i}{\sum_{i=1}^n i^2}\right)^2$ 의 확률분포를 구하여라.

제 6 장

확률변수의 극한

실수의 수열 $\{a_n\}$에 대하여 a_n이 a로 수렴한다는 의미는 a를 중심으로 아무리 작은 구간을 잡아도 충분히 큰 n에 대하여 a_n 이후의 원소는 모두 이 구간에 들어간다는 것이다. 즉 임의의 양수 t에 대하여 다음 조건을 만족하는 n이 존재할 때 $a_n \to a$라 한다.

$$|a_m - a| < t, \quad m \geq n \text{일 때}$$

실수의 수열에 대한 극한의 개념은 확률변수의 수열로 확장될 수 있다. 통계학에서 극한을 생각하는 이유는 모집단에서 추출한 자료의 수가 늘어날 때 통계량이 어떠한 성질을 가지는지 알아보기 위해서이다. 예를 들어 자료의 수 n이 증가할 때 표본평균이 가지는 성질은 확률변수에 대한 수열을 통해 파악할 수 있다.

이 장에서는 확률변수의 수열 $\{X_n\}$에 대하여 X_n이 X로 수렴하는 의미에 대하여 다루고자 한다. 확률변수는 표본공간에서 실수로의 함수이므로 X_n은 $X_n(\omega)$로 나타낼 수 있다. 여기서 ω는 표본공간에 속하는 원소를 나타낸다. $X_n(\omega)$은 실수이므로 기본적으로 $\{X_n(\omega)\}$는 실수의 수열이라 할 수 있다. 따라서 X_n이 X로 수렴한다는 것을 다음과 같이 정의하는 것이 한 가지 방법일 것이다.

$$\text{표본공간의 모든 원소 } \omega \text{에 대하여} \quad X_n(\omega) \longrightarrow X(\omega)$$

실제로 위 정의에서 표본공간의 모든 ω에서 수렴할 필요는 없고 수렴하는 ω를 모아놓은 집합의 확률이 1이면 된다. 이러한 수렴의 정의를 almost sure convergence라 한다. 통계학에서는 위의 정의 외에도 확률변수의 수렴에 대한 다른 몇 가지 정의를 더 사용하고 있다.

임의의 양수 t에 대하여 다음이 성립할 때

$$P(|X_n - X| > t) \longrightarrow 0$$

X_n이 X로 **확률수렴**(convergence in probability)한다고 하며, 기호로 $X_n \xrightarrow{p} X$와 같이 나타낸다. 확률수렴은 X_n과 X의 차이가 일정한 값 이상일 확률이 0으로 수렴한다는 것이다.

6.1 대수의 법칙

$X_1, X_2, \ldots, X_n$ 이 랜덤샘플이라 하자. 표본의 크기 n 이 커질 때 표본평균이 모평균으로 수렴한다는 것을 **대수의 법칙**(law of large numbers)이라 한다. 자료의 수가 n 이라는 것을 나타내기 위해 표본평균을 $\overline{X}_n$ 으로 나타내기도 한다.

정리 6.1 [대수의 법칙] $X_1, X_2, \ldots, X_n$ 이 기댓값 μ, 분산 σ^2 인 확률분포부터의 랜덤샘플일 때 임의의 양수 t 에 대하여 다음이 성립한다.

$$P\left(|\overline{X}_n - \mu| > t\right) \longrightarrow 0$$

즉 $\overline{X}_n \xrightarrow{p} \mu$ 이다.

[**증명**] 정리 4.6에서 주어진 체비세프 부등식에 의하여 다음 부등식이 성립한다.

$$P\left(|\overline{X}_n - \mu| > t\right) \le \frac{Var(\overline{X}_n)}{t^2} \le \frac{\sigma^2}{nt^2}$$

n 이 커질 때 $\dfrac{\sigma^2}{nt^2}$ 은 0으로 수렴하므로 $P\left(|\overline{X}_n - \mu| > t\right) \longrightarrow 0$ 임을 알 수 있다. ■

6.2 분포수렴

확률변수의 수렴에 대한 또 다른 정의는 누적분포함수를 이용한 것이다. X_n 의 누적분포함수를 F_n, X 의 누적분포함수를 F 라 할 때 $F_n(x)$ 이 $F(x)$ 으로 수렴한다는 것을 이용한 정의이다. 여기서 모든 x 에 대하여 $F_n(x) \to F(x)$ 일 필요는 없으며 $F(\cdot)$ 가 연속인 모든 x 에서 수렴하기만 하면 된다.

정의 6.1 $X_1, X_2, \ldots$ 이 누적분포함수 $F_1, F_2, \ldots$ 를 갖는 확률변수의 수열이고, 확률변수 X 의 누적분포함수는 F 라 하자. $F(\cdot)$ 가 연속인 모든 점 x 에서 다음이 성립할 때

$$\lim_{n\to\infty} F_n(x) = F(x)$$

X_n 이 X 로 **분포수렴**(convergence in distribution)한다고 한다. 이를 기호로 $X_n \xrightarrow{d} X$ 와 같이 나타낸다.

위의 분포수렴에 대한 정의에서 모든 x 에 대하여 $F_n(x)$ 이 $F(x)$ 로 수렴한다는 것을 이용하지 않고 $F(\cdot)$ 가 연속인 점 x 에서만 $F_n(x) \longrightarrow F(x)$ 이면 된다는 것을 사용하였다. 이를 알아보기 위해 한 가지 예를 살펴보도록 하자. $Z_1, Z_2, \ldots, Z_n$ 이 $\mathrm{N}(0,1)$ 로부터의 랜덤샘플이라 할 때

표본평균의 수열을 생각하여 보자. $Z_1, Z_2, \ldots, Z_n$ 의 표본평균을 X_n 이라 할 때 X_n 은 기댓값 0, 분산 $\dfrac{1}{n}$ 인 정규분포를 따른다. X_n 의 분산이 0으로 수렴하므로 X_n 은 상수로 수렴한다고 할 수 있다. X_n 의 기댓값이 0이므로 직관적으로 볼 때 X_n 은 0으로 수렴한다고 볼 수 있다.

한편 X_n 의 누적분포함수 $F_n(x)$ 는 다음과 같이 나타낼 수 있다.

$$F_n(x) \;=\; \int_{-\infty}^{x} \frac{\sqrt{n}}{\sqrt{2\pi}} e^{-\frac{n}{2}t^2}\, dt \;=\; \int_{-\infty}^{\sqrt{n}\,x} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}\, dy \quad (\text{단, } y = \sqrt{n}\,t)$$

표본의 크기 n 이 커질 때 $\sqrt{n}\,x$ 의 극한은 다음과 같으므로

$$\sqrt{n}\,x \longrightarrow \begin{cases} -\infty, & x < 0 \text{일 때} \\ 0, & x = 0 \text{일 때} \\ \infty, & x > 0 \text{일 때} \end{cases}$$

X_n 의 누적분포함수의 극한은 다음과 같다.

$$F_n(x) \longrightarrow \begin{cases} 0, & x < 0 \text{일 때} \\ 0.5, & x = 0 \text{일 때} \\ 1, & x > 0 \text{일 때} \end{cases}$$

그러나 이러한 극한을 누적분포함수로 가지는 확률분포는 존재하지 않는다. 2장에서 설명한 것처럼 누적분포함수는 우측에서 연속이지만 위에서 구한 $F_n(x)$ 의 극한은 $x = 0$ 에서 우측에서 연속인 함수가 아니다. 한편 위에서 설명한 것과 같이 직관적으로 볼 때 X_n 은 0으로 수렴한다고 할 수 있다. X 를 항상 0의 값을 가지는 확률변수라 할 때 X 의 누적분포함수 $F(x)$ 는 다음과 같다.

$$F(x) = \begin{cases} 0, & x < 0 \text{일 때} \\ 1, & x \geq 0 \text{일 때} \end{cases}$$

이 함수는 $x = 0$ 을 제외한 점에서는 $F_n(x)$ 의 극한 값과 같다. 즉 다음이 성립하며

$$\lim_{n\to\infty} F_n(x) = F(x), \qquad x \neq 0 \text{일 때}$$

$x = 0$ 은 함수 $F(x)$ 의 불연속점이다.

예 6.1 $X_1, X_2, \ldots, X_n$ 이 Uniform$(0,\, \theta)$ 로부터의 랜덤샘플일 때 $X_{(n)}$ 이 분포수렴하는 극한을 구하여라.

[풀이] $X_{(n)} = \max(X_1, X_2, \ldots, X_n)$의 누적분포함수 $F_n(x)$는 다음과 같다.

$$\begin{aligned} F_n(x) &= P[\max(X_1, X_2, \ldots, X_n) \le x] = P(X_1 \le x)P(X_2 \le x) \cdots P(X_n \le x) \\ &= \left(\frac{x}{\theta}\right)^n, \quad 0 \le x \le \theta \text{일 때} \end{aligned}$$

이 함수의 극한은 다음과 같다.

$$\lim_{n\to\infty} F_n(x) = \begin{cases} 0, & x < \theta \text{일 때} \\ 1, & x \ge \theta \text{일 때} \end{cases}$$

한편 항상 θ의 값을 가지는 확률변수를 X라 할 때 X의 누적분포함수 $F(x)$는 다음과 같으므로

$$F(x) = \begin{cases} 0, & x < \theta \text{일 때} \\ 1, & x \ge \theta \text{일 때} \end{cases}$$

모든 x에 대하여 $\lim_{n\to\infty} F_n(x) = F(x)$이다. 따라서 $X_{(n)} \xrightarrow{d} \theta$이다. ■

위 예는 Uniform$(0, \theta)$로부터의 랜덤샘플의 최댓값 $X_{(n)}$은 θ로 수렴한다는 사실을 보여주고 있다. 다음 예는 $X_{(n)}$이 얼마나 빨리 θ로 수렴하는지 보여주고 있다. $n(\theta - X_{(n)})$이 Exp$\left(\frac{1}{\theta}\right)$로 수렴한다는 사실을 통해 $X_{(n)}$은 $\frac{1}{n}$의 속도로 θ로 수렴함을 보일 수 있다.

예 6.2 $X_1, X_2 \ldots, X_n$이 Uniform$(0, \theta)$로부터의 랜덤샘플일 때 $Y_n = n(\theta - X_{(n)})$이 분포수렴하는 극한을 구하여라.

[풀이] Y_n의 누적분포함수 $F_n(y)$는 다음과 같다.

$$\begin{aligned} F_n(y) &= P\Big(n(\theta - X_{(n)}) \le y\Big) = 1 - P\Big(n(\theta - X_{(n)}) > y\Big) \\ &= 1 - P\left(X_{(n)} < \theta - \frac{y}{n}\right) \\ &= 1 - \left(1 - \frac{y}{n\theta}\right)^n, \quad 0 < y < n\theta \end{aligned}$$

이 함수의 극한은 다음과 같다.

$$\lim_{n\to\infty} F_n(y) = 1 - e^{-\frac{y}{\theta}}, \quad y > 0$$

이는 Exp $\left(\frac{1}{\theta}\right)$ 의 누적분포함수이므로 다음이 성립한다.

$$n(\theta - X_{(n)}) \xrightarrow{d} \text{Exp}\left(\frac{1}{\theta}\right)$$

■

X_n 이 X 로 분포수렴한다는 것을 보이기 위해서는 X_n 의 누적분포함수가 X 의 누적분포함수로 수렴한다는 사실을 보여야 한다. 이 방법 외에 분포수렴을 보일 수 있는 다른 방법은 X_n 의 적률생성함수가 X 의 적률생성함수로 수렴한다는 사실을 보이는 것이다. 다음 정리의 증명은 이 교재에서 다루는 수준을 넘어서기 때문에 생략하겠지만 확률변수의 극한을 구할 때 유용하게 쓰일 수 있다.

정리 6.2 $X_1, X_2, \ldots$ 의 적률생성함수를 $M_1(t), M_2(t), \ldots$ 라 하고 X 의 적률생성함수를 $M(t)$ 라 하자. 모든 t 에 대하여 다음이 성립할 때

$$\lim_{n\to\infty} M_n(t) = M(t)$$

$X_n \xrightarrow{d} X$ 이다.

예 6.3 $X_n \sim \text{Poisson}(\lambda_n)$ 이며 $\lambda_n \to \infty$ 라고 한다. 이때 X_n 을 표준화한 Z_n 이 분포수렴하는 극한을 구하여라.

$$Z_n = \frac{X_n - E(X_n)}{\sqrt{Var(X_n)}} = \frac{X_n - \lambda_n}{\sqrt{\lambda_n}}$$

[풀이] Z_n 의 적률생성함수 $M_{Z_n}(t)$ 는 다음과 같이 나타낼 수 있다.

$$\begin{aligned} M_{Z_n}(t) &= E(e^{tZ_n}) = E\left[e^{t\left(\frac{X_n-\lambda_n}{\sqrt{\lambda_n}}\right)}\right] = E\left(e^{-t\sqrt{\lambda_n}+\frac{t}{\sqrt{\lambda_n}}X_n}\right) \\ &= e^{-t\sqrt{\lambda_n}}E\left(e^{\frac{t}{\sqrt{\lambda_n}}X_n}\right) = e^{-t\sqrt{\lambda_n}}M_{X_n}\left(\frac{t}{\sqrt{\lambda_n}}\right) \end{aligned}$$

위에서 $M_{X_n}(t)$ 는 $\text{Poisson}(\lambda_n)$ 분포를 따르는 X_n 의 적률생성함수이다. $\text{Poisson}(\lambda)$ 의 적률생성함수는 $M(t) = e^{\lambda(e^t-1)}$ 이므로 Z_n 의 적률생성함수는 다음과 같다.

$$M_{Z_n}(t) = e^{-t\sqrt{\lambda_n}}e^{\lambda_n\left(e^{\frac{t}{\sqrt{\lambda_n}}}-1\right)} = e^{-t\sqrt{\lambda_n}+\lambda_n\left(e^{\frac{t}{\sqrt{\lambda_n}}}-1\right)}$$

한편 테일러 전개에 의하여 $e^x = \sum\limits_{k=0}^{\infty} \frac{x^k}{k!}$ 이므로 다음이 성립한다.

$$e^{\frac{t}{\sqrt{\lambda_n}}} = \sum_{k=0}^{\infty} \frac{1}{k!}\left(\frac{t}{\sqrt{\lambda_n}}\right)^k = 1 + \frac{t}{\sqrt{\lambda_n}} + \frac{1}{2}\left(\frac{t}{\sqrt{\lambda_n}}\right)^2 + \sum_{k=3}^{\infty} \frac{1}{k!}\left(\frac{t}{\sqrt{\lambda_n}}\right)^k$$

따라서 Z_n 의 적률생성함수는 다음과 같다.

$$\begin{aligned} M_{Z_n}(t) &= e^{-t\sqrt{\lambda_n}+\lambda_n\left(1+\frac{t}{\sqrt{\lambda_n}}+\frac{1}{2}\left(\frac{t}{\sqrt{\lambda_n}}\right)^2+\sum_{k=3}^{\infty}\frac{1}{k!}\left(\frac{t}{\sqrt{\lambda_n}}\right)^k - 1\right)} \\ &= e^{\frac{t^2}{2}+\sum_{k=3}^{\infty}\frac{1}{k!}\frac{t^k}{\lambda_n^{k/2-1}}} \end{aligned}$$

λ_n 이 커질 때 $\sum\limits_{k=3}^{\infty} \frac{1}{k!}\frac{t^k}{\lambda_n^{k/2-1}}$ 는 0으로 수렴하므로 $M_{Z_n}(t) \;\to\; e^{\frac{t^2}{2}}$ 이다. $e^{\frac{t^2}{2}}$ 은 표준정규분포의 적률생성함수이므로 $Z_n \xrightarrow{d} \mathrm{N}(0,1)$ 이다. ■

예 6.4 $X \sim \mathrm{Poisson}(900)$ 일 때 $P(X > 950)$ 의 근삿값을 구하여라.

[풀이] 앞의 예에 의하여 $\frac{X-900}{\sqrt{900}}$ 이 표준정규분포에 근사하므로 다음이 성립한다.

$$P(X > 950) = P\left(\frac{X-900}{\sqrt{900}} > \frac{950-900}{\sqrt{900}}\right) \approx P(\mathrm{N}(0,1) > 1.67) = 0.04779$$

참고로 포아송분포로부터 직접 구한 $P(X > 950)$ 의 값은 0.04712이다. ■

6.3 중심극한정리

모집단의 확률분포에 상관없이 랜덤샘플의 표본평균을 표준화하면 자료의 수가 증가할 때 표준정규분포에 가까워진다는 것이 중심극한정리 (central limit theorem) 이다.

정리 6.3 [중심극한정리] $X_1, X_2, \ldots, X_n$ 이 기댓값 μ, 분산 σ^2 인 확률분포부터의 랜덤샘플이다. $S_n = \sum_{i=1}^{n} X_i$ 이라 할 때 다음이 성립한다.

$$Z_n = \frac{S_n - n\mu}{\sqrt{n}\sigma} = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \xrightarrow{d} \mathrm{N}(0,\,1)$$

즉 다음 결과를 얻을 수 있다.

$$\lim_{n\longrightarrow\infty} P\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} \le x\right) = P(\mathrm{N}(0,1) \le x)$$

[증명] Z_n의 적률생성함수 $M_{Z_n}(t)$는 다음과 같이 나타낼 수 있다.

$$\begin{aligned} M_{Z_n}(t) &= E\left(e^{\frac{t}{\sqrt{n}}\sum_{i=1}^{n}\frac{(X_i-\mu)}{\sigma}}\right) \ = \ E\left(e^{\sum_{i=1}^{n}\frac{t}{\sqrt{n}}Y_i}\right) \\ &= \left[E\left(e^{\frac{t}{\sqrt{n}}Y}\right)\right]^n \ = \ \left[M_Y\left(\frac{t}{\sqrt{n}}\right)\right]^n \end{aligned} \tag{6.1}$$

위 식에서 $Y_i = \dfrac{X_i-\mu}{\sigma}$이고 M_Y는 Y_i의 적률생성함수이다. $M_Y(s)$를 0을 중심으로 테일러 전개하면 다음과 같다.

$$M_Y(s) \ = \ M_Y(0) + M_Y'(0)s + \frac{1}{2}M_Y''(0)s^2 + \sum_{k=3}^{\infty}\frac{1}{k!}M_Y^{(k)}(0)s^k$$

적률생성함수의 성질에 의하여 $M_Y(0) = 1,\ M_Y'(0) = E(Y) = 0,\ M_Y''(0) = E(Y^2) = 1$이므로 다음이 성립한다.

$$M_Y(s) \ = \ 1 + \frac{1}{2}s^2 + \sum_{k=3}^{\infty}\frac{1}{k!}M_Y^{(k)}(0)s^k$$

이를 식 (6.1)에 대입하면 Z_n의 적률생성함수를 다음과 같이 나타낼 수 있다.

$$\begin{aligned} M_{Z_n}(t) &= \left[1 + \frac{1}{2}\left(\frac{t}{\sqrt{n}}\right)^2 + \sum_{k=3}^{\infty}\frac{1}{k!}M_Y^{(k)}(0)\left(\frac{t}{\sqrt{n}}\right)^k\right]^n \\ &= \left[1 + \frac{1}{n}\left(\frac{1}{2}t^2 + \sum_{k=3}^{\infty}\frac{1}{k!}M_Y^{(k)}(0)\frac{t^k}{n^{k/2-1}}\right)\right]^n \end{aligned}$$

한편 $\lim_{n\to\infty} a_n = a$일 때 $\lim_{n\to\infty}\left(1+\frac{a_n}{n}\right) = e^a$이므로 Z_n의 적률생성함수의 극한은 다음과 같다.

$$M_{Z_n}(t) \ \longrightarrow \ e^{\frac{t^2}{2}}$$

$e^{\frac{t^2}{2}}$은 표준정규분포의 적률생성함수이므로 $Z_n \xrightarrow{d} \mathrm{N}(0,1)$이다. ■

중심극한정리는 서로 독립이며 같은 분포를 가지는 확률변수의 합을 표준화하면 정규분포에 가까이 간다는 것을 말해주고 있다. 즉 같은 분포를 가지며 독립인 n개의 확률변수의 합의 형태로 표현될 수 있는 것은 기댓값 0, 분산 1이 되도록 표준화하였을 때 표준정규분포에 수렴함을 알려주고 있다. 예를 들어 이항분포는 베르누이분포의 합으로 나타낼 수 있으므로 이항분포를 따르는 확률변수를 표준화하면 시행 횟수가 증가할 때 표준정규분포로 수렴한다. 즉 $X_n \sim \mathrm{B}(n, p)$이고 Y_i는 서로 독립이며 Bernoulli(p)를 따르는 확률변수일 때 $X_n = Y_1 + Y_2 + \cdots + Y_n$이므로

다음이 성립한다.

$$\frac{X_n - np}{\sqrt{np(1-p)}} \xrightarrow{d} \mathrm{N}(0,1)$$

또한 앞에서 다룬 예 6.3도 중심극한정리로 설명할 수 있다. $X_n \sim \mathrm{Poisson}(\lambda_n)$ 인 확률변수 X_n 은 예 4.17에 의하여 포아송분포를 따르는 확률변수의 합으로 표현할 수 있다. 즉 $Y_i \sim \mathrm{Poisson}\left(\frac{\lambda_n}{n}\right)$ 이라 할 때 $X_n = Y_1 + Y_2 + \cdots + Y_n$ 라 나타낼 수 있다. 따라서 $\frac{\lambda_n}{n}$ 이 어느 정도 일정한 수준을 유지하면 중심극한정리에 의하여 X_n 을 표준화한 것은 표준정규분포에 수렴하게 된다.

예 6.5 $X \sim \mathrm{B}(100,\ 0.5)$ 일 때 $P(X \geq 60)$ 의 근삿값을 구하여라.

[풀이] 앞에서 설명한 것과 같이 이항분포를 표준화한 것은 시행횟수가 커질 때 표준정규분포에 수렴하게 된다. X 의 기댓값은 50이고 분산은 25이므로 $\frac{X-50}{\sqrt{25}}$ 은 표준정규분포에 가깝게 된다. 따라서 다음 결과를 얻을 수 있다.

$$P(X \geq 60) = P\left(\frac{X-50}{\sqrt{25}} \geq \frac{60-50}{\sqrt{25}}\right) \approx P(\mathrm{N}(0,1) \geq 2) = 0.0228$$

■

연습문제

1. $X_1, X_2, \ldots, X_n$ 이 Uniform$(-\theta, \theta)$ 로부터의 랜덤샘플일 때 다음 물음에 답하여라.

   a) $n(\theta - \max_{1\le i\le n} |X_i|)$ 이 분포수렴하는 극한을 구하여라.

   b) $n(\min_{1\le i\le n} |X_i|)$ 이 분포수렴하는 극한을 구하여라.

2. $X \sim \mathrm{B}(n,p)$ 이고 $np = \lambda$ 일 때 적률생성함수를 이용하여 n 이 커질 때 X 의 확률분포는 Poisson(λ) 로 분포수렴함을 보여라.

3. $X_1, X_2, \ldots, X_{450}$ 이 확률밀도함수 $f(x) = 2x,\ 0 < x < 1$ 인 확률분포로부터의 랜덤샘플일 때 $P(\sum_{i=1}^{450} X_i > 295)$ 의 근삿값을 구하여라.

4. $X \sim$ Gamma(n, λ) 일 때 다음 물음에 답하여라.

   a) $Y = aX + b$ 의 기댓값이 0이고 표준편차가 1이 되도록 하는 상수 a 와 b 를 구하여라.

   b) 위에서 정의한 Y 는 n 이 커질 때 표준정규분포에 분포수렴하는 이유를 설명하여라.

5. $X_1, X_2, \ldots, X_n$ 이 $\mathrm{N}(\mu, \sigma^2)$ 로부터의 랜덤샘플일 때 다음 물음에 답하여라.

   a) $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ 의 적률생성함수를 구하여라.

   b) 위 결과를 이용하여 표본의 크기 n 이 커질 때 S^2 의 극한을 구하여라.

제 7 장

충분통계량

통계학의 주요한 목표는 관측된 자료를 이용하여 우리가 알고자 하는 대상을 규명하고자 하는 것이다. 예를 들어 우리나라 전체 국민의 소득을 알아보기 위해 1000명의 사람을 선택하여 이들의 소득을 조사하였을 때 관측된 자료는 선택된 1000명의 소득이며 우리가 알고자 하는 대상은 전체 국민의 소득이다. 이처럼 우리가 알고자 하는 대상을 **모집단**(population)이라 한다. 모집단을 규명하기 위해 우리가 관측하는 자료는 일반적으로 모집단의 일부이며 이 자료를 **샘플**(sample)이라 한다.

랜덤샘플을 이용하여 모집단에 대한 추론을 할 때 일반적으로 모집단의 모형이나 확률분포에 대한 가정을 하게 된다. 모집단에 대한 가정 하에서 샘플을 이용하여 구체적인 모형이나 확률분포를 추측하게 된다. 예를 들어 전체 국민의 소득이 정규분포를 따른다고 가정하였을 때, 정규분포는 모평균과 모분산에 따라 달라지므로 전체 국민 소득은 여러 정규분포 중의 하나가 될 것이다. 이때 1000명의 소득에 대한 샘플을 이용하여 모평균과 모분산을 추측함으로써 모집단에 대한 확률분포를 알아낼 수 있을 것이다. 모집단의 모형이나 확률분포에 대해 가정했을 때 모집단의 구체적인 모형이나 확률분포를 결정지어주는 값을 **모수**(parameter)라 한다. 예를 들어 모집단이 정규분포를 따른다고 가정하였을 경우는 μ와 σ^2이 모수이며, 모집단이 포아송분포를 따른다고 가정하였을 경우는 λ가 모수이다.

모집단의 모형이나 확률분포에 대한 가정을 하였을 때 모집단을 규명하는 문제는 모수를 추측하는 문제로 귀결된다. 모수에 대해 추측한다는 것은 모수의 값을 한 값으로 결정하는 것일 수도 있고 또는 모수의 값이 속할 가능성이 높은 어떤 구간을 정해주는 것일 수도 있다. 또한 모수에 대한 몇 가지 가설 중에서 어느 가설이 타당한지 결정하는 것일 수도 있다. 이러한 것을 통계학에서는 **추론**(inference)이라 한다. 따라서 자료를 이용하여 모집단의 특성을 파악하는 통계학의 기본적인 목표는 자료를 이용하여 모수에 대해 추론을 하는 것으로 단순화될 수 있다. 이러한 단순화는 기본적으로 모집단에 대한 가정을 통해 이루어진다. 만약 이러한 가정이 적절하지 않다면 추론을 통해 얻은 결론도 타당하지 않을 것이다.

자료를 이용하여 우리가 알고자 하는 모집단을 규명하는 일반적인 과정은 다음과 같이 나타낼 수 있다.

- 모집단의 모형이나 확률분포에 대한 가정을 한다.

- 자료를 이용하여 모집단의 모형이나 확률분포를 결정하는 모수에 대하여 추론을 한다.

예를 들어 앞에서 설명한 1000명의 소득을 이용하여 전체 국민의 소득을 규명하는 과정은 다음과 같이 나타낼 수 있다.

- 전체 국민의 소득은 평균 μ이고 분산이 σ^2인 정규분포를 따른다고 가정한다.
- 1000명의 소득을 이용하여 μ와 σ^2에 대한 추론을 한다.

일반적으로 모집단의 확률분포는 확률밀도함수(확률질량함수)에 의하여 나타낼 수 있으며 이 확률밀도함수(확률질량함수)는 모수에 의하여 결정된다. 일반적으로 모수는 θ로 나타낸다. 모집단의 확률분포가 모수 θ에 의존한다는 것을 명확히 하기 위해 확률밀도함수 또는 확률질량함수를 다음과 같이 나타낸다.

$$f(x|\theta)$$

자료를 이용하여 모집단을 규명하는 것은 모형화를 통하여 모수 θ에 대하여 추론하는 과정이라 할 수 있다.

7.1 충분통계량

앞에서 설명한 것처럼 통계적 추론이란 주어진 자료 $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$로부터 모수 θ에 대한 정보를 이끌어 내는 것을 말한다. 그런데 주어진 자료에는 모수에 대한 정보를 가지고 있는 부분도 있을 수 있고 정보를 가지지 못한 부분도 있을 수 있다. 통계적 추론을 단순화하기 위해서는 주어진 자료에서 모수에 대한 정보를 가진 부분만을 간추릴 필요성이 있다. 이러한 축약(data reduction)을 통해 전체 자료 대신에 축약된 정보만을 이용하여 모수에 대한 추론을 할 수 있다.

일반적으로 $\boldsymbol{X}$에서 $T(\boldsymbol{X})$로 자료를 축약하는 경우[1] 원 자료에 대한 정보의 일부분을 상실하게 되어 θ에 대한 정보 역시 손실을 입을 수 있다. 예를 들어 $T(\boldsymbol{X}) = \sum_{i=1}^n X_i$일 때 자료의 합인 $T(\boldsymbol{X})$만 주어졌다면 원 자료에 대한 정보는 얻을 수 없으므로 모수에 대한 추론의 결과는 달라질 수 있다. 그런데 때에 따라서는 축약된 통계량[2] $T(\boldsymbol{X})$가 원 자료에 대한 정보는 잃어버리지만 모수에 대한 정보는 그대로 가지고 있을 수도 있다. 만약 $T(\boldsymbol{X})$가 모수에 대한 정보를 그대로 가지고 있다면 통계적 추론의 관점에서 볼 때 원 자료 대신에 축약된 통계량 $T(\boldsymbol{X})$만 가지고 있어도 충분하다고 할 수 있다. 이와 같이 어떤 통계량 $T(\boldsymbol{X})$가 모수 θ에 대해 가지고 있는 정보가 원 자료 $\boldsymbol{X}$가 모수에 대해 가지고 있는 정보와 같을 때 이 통계량을 **충분통계량**(sufficient statistic)이라 한다. 참고로 어떤 통계량 $T(\boldsymbol{X})$의 확률분포가 θ에 의존하지 않는다면 $T(\boldsymbol{X})$는 θ에 상관없는 값을 가지게 되므로 $T(\boldsymbol{X})$는 θ에 대한 정보를 가지고 있지 않다고 할 수 있다.

충분통계량에 대한 이해를 돕기 위해 다음과 같은 예를 생각하여 보자. 자료는 (X, Y)이며 X와 Y는 각각 $N(\theta, 1)$를 따르는 확률변수이며 서로 독립이라 하자. 자료 (X, Y)를 이용하여

[1] 원 자료인 $\boldsymbol{X}$ 대신에 $\boldsymbol{X}$의 함수인 $T(\boldsymbol{X})$가 주어졌다면 $\boldsymbol{X}$의 값을 정확히 알 수 없다. 반대로 $\boldsymbol{X}$가 주어지면 $T(\boldsymbol{X})$의 값은 알 수 있다. 따라서 $T(\boldsymbol{X})$는 $\boldsymbol{X}$보다 축약된 정보를 가지고 있다고 할 수 있다. 예를 들어 $T(x) = x^2$인 경우 $x = 2$이면 $T(x) = 4$이지만 $T(x) = 4$가 주어졌다고 할지라도 x의 값이 2와 -2 중 어느 것인지 알 수 없다.

[2] 통계량이란 주어진 자료 $\boldsymbol{X}$에 의해서만 결정되는 것을 말한다.

θ에 대한 추론을 하고자 한다. 먼저 (X, Y)와 $(X + Y, X - Y)$는 서로 일대일 대응 관계에 있으므로 (X, Y)가 θ에 대해 가지고 있는 정보와 $(X + Y, X - Y)$가 θ에 대해 가지고 있는 정보는 같다. 예를 들어 $X = 3, Y = 1$이면 $X + Y = 4, X - Y = 2$임을 알 수 있고, 반대로 $X + Y = 4, X - Y = 2$이면 $X = 3, Y = 1$임을 알 수 있기 때문에 (X, Y)가 가진 정보와 $(X + Y, X - Y)$가 가진 정보는 같다. 한편 $(X - Y) \sim \mathrm{N}(0, 2)$이므로 $(X - Y)$의 확률분포는 θ에 의존하지 않는다. 따라서 $(X - Y)$는 θ에 대한 정보를 가지고 있지 않다. 즉 $(X + Y, X - Y)$가 θ에 대해 가지고 있는 정보는 $(X + Y)$에만 포함되어 있다고 볼 수 있다.

$$(X, Y) \overset{1-1}{\Longleftrightarrow} (X + Y, X - Y) \overset{\text{모수에 대한 정보를 잃지 않음}}{\Longrightarrow} (X + Y)$$

따라서 (X, Y)가 θ에 대해 가지고 있는 정보와 $(X + Y)$가 θ에 대해 가지고 있는 정보는 같다. 이 예에서 (X, Y)가 $X + Y$로 축약되면서 θ에 대한 정보는 잃지 않기 때문에 $X + Y$는 θ에 대한 충분통계량이라 할 수 있다.

위에서 개념적으로 표현된 충분통계량을 구체적으로 정의하면 다음과 같다.

정의 7.1 $T(\boldsymbol{X})$의 값이 주어졌을 때 $\boldsymbol{X}$의 조건부분포가 θ에 의존하지 않을 때 $T(\boldsymbol{X})$를 θ에 대한 충분통계량이라 한다.

위 정의는 $T(\boldsymbol{X})$의 값이 주어 졌을 때 $\boldsymbol{X}$의 조건부분포가 θ에 의존하지 않으므로 $T(\boldsymbol{X})$의 값으로부터 X의 값을 복구하는 과정은 θ에 관련이 없다는 것을 의미한다. $\boldsymbol{X}$에서 $T(\boldsymbol{X})$로 축약되는 과정에서 분명히 자료에 대한 정보를 일정 부분 잃어버리기 때문에 $T(\boldsymbol{X})$로부터 $\boldsymbol{X}$의 값을 알 수는 없다. 그러나 $T(\boldsymbol{X})$의 값이 주어졌을 때 $\boldsymbol{X}$가 여러 가능한 값 중 어떤 하나의 값을 가질 확률분포가 θ에 관련이 없다면 $\boldsymbol{X}$에서부터 $T(\boldsymbol{X})$로 축약되는 과정은 θ에 관련이 없다는 것을 의미한다. 따라서 $T(\boldsymbol{X})$에 포함되어 있는 θ에 대한 정보는 $\boldsymbol{X}$가 가지고 있는 정보와 같다고 할 수 있다. 그림 7.1은 이러한 과정을 나타낸 것이다.

구체적인 예를 통해 위에서 설명한 의미를 살펴보도록 하자. X는 $\mathrm{N}(0, \sigma^2)$를 따르는 확률변수이며 X를 이용하여 분산에 대한 추론을 한다고 한다. 이때 모수는 분산이므로 $\theta = \sigma^2$이라 할 수 있다. $T(X) = X^2$이라 하면 $T(X) = t$일 때 X의 조건부분포는 다음과 같다.

$$X = \begin{cases} \sqrt{t}, & \frac{1}{2}\text{의 확률로} \\ -\sqrt{t}, & \frac{1}{2}\text{의 확률로} \end{cases}$$

이는 X의 확률분포가 0에 대해 대칭이므로 $X = -\sqrt{t}$에서 확률밀도함수의 값과 $X = \sqrt{t}$에서 확률밀도함수의 값이 같다는 것으로부터 쉽게 확인할 수 있다. 따라서 $T(X) = X^2$으로부터 X를 복구하는 과정은 $\theta = \sigma^2$에 관련이 없다. 한편 $T(X) = X^2$이 σ^2에 대해 가지고 있는 정보는 X가 가지고 있는 정보와 같다는 것은 다음과 같이 설명할 수 있다. 분산은 기댓값으로 얼마만큼 퍼져있는지를 측정한 값이므로 0으로부터 같은 거리만큼 떨어진 $\sqrt{t}$가 나타내는 퍼진 정도나 $-\sqrt{t}$가 나타내는 퍼진 정도는 같다고 할 수 있다. $T(X)$의 값이 t로 주어지면 X의 값은 $\sqrt{t}$

또는 $-\sqrt{t}$일 수 있지만 두 값 모두 σ^2에 대한 정보에는 차이가 없다. 따라서 $T(X)$는 σ^2에 대한 정보를 잃어버리지 않았다고 할 수 있다. 위의 설명을 요약하면 $T(X) = X^2$으로부터 X의 값을 복구하는 과정은 모수에 의존하지 않으며 반대로 X에서 $T(X) = X^2$으로 축약되는 과정에서 σ^2에 대한 정보는 잃어버리지 않는다는 것을 의미한다.

예 7.1 $X_1, X_2, \ldots, X_n$이 Bernoulli(θ)로부터의 랜덤샘플일 때 $T(\boldsymbol{X}) = \sum_{i=1}^n X_i$는 θ에 대한 충분통계량임을 보여라.

[풀이] $T(\boldsymbol{X}) = t$일 때 $\boldsymbol{X}$의 조건부분포는 다음과 같다.

$$\begin{aligned} &P\Big(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \,\Big|\, \sum_{i=1}^n X_i = t\Big) \\ &= \frac{P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n,\, \sum_{i=1}^n X_i = t)}{P(\sum_{i=1}^n X_i = t)} \end{aligned}$$

$\sum_{i=1}^n x_i = t$일 때 사건 $\{X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n,\, \sum_{i=1}^n X_i = t\}$는 $\{X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\}$와 같으므로 조건부분포는 다음과 같다.

$$\begin{aligned} &P\Big(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \,\Big|\, \sum_{i=1}^n X_i = t\Big) \\ &= \frac{P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)}{P(\sum_{i=1}^n X_i = t)} \\ &= \frac{\theta^{x_1}(1-\theta)^{1-x_1} \cdots \theta^{x_n}(1-\theta)^{1-x_n}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}} \\ &= \frac{1}{\begin{pmatrix} n \\ t \end{pmatrix}} \end{aligned}$$

이는 θ에 의존하지 않으므로 $T(\boldsymbol{X}) = \sum_{i=1}^n$는 θ에 대한 충분통계량이다. ■

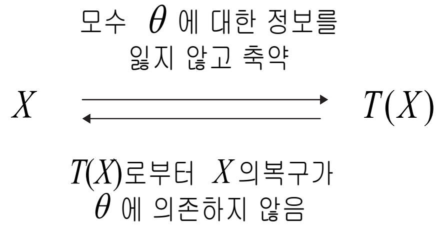

그림 7.1: 충분통계량

모수에 대한 충분통계량은 하나만 있는 것은 아니다. 예를 들어 원 자료는 항상 충분통계량이며 또한 랜덤샘플인 경우 순서통계량도 충분통계량이 된다. (연습문제 7) 이처럼 충분통계량은 여러 개가 있을 수 있으며 가장 많이 축약된 통계량을 **최소충분통계량**(minimal sufficient statistic)이라 한다. 일반적으로 모수의 수가 p이면 최소충분통계량은 p개의 통계량으로 주어지는 경우가 많다.

충분통계량을 구하기 위해 정의 7.1을 이용하는 것은 쉽지 않다. 정의를 이용하여 충분통계량을 구하기 위해서는 먼저 충분통계량이 될 가능성이 있는 통계량을 추측하고 이 통계량이 주어졌을 때 $\boldsymbol{X}$의 조건부분포가 모수에 의존하지 않는다는 사실을 보여야 하기 때문이다. 일반적으로 모수에 대한 충분통계량을 구할 때 자주 이용되는 것이 다음 정리이다.

정리 7.1 [인수분해정리, factorization theorem] 통계량 $T(\boldsymbol{X})$가 θ에 대한 충분통계량일 필요충분조건은 $\boldsymbol{X}$의 확률밀도함수(확률질량함수) $f(x_1, x_2, \ldots, x_n|\theta)$가 $T(x_1, x_2, \ldots, x_n)$와 θ에 의하여 결정되는 한 부분과 $x_1, x_2, \ldots, x_n$에 의하여 결정되는 다른 한 부분의 곱으로 표현될 때이다. 즉 적절한 함수 $g(\cdot)$와 $h(\cdot)$에 대하여 확률밀도함수 $f(x_1, x_2, \ldots, x_n|\theta)$를 다음과 같이 나타낼 수 있을 때 $T(\boldsymbol{X})$가 θ에 대한 충분통계량이다.

$$f(x_1, x_2, \ldots, x_n|\theta) \;=\; g(T(x_1, x_2, \ldots, x_n), \theta)\, h(x_1, x_2, \ldots, x_n)$$

[**증명**] 증명은 이산형 확률분포인 경우만 다루도록 하겠다. 먼저 $\boldsymbol{X}$의 확률질량함수 $f(x_1, x_2, \ldots, x_n|\theta)$가 다음과 같이 표현될 수 있다고 가정하자.

$$f(x_1, x_2, \ldots, x_n|\theta) \;=\; g(T(x_1, x_2, \ldots, x_n), \theta)\, h(x_1, x_2, \ldots, x_n)$$

이때 $T(\boldsymbol{X})$의 확률분포는 다음과 같이 나타낼 수 있다.

$$\begin{aligned} P(T(\boldsymbol{X}) = t) &= \sum_{x_1, \ldots, x_n \,:\, T(\boldsymbol{x}) = t} f(x_1, x_2, \ldots, x_n|\theta) \\ &= \sum_{x_1, \ldots, x_n \,:\, T(\boldsymbol{x}) = t} g(T(x_1, x_2, \ldots, x_n), \theta)\, h(x_1, x_2, \ldots, x_n) \\ &= g(t, \theta) \sum_{x_1, \ldots, x_n \,:\, T(\boldsymbol{x}) = t} h(x_1, x_2, \ldots, x_n) \end{aligned}$$

위에서 $\sum_{x_1, \ldots, x_n \,:\, T(\boldsymbol{x}) = t}$는 $T(\boldsymbol{x}) = t$를 만족하는 모든 $(x_1, \ldots, x_n)$에서 합을 나타낸다. $T(\boldsymbol{X}) = t$로 주어졌을 때 $T(\boldsymbol{x}) = t$를 만족하는 $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$의 영역에서 $\boldsymbol{X}$의

조건부확률은 다음과 같다.

$$
\begin{aligned}
&P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n \,|\, T(\boldsymbol{X}) = t) \\
&= \frac{P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n,\, T(\boldsymbol{X}) = t)}{P(T(\boldsymbol{X}) = t)} \\
&= \frac{P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)}{P(T(\boldsymbol{X}) = t)} \\
&= \frac{g(t,\theta)\, h(x_1, x_2, \dots, x_n)}{g(t,\theta) \sum_{x_1,\dots,x_n \,:\, T(\boldsymbol{x})=t} h(x_1, x_2, \dots, x_n)} \\
&= \frac{h(x_1, x_2, \dots, x_n)}{\sum_{x_1,\dots,x_n \,:\, T(\boldsymbol{x})=t} h(x_1, x_2, \dots, x_n)}
\end{aligned}
$$

이는 θ에 의존하지 않으므로 $T(\boldsymbol{X})$는 θ에 대한 충분통계량이다.

필요조건을 보이기 위해 $T(\boldsymbol{X})$가 θ에 대한 충분통계량이라 가정하자. 이때 함수 $g(\cdot)$와 $h(\cdot)$를 다음과 같이 정의하자.

$$
\begin{aligned}
g(t,\theta) &= P(T(\boldsymbol{X}) = t) \\
h(x_1, x_2, \dots, x_n) &= P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n \,|\, T(\boldsymbol{X}) = t)
\end{aligned}
$$

위에서 $T(\boldsymbol{X}) = t$일 확률은 t와 모수 θ에만 의존하기 때문에 $P(T(\boldsymbol{X}) = t)$는 t와 모수 θ의 함수가 된다. 또한 $T(\boldsymbol{X})$는 θ에 대한 충분통계량이기 때문에 조건부확률 $P(X_1 = x_1, X_2 = x_2, \dots, X_n \,|\, T(\boldsymbol{X}) = t)$는 모수 θ에 의존하지 않는 $(x_1, x_2, \dots, x_n)$만의 함수가 된다. 따라서 $T(\boldsymbol{x}) = t$일 때 $(X_1, X_2, \dots, X_n)$의 확률질량함수는 다음과 같이 표현할 수 있다.

$$
\begin{aligned}
&P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \\
&= P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n,\, T(\boldsymbol{X}) = t) \\
&= P(T(\boldsymbol{X}) = t) P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n \,|\, T(\boldsymbol{X}) = t) \\
&= g(t,\theta)\, h(x_1, x_2, \dots, x_n)
\end{aligned}
$$

■

인수분해정리를 이용하여 충분통계량을 구하기 위해서는 $(X_1, X_2, \dots, X_n)$의 확률밀도함수를 두 부분의 곱으로 분리하여야 한다. $(x_1, x_2, \dots, x_n)$에만 의존하는 부분 $h(\cdot)$와 $(x_1, x_2, \dots, x_n)$과 θ에 동시에 의존하는 나머지 부분 $g(\cdot)$으로 분해해야 한다. $\boldsymbol{x}$와 θ에 의존하는 부분 $g(\cdot)$가 $T(\boldsymbol{x})$와 θ의 함수일 때 $T(\boldsymbol{X})$가 θ에 대한 충분통계량이 된다. 즉 확률밀도함수를 곱으로 분해할 때 $\boldsymbol{x}$와 θ에 의존하는 부분으로부터 충분통계량을 구할 수 있다는 것이다. 이 부분이 $T(\boldsymbol{x})$와 θ의 함수일 때 $T(\boldsymbol{X})$가 θ에 대한 충분통계량이 된다.

예 7.2 $X_1, X_2, \ldots, X_n$ 이 Poisson(θ) 으로부터의 랜덤샘플일 때 $T(\boldsymbol{X}) = \sum_{i=1}^n X_i$ 는 θ 에 대한 충분통계량임을 보여라.

[풀이] 포아송분포로부터의 랜덤샘플 $\boldsymbol{X}$ 의 확률질량함수는 다음과 같다.

$$\begin{aligned} f(x_1, x_2, \ldots, x_n|\theta) &= \prod_{i=1}^n f(x_i|\theta) = \prod_{i=1}^n \frac{\theta^{x_i} e^{-\theta}}{x_i!} \\ &= \theta^{\sum_{i=1}^n x_i} e^{-n\theta} \frac{1}{x_1! x_2! \cdots x_n!} \end{aligned}$$

위에서 $\prod_{i=1}^n f(x_i|\theta)$ 는 각 $f(x_i|\theta)$ 를 곱한 $f(x_1|\theta)f(x_2|\theta)\cdots f(x_n|\theta)$ 를 나타낸다. 위 확률질량함수에서 $\dfrac{1}{x_1!x_2!\cdots x_n!}$ 은 $(x_1, x_2, \ldots, x_n)$ 에만 의존하는 부분이고 $\theta^{\sum_{i=1}^n x_i} e^{-n\theta}$ 는 $(x_1, x_2, \ldots, x_n)$ 와 θ 에 의존하는 부분이다. 이 부분은 $\sum_{i=1}^n x_i$ 와 θ 의 함수이므로 $T(\boldsymbol{X}) = \sum_{i=1}^n X_i$ 는 θ 에 대한 충분통계량이다. ■

예 7.3 $X_1, X_2, \ldots, X_n$ 이 N(μ, σ^2) 으로부터의 랜덤샘플일 때 $T(\boldsymbol{X}) = (\sum_{i=1}^n X_i^2, \sum_{i=1}^n X_i)$ 는 (μ, σ^2) 에 대한 충분통계량임을 보여라.

[풀이] 정규분포의 경우 모수는 $\theta = (\mu, \sigma^2)$ 이다. 이때 $\boldsymbol{X}$ 의 확률밀도함수는 다음과 같이 나타낼 수 있다.

$$\begin{aligned} f(x_1, x_2, \ldots, x_n|\mu, \sigma^2) &= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^n (x_i-\mu)^2}{2\sigma^2}\right) \\ &= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^n x_i^2}{2\sigma^2} + \frac{\mu}{\sigma^2}\sum_{i=1}^n x_i - \frac{n\mu^2}{2\sigma^2}\right) \end{aligned}$$

이 확률밀도함수는 $(\sum_{i=1}^n x_i^2, \sum_{i=1}^n x_i)$ 와 (μ, σ^2) 의 함수이므로 $T(\boldsymbol{X}) = (\sum_{i=1}^n X_i^2, \sum_{i=1}^n X_i)$ 는 (μ, σ^2) 에 대한 충분통계량이다. ■

예 7.4 $X_1, X_2, \ldots, X_n$ 이 Uniform$(0, \theta)$ 로부터의 랜덤샘플일 때 $T(\boldsymbol{X}) = X_{(n)}$ 은 θ 에 대한 충분통계량임을 보여라.

[풀이] $\boldsymbol{X}$의 확률밀도함수는 다음과 같이 나타낼 수 있다.

$$\begin{aligned} f(x_1, x_2, \ldots, x_n|\theta) &= \left(\frac{1}{\theta}\right)^n, \quad 0 < x_i < \theta \\ &= \left(\frac{1}{\theta}\right)^n I(0 < x_{(1)} \le x_{(n)} < \theta) \\ &= \left(\frac{1}{\theta}\right)^n I(0 < x_{(1)}) I(x_{(n)} < \theta) \end{aligned}$$

따라서 $T(\boldsymbol{X}) = X_{(n)}$은 θ에 대한 충분통계량이다. ■

예 7.5 $X_1, X_2, \ldots, X_n$이 확률밀도함수 $f(x|\theta) = \exp(-(x-\theta)),\ x \ge \theta$를 갖는 랜덤샘플일 때 $T(\boldsymbol{X}) = X_{(1)}$은 θ에 대한 충분통계량임을 보여라.

[풀이] $\boldsymbol{X}$의 확률밀도함수는 다음과 같이 나타낼 수 있다.

$$\begin{aligned} f(x_1, x_2, \ldots, x_n|\theta) &= e^{-(x_1-\theta)} e^{-(x_2-\theta)} \cdots e^{-(x_n-\theta)} I(x_{(1)} \ge \theta) \\ &= e^{-\sum_{i=1}^n x_i} e^{n\theta} I(x_{(1)} \ge \theta) \end{aligned}$$

이 확률밀도함수에서 $(x_1, x_2, \ldots, x_n)$와 θ에 의존하는 부분은 $I(x_{(1)} \ge \theta)$이다. 따라서 $X_{(1)}$은 θ에 대한 충분통계량이다. ■

7.2 지수족

확률분포 중에는 확률밀도함수 또는 확률질량함수가 특수한 형태를 가지는 경우가 있다. 이런 형태를 가지는 경우에 일반화된 결과를 얻을 수도 있다. 한 가지 경우는 확률밀도함수 또는 확률질량함수가 다음과 같은 지수함수 형태로 표현될 때이다.

정의 7.2 함수 $c(\cdot)$, $T(\cdot)$, $d(\cdot)$, $S(\cdot)$에 대하여 확률밀도함수(확률질량함수) $f(x|\theta)$가 다음과 같이 나타낼 수 있을 때 이 확률분포의 모임을 지수족(exponential family)이라 한다.

$$f(x|\theta) \;=\; \exp\left(c(\theta)T(x) + d(\theta) + S(x)\right) I_A(x) \tag{7.1}$$

위에서 $I_A(x)$는 $x \in A$일 때 1의 값을 갖고 $x \notin A$일 때 0의 값을 가지는 함수이다. 확률밀도함수가 A의 영역에서만 양의 값을 가질 때 위의 표현을 사용하여 나타낸다.

$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ 이 지수족에 속하는 확률분포로부터의 랜덤샘플일 때 $\boldsymbol{X}$ 의 확률밀도함수는 다음과 같다.

$$f(\boldsymbol{x}|\theta) \;=\; \exp\left(c(\theta)\sum_{i=1}^{n} T(x_i) + nd(\theta) + \sum_{i=1}^{n} S(x_i)\right) I_A(x_1, x_2, \ldots, x_n)$$

이때 정리 7.1에 의하여 $\sum_{i=1}^{n} T(X_i)$ 는 θ 에 대한 충분통계량임을 알 수 있다.

예를 들어 $\mathrm{B}(m, \theta)$ 의 확률질량함수는 다음과 같이 나타낼 수 있다.

$$\begin{aligned} f(x|\theta) &= \binom{m}{x}\theta^x(1-\theta)^{m-x}, \quad x = 0, 1, \ldots, m \\ &= \binom{m}{x}\left(\frac{\theta}{1-\theta}\right)^x (1-\theta)^m, \quad x = 0, 1, \ldots, m \\ &= \exp\left(x\log\left(\frac{\theta}{1-\theta}\right) + m\log(1-\theta) + \log\binom{m}{x}\right) I_{\{0,1,\ldots,m\}}(x) \end{aligned}$$

위 확률질량함수를 지수족의 식 (7.1)과 비교하면 다음 관계를 얻을 수 있다.

$$c(\theta) = \log\left(\tfrac{\theta}{1-\theta}\right), \quad T(x) = x, \quad d(\theta) = m\log(1-\theta)$$
$$S(x) = \log\tbinom{m}{x}, \quad A = \{0, 1, \ldots, m\}$$

따라서 이항분포는 지수족에 속하며 $\sum_{i=1}^{n} T(X_i) = \sum_{i=1}^{n} X_i$ 는 θ 에 대한 충분통계량이다.

확률밀도함수(확률질량함수)가 식 (7.1)과 같이 표현되는 지수족을 확장하여 확률밀도함수가 다음과 같이 나타낼 수 있을 때 이 확률분포의 모임을 k차 지수족이라 한다.

$$f(x|\theta) \;=\; \exp\left(\sum_{j=1}^{k} c_j(\theta)T_j(x) + d(\theta) + S(x)\right) I_A(x) \tag{7.2}$$

예를 들어 $\mathrm{N}(\mu, \sigma^2)$ 의 확률밀도함수는 다음과 같이 나타낼 수 있으므로

$$\begin{aligned} f(x|\mu, \sigma^2) &= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{1}{2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \\ &= \exp\left(-\frac{1}{2\sigma^2}x^2 + \frac{\mu}{\sigma^2}x - \frac{\mu^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)\right) \end{aligned} \tag{7.3}$$

정규분포는 2차 지수족이라 할 수 있다. $\theta = (\mu, \sigma^2)$ 이라고 할 때 위 확률밀도함수를 식 (7.2)와 비교하면 다음 관계를 얻을 수 있다.

$$c_1(\theta) = -\frac{1}{2\sigma^2}, \quad c_2(\theta) = \frac{\mu}{\sigma^2}, \quad d(\theta) = -\frac{\mu^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)$$
$$T_1(x) = x^2, \quad T_2(x) = x, \quad S(x) = 0, \quad A = (-\infty, \infty)$$

$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ 이 k 차 지수족에 속하는 확률분포로부터의 랜덤샘플일 때 정리 7.1에 의하여 다음 통계량이 θ 에 대한 충분통계량임을 알 수 있다.

$$\left(\sum_{i=1}^{n} T_1(X_i), \sum_{i=1}^{n} T_2(X_i), \ldots, \sum_{i=1}^{n} T_k(X_i)\right)$$

예를 들어 $X_1, X_2, \ldots, X_n$ 이 $\mathrm{N}(\mu, \sigma^2)$ 로부터의 랜덤샘플일 때 다음 통계량이 (μ, σ^2) 에 대한 충분통계량이다.

$$\left(\sum_{i=1}^{n} X_i^2,\ \sum_{i=1}^{n} X_i\right)$$

이항분포와 정규분포 외에도 포아송분포, 감마분포 등이 지수족에 속하는 확률분포이다.

7.3 완비통계량

충분통계량과 함께 모수에 대한 추론에서 흔히 이용되는 것이 **완비통계량**(complete statistic)이다.

정의 7.3 통계량 $T(\boldsymbol{X})$ 의 함수 $g(T(\boldsymbol{X}))$ 에 대하여 모든 θ 에서

$$E_\theta[g(T(\boldsymbol{X}))] = 0$$

이면 다음이 성립할 때 $T(\boldsymbol{X})$ 를 완비통계량이라 한다.

$$P_\theta\{\boldsymbol{X} : g(T(\boldsymbol{X})) = 0\} \ = \ 1$$

위에서 기댓값이나 확률을 나타낸 식에서 첨자 θ 를 붙인 이유는 이러한 값이 θ 의 함수라는 것을 명확히 나타내기 위하여 사용한 것이다. 위 정의는 θ 의 모든 값에서 $T(\boldsymbol{X})$ 의 함수 $g(T(\boldsymbol{X}))$ 의 기댓값이 0이면 함수 $g(T(\boldsymbol{X}))$ 의 값은 항상 0일 수밖에 없을 때 통계량 $T(\boldsymbol{X})$ 를 완비통계량이라 한다는 것이다.

$T(\boldsymbol{X})$ 의 함수 $g(T(\boldsymbol{X}))$ 의 기댓값이 모든 θ 에서 0이 된다는 것은 이 부분이 θ 에 대한 정보를 가지고 있지 않음을 나타낸다. 만약 $g(T(\boldsymbol{X}))$ 가 0이 아닌 함수라면 $T(\boldsymbol{X})$ 에는 θ 에 대한 정보를 가지고 있지 않은 부분이 있다는 것을 의미한다. 따라서 이런 관점에서 보면 완비통계량은 θ 에 대한 정보를 가지고 있지 않은 부분을 모두 제거한 가장 간략히 축소된 통계량이라 할 수 있다.

예 7.6 $X_1, X_2, \ldots, X_n$ 이 Bernoulli(θ) 로부터의 랜덤샘플일 때 $T(\boldsymbol{X}) = \sum_{i=1}^{n} X_i$ 는 θ 에 대한 완비통계량임을 보여라.

[풀이] $T(\boldsymbol{X}) = \sum_{i=1}^{n} X_i \ \sim\ \mathrm{B}(n, \theta)$ 이므로 $T(\boldsymbol{X})$ 의 함수 $g(T(\boldsymbol{X}))$ 의 기댓값은 다음과 같이 나타낼 수 있다.

$$
\begin{aligned}
E_\theta[g(T(\boldsymbol{X}))] &= \sum_{k=0}^{n} g(k) \binom{n}{k} \theta^k (1-\theta)^{n-k} \\
&= \sum_{k=0}^{n} g(k) \binom{n}{k} \left(\frac{\theta}{1-\theta}\right)^k (1-\theta)^n
\end{aligned}
$$

이 값이 0과 1 사이의 모든 θ 에 대해 0이 된다는 것은 다음을 뜻한다.

$$
\sum_{k=0}^{n} g(k) \binom{n}{k} \left(\frac{\theta}{1-\theta}\right)^k \ = \ 0
$$

위 식을 $\nu = \dfrac{\theta}{1-\theta}$ 의 다항식으로 볼 때 $\nu > 0$ 인 부분에서 이 다항식의 값은 항상 0이므로 다항식의 계수는 모두 0이 되어야 한다. 따라서 $g(\cdot)$ 는 0부터 n 사이의 정수 값에서 항상 0의 값을 가지는 함수라는 것을 나타낸다. $T(\boldsymbol{X})$ 는 0부터 n 사이의 정수 값만 가지므로 $T(\boldsymbol{X}) = \sum_{i=1}^{n} X_i$ 는 θ 에 대한 완비통계량이다. ■

예 7.7 $X_1, X_2, \ldots, X_n$ 이 Poisson(θ) 로부터의 랜덤샘플일 때 $T(\boldsymbol{X}) = \sum_{i=1}^{n} X_i$ 는 θ 에 대한 완비통계량임을 보여라.

[풀이] $T(\boldsymbol{X}) = \sum_{i=1}^{n} X_i \ \sim\ \mathrm{Poisson}(n\theta)$ 이므로 $T(\boldsymbol{X})$ 의 함수 $g(T(\boldsymbol{X}))$ 의 기댓값이 0이라는 것은 다음을 뜻한다.

$$
\begin{aligned}
E_\theta[g(T(\boldsymbol{X}))] = \sum_{k=0}^{\infty} g(k) \frac{(n\theta)^k e^{-n\theta}}{k!} \ &= \ 0 \\
\Longrightarrow \sum_{k=0}^{\infty} g(k) \frac{(n\theta)^k}{k!} &= 0 \quad (e^{-n\theta} \text{는 0이 아니므로})
\end{aligned}
$$

즉 모든 θ 에 대하여 $E_\theta[g(T(\boldsymbol{X}))] = 0$ 이라는 것은 다음을 뜻한다.

$$
g(0) \ + \ ng(1)\theta \ + \ \frac{n^2 g(2)}{2!}\theta^2 \ + \ \frac{n^3 g(3)}{3!}\theta^3 \ + \ \cdots \ = \ 0
$$

위 식을 θ 에 대한 다항식으로 볼 때 모든 값에서 0이 되려면 각 계수의 값이 0이어야 한다. 즉 다음이 성립한다.

$$
g(0) = 0, \ ng(1) = 0, \ \frac{n^2 g(2)}{2!} = 0, \ \cdots
$$

따라서 $g(\cdot)$는 음이 아닌 정수 값에서 항상 0의 값을 가질 수밖에 없다. 포아송분포는 음이 아닌 정수 값만을 가지므로 $T(\boldsymbol{X}) = \sum_{i=1}^{n} X_i$는 θ에 대한 완비통계량이다. ■

예 7.8 $X_1, X_2, \ldots, X_n$이 Uniform$(0, \theta)$로부터의 랜덤샘플일 때 $T(\boldsymbol{X}) = X_{(n)}$이 θ에 대한 완비통계량임을 보여라.

[풀이] 정리 5.8에 의하여 $T(\boldsymbol{X}) = X_{(n)}$의 확률밀도함수는 다음과 같다.

$$f_T(t) = n\,\frac{t^{n-1}}{\theta^n}, \quad 0 < t < \theta$$

0보다 큰 모든 θ에 대하여 $E[g(X_{(n)})] = 0$이라는 것은 다음을 뜻한다.

$$\begin{aligned} E_\theta[g(T(X))] &= \int_0^\theta g(t)\,n\,\frac{t^{n-1}}{\theta^n}\,dt = 0 \quad (\text{모든 } \theta > 0\text{에 대하여}) \\ \Longrightarrow \int_0^\theta g(t)\,t^{n-1}\,dt &= 0 \quad (\text{모든 } \theta > 0\text{에 대하여}) \end{aligned}$$

위 식을 θ에 대해 미분하면 모든 $\theta > 0$에 대하여 $g(\theta)\theta^{n-1} = 0$이 된다. 즉 0보다 큰 모든 t에 대하여 $g(t) = 0$이므로 $T(\boldsymbol{X}) = X_{(n)}$은 θ에 대한 완비통계량이다. ■

일반적으로 어떤 통계량이 완비통계량임을 보이는 것은 쉽지 않으며 지수족에 속하는 분포에 대해서는 다음 사실을 이용할 수 있다.

정리 7.2 $X_1, X_2, \ldots, X_n$이 다음과 같은 확률밀도함수 $f(x|\theta)$를 가지는 지수족으로부터의 랜덤샘플일 때 $\sum_{i=1}^{n} T(X_i)$는 θ에 대한 완비통계량이다.

$$f(x|\theta) = \exp\left(c(\theta)T(x) + d(\theta) + S(x)\right) I_A(x) \tag{7.4}$$

통계량 $T(\boldsymbol{X})$가 θ에 대한 충분통계량이며 또한 완비통계량일 때 이 통계량을 **완비충분통계량**(complete and sufficient statistic)이라 한다. 예를 들어 $X_1, X_2, \ldots, X_n$이 Bernoulli(θ) 로부터의 랜덤샘플일 때, 예 7.1과 예 7.6에 의하여 $T(\boldsymbol{X}) = \sum_{i=1}^{n} X_i$는 θ에 대한 완비충분통계량임을 알 수 있다. 식 (7.4)와 같은 확률밀도함수를 가지는 지수족에 대하여 $\sum_{i=1}^{n} T(X_i)$는 충분통계량이며 또한 완비통계량이므로 $\sum_{i=1}^{n} T(X_i)$는 θ에 대한 완비충분통계량이다.

예 7.9 $(X_1, X_2, \ldots, X_n)$ 이 Poisson(θ) 로부터의 랜덤샘플일 때 $\sum_{i=1}^n X_i$ 가 θ 에 대한 완비충분통계량임을 보여라.

[풀이] Poisson(θ) 의 확률질량함수는 다음과 같이 나타낼 수 있다.

$$\begin{aligned} f(x|\theta) &= \frac{\theta^x e^{-\theta}}{x!}, \quad x \in \{0, 1, \ldots\} \\ &= \exp(x \log\theta - \theta - \log x!) I_{\{0,1,\ldots\}}(x) \end{aligned}$$

지수족의 성질에 의하여 $\sum_{i=1}^n X_i$ 가 θ 에 대한 완비충분통계량이다. ■

지수족에 대한 정리 7.2는 확률밀도함수가 식 (7.2) 로 표현되는 k 차 지수족에 대해서도 성립한다. 즉 $X_1, X_2, \ldots, X_n$ 이 다음과 같은 확률밀도함수를 가지는 k 차 지수족으로부터의 랜덤샘플일 때

$$f(x|\theta) = \exp\left(\sum_{j=1}^k c_j(\theta) T_j(x) + d(\theta) + S(x)\right) I_A(x)$$

다음 통계량이 θ 에 대한 완비충분통계량이다.

$$\left(\sum_{i=1}^n T_1(X_i), \sum_{i=1}^n T_2(X_i), \ldots, \sum_{i=1}^n T_k(X_i)\right)$$

예를 들어 정규분포의 확률밀도함수는 식 (7.3) 으로 표현되므로 $(X_1, X_2, \ldots, X_n)$ 이 N(μ, σ^2) 으로부터의 랜덤샘플일 때 다음 통계량이 $\theta = (\mu, \sigma^2)$ 에 대한 완비충분통계량이다.

$$\left(\sum_{i=1}^n X_i^2, \sum_{i=1}^n X_i\right)$$

연습문제

1. 감마분포의 확률밀도함수를 지수족의 형태로 나타내고, 이를 이용하여 모수 $\theta = (\alpha,\ \lambda)$ 에 대한 충분통계량을 구하여라.

2. $X_1, X_2, \ldots, X_n$ 이 확률밀도함수 $f(x|\mu, \sigma) = \dfrac{1}{\sigma}\, e^{-\frac{(x-\mu)}{\sigma}},\ x > \mu$ 인 랜덤샘플일 때 $(\mu,\ \sigma)$ 에 대한 충분통계량을 구하여라.

3. $X_1, X_2, \ldots, X_n$ 이 확률밀도함수 $f(x|\theta) = \dfrac{1}{\Gamma(\theta)2^\theta}\, x^{\theta-1} e^{-\frac{x}{2}},\ x > 0$ 인 Gamma$(\theta, \frac{1}{2})$ 로부터의 랜덤샘플이다.

   a) θ 에 대한 충분통계량을 구하여라.

   b) 위에서 구한 통계량이 완비통계량인지 확인하여라.

4. $X_1, X_2, \ldots, X_n$ 이 Uniform$(\theta,\ 2\theta)$ (단, $\theta > 0$) 로부터의 랜덤샘플일 때 다음 물음에 답하여라.

   a) $(X_{(1)},\ X_{(n)})$ 이 θ 에 대한 충분통계량임을 보여라.

   b) $(X_{(1)},\ X_{(n)})$ 이 θ 에 대한 완비통계량인지 확인하여라.

5. $X_1, X_2, \ldots, X_n$ 이 N$(0,\ \sigma^2)$ 로부터의 랜덤샘플일 때 σ^2 에 대한 완비충분통계량을 구하여라.

6. $X_1, X_2, \ldots, X_n$ 이 N$(1,\ \sigma^2)$ 로부터의 랜덤샘플일 때 σ^2 에 대한 완비충분통계량을 구하여라.

7. $X_1, X_2, \ldots, X_n$ 이 랜덤샘플일 때 확률분포와 상관없이 순서통계량 $(X_{(1)}, X_{(2)}, \ldots,\ X_{(n)})$ 은 항상 충분통계량임을 보여라.

8. $X_1, X_2, \ldots, X_n$ 이 N$(\theta,\ a\theta^2)$ (단, a 는 양의 상수) 로부터의 랜덤샘플일 때 다음 물음에 답하여라.

   a) $T(\boldsymbol{X}) = (\overline{X},\ S^2)$ 이 θ 에 대한 충분통계량임을 보여라.

   b) $T(\boldsymbol{X}) = (\overline{X},\ S^2)$ 이 θ 에 대한 완비통계량은 아님을 보여라.

9. $X_1, X_2, \ldots, X_n$ 이 확률질량함수 $f(x|\theta) = \left(\frac{\theta}{2}\right)^{|x|} (1-\theta)^{1-|x|}, \quad x = -1, 0, 1$ 인 랜덤샘플일 때 다음 물음에 답하여라.

   a) θ 에 대한 충분통계량을 구하여라.

   b) $T(\boldsymbol{X}) = \sum_{i=1}^n X_i$ 이 θ 에 대한 완비통계량인지 확인하여라.

제 8 장

모수의 추정

통계적 추론은 주어진 자료를 이용하여 모집단의 특성을 결정지어주는 모수 θ에 대한 정보를 이끌어내는 것을 말한다. 통계적 추론에는 자료를 이용하여 모수를 한 값으로 추측하는 **점추정**(point estimation)과 모수에 대한 가설의 세우고 이 가설의 타당성을 판단하는 **가설검정**(hypothesis testing) 등이 있다. 이 장에서는 모수를 한 값으로 추측하는 점추정에 대해 살펴보도록 하자.

모수 θ를 하나의 값으로 추측하는 통계량을 $\widehat{\theta}$로 나타내며 이를 **추정량**(estimator)이라 한다. 추정량은 자료 $X_1, X_2, \ldots, X_n$에 의하여 결정되므로 추정량을 다음과 같이 나타내기도 한다.

$$\widehat{\theta} = \widehat{\theta}(X_1, X_2, \ldots, X_n)$$

자료를 이용하여 구한 추정량의 값을 **추정값**(estimate)이라 한다. 추정값을 $\widehat{\theta}$로 나타내기도 한다.

일반적으로 사용되는 모수의 추정 방법으로는 적률추정법, 최대가능도추정법 등이 있다.

8.1 적률추정법

$X_1, X_2, \ldots, X_n$이 확률밀도함수 $f(x\,|\,\theta)$를 갖는 모집단으로부터의 랜덤샘플이라 하자. 이때 $E(X_i^k)$를 모집단의 k차 적률 또는 k차 모적률이라 하며 μ_k로 나타낸다. 한편 자료 $X_1, X_2, \ldots, X_n$에 대한 X_i^k의 평균을 샘플의 k차 적률 또는 k차 샘플적률이라 하며 $\overline{X^k}$로 나타낸다.

- k차 모적률 (population moment) : $\mu_k = E(X^k)$
- k차 샘플적률 (sample moment) : $\overline{X^k} = \frac{1}{n}\sum_{i=1}^n X_i^k$

모적률 μ_k는 모수 θ에 의하여 결정되므로 θ의 함수이다. 이 함수를 $h_k(\theta)$라 할 때 모적률 μ_k는 다음과 같이 나타낼 수 있다.

$$\mu_k = h_k(\theta)$$

적률추정법의 기본적인 아이디어는 다음과 같다. 대수의 법칙 (정리 6.1)에 의하면 자료의 수가 증가하면 랜덤샘플로부터 구한 X_i^k의 평균은 X_i^k의 기댓값으로 수렴하게 된다. 즉 자료의 수가

증가함에 따라 샘플적률은 모적률에 수렴하게 된다.

$$\overline{X^k} \;=\; \frac{1}{n}\sum_{i=1}^{n} X_i^k \quad \xrightarrow{p} \quad E(X_i^k) \;=\; \mu_k$$

따라서 자료의 수가 충분히 크면 모적률과 샘플적률의 차이는 크지 않을 것으로 기대할 수 있다. 즉 $h_k(\theta) \approx \overline{X^k}$ 일 것이다. 이를 이용하여 θ를 추정하는 한 가지 방법은 다음 식을 만족하는 해를 θ의 추정값으로 정하는 것이다.

$$h_k(\theta) \;=\; \overline{X^k} \tag{8.1}$$

이를 적률추정법 (method of moments estimation)이라 하며, 식 (8.1) 을 만족하는 해 $\widehat{\theta}$를 **적률추정량**(moment estimator)이라 한다. 즉 적률추정량 $\widehat{\theta}$는 아래의 식을 만족하게 된다.

$$h_k(\widehat{\theta}\,) \;=\; \overline{X^k} \tag{8.2}$$

적률추정법에서 고려할 한 가지 사항은 몇 차 적률을 사용할 것인가의 문제이다. 일반적으로 k가 클수록 $Var(X^k)$는 커지는 경향이 있다. 따라서 k의 값이 커질수록 샘플적률의 변동이 커지기 때문에 가능한 작은 k를 이용하는 것이 적절하다고 할 수 있다. 모수의 수가 하나인 경우는 대부분 1차 적률을 사용하여 모수를 추정하게 된다.

예 8.1 $X_1, X_2, \ldots, X_n$이 Binomial(m, θ)로부터의 랜덤샘플일 때 θ의 적률추정량을 구하여라.

[풀이] 이항분포의 경우는 1차 모적률과 샘플적률을 이용하여 추정량을 구할 수 있다. 1차 모적률은 $\mu = m\theta$이고 1차 샘플적률은 표본평균이므로 적률추정법에 의하여 θ를 추정하는 식 (8.2) 는 다음과 같이 나타낼 수 있다.

$$m\widehat{\theta} \;=\; \overline{X}$$

따라서 적률추정량은 $\widehat{\theta} = \dfrac{\overline{X}}{m}$ 이다. ■

예 8.2 $X_1, X_2, \ldots, X_n$이 Poisson(θ)로부터의 랜덤샘플일 때 θ의 적률추정량을 구하여라.

[풀이] 포아송분포의 1차 모적률은 $\mu = \theta$이므로 θ를 추정하는 식은 다음과 같이 나타낼 수 있다.

$$\widehat{\theta} \;=\; \overline{X}$$

따라서 적률추정량은 $\widehat{\theta} = \overline{X}$ 이다. ■

예 8.3 $X_1, X_2, \ldots, X_n$ 이 Gamma(2, θ) 로부터의 랜덤샘플일 때 θ 의 적률추정량을 구하여라.

[풀이] Gamma(2, θ) 의 1차 모적률은 기댓값인 $\mu = \dfrac{2}{\theta}$ 이므로 θ 를 추정하는 식은 다음과 같이 나타낼 수 있다.

$$\frac{2}{\widehat{\theta}} = \overline{X}$$

따라서 적률추정량은 $\widehat{\theta} = \dfrac{2}{\overline{X}}$ 이다. ■

예 8.4 $X_1, X_2, \ldots, X_n$ 이 확률밀도함수 $f(x|\theta) = \exp(-(x-\theta)),\ x \geq \theta$ 를 갖는 랜덤샘플일 때 θ 의 적률추정량을 구하여라.

[풀이] 주어진 확률분포의 기댓값은 다음과 같으므로

$$\begin{aligned} E(X) &= \int_\theta^\infty xe^{-(x-\theta)}\,dx \;=\; \theta + \int_\theta^\infty (x-\theta)e^{-(x-\theta)}\,dx \\ &= \theta + \int_0^\infty ye^{-y}\,dy \;=\; \theta + 1 \end{aligned}$$

1차 적률을 이용하여 적률추정량 $\widehat{\theta}$ 를 구하는 식은 다음과 같다.

$$\widehat{\theta} + 1 = \overline{X}$$

따라서 적률추정량은 $\widehat{\theta} = \overline{X} - 1$ 이다. ■

위의 예에서 살펴본 것과 같이 모수가 1개인 경우는 일반적으로 1차 적률을 사용하여 모수를 추정하게 된다. 그러나 1차 모적률이 θ 와 무관한 값이 나온다면 이를 이용하여 모수의 적률추정량을 구할 수 없다. 이러한 경우에는 2차 적률을 사용하여 추정량을 구할 수 있다. 예를 들어 $X_1, X_2, \ldots, X_n$ 이 확률밀도함수 $f(x|\theta) = \frac{1}{2\theta}e^{-\frac{|x|}{\theta}},\ -\infty < x < \infty$ 를 갖는 랜덤샘플일 때 θ 의 적률추정량을 구하여보자. 이 확률분포의 기댓값은 $E(X) = 0$ 이므로 1차 적률을 사용하여 모수를

추정할 수 없다. 2차 모적률을 구하면 다음과 같으므로

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{\infty} \frac{x^2}{2\theta} e^{-\frac{|x|}{\theta}}\, dx = \int_0^{\infty} \frac{x^2}{\theta} e^{-\frac{x}{\theta}}\, dx \\ &= \theta^2 \int_0^{\infty} t^2 e^{-t}\, dt \quad (\text{단}, t = \tfrac{x}{\theta}) \\ &= 2\theta^2 \end{aligned}$$

2차 적률을 이용하여 θ를 추정하는 식은 다음과 같이 나타낼 수 있다.

$$2\widehat{\theta}^2 = \overline{X^2}$$

따라서 적률추정량은 $\widehat{\theta} = \sqrt{\frac{1}{2n}\sum_{i=1}^n X_i^2}$ 이다.

지금까지는 모수의 수가 하나인 경우를 다루었다. 따라서 미지수 θ는 하나의 방정식 (8.2)로부터 $\widehat{\theta}$를 구할 수 있었다. 그러나 모수가 두 개 이상이라면 미지수가 두 개 이상이라 할 수 있으므로 모수를 추정하기 위해서는 모수의 수만큼 방정식을 필요로 한다. 확률분포의 모수가 $(\theta_1, \theta_2, \ldots, \theta_r)$로 r개인 경우를 생각하여 보자. 이때 추정해야 할 모수의 수는 r이므로 $(\theta_1, \theta_2, \ldots, \theta_r)$를 추정하는 방정식은 r개가 필요하며, 이는 1차부터 r차까지 적률을 사용하여 방정식을 세울 수 있다. 1차부터 r차까지 모적률은 모수 $(\theta_1, \theta_2, \ldots, \theta_r)$의 함수이다. 이 함수를 $h_1(\cdot)$, $h_2(\cdot)$, $\ldots$, $h_r(\cdot)$라 할 때 모적률 μ_1, μ_2, $\ldots$, μ_r은 다음과 같이 나타낼 수 있다.

$$\begin{aligned} \mu_1 &= h_1(\theta_1, \theta_2, \ldots, \theta_r) \\ \mu_2 &= h_2(\theta_1, \theta_2, \ldots, \theta_r) \\ &\vdots \\ \mu_r &= h_r(\theta_1, \theta_2, \ldots, \theta_r) \end{aligned}$$

이 모적률이 1차부터 r차까지의 샘플적률과 같다고 놓은 방정식의 해가 적률추정량이 된다. 따라서 적률추정량 $\widehat{\theta}_1,\ \widehat{\theta}_2, \ldots, \widehat{\theta}_r$는 다음 식을 만족하게 된다.

$$\begin{aligned} h_1(\widehat{\theta}_1, \widehat{\theta}_2, \ldots, \widehat{\theta}_r) &= \overline{X^1} \\ h_2(\widehat{\theta}_1, \widehat{\theta}_2, \ldots, \widehat{\theta}_r) &= \overline{X^2} \\ &\vdots \\ h_r(\widehat{\theta}_1, \widehat{\theta}_2, \ldots, \widehat{\theta}_r) &= \overline{X^r} \end{aligned}$$

예 8.5 $X_1, X_2, \ldots, X_n$이 $N(\mu, \sigma^2)$로부터의 랜덤샘플일 때 μ와 σ^2의 적률추정량을 구하여라.

[풀이] 정규분포는 모수가 $\theta_1 = \mu,\ \theta_2 = \sigma^2$ 로 2개인 경우이다. 정규분포의 1차와 2차 모적률을 구하면 다음과 같으므로

$$\begin{aligned} E(X) &= \mu \\ E(X^2) &= \mu^2 + \sigma^2 \end{aligned}$$

μ와 σ^2을 추정하는 방정식은 다음과 같다.

$$\begin{aligned} \widehat{\mu} &= \overline{X} \\ \widehat{\mu}^2 + \widehat{\sigma}^2 &= \overline{X^2} \end{aligned}$$

이로부터 μ와 σ^2의 적률추정량은 다음과 같다.

$$\begin{aligned} \widehat{\mu} &= \overline{X} \\ \widehat{\sigma}^2 &= \overline{X^2} - (\overline{X})^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\overline{X})^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \end{aligned}$$

■

예 8.6 $X_1, X_2, \ldots, X_n$이 Gamma$(\alpha,\ \lambda)$로부터의 랜덤샘플일 때 α와 λ의 적률추정량을 구하여라.

[풀이] 감마분포의 1차와 2차 모적률을 구하면 다음과 같으므로

$$\begin{aligned} E(X) &= \frac{\alpha}{\lambda} \\ E(X^2) &= \frac{\alpha(\alpha+1)}{\lambda^2} \end{aligned}$$

α와 λ를 추정하는 방정식은 다음과 같다.

$$\begin{aligned} \frac{\widehat{\alpha}}{\widehat{\lambda}} &= \overline{X} \\ \frac{\widehat{\alpha}(\widehat{\alpha}+1)}{\widehat{\lambda}^2} &= \overline{X^2} \end{aligned}$$

첫 번째 식으로부터 $\widehat{\alpha} = \widehat{\lambda}\,\overline{X}$이며 이를 두 번째 식에 대입하면 다음을 얻을 수 있다.

$$(\overline{X})^2 + \frac{\overline{X}}{\widehat{\lambda}} = \overline{X^2}$$

따라서 α와 λ의 적률추정량은 다음과 같다.

$$\begin{aligned}\widehat{\lambda} &= \frac{\overline{X}}{\overline{X^2} - (\overline{X})^2} \\ \widehat{\alpha} &= \frac{(\overline{X})^2}{\overline{X^2} - (\overline{X})^2}\end{aligned}$$

■

적률추정법은 일반적으로 많이 이용되는 추정 방법은 아니다. 적률추정법은 모수의 추정 방법이 직관적이지만 추정량이 가지는 성질이 다른 추정 방법에 비해 우수하지 않을 수도 있기 때문이다. 예 (8.4)에서 $\theta > 0$임을 알고 있다고 하자. 이때 표본평균이 0.5라면 적률추정법에 의한 $\theta 0$의 추정값은 $\widehat{\theta} = \overline{X} - 1 = -0.5$이다. 이는 θ가 0보다 크다는 사실과 어긋난다.

8.2 최대가능도추정법

모수의 추정 방법으로 가장 널리 이용되는 방법이 최대가능도추정법(method of maximum likelihood estimation)이다. 최대가능도추정법은 기본적으로 주어진 자료에 대하여 확률밀도함수 또는 확률질량함수를 가장 크게 하는 θ의 값을 모수의 추정값으로 정하는 방법이다.

$X_1, X_2, \ldots, X_n$이 확률밀도함수(확률질량함수) $f(x|\theta)$를 갖는 모집단으로부터의 랜덤샘플이라 하자. 이때 $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$의 결합확률밀도함수는 다음과 같다.

$$f(\boldsymbol{x} \,|\, \theta) \;=\; f(x_1 \,|\, \theta) f(x_2 \,|\, \theta) \cdots f(x_n \,|\, \theta)$$

이 확률밀도함수는 주어진 θ에 대하여 $\boldsymbol{x}$를 관측할 확률밀도를 나타낸다. 확률밀도함수는 고정된 모수 θ의 값에서 $\boldsymbol{x}$의 함수라 할 수 있다. 최대가능도추정법의 기본적인 아이디어는 이러한 관점에서 벗어나서 결합확률밀도함수를 모수 θ의 함수로 생각한다는데 있다. 최대가능도추정법에서는 $\boldsymbol{X} = \boldsymbol{x}$를 관측한 후 $\boldsymbol{x}$는 고정되어 있는 것으로 생각하고 확률밀도함수 $f(\boldsymbol{x}|\theta)$를 θ의 함수로 간주하게 된다. 이처럼 확률밀도함수를 $\boldsymbol{x}$의 함수가 아닌 θ의 함수로 생각할 때, 이를 **가능도**(likelihood)라 하며 $L(\theta)$로 나타낸다. 즉 가능도는 다음과 같다.

$$L(\theta) \;=\; f(x_1 \,|\, \theta) f(x_2 \,|\, \theta) \cdots f(x_n \,|\, \theta) \tag{8.3}$$

기본적으로 가능도는 확률밀도함수 또는 확률질량함수를 나타내지만 이 함수를 보는 관점이 $\boldsymbol{x}$의 함수가 아니라 θ의 함수로 간주한다는데 차이점이 있다. 위 가능도 (8.3)을 최대로 하는 θ를 모수의 추정값으로 정하는 방법을 최대가능도추정법이라 하며, 이 추정량 $\widehat{\theta}$를 **최대가능도추정량**(maximum likelihood estimator)이라 한다. 즉 최대가능도추정량 $\widehat{\theta}$는 다음과 같이 정의할 수 있다.

$$\widehat{\theta} = \underset{\theta}{\operatorname{argmax}}\ L(\theta)$$

위 식에서 $\text{argmax}_\theta L(\theta)$는 함수 $L(\theta)$를 최대로 하는 θ를 말한다.

예 8.7 모수 θ는 θ_1, θ_2 또는 θ_3 중 하나의 값을 갖는다고 한다. 각 모수의 값에서 이산형 확률변수 X의 확률분포가 다음과 같을 때 X를 이용하여 모수 θ의 최대가능도추정량을 구하여라.

x	1	2	3	4	5	6	7
$f(x\|\theta_1)$	0.4	0.2	0.1	0.1	0.1	0.05	0.05
$f(x\|\theta_2)$	0.05	0.05	0.1	0.1	0.2	0.2	0.3
$f(x\|\theta_3)$	0.2	0.1	0.2	0.15	0.15	0.1	0.1

[풀이] 최대가능도추정량은 각 관측 값에서 가능도(확률질량함수)를 최대로 하는 모수의 값이다. 예를 들어 관측 값이 $X = 1$이라면 $f(1|\theta_1) = 0.4$, $f(1|\theta_2) = 0.05$, $f(1|\theta_3) = 0.2$이므로 θ_1에서 최대가 된다. 따라서 관측 값이 $X = 1$일 때 θ의 최대가능도추정량은 $\widehat{\theta} = \theta_1$이다. 마찬가지로 $X = x$일 때 최대가능도추정량 $\widehat{\theta}$는 θ_1, θ_2, θ_3 중에서 다음을 만족하는 θ이다.

$$f(x|\widehat{\theta}\,) \;=\; \max(f(x|\theta_1), f(x|\theta_2), f(x|\theta_3))$$

각 관측 값에서 최대가능도추정량을 구하면 다음과 같다.

x	1	2	3	4	5	6	7
$\widehat{\theta}$	θ_1	θ_1	θ_3	θ_3	θ_2	θ_2	θ_2

■

일반적으로 가능도 $L(\theta)$를 최대로 하는 θ를 구하는 것보다 가능도에 로그를 취한 로그가능도(log-likelihood)를 최대로 하는 θ를 구하는 것이 편리한 경우가 많다. 로그함수는 증가함수이므로 가능도 $L(\theta)$와 로그가능도 $\log L(\theta)$를 최대로 하는 θ의 값은 같다. 따라서 가능도 $L(\theta)$ 대신에 $\log L(\theta)$를 이용하여 최대가능도추정량을 구할 수 있다. 일반적으로 로그가능도를 $l(\theta)$로 나타내며 최대가능도추정량은 다음과 같이 나타낼 수 있다.

$$\begin{aligned}
\widehat{\theta} &= \underset{\theta}{\text{argmax}} \log L(\theta) \\
&= \underset{\theta}{\text{argmax}}\; l(\theta) \\
&= \underset{\theta}{\text{argmax}} \sum_{i=1}^{n} \log f(x_i \,|\, \theta)
\end{aligned}$$

로그가능도는 일반적으로 오목함수(위로 볼록한 함수, concave function)이며 θ에 대해 미분가능한 함수인 경우가 많다. 그림 8.1은 전형적인 로그가능도함수의 한 예를 보여주고 있다. 이 그림에서 주어진 로그가능도는 어느 점까지는 증가하고 그 이후로는 감소하는 위로 볼록한 형태를 보여주고 있다. 일반적으로 두 번 미분한 함수의 값이 항상 음수일 때 이 함수는 오목함수가 된다. 오목함수는 함수를 최대로 하는 값이 하나 존재하며 이 함수가 미분가능하면 미분한 값이 0이 되는 점에서 최대가 된다. 따라서 로그가능도를 최대로 하는 θ를 구하는 일반적인 방법은 $l(\theta)$를 미분한 값이 0이 되는 해를 구하는 것이다. 즉 θ의 최대가능도추정량 $\widehat{\theta}$는 다음을 을 만족하는 해이다.

$$\frac{d}{d\theta}\, l(\theta) \;=\; 0$$

예 8.8 $X_1, X_2, \ldots, X_n$이 Bernoulli(θ)로부터의 랜덤샘플일 때 θ의 최대가능도추정량을 구하여라.

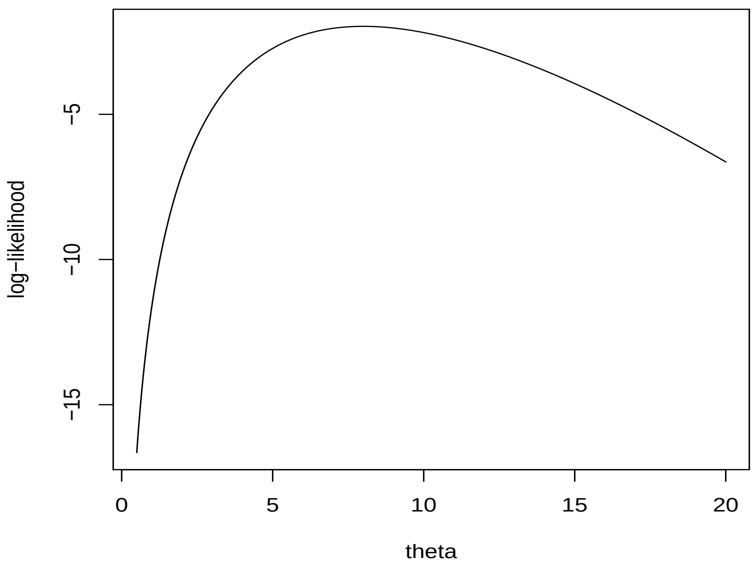

그림 8.1: 로그가능도함수의 예

[풀이] $X_1, X_2, \ldots, X_n$ 에 대한 로그가능도는 다음과 같다.

$$\begin{aligned} l(\theta) &= \sum_{i=1}^{n} \log f(x_i|\theta) = \sum_{i=1}^{n} \Big(x_i \log\theta + (1-x_i)\log(1-\theta) \Big) \\ &= (\log\theta)\sum_{i=1}^{n} x_i + (\log(1-\theta))\sum_{i=1}^{n}(1-x_i) \end{aligned}$$

이 함수를 θ 에 대해 미분하면 다음을 얻을 수 있다.

$$\frac{d}{d\theta}\, l(\theta) = \frac{\sum_{i=1}^{n} x_i}{\theta} - \frac{n - \sum_{i=1}^{n} x_i}{1-\theta} \tag{8.4}$$

$$\frac{d^2}{d\theta^2}\, l(\theta) = -\frac{\sum_{i=1}^{n} x_i}{\theta^2} - \frac{n - \sum_{i=1}^{n} x_i}{(1-\theta)^2} \tag{8.5}$$

식 (8.5) 로부터 $\frac{d^2}{d\theta^2}\, l(\theta) < 0$ 임을 알 수 있으며 이는 로그가능도 $l(\theta)$ 가 오목함수임을 나타낸다. 따라서 최대가능도추정량 $\widehat{\theta}$ 는 $\frac{d}{d\theta}\, l(\theta) = 0$ 을 만족하는 해이다. 식 (8.4) 로부터 θ 의 최대가능도추정량은 $\widehat{\theta} = \overline{X}$ 이다. ■

예 8.9 $X_1, X_2, \ldots, X_n$ 이 Poisson(θ) 로부터의 랜덤샘플일 때 θ 의 최대가능도추정량을 구하여라.

[풀이] $X_1, X_2, \ldots, X_n$ 에 대한 로그가능도는 다음과 같다.

$$\begin{aligned} l(\theta) &= \sum_{i=1}^{n} \log f(x_i|\theta) = \sum_{i=1}^{n} \Big(x_i \log\theta - \theta - \log(x_i!) \Big) \\ &= (\log\theta)\sum_{i=1}^{n} x_i - n\theta - \sum_{i=1}^{n} \log(x_i!) \end{aligned}$$

이 함수를 θ 에 대해 미분하면 다음을 얻을 수 있다.

$$\begin{aligned} \frac{d}{d\theta}\, l(\theta) &= \frac{\sum_{i=1}^{n} x_i}{\theta} - n \\ \frac{d^2}{d\theta^2}\, l(\theta) &= -\frac{\sum_{i=1}^{n} x_i}{\theta^2} \end{aligned}$$

따라서 로그가능도는 오목함수임을 확인할 수 있다. θ 의 최대가능도추정량은 $\frac{d}{d\theta}\, l(\theta) = 0$ 을 만족하는 해이므로 $\widehat{\theta} = \overline{X}$ 이다. ■

위의 두 예에서 다룬 최대가능도추정량은 적률추정량과 같지만 일반적으로 두 추정량이 같은 것은 아니다. 다음은 최대가능도추정량이 적률추정량과 다른 예를 보여주고 있다.

예 8.10 $X_1, X_2, \ldots, X_n$ 이 확률밀도함수 $f(x|\theta) = \frac{1}{2\theta} e^{-\frac{|x|}{\theta}}, \ -\infty < x < \infty$ 를 가지는 랜덤샘플일 때 θ 의 최대가능도추정량을 구하여라.

[풀이] $X_1, X_2, \ldots, X_n$ 에 대한 로그가능도는 다음과 같으며

$$l(\theta) \ = \ -n\log(2\theta) - \frac{\sum_{i=1}^n |x_i|}{\theta}$$

로그가능도를 미분한 함수는 다음과 같다.

$$\frac{d}{d\theta} l(\theta) \ = \ -\frac{n}{\theta} + \frac{\sum_{i=1}^n |x_i|}{\theta^2}$$

따라서 최대가능도추정량은 $\widehat{\theta} = \frac{1}{n}\sum_{i=1}^n |X_i|$ 이다. 이는 앞에서 구한 적률추정량 $\widehat{\theta} = \sqrt{\frac{1}{2n}\sum_{i=1}^n X_i^2}$ 과 다름을 알 수 있다. ■

모수가 $\theta_1, \ \theta_2$ 로 두 개인 경우는 모수가 한 개일 때 사용한 방법을 확장하여 최대가능도추정량을 구할 수 있다. 로그가능도 $l(\theta_1, \ \theta_2)$ 가 미분가능하고 오목함수일 때 로그가능도를 최대로 하는 θ_1 과 θ_2 는 $l(\theta_1, \ \theta_2)$ 를 θ_1 과 θ_2 로 미분한 값이 0이 되는 해이다. $l(\theta_1, \ \theta_2)$ 가 오목함수인지 판단하는 한 가지 방법은 아래와 같이 로그가능도를 두 번 미분한 행렬이 음의 정부호 행렬 (negative definite matrix)[1] 인지 확인하는 것이다.

$$\begin{pmatrix} \dfrac{d^2}{d\theta_1^2} l(\theta_1, \theta_2) & \dfrac{d^2}{d\theta_1 \, d\theta_2} l(\theta_1, \theta_2) \\ \dfrac{d^2}{d\theta_2 \, d\theta_1} l(\theta_1, \theta_2) & \dfrac{d^2}{d\theta_2^2} l(\theta_1, \theta_2) \end{pmatrix}$$

모든 로그가능도가 그러한 것은 아니지만 대부분의 로그가능도는 미분가능하며 오목함수이다. 따라서 로그가능도를 최대로 하는 θ_1 과 θ_2 를 구하는 일반적인 방법은 $l(\theta_1, \ \theta_2)$ 를 미분한 값이 0 이 되는 해를 구하는 것이다. 즉 θ_1 과 θ_2 의 최대가능도추정량 $\widehat{\theta}_1$ 과 $\widehat{\theta}_2$ 는 다음을 만족하는 해이다.

$$\begin{aligned} \frac{d}{d\theta_1} l(\theta_1, \ \theta_2) \ &= \ 0 \\ \frac{d}{d\theta_2} l(\theta_1, \ \theta_2) \ &= \ 0 \end{aligned}$$

다음 예는 모수가 두 개인 정규분포의 최대가능도추정량을 구하는 방법을 보여주고 있다.

예 8.11 $X_1, X_2, \ldots, X_n$ 이 $N(\mu, \sigma^2)$ 로부터의 랜덤샘플일 때 $\theta = (\mu, \ \sigma^2)$ 의 최대가능도추정량을 구하여라.

[1] 음의 정부호 행렬은 실수의 음수에 해당하는 개념으로 0이 아닌 모든 벡터 $\boldsymbol{x}$ 에 대하여 $\boldsymbol{x}^t A \boldsymbol{x} \ < \ 0$ 일 때 행렬 A 를 음의 정부호 행렬이라 한다.

[풀이] $X_1, X_2, \ldots, X_n$ 의 결합확률밀도함수는 다음과 같으므로

$$\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x_i-\mu)^2}{2\sigma^2}} = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

로그가능도는 다음과 같다.

$$l(\mu,\, \sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

로그가능도를 μ 와 σ^2 에 대해 미분하면 다음을 얻을 수 있다.

$$\begin{aligned} \frac{d}{d\mu}\, l(\mu,\sigma^2) &= \frac{\sum_{i=1}^{n}(x_i-\mu)}{\sigma^2} \\ \frac{d}{d\sigma^2}\, l(\mu,\sigma^2) &= -\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2(\sigma^2)^2} \end{aligned}$$

한 가지 주의할 사항은 위의 로그가능도를 미분할 때 σ^2 을 하나의 문자로 간주하여 미분해야 한다. μ 와 σ^2 의 최대가능도추정량은 $\frac{d}{d\mu}\, l(\mu,\sigma^2) = 0$ 과 $\frac{d}{d\sigma^2}\, l(\mu,\sigma^2) = 0$ 을 만족하는 해이므로 다음과 같다.

$$\widehat{\mu} = \overline{X}, \quad \widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

■

지금까지는 가능도가 미분가능한 경우를 다루었다. 따라서 미분을 통하여 최대가능도추정량을 구할 수 있었다. 그러나 모든 가능도가 미분가능하지는 않다. 다음은 가능도가 미분가능하지 않은 예를 보여주고 있다.

예 8.12 $X_1, X_2, \ldots, X_n$ 이 확률밀도함수 $f(x|\theta) = \exp(-(x-\theta)),\ x \geq \theta$ 를 갖는 랜덤샘플일 때 θ 의 최대가능도추정량을 구하여라.

[풀이] $X_1, X_2, \ldots, X_n$ 의 확률밀도함수는 다음과 같다.

$$\begin{aligned} & f(x_1, x_2, \ldots, x_n \,|\, \theta) \\ &= e^{-(x_1-\theta)} e^{-(x_2-\theta)} \cdots e^{-(x_n-\theta)} I(x_1 \geq \theta) I(x_2 \geq \theta) \cdots I(x_n \geq \theta) \\ &= e^{-(x_1-\theta)} e^{-(x_2-\theta)} \cdots e^{-(x_n-\theta)} I(x_{(1)} \geq \theta) \end{aligned}$$

이를 θ의 함수로 나타낸 가능도의 그래프는 그림 8.2와 같은 형태를 가진다. 이 가능도는 $\theta \le x_{(1)}$ 인 영역에서는 증가함수이고 $\theta > x_{(1)}$ 인 영역에서는 0의 값을 가지므로 $\theta = x_{(1)}$ 에서 최대가 된다. 따라서 θ의 최대가능도추정량은 $\widehat{\theta} = X_{(1)}$ 이다. ■

최대가능도추정량이 가지는 성질 중의 하나는 변환된 모수의 최대가능도추정량은 원래 모수의 최대가능도추정량을 그대로 변환하면 된다는 사실이다. 즉 θ의 최대가능도추정량을 $\widehat{\theta}$라 할 때 $h(\theta)$의 최대가능도추정량은 $h(\widehat{\theta})$가 된다. 즉 최대가능도추정법에 의한 추정량은 다음 관계가 성립한다.

$$\widehat{h(\theta)} \;=\; h(\widehat{\theta})$$

예를 들어 θ의 최대가능도추정량이 $\overline{X}$라면 θ^2의 최대가능도추정량은 $(\overline{X})^2$이 된다. 일반적인 추정량은 이러한 성질을 가지고 있지 않다.

정리 8.1 [Invariance property] θ의 최대가능도추정량이 $\widehat{\theta}$일 때 $h(\theta)$의 최대가능도추정량은 $h(\widehat{\theta})$ 이다.

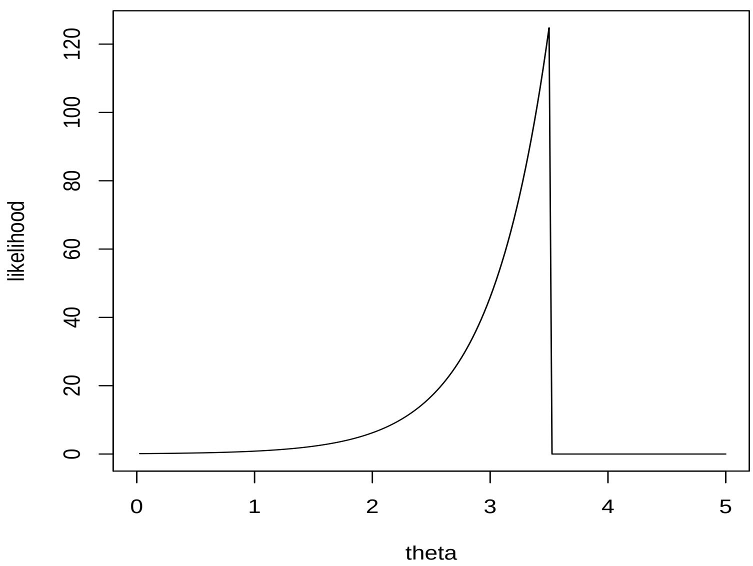

그림 8.2: 미분가능하지 않은 가능도의 예

[증명] 여기서는 증명을 간단히 하기 위해 $h(\theta)$가 일대일 함수라고 가정을 하겠다. $\eta = h(\theta)$라 하면 θ는 $h^{-1}(\eta)$로 나타낼 수 있다. 로그가능도를 η로 나타낸 함수를 $m(\eta)$라 할 때 $m(\eta)$는 다음과 같이 나타낼 수 있다.

$$m(\eta) \ = \ l(\theta) \ = \ l(h^{-1}(\eta))$$

η의 최대가능도추정량은 $m(\eta)$를 최대로 하는 η이므로 $\widehat{\eta}$는 다음과 같다.

$$\widehat{\eta} \ = \ \underset{\eta}{\text{argmax}} \ m(\eta) \ = \ \underset{\eta}{\text{argmax}} \ l(h^{-1}(\eta))$$

한편 $l(\theta)$는 $\theta = \widehat{\theta}$일 때 최대가 되므로 $m(\eta)$는 $h^{-1}(\eta) = \widehat{\theta}$일 때 최대가 된다. 즉 $\eta = h(\widehat{\theta})$일 때 $m(\eta)$는 최대가 된다. 따라서 η의 최대가능도추정량은 $h(\widehat{\theta})$가 된다. $h(\theta)$가 일대일 함수가 아닌 경우는 $h^{-1}(\eta)$를 $h(\theta) = \eta$인 θ의 집합으로 간주하고 위의 증명 과정을 이용하면 $h(\theta)$의 최대가능도추정량이 $h(\widehat{\theta})$임을 보일 수 있다. ■

정리 8.1의 증명 과정은 예를 통해 살펴보면 이해가 빠를 수도 있다. $X_1, X_2, \ldots, X_n$이 Poisson(θ)로부터의 랜덤샘플일 때 $\eta = \theta^2$의 최대가능도추정량을 구하는 문제를 생각하여 보자. 포아송분포로부터의 랜덤샘플 $X_1, X_2, \ldots, X_n$에 대한 로그가능도는 다음과 같다.

$$l(\theta) \ = (\log\theta)\sum_{i=1}^{n} x_i - n\theta - \sum_{i=1}^{n}\log(x_i!)$$

$\theta = \sqrt{\eta}$이므로 η의 함수로 나타낸 로그가능도 $l(\eta)$는 다음과 같다.

$$l(\eta) \ = \ \left(\frac{1}{2}\log\eta\right)\sum_{i=1}^{n} x_i - n\sqrt{\eta} - \sum_{i=1}^{n}\log(x_i!)$$

η의 최대가능도추정량을 구하기 위해 $l(\eta)$를 미분하면 다음을 얻을 수 있다.

$$\frac{d}{d\eta}l(\eta) \ = \ \frac{\sum_{i=1}^{n} x_i}{2\eta} - \frac{n}{2\sqrt{\eta}}$$

위 식이 0이 될 때는 다음을 만족할 때이다.

$$\frac{1}{n}\sum_{i=1}^{n} x_i \ = \ \sqrt{\eta} \tag{8.6}$$

따라서 η의 최대가능도추정량은 다음과 같다.

$$\widehat{\eta} \ = \ (\overline{X})^2$$

이는 $\widehat{\theta} = \overline{X}$ 이므로 앞의 정리가 성립함을 확인할 수 있다. 이 예에서 $\theta = \sqrt{\eta}$이고 $\widehat{\theta} = \overline{X}$ 이므로 위의 식 (8.6) 은 정리 8.1의 증명에서 $\widehat{\theta} = h^{-1}(\eta)$ 를 나타내는 식이라 할 수 있다.

8.3 추정량의 성질

모수 θ의 추정 방법은 여러 가지가 있을 수 있다. 중요한 문제는 여러 가지 추정법 중에서 최적의 추정 방법을 구하는 것이다. 최적의 추정량을 구하기 위해서는 먼저 추정량이 가지는 성질을 알아야 할 것이다.

정의 8.1 (i) $E(\widehat{\theta}) = \theta$일 때 $\widehat{\theta}$를 **비편향추정량**(unbiased estimator) 또는 **불편추정량**이라 한다.

(ii) 표본의 크기 n이 증가함에 따라 $\widehat{\theta} \xrightarrow{p} \theta$일 때 $\widehat{\theta}$를 **일치추정량**(consistent estimator)이라 한다.

θ의 추정량 $\widehat{\theta}$에 대하여 $(E(\widehat{\theta}) - \theta)$를 $\widehat{\theta}$의 **편향**(bias) 또는 **편차**라 하고 $Bias(\widehat{\theta})$로 나타낸다. 비편향추정량은 편향이 0인 추정량을 말한다. 즉 추정량의 기댓값이 추정하는 모수의 값과 같을 때 비편향추정량이라 한다. 일치추정량은 표본의 크기가 증가할 때 추정량이 모수의 값으로 수렴하는 추정량을 말한다. 일반적으로 비편향추정량이 일치추정량일 필요는 없으며 또한 일치추정량이 비편향추정량일 필요도 없다.

예 8.13 $X_1, X_2, \ldots, X_n$ 이 기댓값 μ, 분산 $\sigma^2 > 0$인 모집단으로부터의 랜덤샘플이라 한다.

(a) X_1은 μ의 비편향추정량이지만 일치추정량은 아님을 보여라.

(b) $\overline{X}$는 μ의 비편향추정량이며 또한 일치추정량임을 보여라.

(c) $\dfrac{n}{n+1}\overline{X}$는 μ의 비편향추정량은 아니지만 일치추정량임을 보여라.

[풀이]

(a) $E(X_1) = \mu$이므로 X_1은 μ의 비편향추정량임을 알 수 있다. 한편 표본의 크기 n에 상관없이 X_1은 항상 같은 확률분포를 가지며 이는 μ로 수렴하지 않는다. 따라서 X_1은 일치추정량이 아니다.

(b) $E(\overline{X}) = \mu$이므로 $\overline{X}$는 μ의 비편향추정량이며, 정리 6.1의 대수의 법칙에서 보인 것과 같이 표본평균은 모평균으로 수렴하므로 $\overline{X}$는 μ의 일치추정량이다.

(c) $E\left(\frac{n}{n+1}\overline{X}\right) = \frac{n}{n+1}\mu \neq \mu$이므로 $\frac{n}{n+1}\overline{X}$는 μ의 비편향추정량은 아니다. 한편 $\frac{n}{n+1}$은 1로 수렴하며 $\overline{X}$는 μ로 수렴하므로 $\frac{n}{n+1}\overline{X}$는 μ로 수렴하게 된다. 따라서 $\frac{n}{n+1}\overline{X}$는 μ의 일치추정량이다.

■

추정량이 가져야 할 바람직한 성질 중의 하나는 일치추정량이어야 한다는 것이다. 어느 추정량이 일치추정량이 아니라면 아무리 표본의 크기를 크게 하여도 모수의 참값에 가까이 가지 않는다는 것을 의미하므로 이 추정량은 좋은 추정량이 아니라고 할 수 있다. 앞의 예에서 모평균의 추정량으로 $\frac{n}{n+1}\overline{X}$와 X_1 중 어느 추정량이 더 의미가 있는지 살펴보도록 하자. X_1은 비편향추정량이므로 추정량의 기댓값은 μ이지만 표본의 크기가 증가하여도 이 추정량은 모평균으로 가까이 가지 않는다. 한편 $\frac{n}{n+1}\overline{X}$는 편향이 $-\frac{1}{n+1}\mu$이므로 비편향추정량은 아니다. 그러나 표본의 크기가 증가하면 이 추정량은 μ로 수렴하게 된다. 즉 표본의 크기가 충분히 크면 $\frac{n}{n+1}\overline{X}$는 추정하는 모평균에 충분히 가까이 있음을 알 수 있다. 따라서 μ에 대한 추정량으로 X_1보다는 $\frac{n}{n+1}\overline{X}$이 더 적절하다고 판단할 수 있다.

모수 θ의 추정량 $\widehat{\theta}$가 일치추정량임을 확인할 수 있는 한 가지 방법은 정리 4.6의 체비세프 부등식을 이용하는 것이다. 체비세프 부등식은 확률변수 X의 기댓값이 μ일 때 양의 상수 t에 대하여 다음이 성립한다.

$$P(\,|X-\mu| \geq t\,) \;\leq\; \frac{\sigma^2}{t^2}$$

그런데 체비세프 부등식의 증명 과정을 이용하면 모든 확률변수 X에 대하여 다음 부등식이 성립함을 보일 수 있다.

$$P(\,|X| \geq t\,) \;\leq\; \frac{E(X^2)}{t^2} \tag{8.7}$$

위의 부등식 (8.7)을 마르코프 부등식(Markov inequality)이라 한다. 이 부등식을 $(\widehat{\theta}-\theta)$에 적용하면 다음 부등식을 얻을 수 있다.

$$P(\,|\widehat{\theta}-\theta| \geq t\,) \;\leq\; \frac{E(\widehat{\theta}-\theta)^2}{t^2}$$

한편 $E(\widehat{\theta}-\theta)^2$은 다음과 같이 나타낼 수 있다.

$$\begin{aligned}
&E(\widehat{\theta}-\theta)^2 \\
&\quad= E\left(\widehat{\theta}-E(\widehat{\theta}\,)+E(\widehat{\theta}\,)-\theta\right)^2 \\
&\quad= E(\widehat{\theta}-E(\widehat{\theta}\,))^2+2E(\widehat{\theta}-E(\widehat{\theta}\,))(E(\widehat{\theta}\,)-\theta)+(E(\widehat{\theta}\,)-\theta)^2 \\
&\quad= E(\widehat{\theta}-E(\widehat{\theta}\,))^2+(E(\widehat{\theta}\,)-\theta)^2 \\
&\quad= Var(\widehat{\theta}\,)+(Bias(\widehat{\theta}\,))^2
\end{aligned}$$

따라서 $Var(\widehat{\theta}\,)$와 $Bias(\widehat{\theta}\,)$ 모두 0으로 수렴한다면 $\widehat{\theta}$는 일치추정량이 된다.

예 8.14 $X_1, X_2, \dots, X_n$ 이 Poisson(θ) 로부터의 랜덤샘플일 때 $\widehat{\theta} = \overline{X}$ 는 θ 의 일치추정량임을 보여라.

[풀이] $\overline{X}$ 의 기댓값은 $E(\overline{X}) = \theta$ 이므로 $\overline{X}$ 의 편향은 0이고 분산 $Var(\overline{X}) = \frac{\theta}{n}$ 는 0으로 수렴하므로 $\overline{X}$ 는 θ 의 일치추정량이다. ■

예 8.15 $X_1, X_2, \dots, X_n$ 이 $N(\mu, \sigma^2)$ 으로부터의 랜덤샘플이라 한다.

(a) μ 의 최대가능도추정량이 일치추정량임을 보여라.

(b) σ^2 의 최대가능도추정량이 일치추정량임을 보여라.

[풀이]

(a) μ 의 최대가능도추정량 $\widehat{\mu} = \overline{X}$ 에 대하여 다음이 성립하므로

$$E(\overline{X}) = \mu, \quad Var(\overline{X}) = \frac{\sigma^2}{n}$$

편향은 0이고 분산은 0으로 수렴한다. 따라서 $\overline{X}$ 는 μ 의 일치추정량이다.

(b) σ^2 의 최대가능도추정량은 다음과 같다.

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

한편 $\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \overline{X})^2 \sim \chi^2_{n-1}$ 이므로 다음을 얻을 수 있다.

$$\begin{aligned} E\left(\frac{1}{\sigma^2}\textstyle\sum_{i=1}^{n}(X_i - \overline{X})^2\right) &= (n-1) \\ Var\left(\frac{1}{\sigma^2}\textstyle\sum_{i=1}^{n}(X_i - \overline{X})^2\right) &= 2(n-1) \end{aligned}$$

따라서 $\widehat{\sigma}^2$ 의 편향과 분산은 다음과 같다.

$$Bias(\widehat{\sigma}^2) = E\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2\right) - \sigma^2 = \frac{n-1}{n}\sigma^2 - \sigma^2 = -\frac{1}{n}\sigma^2 \quad (8.8)$$

$$\begin{aligned} Var(\widehat{\sigma}^2) &= Var\left(\frac{\sigma^2}{n}\cdot\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{\sigma^2}\right) \\ &= \frac{(\sigma^2)^2}{n^2}Var\left(\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{\sigma^2}\right) = \frac{2(n-1)}{n^2}(\sigma^2)^2 \end{aligned} \quad (8.9)$$

$\widehat{\sigma}^2$ 의 편향과 분산 모두 0 으로 수렴하므로 $\widehat{\sigma}^2$ 은 σ^2 의 일치추정량이다.

■

8.4 추정량의 효율과 크래머-라오 부등식

이 절에서는 추정량의 비교 방법을 알아보도록 하자. 모수 θ에 대한 여러 추정 방법 중에서 어느 것이 우수한지 판단하기 위해서는 추정량의 타당한 비교 기준을 먼저 설정해야 할 것이다. 기본적으로 추정량이 θ에 가까울수록 좋은 추정량이라 할 수 있다. 한 가지 문제점은 θ의 값을 알지 못하기 때문에 추정량이 θ에 얼마만큼 가까운지 정확히 판단할 수는 없다. 따라서 추정량의 비교는 확률적으로 할 수밖에 없다. 즉 어떤 추정량이 다른 추정량에 비해 θ에 가까이 있을 확률이 높다면 더 우수한 추정량이라 할 수 있다. 만약 비교하는 추정량이 모두 비편향추정량이라면 분산을 이용하여 비교할 수 있다. 비편향추정량의 기댓값은 모수의 참값과 같으므로 분산이 작을수록 추정량이 모수의 참값에 가까이 있을 것으로 기대되기 때문이다.

정의 8.2 θ의 두 비편향추정량 $\widehat{\theta}$와 $\tilde{\theta}$에 대하여 다음과 같은 분산의 비율을

$$\mathrm{eff}\,(\widehat{\theta}, \tilde{\theta}) \;=\; \frac{Var(\tilde{\theta})}{Var(\widehat{\theta}\,)}$$

$\tilde{\theta}$에 대한 $\widehat{\theta}$의 효율(efficiency)이라 한다.

두 비편향추정량 $\widehat{\theta}$와 $\tilde{\theta}$에 대하여 $\mathrm{eff}\,(\widehat{\theta}, \tilde{\theta}\,) \;>\; 1$ 이면 $\widehat{\theta}$의 분산이 $\tilde{\theta}$의 분산보다 작다는 것을 의미하므로 $\widehat{\theta}$가 더 좋은 추정량이라 할 수 있다.

예 8.16 $X_1, X_2, \ldots, X_n$ 이 평균 μ, 분산 σ^2 인 모집단으로부터의 랜덤샘플일 때 μ의 비편향추정량 $\tilde{\mu} = X_1$ 에 대한 $\widehat{\mu} = \overline{X}$ 의 효율을 구하여라.

[풀이] $\tilde{\mu} = X_1$ 에 대한 $\widehat{\mu} = \overline{X}$ 의 효율은 다음과 같으므로

$$\mathrm{eff}\,(\overline{X}, X_1) \;=\; \frac{Var(X_1)}{Var(\overline{X}\,)} \;=\; \frac{\sigma^2}{\sigma^2/n} \;=\; n$$

X_1 보다 $\overline{X}$ 가 더 좋은 추정량이라 할 수 있다. ■

추정량의 비교 기준으로 위의 $\mathrm{eff}(\widehat{\theta}, \tilde{\theta}\,)$를 사용하는 것은 두 추정량이 모두 비편향추정량인 경우는 적절하지만 비편향추정량이 아닌 경우는 적절하지 않다. 일반적인 추정량에 대해서는 기댓값이 θ가 아닐 수 있으므로 추정량의 편향과 분산을 함께 고려하여야 한다. 추정량의 편향과

분산이 작을수록 좋은 추정량이라 할 수 있다. 이 두 가지를 함께 고려한 것이 **평균제곱오차**(mean squared error)이다. 추정량 $\widehat{\theta}$와 모수 θ의 차이의 제곱에 대한 기댓값을 평균제곱오차라 하며 $MSE(\widehat{\theta})$로 나타낸다. 즉 추정량 $\widehat{\theta}$의 평균제곱오차는 다음과 같이 정의할 수 있다.

$$MSE(\widehat{\theta}) = E(\widehat{\theta} - \theta)^2$$

이 값이 작을수록 추정량이 모수에 가까이 있다고 할 수 있으므로 평균제곱오차가 작을수록 더 좋은 추정량이라 할 수 있다. 한편 $E(\widehat{\theta} - \theta)^2$는 다음과 같이 나타낼 수 있다.

$$\begin{aligned} E(\widehat{\theta} - \theta)^2 &= E[(\widehat{\theta} - E(\widehat{\theta})) + (E(\widehat{\theta}) - \theta)]^2 = E(\widehat{\theta} - E(\widehat{\theta}))^2 + (E(\widehat{\theta}) - \theta)^2 \\ &= Var(\widehat{\theta}) + (Bias(\widehat{\theta}))^2 \end{aligned}$$

즉 평균제곱오차는 추정량의 분산과 편향의 제곱의 합이다.

예 8.17 $X_1, X_2, \ldots, X_n$이 $N(\mu, \sigma^2)$으로부터의 랜덤샘플일 때 다음과 같이 주어진 μ와 σ^2의 추정량의 평균제곱오차를 구하여라.

$$\widehat{\mu} = \overline{X}, \quad \widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

[풀이] $\overline{X}$는 μ의 비편향추정량이고 분산은 $\frac{\sigma^2}{n}$이므로 $\overline{X}$의 평균제곱오차는 다음과 같다.

$$MSE(\widehat{\mu}) = Var(\widehat{\mu}) = \frac{\sigma^2}{n}$$

σ^2의 추정량 $\widehat{\sigma}^2$의 편향과 분산에 대한 식 (8.8)과 (8.9)으로부터 $\widehat{\sigma}^2$의 평균제곱오차는 다음과 같다.

$$\begin{aligned} MSE(\widehat{\sigma}^2) &= Var(\widehat{\sigma}^2) + (Bias(\widehat{\sigma}^2))^2 \\ &= \frac{2(n-1)}{n^2}(\sigma^2)^2 + \left(-\frac{\sigma^2}{n}\right)^2 = \frac{(2n-1)}{n^2}(\sigma^2)^2 \end{aligned}$$

■

추정량의 평균제곱오차를 이용하여 추정량이 일치추정량인지 확인할 수도 있다. 마르코프 부등식에 의하여 다음 부등식이 성립하므로

$$P(|\widehat{\theta} - \theta| \geq t) \leq \frac{E(\widehat{\theta} - \theta)^2}{t^2} = \frac{MSE(\widehat{\theta})}{t^2}$$

추정량의 평균제곱오차가 0으로 수렴하면 이 추정량은 일치추정량이다. 앞의 예에서 $\widehat{\sigma}^2$의 평균제곱오차인 $\frac{(2n-1)}{n^2}(\sigma^2)^2$은 n이 증가할 때 0으로 수렴하므로 $\widehat{\sigma}^2$은 σ^2의 일치추정량임을 알 수 있다.

지금부터는 최적의 추정량을 구하는 방법에 대해 알아보도록 하자. 추정량의 비교 기준으로 평균제곱오차를 사용할 수 있으므로 이 값을 최소로 하는 추정량이 최적의 추정량이라 할 수 있다. 그런데 추정량 $\widehat{\theta}$의 평균제곱오차는 모수 θ의 값에 따라 달라지는 함수이다.

$$MSE(\widehat{\theta}\,) \;=\; E(\widehat{\theta}-\theta)^2$$

따라서 θ의 함수로 주어지는 추정량의 평균제곱오차를 비교하는 것은 쉽지 않다.

예를 들어 베르누이분포에서 θ의 최대가능도추정량 $\widehat{\theta}=\overline{X}$와 자료와 상관없이 θ의 값을 항상 $\dfrac{1}{2}$로 추정한 $\tilde{\theta}=\dfrac{1}{2}$ 중 어느 추정량이 더 좋은지 살펴보도록 하자. $\widehat{\theta}=\overline{X}$는 θ의 비편향추정량이므로 평균제곱오차는 다음과 같으며

$$MSE(\widehat{\theta}) \;=\; Var(\overline{X}) \;=\; \frac{\theta(1-\theta)}{n}$$

$\tilde{\theta}$의 평균제곱오차는 다음과 같다.

$$MSE(\tilde{\theta}) \;=\; E\left(\frac{1}{2}-\theta\right)^2 \;=\left(\frac{1}{2}-\theta\right)^2$$

두 추정량을 비교하면 $\theta=\dfrac{1}{2}$인 근처를 제외한 영역에서는 $\widehat{\theta}$의 평균제곱오차가 더 작다. 특히 표본의 크기 n이 증가할수록 $\widehat{\theta}$의 평균제곱오차가 작아지는 영역은 더욱 넓어진다. 그러나 $\theta=\dfrac{1}{2}$일 때 $\tilde{\theta}$의 평균제곱오차는 0이므로 n이 아무리 커질지라도 $\widehat{\theta}$의 평균제곱오차가 더 작아질 수는 없다. 즉 n이 충분히 크면 $\theta=\dfrac{1}{2}$인 경우를 제외하고는 $\widehat{\theta}=\overline{X}$의 평균제곱오차가 더 작다고 할 수 있다. 그러나 θ의 값을 알지 못하기 때문에 $\widehat{\theta}$가 $\tilde{\theta}$보다 항상 더 좋은 추정량이라고 주장하기는 힘들다.

위에서 설명한 것처럼 θ의 함수로 주어지는 평균제곱오차를 이용하여 모든 추정량을 비교하는 것은 쉽지 않다. 이러한 점 때문에 모든 추정량 중에서 가장 좋은 추정량을 찾는다는 것은 거의 불가능하다고 생각할 수 있다. 한 가지 대안은 특정한 성질을 가지는 추정량으로 한정하고 이 중에서 최적의 추정량을 구하는 것이다. 여기에서는 추정량의 편향이 0인 비편향추정량 중에서 최적의 추정량을 구하는 문제를 생각하여 보자. 비편향추정량의 경우는 평균제곱오차가 분산과 같으므로 분산을 최소로 하는 것이 최적의 추정량이라 할 수 있다.

비편향추정량 중에서 분산을 최소로 하는 추정량을 구하는 방법을 알아보도록 하자. 한가지 알려진 사실은 적절한 조건 하에서 비편향추정량의 분산의 하한이 존재한다는 것이다. 즉 비편향추정량의 분산은 어떤 값보다 작을 수는 없다. 이 하한은 모집단의 확률분포에 의하여 결정된다.

확률밀도함수와 확률질량함수에 대하여 다음 정리를 얻을 수 있다.

정리 8.2 확률변수 X의 확률밀도함수(확률질량함수)가 $f(x|\theta)$일 때 적절한 조건하에서 다음이

성립한다.

$$E\left(\tfrac{d}{d\theta}\log f(X\,|\,\theta)\right) \;=\; 0, \tag{8.10}$$

$$E\left(\tfrac{d}{d\theta}\log f(X\,|\,\theta)\right)^2 \;=\; -E\left(\frac{d^2}{d\theta^2}\log f(X\,|\,\theta)\right) \tag{8.11}$$

[증명] 증명은 연속형 확률분포인 경우만 다루도록 하겠다. 먼저 $\frac{d}{d\theta}\log f(X\,|\,\theta)$의 기대값은 다음과 같이 구할 수 있다.

$$\begin{aligned}
&E\left(\frac{d}{d\theta}\log f(X\,|\,\theta)\right)\\
&= \int_{-\infty}^{\infty}\left(\frac{d}{d\theta}\log f(x\,|\,\theta)\right) f(x|\theta)\;dx\\
&= \int_{-\infty}^{\infty}\frac{\frac{d}{d\theta}f(x\,|\,\theta)}{f(x\,|\,\theta)}\,f(x\,|\,\theta)\;dx \;=\; \int_{-\infty}^{\infty}\left(\frac{d}{d\theta}f(x\,|\,\theta)\right)\;dx\\
&= \frac{d}{d\theta}\left(\int_{-\infty}^{\infty}f(x\,|\,\theta)\;dx\right) \;=\; \frac{d}{d\theta}\,(1) \;=\; 0
\end{aligned}$$

한편 로그 확률밀도함수를 θ에 대하여 두 번 미분하면 다음을 얻을 수 있다.

$$\begin{aligned}
&\frac{d^2}{d\theta^2}\log f(x\,|\,\theta) \;=\; \frac{d}{d\theta}\left(\frac{\frac{d}{d\theta}f(x\,|\,\theta)}{f(x\,|\,\theta)}\right)\\
&= \frac{\left(\frac{d^2}{d\theta^2}f(x\,|\,\theta)\right)f(x\,|\,\theta)-\left(\frac{d}{d\theta}f(x\,|\,\theta)\right)^2}{(f(x\,|\,\theta))^2}\\
&= \frac{\frac{d^2}{d\theta^2}f(x\,|\,\theta)}{f(x\,|\,\theta)}-\left(\frac{\frac{d}{d\theta}f(x\,|\,\theta)}{f(x\,|\,\theta)}\right)^2 \;=\; \frac{\frac{d^2}{d\theta^2}f(x\,|\,\theta)}{f(x\,|\,\theta)}-\left(\frac{d}{d\theta}\log f(x\,|\,\theta)\right)^2
\end{aligned}$$

이다. 따라서 식 (8.11)의 우측 값은 다음과 같다.

$$\begin{aligned}
&-E\left(\frac{d^2}{d\theta^2}\log f(X\,|\,\theta)\right)\\
&\quad= -E\left(\frac{\frac{d^2}{d\theta^2}f(X\,|\,\theta)}{f(X\,|\,\theta)}\right) \;+\; E\left(\frac{d}{d\theta}\log f(X\,|\,\theta)\right)^2
\end{aligned}$$

한편 다음이 성립하므로

$$\begin{aligned} E\left(\frac{\frac{d^2}{d\theta^2}f(X\,|\,\theta)}{f(X\,|\,\theta)}\right) &= \int_{-\infty}^{\infty}\left(\frac{\frac{d^2}{d\theta^2}f(x\,|\,\theta)}{f(x\,|\,\theta)}\right)f(x\,|\,\theta)\;dx \\ &= \int_{-\infty}^{\infty}\frac{d^2}{d\theta^2}f(x\,|\,\theta)\;dx \;=\; \frac{d^2}{d\theta^2}\left(\int_{-\infty}^{\infty}f(x\,|\,\theta)\;dx\right) \\ &= \frac{d^2}{d\theta^2}\,(1) \;=\; 0 \end{aligned}$$

$E\left(\frac{d}{d\theta}\log f(X\,|\,\theta)\right)^2 = -E\left(\frac{d^2}{d\theta^2}\log f(X\,|\,\theta)\right)$ 임을 알 수 있다. ■

앞의 정리에서 로그가능도는 미분가능하고 미분과 적분의 순서를 바꾸어 계산을 하였다. 이와 같이 적절한 조건이란 로그가능도의 미분가능성과 적분과 미분의 순서를 바꿀 수 있는 조건 등을 말한다.

확률밀도함수 또른 확률질량함수가 $f(x|\theta)$ 인 확률분포를 따르는 확률변수 X 에 대하여 $\left(\frac{d}{d\theta}\log f(X\,|\,\theta)\right)^2$ 의 기대값을 **피셔 정보수**(Fisher information number) 또는 간단히 **정보수**라 하며 기호로 $I(\theta)$ 와 같이 나타낸다. 즉 피셔 정보수는 다음과 같다.

$$I(\theta) \;=\; E\left(\frac{d}{d\theta}\log f(X\,|\,\theta)\right)^2 \;=\; -E\left(\frac{d^2}{d\theta^2}\log f(X\,|\,\theta)\right)$$

정리 8.2를 포아송분포를 통하여 확인하여 보자. 포아송분포의 확률질량함수는 $f(x\,|\,\theta) = \frac{\theta^x\,e^{-\theta}}{x!}$ 이므로 $\log f(x\,|\,\theta) = x\log\theta - \theta - \log(x!)$ 이다. 이를 θ 에 대해 미분하면 다음과 같다.

$$\frac{d}{d\theta}\log f(x\,|\,\theta) \;=\; \frac{x}{\theta} - 1$$

따라서 다음이 성립함을 알 수 있다.

$$E\left(\frac{d}{d\theta}\log f(X\,|\,\theta)\right) \;=\; E\left(\frac{x}{\theta} - 1\right) \;=\; 0$$

또한 $\left(\frac{d}{d\theta}\log f(X\,|\,\theta)\right)^2$ 의 기대값은 다음과 같다.

$$E\left(\frac{d}{d\theta}\log f(X\,|\,\theta)\right)^2 \;=\; E\left(\frac{X}{\theta} - 1\right)^2 \;=\; \frac{E(X-\theta)^2}{\theta^2} \;=\; \frac{1}{\theta}$$

한편 $\log f(x\,|\,\theta)$ 를 θ 에 대해 두 번 미분하면 다음과 같으므로

$$\frac{d^2}{d\theta^2}\log f(x\,|\,\theta) \;=\; -\tfrac{x}{\theta^2}$$

$-\frac{d^2}{d\theta^2}\log f(X\,|\,\theta)$ 의 기대값은 다음과 같다.

$$-E\left(\frac{d^2}{d\theta^2}\log f(X\,|\,\theta)\right) \;=\; E\left(\frac{X}{\theta^2}\right) \;=\; \frac{1}{\theta}$$

따라서 $E\left(\frac{d}{d\theta}\log f(X\,|\,\theta)\right)^2 \;=\; -E\left(\frac{d^2}{d\theta^2}\log f(X|\theta)\right)$ 임을 확인할 수 있다.

예 8.18 Bernoulli(θ) 에 대한 피셔 정보수를 구하여라.

[풀이] 베르누이분포의 확률질량함수는 $f(x\,|\,\theta) = \theta^x(1-\theta)^{1-x}$ 이므로 다음을 얻을 수 있다.

$$\frac{d^2}{d\theta^2}\log f(x\,|\,\theta) \;=\; -\frac{x}{\theta^2} - \frac{1-x}{(1-\theta)^2}$$

따라서 베르누이분포의 피셔 정보수는 다음과 같다.

$$I(\theta) \;=\; E\left(\frac{X}{\theta^2} + \frac{1-X}{(1-\theta)^2}\right) \;=\; \frac{1}{\theta(1-\theta)}$$

■

정리 8.3 임의의 두 확률변수 U 와 V 에 대하여 다음 부등식이 성립한다.

$$(Cov(U,V))^2 \;\le\; Var(U)\,Var(V)$$

[증명] 분산은 항상 0보다 크거나 같으므로 모든 실수 a 에 대하여 다음이 성립한다.

$$Var(U+aV) = Var(U) + 2aCov(U,V) + a^2Var(V) \;\ge\; 0$$

위의 식을 a 에 대한 이차다항식으로 볼 때 항상 0보다 크거나 같기 위해서는 판별식이 0보다 작거나 같아야 한다. 따라서 다음을 얻을 수 있다.

$$(Cov(U,V))^2 - Var(U)\,Var(V) \;\le\; 0$$

■

위의 정리를 이용하여 크래머-라오 부등식을 보일 수 있다.

정리 8.4 [크래머-라오 부등식, Cramer-Rao inequality] 랜덤샘플 $X_1, X_2, \ldots, X_n$ 을 이용한 θ의 비편향추정량 $T = T(\boldsymbol{X})$에 대하여 적절한 조건 하에서 다음 부등식이 성립한다.

$$Var(T) \;\geq\; \frac{1}{nI(\theta)}$$

위 부등식에서 $\dfrac{1}{nI(\theta)}$ 을 크래머-라오 하한(Cramer-Rao lower bound)이라 한다.

[증명] 정리 8.3에서 U와 V를 다음과 같이 정의하자.

$$\begin{aligned} U &= T(\boldsymbol{X}) \\ V &= \frac{d}{d\theta}\log f(\boldsymbol{X}\,|\,\theta) \;=\; \sum_{i=1}^{n}\frac{d}{d\theta}\log f(X_i\,|\,\theta) \end{aligned}$$

$\dfrac{d}{d\theta}\log f(X_i\,|\,\theta)$는 서로 독립이므로 V의 분산은 다음과 같다.

$$\begin{aligned} Var(V) &= Var\left(\sum_{i=1}^{n}\frac{d}{d\theta}\log f(X_i\,|\,\theta)\right) \\ &= nVar\left(\frac{d}{d\theta}\log f(X\,|\,\theta)\right) \;=\; nE\left(\frac{d}{d\theta}\log f(X\,|\,\theta)\right)^2 \;=\; nI(\theta) \end{aligned}$$

한편 $E\left(\frac{d}{d\theta}\log f(\boldsymbol{X}\,|\,\theta)\right) = 0$이므로 U와 V의 공분산은 다음과 같다.

$$\begin{aligned} Cov(U, V) &= Cov\left(T(\boldsymbol{X}),\; \frac{d}{d\theta}\log f(\boldsymbol{X}\,|\,\theta)\right) \\ &= E\left[T(\boldsymbol{X})\;\left(\frac{d}{d\theta}\log f(\boldsymbol{X}\,|\,\theta)\right)\right] \\ &= \int T(\boldsymbol{x})\left(\frac{\frac{d}{d\theta}f(\boldsymbol{x}\,|\,\theta)}{f(\boldsymbol{x}\,|\,\theta)}\right) f(\boldsymbol{x}\,|\,\theta)\;d\boldsymbol{x} \\ &= \int T(\boldsymbol{x})\frac{d}{d\theta}f(\boldsymbol{x}\,|\,\theta)\;d\boldsymbol{x} \\ &= \frac{d}{d\theta}\left(\int T(\boldsymbol{x})f(\boldsymbol{x}\,|\,\theta)\;d\boldsymbol{x}\right) \\ &= \frac{d}{d\theta}(\theta) \;=\; 1 \end{aligned}$$

따라서 정리 8.3로부터 다음 부등식을 얻을 수 있다.

$$Var(T) \geq \frac{\left[Cov\left(T, \sum_{i=1}^{n} \frac{d}{d\theta} \log f(X_i|\theta)\right)\right]^2}{Var\left(\sum_{i=1}^{n} \frac{d}{d\theta} \log f(X_i|\theta)\right)} = \frac{1}{nI(\theta)}$$

■

크래머-라오 부등식은 비편향추정량이 가질 수 있는 최소의 분산을 보여주고 있다. 위 정리에 의하여 $T(\boldsymbol{X})$가 θ의 비편향추정량이고 $Var(T(\boldsymbol{X})) = \frac{1}{nI(\theta)}$이라면 $T(\boldsymbol{X})$는 θ의 비편향추정량 중에서 분산이 가장 작은 추정량이다.

예 8.19 $X_1, X_2, \ldots, X_n$이 Poisson(θ)로부터의 랜덤샘플일 때 θ의 추정량 $\widehat{\theta} = \overline{X}$는 비편향추정량 중에서 분산이 가장 작은 추정량임을 보여라.

[풀이] $\overline{X}$는 θ의 비편향추정량이고 $Var(\overline{X}) = \dfrac{\theta}{n} = \dfrac{1}{nI(\theta)}$이므로 $\overline{X}$는 θ의 비편향추정량 중에서 분산이 가장 작은 추정량이다. ■

예 8.20 $X_1, X_2, \ldots, X_n$이 Bernoulli(θ)로부터의 랜덤샘플일 때 $\widehat{\theta} = \overline{X}$는 비편향추정량 중에서 분산이 가장 작은 추정량임을 보여라.

[풀이] $\overline{X}$는 θ의 비편향추정량이고 $Var(\overline{X}) = \dfrac{\theta(1-\theta)}{n} = \dfrac{1}{nI(\theta)}$이므로 $\overline{X}$는 θ의 비편향추정량 중에서 분산이 가장 작은 추정량이다. ■

8.5 최소분산비편향추정량

비편향추정량의 비교기준은 분산이라 할 수 있다. 비편향추정량 중에서 분산이 가장 작은 추정량을 **최소분산비편향추정량**(minimum variance unbiased estimator, best unbiased estimator)이라 한다. 또는 모든 모수의 값에서 분산을 최소로 한다는 의미에서 **균일최소분산비편향추정량**(uniformly minimum variance unbiased estimator)이라 하기도 한다. 최소분산비편향추정량을 찾는 한 가지 방법은 정리 8.4를 이용하는 것이다. 어떤 비편향추정량의 분산이 크래머-라오 하한과 같다면 정리 8.4에 의하여 이 추정량은 최소분산비편향추정량이 된다. 그러나 이러한 방법으로 최소분산비편향추정량을 구하는 것은 쉽지 않다. 먼저 비편향추정량을 추측하고 이 추정량의 분산이 크래머-라오 하한과 같다는 것을 보여야 하기 때문이다. 이 절에서는 최소분산비편향추정량을 구하는 다른 방법을 알아보도록 하자.

다음 정리는 충분통계량에 대한 조건부 기댓값을 취함으로써 더 좋은 추정량을 만들 수 있음을 보여주고 있다.

정리 8.5 [라오-블랙웰 정리, Rao-Blackewell theorem] $S(\boldsymbol{X})$는 θ의 추정량이며 $T(\boldsymbol{X})$는 θ에 대한 충분통계량이라 하자. $S^*(\boldsymbol{X}) = E(S(\boldsymbol{X}) \,|\, T(\boldsymbol{X}))$일 때 다음이 성립한다.

$$E(S^*(\boldsymbol{X})) = E(S(\boldsymbol{X})), \quad Var(S^*(\boldsymbol{X})) \le Var(S(\boldsymbol{X}))$$

[증명] 정리 4.9와 정리 4.10에 의하여 $S^*(\boldsymbol{X})$의 기댓값은 다음과 같고

$$E(S^*(\boldsymbol{X})) \;=\; E(E(S(\boldsymbol{X})|T(\boldsymbol{X}))) \;=\; E(S(\boldsymbol{X}))$$

분산은 다음과 같다.

$$\begin{aligned} Var(S(\boldsymbol{X})) &= E(Var(S(\boldsymbol{X})|T(\boldsymbol{X}))) + Var(E(S(\boldsymbol{X})|T(\boldsymbol{X}))) \\ &= E(Var(S(\boldsymbol{X})|T(\boldsymbol{X}))) + Var(S^*(\boldsymbol{X})) \end{aligned}$$

따라서 $Var(S^*(\boldsymbol{X})) \le Var(S(\boldsymbol{X}))$이다. ■

위 정리에 의하면 $S(\boldsymbol{X})$가 θ의 비편향추정량일 때 $S^*(\boldsymbol{X})$도 θ의 비편향추정량이며 분산은 $S(\boldsymbol{X})$의 분산보다 작다. 따라서 $S^*(\boldsymbol{X})$와 $S(\boldsymbol{X})$의 평균제곱오차에 대하여 다음 부등식이 성립하므로

$$E(S^*(\boldsymbol{X}) - \theta)^2 \le E(S(\boldsymbol{X}) - \theta)^2$$

$S(\boldsymbol{X})$보다 $S^*(\boldsymbol{X})$가 더 좋은 추정량이라 할 수 있다.

앞의 정리에서 한 가지 주목할 사항은 $T(\boldsymbol{X})$가 충분통계량일 필요가 있는가에 관한 것이다. 증명과정을 보면 $T(\boldsymbol{X})$가 충분통계량이라는 조건은 어디에서도 필요하지 않은 것처럼 보일 수 있다. 그러나 $T(\boldsymbol{X})$가 충분통계량이 아니라면 $S^*(\boldsymbol{X})$가 모수 θ에 의존하는 값이 되므로 $S^*(\boldsymbol{X})$를 추정량이라 할 수 없다. 만약 $T(\boldsymbol{X})$가 충분통계량이면 $T(\boldsymbol{X})$가 주어졌을 때 $\boldsymbol{X}$의 조건부분포는 모수에 의존하지 않는다. 따라서 조건부기댓값 $S^*(\boldsymbol{X})$도 모수에 의존하지 않는 통계량이므로 추정량이라 할 수 있다.

위 정리가 의미하는 것은 θ의 추정량은 항상 충분통계량의 함수이어야 한다는 것이다. 그렇지 않다면 앞의 정리에 의하여 이 추정량보다 더 좋은 추정량을 구할 수 있기 때문이다.

예 8.21 $X_1, X_2, \ldots, X_n$이 Poisson(θ)로부터의 랜덤샘플이라 한다.

(a) θ의 비편향추정량 $S(\boldsymbol{X}) = X_1$의 분산을 구하여라.

(b) θ에 대한 충분통계량 $T(\boldsymbol{X}) = \sum_{i=1}^n X_i$를 이용하여 $S^*(\boldsymbol{X}) \;=\; E(S(\boldsymbol{X})\,|\,T(\boldsymbol{X}))$를 구하여라.

(c) (b)에서 구한 $S^*(\boldsymbol{X})$가 비편향추정량임을 보이고, $S^*(\boldsymbol{X})$의 분산을 구하여 $S(\boldsymbol{X})$의 분산과 비교하여라.

[풀이]

(a) X_1은 포아송분포를 따르므로 $S(\boldsymbol{X}) = X_1$의 분산은 θ이다.

(b) $\sum_{i=1}^n X_i \sim \text{Poisson}(n\theta)$라는 사실을 이용하여 $T(\boldsymbol{X}) = t$일 때 X_1의 조건부확률을 구하면 다음과 같다.

$$\begin{aligned}
P\Big(X_1 = x_1 \,\Big|\, \sum_{i=1}^n X_i = t\Big) &= \frac{P(X_1 = x_1,\, \sum_{i=1}^n X_i = t)}{P(\sum_{i=1}^n X_i = t)} \\
&= \frac{P(X_1 = x_1,\, \sum_{i=2}^n X_i = t - x_1)}{P(\sum_{i=1}^n X_i = t)} = \frac{\dfrac{\theta^{x_1}e^{-\theta}}{x_1!}\,\dfrac{((n-1)\theta)^{t-x_1}e^{-(n-1)\theta}}{(t-x_1)!}}{\dfrac{(n\theta)^t e^{-n\theta}}{t!}} \\
&= \frac{t!}{x_1!(t-x_1)!}\left(\frac{1}{n}\right)^{x_1}\left(1-\frac{1}{n}\right)^{t-x_1}
\end{aligned}$$

즉 $\sum_{i=1}^n X_i = t$일 때 X_1의 조건부분포는 $\mathrm{B}\left(t, \frac{1}{n}\right)$임을 알 수 있다. $\mathrm{B}\left(t, \frac{1}{n}\right)$의 기댓값은 $\frac{t}{n}$이므로 다음을 얻을 수 있다.

$$S^*(\boldsymbol{X}) \;=\; E(S(\boldsymbol{X})\,|\,T(\boldsymbol{X})) \;=\; E\Big(X_1 \,\Big|\, \sum_{i=1}^n X_i\Big) \;=\; \frac{1}{n}\sum_{i=1}^n X_i \;=\; \overline{X}$$

(c) $S^*(\boldsymbol{X}) = \frac{1}{n}\sum_{i=1}^n X_i$의 기댓값은 θ이므로 $S^*(\boldsymbol{X})$는 θ의 비편향추정량이고 $S^*(\boldsymbol{X})$의 분산은 다음과 같다.

$$Var(\overline{X}) = \frac{\theta}{n}$$

이는 $S(\boldsymbol{X})$의 분산보다 작다.

■

정리 8.5에 의하면 θ의 추정량 $S(\boldsymbol{X})$보다 우수한 추정량 $S^*(\boldsymbol{X}) = E(S(\boldsymbol{X})|T(\boldsymbol{X}))$를 만들 수 있다. 또 $S_0(\boldsymbol{X})$를 θ의 다른 추정량이라 할 때 $S_0^*(\boldsymbol{X}) = E(S_0(\boldsymbol{X})|T(\boldsymbol{X}))$는 $S_0(\boldsymbol{X})$보다 더

우수한 추정량이 된다. 그렇다면 $S^*(\boldsymbol{X})$와 $S_0^*(\boldsymbol{X})$ 중에서 어느 추정량이 더 우수할까? 이에 대한 답을 하기 위해서는 완비통계량의 개념이 필요하다. 다음 정리는 완비충분통계량을 이용하여 최소분산비편향추정량을 구하는 방법을 보여주고 있다.

정리 8.6 [레만-쉐페 정리, Lehmann-Scheffe theorem] $S(\boldsymbol{X})$는 θ의 비편향추정량이며 $T(\boldsymbol{X})$는 θ에 대한 완비충분통계량일 때 $S^*(\boldsymbol{X}) = E(S(\boldsymbol{X})\,|\,T(\boldsymbol{X}))$는 θ의 최소분산비편향추정량이다.

[증명] $T(\boldsymbol{X})$는 θ에 대한 충분통계량이므로 정리 8.5에 의하여 $S^*(\boldsymbol{X})$는 θ의 비편향추정량이며 $Var(S^*(\boldsymbol{X})) \le Var(S(\boldsymbol{X}))$임을 알 수 있다. $U(\boldsymbol{X})$를 θ의 임의의 비편향추정량이라 하자. 이때 $U^*(\boldsymbol{X}) = E(U(\boldsymbol{X})|T(\boldsymbol{X}))$는 정리 8.5에 의하여 θ의 비편향추정량이며 $Var(U^*(\boldsymbol{X})) \le Var(U(\boldsymbol{X}))$이다. $S^*(\boldsymbol{X})$와 $U^*(\boldsymbol{X})$는 모두 θ의 비편향추정량이므로 모든 θ에 대하여 다음이 성립한다.

$$E(S^*(\boldsymbol{X}) - U^*(\boldsymbol{X})) = 0$$

한편 $S^*(\boldsymbol{X})$와 $U^*(\boldsymbol{X})$는 모두 완비통계량 $T(\boldsymbol{X})$의 함수이므로 위 식은 $S^*(\boldsymbol{X}) = U^*(\boldsymbol{X})$임을 의미한다. 따라서 다음 부등식을 얻을 수 있다.

$$Var(S^*(\boldsymbol{X})) = Var(U^*(\boldsymbol{X})) \le Var(U(\boldsymbol{X}))$$

즉 $S^*(\boldsymbol{X})$의 분산은 임의의 비편향추정량의 분산보다 작으므로 $S^*(\boldsymbol{X})$는 최소분산비편향추정량이다. ■

$T(\boldsymbol{X})$의 함수 $h(T(\boldsymbol{X}))$에 대해서는 $E(h(T(\boldsymbol{X}))|T(\boldsymbol{X})) = h(T(\boldsymbol{X}))$이다. 이는 $T(\boldsymbol{X})$의 값이 주어지면 $h(T(\boldsymbol{X}))$의 값도 하나로 결정되기 때문이다. 따라서 완비충분통계량 $T(\boldsymbol{X})$ 의 함수인 $h(T(\boldsymbol{X}))$가 θ의 비편향추정량이라면 정리 8.6에 의하여 $h(T(\boldsymbol{X}))$가 최소분산비편향추정량이 된다.

지금까지의 결과를 요약하면 θ에 대한 최소분산비편향추정량을 구하는 방법은 다음과 같이 정리할 수 있다.

- 비편향추정량 $S^*(\boldsymbol{X})$의 분산이 크래머-라오 하한과 같을 때 $S^*(\boldsymbol{X})$는 θ의 최소분산비편향추정량이다.
- $T(\boldsymbol{X})$가 완비충분통계량이고 $h(T(\boldsymbol{X}))$는 θ의 비편향추정량일 때 $h(T(\boldsymbol{X}))$는 θ의 최소분산비편향추정량이다.
- $T(\boldsymbol{X})$가 완비충분통계량이고 $S(\boldsymbol{X})$는 θ의 비편향추정량일 때 다음과 같이 정의된 $S^*(\boldsymbol{X})$는 θ의 최소분산비편향추정량이다.

$$S^*(\boldsymbol{X}) = E(S(\boldsymbol{X})|T(\boldsymbol{X}))$$

위의 세 가지 방법 중에서 두 번째는 조건부 기댓값을 구할 필요가 없으므로 가장 간편하다고 할 수 있다. 즉 $T(\boldsymbol{X})$가 완비충분통계량일 때 $T(\boldsymbol{X})$의 함수로서 기댓값이 θ인 추정량을 찾을 수 있다면 이 추정량이 θ의 최소분산비편향추정량이 된다.

예 8.22 $X_1, X_2, \ldots, X_n$이 $\mathrm{B}(m, \theta)$로부터의 랜덤샘플일 때 θ의 최소분산비편향추정량을 구하여라.

[풀이] 이항분포를 따르는 랜덤샘플에 대해 $T(\boldsymbol{X}) = \sum_{i=1}^n X_i$가 θ에 대한 완비충분통계량이다. 완비충분통계량의 함수인 $\frac{1}{m}\overline{X} = \frac{1}{mn}T(\boldsymbol{X})$는 θ의 비편향추정량이므로 $\frac{1}{m}\overline{X}$가 θ의 최소분산비편향추정량이다. ■

예 8.23 $X_1, X_2, \ldots, X_n$이 $\mathrm{Poisson}(\theta)$로부터의 랜덤샘플이라 한다.

(a) θ의 최소분산비편향추정량을 구하여라.

(b) θ^2의 최소분산비편향추정량을 구하여라.

[풀이]

(a) 포아송분포를 따르는 랜덤샘플에 대해 $T(\boldsymbol{X}) = \sum_{i=1}^n X_i$가 θ에 대한 완비충분통계량이다. $E\left(\frac{1}{n}T(\boldsymbol{X})\right) = \theta$이므로 $\frac{1}{n}T(\boldsymbol{X}) = \overline{X}$가 θ의 최소분산비편향추정량이다.

(b) 완비충분통계량 $T(\boldsymbol{X}) = \sum_{i=1}^n X_i$의 확률분포는 $\mathrm{Poisson}(n\theta)$이므로 $T(\boldsymbol{X})$의 기댓값은 $E(T(\boldsymbol{X})) = n\theta$이고, $(T(\boldsymbol{X}))^2$의 기댓값은 $E(T(\boldsymbol{X}))^2 = n^2\theta^2 + n\theta$이다. 따라서 $\frac{1}{n}((T(\boldsymbol{X}))^2 - T(\boldsymbol{X}))$의 기댓값은 다음과 같다.

$$E\left(\frac{(T(\boldsymbol{X}))^2 - T(\boldsymbol{X})}{n^2}\right) = \theta^2$$

즉 $\frac{1}{n^2}((T(\boldsymbol{X}))^2 - T(\boldsymbol{X}))$는 완비충분통계량의 함수이며 θ^2의 비편향추정량이므로 다음 추정량이 θ^2의 최소분산비편향추정량이다.

$$\frac{1}{n^2}\Big((T(\boldsymbol{X}))^2 - T(\boldsymbol{X})\Big) \; = \; \frac{1}{n^2}\left(\left(\sum_{i=1}^n X_i\right)^2 - \sum_{i=1}^n X_i\right)$$

■

예 8.24 $X_1, X_2, \ldots, X_n$ 이 Uniform$(0, \theta)$ 로부터의 랜덤샘플일 때 θ 의 최소분산비편향추정량을 구하여라.

[풀이] Uniform$(0, \theta)$ 에서 나온 랜덤샘플에 대해 $T(\boldsymbol{X}) = X_{(n)}$ 이 θ 에 대한 완비충분통계량이다. $X_{(n)}$ 의 기댓값이 다음과 같으므로

$$\begin{aligned} E_\theta(X_{(n)}) &= \int_0^\theta t\, n \frac{t^{n-1}}{\theta^n}\, dt \\ &= \frac{n}{n+1}\theta \end{aligned}$$

$E\left(\frac{n+1}{n} X_{(n)}\right) = \theta$ 이다. 즉 완비충분통계량의 함수인 $\frac{n+1}{n} X_{(n)}$ 이 θ 의 비편향추정량이므로 θ 의 최소분산비편향추정량은 $\frac{n+1}{n} X_{(n)}$ 이다. ■

예 8.25 $X_1, X_2, \ldots, X_n$ 이 N$(\mu,\, \sigma^2)$ 으로부터의 랜덤샘플이라 한다.

(a) μ 의 최소분산비편향추정량을 구하여라.

(b) σ^2 의 최소분산비편향추정량을 구하여라.

[풀이]

(a) 정규분포를 따르는 랜덤샘플에 대하여 $T(\boldsymbol{X}) = (\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ 이 $\theta = (\mu, \sigma^2)$ 의 완비충분통계량이다. 한편 $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ 는 μ 의 비편향추정량이며 완비충분통계량 $T(\boldsymbol{X})$ 의 함수이므로 μ 의 최소분산비편향추정량이다.

(b) 표본분산 S^2 은 다음과 같이 나타낼 수 있으므로

$$\begin{aligned} S^2 &= \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2 \\ &= \frac{1}{n-1}\left(\sum_{i=1}^n X_i^2 - n\left(\frac{1}{n}\sum_{i=1}^n X_i\right)^2\right) \end{aligned}$$

S^2 은 완비충분통계량 $T(\boldsymbol{X})$ 의 함수이며 정리 5.1에 의하여 σ^2 의 비편향추정량이다. 따라서 표본분산 S^2 이 σ^2 의 최소분산비편향추정량이다.

■

지금까지 다룬 예에서는 두 번째 방법을 사용하여 최소분산불편추정량을 구하였다. 베르누이 분포에 대해 세 번째 방법을 사용하여 θ의 최소분산비편향추정량을 구하는 방법을 알아보도록 하자. $X_1, X_2, \ldots, X_n$이 Bernoulli(θ)로부터의 랜덤샘플일 때 $T(\boldsymbol{X}) = \sum_{i=1}^n X_i$는 θ에 대한 완비충분통계량이다. 한편 $S(\boldsymbol{X}) = X_1$는 θ의 비편향추정량이므로 θ의 최소분산비편향추정량은 $E(X_1 | \sum_{i=1}^n X_i)$라 할 수 있다. 지금부터는 이 조건부기댓값을 구하는 과정을 살펴보도록 하자. $\sum_{i=1}^n X_i \sim \mathrm{B}(n, \theta)$이므로 $T(\boldsymbol{X}) = t$일 때 X_1의 조건부분포는 다음과 같다.

$$
\begin{aligned}
P\Big(X_1 = x_1 \Big| \sum_{i=1}^n X_i = t\Big) &= \frac{P(X_1 = x_1,\, \sum_{i=1}^n X_i = t)}{P(\sum_{i=1}^n X_i = t)} \\
&= \frac{P(X_1 = x_1,\, \sum_{i=2}^n X_i = t - x_1)}{P(\sum_{i=1}^n X_i = t)} \\
&= \frac{\theta^{x_1}(1-\theta)^{1-x_1}\dbinom{n-1}{t-x_1}\theta^{t-x_1}(1-\theta)^{(n-1)-(t-x_1)}}{\dbinom{n}{t}\theta^t(1-\theta)^{n-t}} \\
&= \frac{\dbinom{n-1}{t-x_1}}{\dbinom{n}{t}} \quad = \begin{cases} \dfrac{\dbinom{n-1}{t-1}}{\dbinom{n}{t}} = \dfrac{t}{n}, & x_1 = 1 \text{일 때} \\[2ex] \dfrac{\dbinom{n-1}{t}}{\dbinom{n}{t}} = \dfrac{n-t}{n}, & x_1 = 0 \text{일 때} \end{cases}
\end{aligned}
$$

즉 $T(\boldsymbol{X}) = t$일 때 X_1의 조건부분포는 다음과 같다.

$$
\begin{aligned}
P\Big(X_1 = 1 \Big| \textstyle\sum_{i=1}^n X_i = t\Big) &= \frac{t}{n} \\
P\Big(X_1 = 0 \Big| \textstyle\sum_{i=1}^n X_i = t\Big) &= 1 - \frac{t}{n}
\end{aligned}
$$

따라서 조건부기댓값은 다음과 같으므로

$$
\begin{aligned}
E\Big(X_1 \Big| \textstyle\sum_{i=1}^n X_i = t\Big) &= 1 \cdot P\Big(X_1 = 1 \Big| \displaystyle\sum_{i=1}^n X_i = t\Big) + 0 \cdot P\Big(X_1 = 0 \Big| \displaystyle\sum_{i=1}^n X_i = t\Big) \\
&= \frac{t}{n}
\end{aligned}
$$

θ의 최소분산비편향추정량은 $E(X_1 | \sum_{i=1}^n X_i) = \frac{1}{n}\sum_{i=1}^n X_i$이다.

8.6 최대가능도추정량의 점근적 성질

표본의 크기가 커질 때 추정량이 가지는 성질을 점근적 성질 (asymptotic property) 이라 한다. 최대가능도추정량이 널리 쓰이는 중요한 이유 중의 하나는 이 추정량이 아주 좋은 점근적 성질을 가지고 있다는데 있다. 점근적 의미에서 보면 최대가능도추정량은 최적의 추정량이라 할 수 있다.

최대가능도추정량이 가지는 점근적 성질을 알아보기 위해 이 절에서는 로그가능도는 미분가능하고 최대가능도추정량은 로그가능도를 미분하여 0이 되는 해로 주어진다고 가정을 하겠다. 또한 로그가능도에 대한 미분과 적분의 순서를 바꿀 수 있다고 가정을 하자. 이 교재에서 다루는 대부분의 확률분포는 지수족에 속하며 이러한 가정을 만족한다. 참고로 이러한 가정을 만족하지 않는 대표적인 확률분포가 Uniform$(0,\,\theta)$ 이다. Uniform$(0,\,\theta)$ 와 같이 확률분포가 값을 가지는 영역이 모수 θ 에 따라 달라지는 경우는 가능도가 θ 에 대해 미분가능하지 않다. Uniform$(0,\,\theta)$ 의 확률밀도함수를 보면 0부터 θ 에서만 0이 아닌 상수 값을 가지며 다른 부분에서는 0이므로 가능도는 θ 에서 미분가능하지 않다.

최대가능도추정량이 가지는 첫 번째 점근적 성질은 일치추정량이 된다는 것이다. 즉 최대가능도추정량은 표본의 크기가 증가하면 추정하는 모수로 수렴하게 된다. 로그가능도를 θ 의 함수로 나타낼 때 모수의 참값 θ 와 혼동되는 것을 방지하기 위하여 이 절에서는 모수의 참값을 θ_0 로 나타내도록 하자. 최대가능도추정량 $\widehat{\theta}$ 는 로그가능도를 미분한 값이 0이 되는 해이므로 $\dfrac{d}{d\theta}\, l(\theta)\Big|_{\theta=\widehat{\theta}} = 0$ 이다.

$X_1, X_2, \ldots, X_n$ 이 확률밀도함수 (확률질량함수) $f(x\,|\,\theta\,)$ 를 가지는 랜덤샘플일 때 로그가능도는 다음과 같으므로

$$l(\theta) \;=\; \sum_{i=1}^{n} \log f(X_i\,|\,\theta)$$

최대가능도추정량 $\widehat{\theta}$ 에 대하여 다음이 성립한다.

$$\frac{d}{d\theta} \sum_{i=1}^{n} \log f(X_i\,|\,\theta)\bigg|_{\theta=\widehat{\theta}} \;=\; 0 \tag{8.12}$$

여기서 로그가능도를 나타낼 때 x_i 대신에 확률변수를 나타내는 X_i 를 사용한 이유는 후에 위 식의 기댓값을 구하기 때문이다. 위의 식 (8.12) 을 n 으로 나누고 미분과 합의 순서를 바꾸면 다음 식을 얻을 수 있다.

$$\frac{1}{n} \sum_{i=1}^{n} \frac{d}{d\theta} \log f(X_i|\,\theta)\bigg|_{\theta=\widehat{\theta}} \;=\; 0 \tag{8.13}$$

한편 대수의 법칙에 의하여 다음이 성립한다.

$$\frac{1}{n} \sum_{i=1}^{n} \frac{d}{d\theta} \log f(X_i\,|\,\theta) \;\longrightarrow\; E\left(\frac{d}{d\theta} \log f(X\,|\,\theta)\right) \tag{8.14}$$

최대가능도추정량은 (8.13)에 의하여 위의 좌측 식을 0으로 하는 해이다. 또한 모수의 참값 θ_0는 우측 식을 0으로 하는 해임을 보일 수 있다. 이는 $\theta = \theta_0$에서 $E\left(\frac{d}{d\theta}\log f(X\,|\,\theta)\right)$의 값이 0이 된다는 것을 보임으로써 확인할 수 있다. 적분과 미분의 순서를 바꿀 수 있을 때 $\theta = \theta_0$에서 구한 $E\left(\frac{d}{d\theta}\log f(X\,|\,\theta)\right)$의 값은 다음과 같다.

$$\begin{aligned}
&E\left(\frac{d}{d\theta}\log f(X\,|\,\theta)\bigg|_{\theta=\theta_0}\right) \\
&= \int_{-\infty}^{\infty}\left(\frac{d}{d\theta}\log f(x\,|\,\theta)\bigg|_{\theta=\theta_0}\right) f(x|\theta_0)\; dx \\
&= \int_{-\infty}^{\infty}\frac{\frac{d}{d\theta}f(x\,|\,\theta)\Big|_{\theta=\theta_0}}{f(x\,|\,\theta_0)} f(x\,|\,\theta_0)\; dx \;=\; \int_{-\infty}^{\infty}\left(\frac{d}{d\theta}f(x\,|\,\theta)\bigg|_{\theta=\theta_0}\right) dx \\
&= \frac{d}{d\theta}\left(\int_{-\infty}^{\infty} f(x\,|\,\theta)\; dx\right)\bigg|_{\theta=\theta_0} \;=\; \frac{d}{d\theta}(1)\bigg|_{\theta=\theta_0} \;=\; 0
\end{aligned} \tag{8.15}$$

따라서 θ_0는 다음 방정식을 만족하는 해이다.

$$E\left(\frac{d}{d\theta}\log f(X\,|\,\theta)\right) \;=\; 0$$

위에서 설명한 것을 정리하면 식 (8.14)의 좌측을 0으로 하는 해는 최대가능도추정량인 $\widehat{\theta}$이고 우측을 0으로 하는 해는 모수의 참값인 θ_0이다. 좌측 식이 우측 식으로 수렴하므로 좌측 식의 해인 $\widehat{\theta}$는 우측 식의 해인 θ_0로 수렴할 것으로 기대할 수 있다. 따라서 다음 정리를 얻을 수 있다.

정리 8.7 $X_1, X_2, \ldots, X_n$이 확률밀도함수 $f(x\,|\,\theta)$를 가지는 랜덤샘플일 때 적절한 조건 하에서 최대가능도추정량 $\widehat{\theta}$는 θ의 일치추정량이다. 즉 $\widehat{\theta} \;\xrightarrow{p}\; \theta$이다.

최대가능도추정량이 가지는 두 번째 점근적 성질은 확률분포에 대한 것이다. 정리 8.7에 의하여 최대가능도추정량 $\widehat{\theta}$는 모수의 참값 θ_0로 수렴함을 알 수 있다. 따라서 $(\widehat{\theta} - \theta_0)$는 0으로 수렴하게 된다. 그런데 $\sqrt{n}$을 곱해준 $\sqrt{n}(\widehat{\theta} - \theta_0)$는 정규분포로 분포수렴함을 보일 수 있다. 따라서 $\widehat{\theta}$는 $\dfrac{1}{\sqrt{n}}$의 속도로 θ_0에 수렴한다고 할 수 있다. 표본의 크기 n이 커질 때 $\sqrt{n}(\widehat{\theta} - \theta_0)$이 수렴하는 확률분포를 점근분포(asymptotic distribution) 또는 극한분포(limiting distribution)라 한다.

로그가능도를 미분한 함수 $l'(\theta)$를 θ_0에서 테일러 전개하면 다음과 같이 나타낼 수 있다.

$$l'(\theta) \;=\; l'(\theta_0) \;+\; l''(\theta_0)(\theta - \theta_0) \;+\; \cdots$$

위의 테일러 전개를 이용하여 $\theta = \widehat{\theta}$에서 $l'(\theta)$의 값을 구하면 다음과 같다.

$$l'(\widehat{\theta}) \;=\; l'(\theta_0) \;+\; l''(\theta_0)(\widehat{\theta} - \theta_0) \;+\; \cdots \tag{8.16}$$

정리 8.7에서 보인 것처럼 $\widehat{\theta}$는 θ_0에 수렴하므로 표본의 크기가 충분히 크면 $(\widehat{\theta}-\theta)$는 작은 값이 된다. 따라서 위의 테일러 전개한 식 (8.16)에서 뒷부분의 항은 큰 역할을 하지 못한다. 처음 두 항만 이용하면 $l'(\widehat{\theta}) \;\approx\; l'(\theta_0) \;+\; l''(\theta_0)(\widehat{\theta}-\theta_0)$로 나타낼 수 있다. 한편 $l'(\widehat{\theta}\,) = 0$이므로 식 (8.16)을 다르게 표현하면 다음과 같다.

$$(\widehat{\theta}-\theta_0) \;\approx\; -\frac{l'(\theta_0)}{l''(\theta_0)}$$

이로부터 $\sqrt{n}(\widehat{\theta}-\theta_0)$는 다음과 같이 나타낼 수 있다.

$$\sqrt{n}(\widehat{\theta}-\theta_0) \;\approx\; \frac{\frac{1}{\sqrt{n}}\,l'(\theta_0)}{-\frac{1}{n}\,l''(\theta_0)} \tag{8.17}$$

$\sqrt{n}(\widehat{\theta}-\theta_0)$의 점근적 확률분포를 구하는 한 가지 방법은 식 (8.17)의 분자의 점근분포와 분모가 수렴하는 값을 찾는 것이다. 다음 정리는 위 식의 분자의 점근분포를 보여주고 있다.

정리 8.8 적절한 조건 하에서 다음이 성립한다.

$$\frac{1}{\sqrt{n}}\,l'(\theta_0) \;\xrightarrow{d}\; \mathrm{N}(0,\,I(\theta_0))$$

[증명] $l(\theta) \;=\; \sum_{i=1}^n \log f(X_i\,|\,\theta)$이므로 $l'(\theta_0)$는 다음과 같이 나타낼 수 있다.

$$l'(\theta_0) \;=\; \sum_{i=1}^n \frac{d}{d\theta}\log f(X_i\,|\,\theta)\bigg|_{\theta=\theta_0}$$

위 식에서 $Y_i = \frac{d}{d\theta}\log f(X_i\,|\,\theta)\Big|_{\theta=\theta_0}$ 라 할 때, 식 (8.15)에 의하여 $E(Y_i) = 0$이다. 한편 Y_i의 분산은 다음과 같다.

$$Var(Y_i) \;=\; E(Y_i^2) \;=\; E\left(\frac{d}{d\theta}\log f(X\,|\,\theta)\bigg|_{\theta=\theta_0}\right)^2 \;=\; I(\theta_0)$$

Y_i 들은 서로 독립이며 같은 분포를 가지므로 중심극한정리에 의하여 다음이 성립한다.

$$\frac{1}{\sqrt{n}}\,l'(\theta_0) = \sqrt{n}\,\overline{Y} \;\xrightarrow{d}\; \mathrm{N}(0,\,I(\theta_0))$$

■

정리 8.8을 이해하기 위하여 포아송분포의 경우를 생각하여 보자. Poisson(θ)의 확률질량함수는 $f(x\,|\,\theta) = \dfrac{\theta^x e^{-\theta}}{x!}$ 이므로 $\log f(x\,|\,\theta) = x\log\theta - \theta - \log(x!)$ 이다. 이를 θ_0 에서 미분하면 다음과 같다.

$$\frac{d}{d\theta}\log f(x\,|\,\theta)\Big|_{\theta=\theta_0} = \frac{x}{\theta_0} - 1$$

따라서 $l'(\theta_0)$ 는 다음과 같다.

$$\begin{aligned} l'(\theta_0) &= \sum_{i=1}^{n} \frac{d}{d\theta}\log f(X_i\,|\,\theta)\Big|_{\theta=\theta_0} \\ &= \sum_{i=1}^{n}\left(\frac{X_i}{\theta_0} - 1\right) \end{aligned}$$

즉 $l'(\theta_0)$ 는 기댓값은 $E\left(\frac{X_i}{\theta_0} - 1\right) = 0$ 이고 분산은 $Var\left(\frac{X_i}{\theta_0} - 1\right) = \frac{1}{\theta_0}$ 인 서로 독립인 항의 합으로 표현할 수 있으므로 중심극한정리에 의하여 다음이 성립한다.

$$\frac{1}{\sqrt{n}}\,l'(\theta_0) \xrightarrow{d} \mathrm{N}\left(0,\,\frac{1}{\theta_0}\right) \tag{8.18}$$

한편 포아송분포에 대한 피셔 정보수는 다음과 같으므로

$$\begin{aligned} I(\theta_0) &= E\left(\frac{d}{d\theta}\log f(X\,|\,\theta)\Big|_{\theta=\theta_0}\right)^2 \\ &= E\left(\frac{X}{\theta_0} - 1\right)^2 = \frac{1}{\theta_0} \end{aligned}$$

정리 8.8이 성립함을 확인할 수 있다.

식 (8.17) 에서 분모가 수렴하는 값은 대수의 법칙을 이용하여 구할 수 있다. 분모에 주어진 식은 다음과 같이 나타낼 수 있으며

$$-\frac{1}{n}\,l''(\theta_0) = -\frac{1}{n}\sum_{i=1}^{n}\frac{d^2}{d\theta^2}\log f(X_i\,|\,\theta)\Big|_{\theta=\theta_0}$$

$\dfrac{d^2}{d\theta^2}\log f(X_i\,|\,\theta)\Big|_{\theta=\theta_0}$ 는 서로 독립이며 같은 분포를 가지므로 대수의 법칙에 의하여 다음 정리를 얻을 수 있다.

정리 8.9 적절한 조건 하에서 다음이 성립한다.

$$-\frac{1}{n}\,l''(\theta_0) \;\xrightarrow{p}\; -E\left(\frac{d^2}{d\theta^2}\log f(X\,|\,\theta)\bigg|_{\theta=\theta_0}\right) \;=\; I(\theta_0)$$

위 정리도 포아송분포의 예를 통하여 확인할 수 있다. 포아송분포에 대하여 $\log f(x\,|\,\theta)$를 θ_0에서 두 번 미분하면 다음과 같으므로

$$\frac{d^2}{d\theta^2}\log f(x\,|\,\theta)\bigg|_{\theta=\theta_0} \;=\; -\frac{x}{\theta_0^2}$$

식 (8.17)의 분모는 다음과 같이 나타낼 수 있다.

$$-\frac{1}{n}\,l''(\theta_0) \;=\; \frac{1}{n}\sum_{i=1}^{n}\frac{X_i}{\theta_0^2}$$

즉 $-\frac{1}{n}\,l''(\theta_0)$는 서로 독립이며 같은 분포를 가지는 $\frac{X_i}{\theta_0^2}$의 평균이므로 대수의 법칙에 의하여 다음이 성립한다.

$$-\frac{1}{n}\,l''(\theta_0) \;\xrightarrow{p}\; E\left(\frac{X}{\theta_0^2}\right) \;=\; \frac{1}{\theta_0}$$

한편 다음이 성립하므로

$$-E\left(\frac{d^2}{d\theta^2}\log f(X\,|\,\theta)\bigg|_{\theta=\theta_0}\right) \;=\; E\left(\frac{X}{\theta_0^2}\right) \;=\; \frac{1}{\theta_0}$$

포아송분포에 대하여 $-\frac{1}{n}\,l''(\theta_0) \;\xrightarrow{p}\; -E\left(\frac{d^2}{d\theta^2}\log f(X\,|\,\theta)\Big|_{\theta=\theta_0}\right) = I(\theta_0)$임을 확인할 수 있다.

식 (8.17)와 정리 8.8, 정리 8.9, 정리 8.2를 종합하면 최대가능도추정량의 점근분포를 얻을 수 있다.

정리 8.10 최대가능도추정량 $\widehat{\theta}$에 대하여 적절한 조건 하에서 다음이 성립한다.

$$\sqrt{n}(\widehat{\theta}-\theta_0) \;\xrightarrow{d} \mathrm{N}\left(0,\;\frac{1}{I(\theta_0)}\right)$$

최대가능도추정량 $\widehat{\theta}$은 일치추정량이고 정리 8.10에 주어진 점근분포로부터 최대가능도추정

량의 점근분산은 크래머-라오 하한과 같다. 즉 점근적인 관점에서 볼 때 최대가능도추정량은 비편향추정량이고 분산은 크래머-라오 하한과 같으므로 최소의 분산을 가지는 최적의 추정량이라 할 수 있다.

예 8.26 $X_1, X_2, \ldots, X_n$ 이 Poisson(θ) 로부터의 랜덤샘플일 때 θ 의 최대가능도추정량 $\widehat{\theta} = \overline{X}$ 에 대하여 $\sqrt{n}(\widehat{\theta} - \theta)$ 의 점근분포를 구하여라.

[풀이] 포아송분포의 피셔 정보수는 $I(\theta) = \dfrac{1}{\theta}$ 이므로 정리 8.10에 의하여 다음을 얻을 수 있다.

$$\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} N(0,\, \theta)$$

■

최대가능도추정량 $\widehat{\theta}$에 대한 점근분포로부터 $\widehat{\theta}$의 분산은 점근적으로 $\frac{1}{nI(\theta)}$ 임을 알 수 있다. 어떤 확률분포에 대해서는 $\widehat{\theta}$의 정확한 분산이 $\frac{1}{nI(\theta)}$ 인 경우도 있다. 예를 들어 Poisson(θ) 로부터의 랜덤샘플에 대하여 θ의 최대가능도추정량은 $\widehat{\theta} = \overline{X}$ 이며 이 추정량의 분산은 $Var(\overline{X}) = \frac{\theta}{n}$ 이다. Poisson(θ) 에 대한 피셔 정보수는 $I(\theta) = \frac{1}{\theta}$ 이므로 $\frac{1}{nI(\theta)} = \frac{\theta}{n}$ 는 $\widehat{\theta}$의 정확한 분산과 일치함을 알 수 있다.

최대가능도추정량의 점근분포에 대한 정리는 모수가 두 개 이상인 경우로도 확장될 수 있다. 예를 들어 모수가 $\theta = (\theta_1, \theta_2)$ 로 두 개인 경우 최대가능도추정량 $(\widehat{\theta}_1, \widehat{\theta}_2)$ 의 점근분포는 다음과 같은 이변량정규분포가 된다.

정리 8.11 모수 $\theta = (\theta_1,\, \theta_2)$ 의 최대가능도추정량 $(\widehat{\theta}_1,\, \widehat{\theta}_2)$ 에 대하여 적절한 조건 하에서 다음이 성립한다.

$$\sqrt{n}\begin{pmatrix} \widehat{\theta}_1 - \theta_1 \\ \widehat{\theta}_2 - \theta_2 \end{pmatrix} \xrightarrow{d} \mathrm{N}_2\left(0,\; I^{-1}(\theta)\right)$$

위에서 $I(\theta)$ 는 다음과 같으며

$$I(\theta) \;=\; -\begin{pmatrix} E\left(\frac{d^2}{d\theta_1^2}\log f(X\,|\,\theta_1,\theta_2)\right) & E\left(\frac{d^2}{d\theta_1 d\theta_2}\log f(X\,|\,\theta_1,\theta_2)\right) \\ E\left(\frac{d^2}{d\theta_2 d\theta_1}\log f(X\,|\,\theta_1,\theta_2)\right) & E\left(\frac{d^2}{d\theta_2^2}\log f(X\,|\,\theta_1,\theta_2)\right) \end{pmatrix}$$

이를 **피셔 정보행렬**(Fisher information matrix) 또는 간단히 **정보행렬**이라 한다.

예 8.27 $X_1, X_2, \ldots, X_n$ 이 $\mathrm{N}(\mu,\, \sigma^2)$ 으로부터의 랜덤샘플이라 한다. 이때 μ 와 σ^2 의 최대가능도추정량에 대한 점근분포를 구하여라.

[풀이] $N(\mu, \sigma^2)$ 의 로그 확률밀도함수는 다음과 같으므로

$$\log f(x|\mu, \sigma^2) \;=\; -\frac{1}{2}\log(2\pi\sigma^2) - \frac{(x-\mu)^2}{2\sigma^2}$$

이를 미분하면 다음을 얻을 수 있다.

$$\begin{aligned}
\frac{d^2}{d\mu^2}\log f(x\,|\,\mu, \sigma^2) &= -\frac{1}{\sigma^2} \\
\frac{d^2}{d\mu\, d\sigma^2}\log f(x\,|\,\mu, \sigma^2) &= -\frac{x-\mu}{(\sigma^2)^2} \\
\frac{d^2}{d(\sigma^2)^2}\log f(x\,|\,\mu, \sigma^2) &= \frac{1}{2(\sigma^2)^2} - \frac{(x-\mu)^2}{(\sigma^2)^3}
\end{aligned}$$

$E(X-\mu) = 0,\; E(X-\mu)^2 = \sigma^2$ 이므로 $N(\mu, \sigma^2)$ 에 대한 정보행렬은 다음과 같다.

$$I(\theta) \;=\; \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2(\sigma^2)^2} \end{pmatrix}$$

따라서 최대가능도추정량 $(\widehat{\mu},\, \widehat{\sigma}^2)$ 에 대하여 다음이 성립한다.

$$\sqrt{n}\begin{pmatrix} \widehat{\mu} - \mu \\ \widehat{\sigma}^2 - \sigma^2 \end{pmatrix} \;\xrightarrow{d}\; \mathrm{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix},\, \begin{pmatrix} \sigma^2 & 0 \\ 0 & 2(\sigma^2)^2 \end{pmatrix}\right)$$

■

최대가능도추정량의 점근분포를 이용하면 모수 θ 에 대한 근사적 신뢰구간을 구할 수 있다. 즉 다음이 성립하므로

$$\sqrt{n}(\widehat{\theta} - \theta_0) \;\xrightarrow{d}\; N\left(0,\, \frac{1}{I(\theta_0)}\right)$$

n 이 충분히 크면 다음을 얻을 수 있다.

$$P\left(-Z_{\frac{\alpha}{2}} \le \sqrt{nI(\theta_0)}\,(\widehat{\theta} - \theta_0) \le Z_{\frac{\alpha}{2}}\right) \;\approx\; 1-\alpha$$

위에서 Z_α 는 $P(N(0,1) \ge Z_\alpha) = \alpha$ 인 값이다. 따라서 θ_0 의 $100(1-\alpha)\%$ 신뢰구간은 다음과

같이 구할 수 있다.

$$\left(\widehat{\theta}-\frac{Z_{\frac{\alpha}{2}}}{\sqrt{nI(\theta_0)}},\quad \widehat{\theta}+\frac{Z_{\frac{\alpha}{2}}}{\sqrt{nI(\theta_0)}}\right)$$

위 신뢰구간의 문제점은 θ_0의 값을 알지 못하므로 $I(\theta_0)$를 직접 사용할 수 없다는 것이다. 일반적으로 사용되는 두 가지 방법은 다음과 같다.

- 피셔 정보수에서 알 수 없는 값인 θ_0 대신 이의 추정값 $\widehat{\theta}$를 이용한다. 즉 $I(\theta_0)$ 대신 $I(\widehat{\theta})$를 사용하여 신뢰구간을 구할 수 있다. 즉 θ_0의 $100(1-\alpha)\%$ 신뢰구간은 다음과 같다.

$$\left(\widehat{\theta}-\frac{Z_{\frac{\alpha}{2}}}{\sqrt{nI(\widehat{\theta})}},\quad \widehat{\theta}+\frac{Z_{\frac{\alpha}{2}}}{\sqrt{nI(\widehat{\theta})}}\right)$$

- $I(\theta_0)$ 대신 $-\frac{1}{n}\sum_{i=1}^{n}\frac{d^2}{d\theta^2}\log f(X_i\,|\,\theta)\Big|_{\theta=\widehat{\theta}}$ 를 사용하여 신뢰구간을 구할 수 있다. 즉 θ_0의 $100(1-\alpha)\%$ 신뢰구간은 다음과 같다.

$$\left(\widehat{\theta}-\frac{Z_{\frac{\alpha}{2}}}{\sqrt{-\sum\limits_{i=1}^{n}\frac{d^2}{d\theta^2}\log f(X_i\,|\,\theta)\Big|_{\theta=\widehat{\theta}}}},\quad \widehat{\theta}+\frac{Z_{\frac{\alpha}{2}}}{\sqrt{-\sum\limits_{i=1}^{n}\frac{d^2}{d\theta^2}\log f(X_i\,|\,\theta)\Big|_{\theta=\widehat{\theta}}}}\right)$$

위의 두 번째 방법을 사용하는 근거는 정리 8.9에 의하여 다음이 성립하고

$$-\frac{1}{n}\sum_{i=1}^{n}\frac{d^2}{d\theta^2}\log f(X_i\,|\,\theta)\Big|_{\theta=\theta_0}\ \xrightarrow{p}\ I(\theta_0)$$

$\widehat{\theta}\ \xrightarrow{p}\ \theta_0$ 이기 때문이다.

예 8.28 $X_1, X_2, \ldots, X_n$ 이 Poisson(θ)로부터의 랜덤샘플일 때 θ에 대한 신뢰구간을 구하여라.

[풀이] θ의 최대가능도추정량은 $\widehat{\theta}=\overline{X}$ 이며 로그 확률밀도함수와 이를 두 번 미분한 함수는 다음과 같다.

$$\begin{aligned}\log f(x\,|\,\theta) &= x\log\theta-\theta-\log(x!)\\ \frac{d^2}{d\theta^2}\log f(x\,|\,\theta) &= -\frac{x}{\theta^2}\end{aligned}$$

포아송분포의 피셔 정보수는 $I(\theta) = \dfrac{1}{\theta}$ 이므로 첫 번째 방법을 사용한 θ에 대한 신뢰구간은 다음과 같다.

$$\begin{aligned} & \left(\widehat{\theta} - Z_{\frac{\alpha}{2}} \sqrt{\frac{\widehat{\theta}}{n}} \ , \quad \widehat{\theta} + Z_{\frac{\alpha}{2}} \sqrt{\frac{\widehat{\theta}}{n}} \right) \\ = & \left(\overline{X} - Z_{\frac{\alpha}{2}} \sqrt{\frac{\overline{X}}{n}} \ , \quad \overline{X} + Z_{\frac{\alpha}{2}} \sqrt{\frac{\overline{X}}{n}} \right) \end{aligned}$$

한편 다음이 성립하므로

$$\begin{aligned} & -\frac{1}{n} \sum_{i=1}^{n} \frac{d^2}{d\theta^2} \log f(X_i \,|\, \theta) \Big|_{\theta=\widehat{\theta}} \\ & = \frac{1}{n} \sum_{i=1}^{n} \frac{X_i}{(\widehat{\theta}\,)^2} = \frac{1}{n} \sum_{i=1}^{n} \frac{X_i}{\overline{X}^2} = \frac{1}{\overline{X}} \end{aligned}$$

두 번째 방법을 사용한 신뢰구간은 첫 번째 신뢰구간과 일치한다. ■

예 8.29 $X_1, X_2, \ldots, X_n$ 이 Bernoulli(θ) 로부터의 랜덤샘플일 때 θ에 대한 신뢰구간을 구하여라.

[풀이] θ의 최대가능도추정량은 $\widehat{\theta} = \overline{X}$ 이며 로그 확률밀도함수와 이를 두 번 미분한 함수는 다음과 같다.

$$\begin{aligned} \log f(x|\theta) &= x \log\theta + (1-x)\log(1-\theta) \\ \frac{d^2}{d\theta^2} \log f(x|\theta) &= -\frac{x}{\theta^2} - \frac{1-x}{(1-\theta)^2} = -\frac{x - 2\theta x + \theta^2}{\theta^2(1-\theta)^2} \end{aligned}$$

예 8.18에 의하여 Bernoulli(θ) 의 피셔 정보수는 $I(\theta) = \dfrac{1}{\theta(1-\theta)}$ 이므로 첫 번째 방법을 사용한 θ에 대한 신뢰구간은 다음과 같다.

$$\begin{aligned} & \left(\widehat{\theta} - Z_{\frac{\alpha}{2}} \sqrt{\frac{\widehat{\theta}\,(1-\widehat{\theta}\,)}{n}} \ , \quad \widehat{\theta} + Z_{\frac{\alpha}{2}} \sqrt{\frac{\widehat{\theta}\,(1-\widehat{\theta}\,)}{n}} \right) \\ = & \left(\overline{X} - Z_{\frac{\alpha}{2}} \sqrt{\frac{\overline{X}(1-\overline{X})}{n}} \ , \quad \overline{X} + Z_{\frac{\alpha}{2}} \sqrt{\frac{\overline{X}(1-\overline{X})}{n}} \right) \qquad (8.19) \end{aligned}$$

한편 베르누이 확률변수에 대하여 $X_i = X_i^2$ 이므로 다음이 성립한다.

$$
\begin{aligned}
-\frac{1}{n}\sum_{i=1}^{n}\frac{d^2}{d\theta^2}\log f(X_i\,|\,\theta)\bigg|_{\theta=\widehat{\theta}} &= \frac{1}{n}\sum_{i=1}^{n}\frac{X_i - 2\widehat{\theta}X_i + (\widehat{\theta}\,)^2}{(\widehat{\theta}\,)^2(1-\widehat{\theta}\,)^2} \\
= \frac{1}{n}\sum_{i=1}^{n}\frac{X_i^2 - 2\widehat{\theta}X_i + (\widehat{\theta}\,)^2}{(\widehat{\theta}\,)^2(1-\widehat{\theta}\,)^2} &= \frac{1}{n}\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{\overline{X}^2(1-\overline{X})^2}
\end{aligned}
$$

따라서 두 번째 방법을 사용한 신뢰구간은 다음과 같다.

$$
\left(\overline{X} - Z_{\frac{\alpha}{2}}\sqrt{\frac{\overline{X}^2(1-\overline{X})^2}{\sum_{i=1}^{n}(X_i-\overline{X})^2}}\;,\quad \overline{X} + Z_{\frac{\alpha}{2}}\sqrt{\frac{\overline{X}^2(1-\overline{X})^2}{\sum_{i=1}^{n}(X_i-\overline{X})^2}}\,\right) \tag{8.20}
$$

■

연습문제

1. $X_1, X_2, \ldots, X_n$ 이 확률밀도함수 $f(x|\theta) = \dfrac{1}{\Gamma(\theta)2^\theta}\, x^{\theta-1}e^{-\frac{x}{2}},\ x > 0$ 인 Gamma$(\theta, \frac{1}{2})$ 로부터의 랜덤샘플일 때 θ 의 적률추정량을 구하여라.

2. 모수 θ 는 θ_1, θ_2 또는 θ_3 중 하나의 값을 갖는다고 한다. 각 모수의 값에서 이산형 확률변수 X 의 확률분포가 다음과 같다고 한다.

x	-3	-2	-1	0	1	2	3
$f(x\|\theta_1)$	0.3	0.15	0.05	0.05	0.05	0.1	0.3
$f(x\|\theta_2)$	0.1	0.1	0.25	0.3	0.15	0.05	0.05
$f(x\|\theta_3)$	0.05	0.05	0.15	0.1	0.35	0.15	0.15

   a) 각 X 의 값에서 θ 의 최대가능도추정량을 구하여라.
   b) 위와 같은 확률분포를 가지는 서로 독립인 확률변수 X_1 과 X_2 의 관측 값이 -1 과 2 일 때 θ 의 최대가능도추정량을 구하여라.

3. $X_1, X_2, \ldots, X_n$ 이 확률밀도함수 $f(x|\mu, \sigma) = \dfrac{1}{\sigma}\, e^{-\frac{(x-\mu)}{\sigma}},\ x > \mu$ 를 가지는 랜덤샘플이다.
   a) μ 와 σ 의 적률추정량을 구하여라.
   b) μ 와 σ 의 최대가능도추정량을 구하여라.

4. $X_1, X_2, \ldots, X_n$ 이 Uniform$(0, \theta)$ 로부터의 랜덤샘플일 때 θ 의 두 추정량이 아래와 같을 때

$$T_1(\boldsymbol{X}) = \frac{n+1}{n} X_{(n)},\ T_2(\boldsymbol{X}) = X_{(n)}$$

   다음 물음에 답하여라.
   a) 두 추정량의 편향을 구하여라.
   b) 두 추정량의 분산을 구하여라.
   c) 두 추정량의 평균제곱오차를 구하여라.
   d) 어느 추정량이 더 좋다고 할 수 있는가?

5. $X \sim \mathrm{N}(\theta, \sigma_1^2),\ Y \sim \mathrm{N}(\theta, \sigma_2^2)$ 이며 서로 독립이다. σ_1^2 와 σ_2^2 는 알려진 상수이며 X 와 Y 를 이용하여 θ 를 추정하려고 한다.
   a) $T_a = aX + (1-a)Y$ (단, $0 \le a \le 1$) 는 θ 의 비편향추정량임을 보여라.
   b) 위 형태를 가지는 추정량 T_a 중에서 분산을 최소로 하는 상수 a 를 구하여라.

6. $X \sim \text{Poisson}(\theta)$, $Y \sim \text{Poisson}(2\theta)$ 이며 서로 독립이다. 이때 X 와 Y 를 이용하여 θ 를 추정하려고 한다. θ 의 추정량 $\widehat{\theta} = aX + bY$ 에 대하여 다음 물음에 답하여라.

   a) $\widehat{\theta}$ 가 비편향추정량이 되기 위한 조건을 구하여라.

   b) $\widehat{\theta}$ 의 평균제곱오차를 구하여라.

   c) $\widehat{\theta} = aX + bY$ 형태의 비편향추정량 중에서 평균제곱오차가 가장 작은 추정량을 구하여라.

7. $X_1, X_2, \ldots, X_n$ 이 $\mathrm{N}(\theta, \theta^2)$ 로부터의 랜덤샘플이다. 단, $\theta > 0$ 이다.

   a) 상수 c 에 대하여 $T_c(\boldsymbol{X}) = c\overline{X} + (1-c)aS$ 가 θ 의 비편향추정량임을 보여라. 단, $a = \dfrac{\sqrt{n-1}\Gamma(\frac{n-1}{2})}{\sqrt{2}\Gamma(\frac{n}{2})}$ 이다.

   b) $T_c(\boldsymbol{X}) = c\overline{X} + (1-c)aS$ 형태의 추정량 중에서 분산을 최소로 하는 상수 c 를 구하여라.

   c) $T_{c_1, c_2}(\boldsymbol{X}) = c_1\overline{X} + c_2 aS$ 형태로 주어지는 추정량의 평균제곱오차를 구하여라.

8. $X_1, X_2, \ldots, X_n$ 이 확률밀도함수 $f(x|\theta) = \dfrac{1}{\theta} e^{-\frac{x}{\theta}},\ x > 0$ 을 가지는 랜덤샘플이다. θ 의 추정량 $T_c(\boldsymbol{X}) = c\overline{X}$ 에 대하여 다음 물음에 답하여라.

   a) $T_c(\boldsymbol{X})$ 의 평균제곱오차를 구하여라.

   b) 위의 평균제곱오차를 최소로 하는 상수 c 를 구하여라.

9. $X_1, X_2, \ldots, X_n$ 이 $\mathrm{N}(0, \sigma^2)$ 로부터의 랜덤샘플이다.

   a) σ^2 의 최대가능도추정량을 구하여라.

   b) $(\sigma^2)^2$ 의 최대가능도추정량을 구하여라.

   c) σ^2 에 대한 피셔 정보수를 구하여라.

10. $X_1, X_2, \ldots, X_n$ 이 확률밀도함수 $f(x|\theta) = \dfrac{1}{\theta} e^{-\frac{x}{\theta}},\ x > 0$ 을 가지는 랜덤샘플이다.

    a) θ 의 적률추정량을 구하여라.

    b) θ 의 최대가능도추정량을 구하여라.

    c) 위에서 구한 최대가능도추정량의 분산을 구하여라.

    d) 피셔 정보수를 이용하여 최대가능도추정량의 점근분산을 구하여라

11. $X_1, X_2, \ldots, X_n$ 이 확률밀도함수 $f(x|\theta) = \dfrac{1}{\theta^2} xe^{-\frac{x}{\theta}},\ x > 0$ 인 $\text{Gamma}(2, \frac{1}{\theta})$ 로부터의 랜덤샘플이다.

    a) θ 의 최대가능도추정량을 구하여라.

    b) θ 에 대한 피셔 정보수를 구하여라.

    c) 최대가능도추정량의 점근분산을 구하여라.

12. $X_1, X_2, \ldots, X_n$ 이 확률밀도함수 $f(x|\theta) = 2\theta x e^{-\theta x^2},\ x > 0$ 을 가지는 랜덤샘플이다.

    a) θ 의 최대가능도추정량 $\widehat{\theta}$ 을 구하여라.

    b) $\sqrt{n}(\widehat{\theta} - \theta)$ 의 점근분포를 구하여라.

    c) θ 에 대한 근사적인 95% 신뢰구간을 구하여라.

13. $X_1, X_2, \ldots, X_n$ 이 Bernoulli(θ) 로부터의 랜덤샘플일 때 $X_1X_2X_3X_4$ 가 θ^4 의 불평추정량임을 이용하여 θ^4 의 최소분산비편향추정량을 구하여라.

14. B(m, θ) 로부터의 랜덤샘플 $X_1, X_2, \ldots, X_n$ 을 이용하여 θ^m 의 최소분산비편향추정량을 구하려 한다.

    a) θ 에 대한 완비충분통계량을 구하여라.

    b) $Y = \begin{cases} 1, & X_1 = m \text{ 일 때} \\ 0, & X_1 \neq m \text{ 일 때} \end{cases}$ 는 θ^m 의 비편향추정량임을 보여라.

    c) θ^m 의 최소분산비편향추정량을 구하여라.

15. Poisson(θ) 로부터의 랜덤샘플 $X_1, X_2, \ldots, X_n$ 을 이용하여 $P(X_1 = 1) = \theta e^{-\theta}$ 의 최소분산비편향추정량을 구하려 한다.

    a) $Y = \begin{cases} 1, & X_1 = 1 \text{ 일 때} \\ 0, & X_1 \neq 1 \text{ 일 때} \end{cases}$ 는 $\theta e^{-\theta}$ 의 비편향추정량임을 보여라.

    b) $\theta e^{-\theta}$ 의 최소분산비편향추정량을 구하여라.

16. $X_1, X_2, \ldots, X_n$ 이 N$(0, \sigma^2)$ 로부터의 랜덤샘플이다.

    a) σ^2 의 최소분산비편향추정량을 구하여라.

    b) $(\sigma^2)^2$ 의 최소분산비편향추정량을 구하여라.

17. $X_k \sim$ N$(k\theta, 1),\ k = 1, 2, \ldots, n$ 이며 서로 독립일 때 다음 물음에 답하여라.

    a) 가능도를 구하여라.

    b) θ 의 최대가능도추정량을 구하여라.

18. 베르누이분포로부터의 랜덤샘플을 이용하여 예 8.29에서 구한 θ 의 두 신뢰구간 (8.19) 과 (8.20) 는 표본의 크기가 커질 때 큰 차이가 나지 않는 이유를 설명하여라.

제 9 장

가설검정

자료를 이용한 모집단에 대한 추론으로 모수의 점추정 외에 **가설검정**(hypothesis testing)이 있다. 가설검정이란 모수에 대한 가설을 세우고 자료를 이용하여 가설이 타당한지 판단하는 방법을 말한다. 가설검정을 간단히 검정이라고 하기도 한다.

가설검정에 대한 이해를 돕기 위해 간단한 예를 생각해 보자. 이산형 확률변수 X의 확률질량함수는 모수 θ의 값에 따라 다음과 같이 주어진다고 가정하자. 모수는 θ_0와 θ_1의 두 가지만 가능하다고 하자.

	-3	-2	-1	0	1	2	3
$f(x\,\vert\,\theta_0)$	0.6	0.3	0.1	0.0	0.0	0.0	0.0
$f(x\,\vert\,\theta_1)$	0.0	0.0	0.0	0.1	0.2	0.3	0.4

이 경우 모수에 대해 다음과 같은 두 가지 가설을 세울 수 있다.

- 가설 0: 모수의 참값은 θ_0이다.
- 가설 1: 모수의 참값은 θ_1이다.

X를 관측한 후 모수에 대한 위의 두 가설 중에서 어느 가설이 옳은지 결정하는 방법으로 다음과 같은 방법을 사용한다고 하자.

$$X \in \{-3, -2, -1\} \text{ 이면 가설 0이 옳다고 판단한다.}$$
$$X \in \{0,\ 1,\ 2,\ 3\} \text{ 이면 가설 1이 옳다고 판단한다.}$$

위의 방법을 사용한다면 X를 이용하여 항상 모수의 가설에 대한 올바른 판단을 할 수 있다. 즉 모수의 참값이 θ_0일 때는 X는 -3, -2, -1 중에서 하나의 값을 가지며 이 경우 가설 0이 옳다고 판단을 내리므로 올바른 판단을 할 수 있다. 마찬가지로 모수의 참값이 θ_1인 경우에도 가설 1이 옳다고 올바른 판단을 하게 된다. 그렇지만 위의 경우는 특수한 경우이며 일반적으로 모수에 대해서 항상 올바른 판단을 내릴 수 있는 것은 아니다.

앞의 확률분포 대신 X의 확률분포가 표 9.1와 같다고 하자.

표 9.1: X의 확률분포

	-3	-2	-1	0	1	2	3
$f(x\|\theta_0)$	0.4	0.2	0.1	0.1	0.1	0.05	0.05
$f(x\|\theta_1)$	0.05	0.05	0.1	0.1	0.2	0.2	0.3

X가 표 9.1와 같은 확률분포를 가진다면 항상 올바른 판단을 내릴 수 있는 가설검정 방법은 없다. 예를 들어 앞의 방법을 사용한다면 X의 값이 -3, -2 또는 -1일 때 모수는 θ_0라고 결정을 내리지만 모수의 참값은 θ_1일 수도 있기 때문이다. 일반적으로 가설검정에서는 다음과 같은 두 가지 오류를 범할 수 있다.

- 첫 번째 종류의 오류 : 가설 0이 참일 때 가설 1이 옳다고 판단하는 오류
- 두 번째 종류의 오류 : 가설 1이 참일 때 가설 0이 옳다고 판단하는 오류

모수의 점추정 방법으로 여러 추정법이 있는 것과 같이 가설검정에도 여러 가지 방법이 있을 수 있다. 이러한 여러 방법 중에서 가장 좋은 검정법을 찾는 것이 문제가 된다. 가설검정의 비교기준으로는 위에서 언급한 두 가지 오류를 범할 확률을 이용할 수 있다.

검정 방법의 비교를 위해 다음 몇 가지 검정법을 생각하여 보자.

- 검정법 1 : $X \in \{-3, -2, -1\}$이면 가설 0이 옳고 $X \in \{0, 1, 2, 3\}$이면 가설 1이 옳다고 판단한다.
- 검정법 2 : $X \in \{0, 1, 2, 3\}$이면 가설 0이 옳고 $X \in \{-3, -2, -1\}$이면 가설 1이 옳다고 판단한다.
- 검정법 3 : $X \in \{-3, -2\}$이면 가설 0이 옳고 $X \in \{-1, 0, 1, 2, 3\}$이면 가설 1이 옳다고 판단한다.

위의 각 검정법에 대해 두 가지 오류를 범할 확률은 다음과 같다.

검정법	첫 번째 종류의 오류를 범할 확률	두 번째 종류의 오류를 범할 확률
검정법 1	0.3	0.2
검정법 2	0.7	0.8
검정법 3	0.4	0.1

예를 들어 검정법 1에 대하여 첫 번째 오류를 범할 확률은 모수가 θ_0 일 때 X 가 $\{0, 1, 2, 3\}$ 에 속할 확률이므로 0.3이다. 또 검정법 1에 대하여 두 번째 오류를 범할 확률은 모수가 θ_1 일 때 X 가 $\{-3, -2, -1\}$ 에 속할 확률이므로 0.2이다. 검정법 1의 두 가지 오류를 범할 확률이 모두 검정법 2에 비해 작으므로 검정법 1이 검정법 2보다 우수하다고 판단할 수 있다. 그렇지만 검정법 1과 검정법 3을 비교하면 어느 검정법이 더 우수한지 판단하는 것이 쉽지 않다. 첫 번째 종류의 오류를 범할 확률은 검정법 1이 작지만 두 번째 종류의 오류를 범할 확률은 검정법 3이 작기 때문에 어느 검정법이 더 좋은지 판단하는 것은 쉽지 않다.

앞장에서 모수의 추정량의 비교기준으로 평균제곱오차를 사용한 것처럼 가설검정에서는 두 가지 오류를 범할 확률을 이용하여 검정 방법을 비교할 수 있다. 최적의 검정법이란 가능하면 위의 두 가지 오류를 범할 확률을 작게 하는 검정법일 것이다. 그렇지만 일반적으로 위의 두 가지 오류를 동시에 작게 하는 검정법은 존재하지 않는다. 두 가지 오류를 범할 확률을 동시에 고려한 검정법의 비교는 쉽지 않기 때문에 먼저 최적의 검정법을 구하기 위한 기준을 정해야 한다.

9.1 용어

가설검정에서는 모집단에 대한 가설을 세우고 이 가설이 옳은지 판단하게 된다. 통계적 가설검정은 관측된 자료를 이용하여 가설을 입증하는 과정이라고 볼 수 있다. 가설검정을 통해 입증하고자 하는 가설을 **대립가설**(alternative hypothesis)이라 하며 H_1 또는 H_A 로 나타낸다. 대립가설이 옳은지 판단하는 한 가지 방법은 대립가설에 반대되는 가설을 가정하고 이 가정 하에서 주어진 자료를 얻을 확률을 이용하는 것이다. 이때 대립가설에 반대되는 가설을 **귀무가설**(null hypothesis)이라 하고 H_0 로 나타낸다. 기본적인 통계적 가설검정의 아이디어는 귀무가설이 참이라는 가정 하에서 주어진 자료보다 더 극단적인 값을 얻을 확률이 매우 작다면 귀무가설이 참이라는 가설을 기각하는 것이다.

$\boldsymbol{X}$ 의 확률밀도함수 또는 확률질량함수를 $f(\boldsymbol{x}|\theta)$ 라 하자. 모수 θ 에 대한 귀무가설과 대립가설은 다음과 같이 나타낼 수 있다.

$$\begin{aligned} H_0 &: \ \theta \in \Theta_0 \\ H_1 &: \ \theta \in \Theta_1 \end{aligned}$$

단, $\Theta_0 \cap \Theta_1 = \emptyset$ 이다. 즉 대립가설이 참이면 모수 θ 는 집합 Θ_1 에 속하고 귀무가설이 참이면 θ 는 집합 Θ_0 에 속하게 된다.

귀무가설이나 대립가설 하에서 $\boldsymbol{X}$ 의 확률분포가 하나로 결정될 때 이 가설을 **단순가설**(simple hypothesis)이라 하며, 확률분포가 하나로 결정되지 않을 때 **복합가설**(composite hypothesis)이라 한다. 예를 들어 $\mathrm{N}(\theta,\, 1)$ 을 따르는 확률변수 $\boldsymbol{X}$ 에 대하여 귀무가설이 $H_0: \ \theta = 1$ 이면 귀무가설이 참일 때 $\boldsymbol{X}$ 의 확률분포가 $\mathrm{N}(1,\, 1)$ 로 결정되므로 귀무가설은 단순가설이다. 한편 대립가설이 $H_1: \ \theta > 2$ 이라면 대립가설이 참일 때 $\boldsymbol{X}$ 의 확률분포는 $\mathrm{N}(3,\, 1)$ 일 수도 있고 $\mathrm{N}(5,\, 1)$ 일 수도 있는 것처럼 확률분포가 하나로 결정되지 않으므로 대립가설은 복합가설이다.

가설검정은 자료가 어떠한 값을 가질 때 귀무가설을 기각할 것인지에 따라 결정된다. 귀무가설을 기각하는 자료 값의 영역을 **기각역**(rejection region)이라 하며 R 로 나타낸다. 관측한 자료가

기각역 R에 속하면 귀무가설을 기각하고 기각역에 속하지 않으면 귀무가설을 기각하지 않게 된다.

앞에서 설명한 것처럼 검정에서는 두 가지 오류가 있다. 귀무가설이 참일 때 귀무가설을 기각하는 오류를 **제1종 오류**(type I error)라 하며, 대립가설이 참일 때 귀무가설을 기각하지 못하는 오류를 **제2종 오류**(type II error)라 한다. 일반적으로 제1종 오류를 범할 확률을 α로 나타내며 제2종 오류를 범할 확률을 β로 나타낸다. 즉 주어진 검정법의 기각역이 R일 때 α와 β는 다음과 같다.

$$\begin{aligned}\alpha &= P(\boldsymbol{X} \in R \,|\, H_0)\\ \beta &= P(\boldsymbol{X} \in R^c \,|\, H_1)\end{aligned}$$

위에서 $P(\cdot|H_0)$와 $P(\cdot|H_1)$는 각각 H_0와 H_1가 참일 때의 확률을 나타낸다.

예를 들어 X가 표 9.1에 주어진 확률분포를 가질 때 귀무가설과 대립가설이 다음과 같은 가설검정에서

$$H_0 : \ \theta = \theta_0 \quad \text{대} \quad H_1 : \ \theta = \theta_1$$

기각역이 $\{0, 1, 2, 3\}$인 검정법의 제1종 오류를 범할 확률과 제2종 오류를 범할 확률은 다음과 같다.

$$\alpha = 0.3, \ \beta = 0.2$$

대립가설이 참일 때 귀무가설을 기각할 확률을 **검정력**(power)이라 한다. 즉 검정력은 $P(\boldsymbol{X} \in R|H_1)$이며 이는 $1-\beta$이다. 위에서 다룬 예의 가설검정에서 주어진 검정법의 검정력은 0.8이 된다. 참고로 복합가설인 경우는 귀무가설이나 대립가설 하에서 모수 θ의 값이 하나로 결정되지 않으므로 α와 β는 θ의 함수가 된다.

기각역을 이용한 검정법은 다음과 같은 함수 $\phi(\boldsymbol{x})$를 사용하여 나타낼 수 있다. $\boldsymbol{X} = \boldsymbol{x}$일 때 $\phi(\boldsymbol{x})$는 0 또는 1의 값을 가지는 함수이며 검정법의 기각역이 R일 때 $\phi(\boldsymbol{x})$는 다음과 같다고 하자.

$$\phi(\boldsymbol{x}) \ = \ \begin{cases} 1, & \boldsymbol{x} \in R \text{일 때} \\ 0, & \boldsymbol{x} \notin R \text{일 때} \end{cases}$$

즉 $\boldsymbol{x} \in R$일 때 $\phi(\boldsymbol{x}) = 1$이며 $\boldsymbol{x} \notin R$일 때 $\phi(\boldsymbol{x}) = 0$이 된다. 관측 값 $\boldsymbol{x}$에서 $\phi(\boldsymbol{x})$의 값이 1이라는 것은 귀무가설을 기각한다는 것을 의미하며 $\phi(\boldsymbol{x})$의 값이 0이라는 것은 귀무가설을 기각하지 못한다는 것을 의미한다. 앞의 예에서 기각역이 $\{0, 1, 2, 3\}$일 때 $\phi(X)$는 다음과 같이 나타낼 수 있다.

$$\phi(x) \ = \ \begin{cases} 1, & x \ \in \ \{0, 1, 2, 3\} \text{일 때} \\ 0, & x \ \in \ \{-3, -2, -1\} \text{일 때} \end{cases}$$

귀무가설이 참일 때 $\phi(\boldsymbol{X})$의 기댓값을 구하면 다음과 같다.

$$\begin{aligned} E(\phi(\boldsymbol{X})\,|\,H_0) &= 1\times P(\boldsymbol{X}\in R\,|\,H_0)+0\times P(\boldsymbol{X}\in R^c\,|\,H_0) \\ &= P(\boldsymbol{X}\in R\,|\,H_0) \;=\; \alpha \end{aligned}$$

즉 귀무가설이 참일 때 $\phi(\boldsymbol{X})$의 기댓값은 제1종 오류를 범할 확률이 된다. 한편 대립가설이 참일 때 $\phi(\boldsymbol{X})$의 기댓값을 구하면 다음과 같다.

$$\begin{aligned} E(\phi(\boldsymbol{X})\,|\,H_1) &= 1\times P(\boldsymbol{X}\in R\,|\,H_1)+0\times P(\boldsymbol{X}\in R^c\,|\,H_1) \\ &= P(\boldsymbol{X}\in R\,|\,H_1) \;=\; 1-P(\boldsymbol{X}\in R^c\,|\,H_1) \;=\; 1-\beta \end{aligned}$$

즉 대립가설이 참일 때 $\phi(\boldsymbol{X})$의 기댓값은 검정력이 된다.

가설검정에서 기각역 대신에 함수 $\phi(\boldsymbol{x})$를 사용함으로써 주어진 관측 값에서 귀무가설을 기각하거나 기각하지 못하는 단순한 검정법에서 벗어날 수 있다. 앞에서 설명한 것처럼 관측 값 $\boldsymbol{x}$에서 $\phi(\boldsymbol{x})=1$은 귀무가설을 기각하는 것을 의미하며 $\phi(\boldsymbol{x})=0$은 귀무가설을 기각하지 못한다는 것을 나타낸다. 이러한 개념을 확장하여 0과 1 사이의 값을 가지는 함수 $\phi(\boldsymbol{x})$의 값을 귀무가설을 기각할 확률로 해석할 수 있다. 이러한 함수 $\phi(\boldsymbol{x})$를 **검정함수**(test function)라 한다. 예를 들어 $\boldsymbol{X}=\boldsymbol{x}$에서 검정함수 값이 $\phi(\boldsymbol{x})=0.5$라면 이 관측 값에서는 귀무가설을 기각할 확률이 0.5가 된다. 이러한 경우는 앞면이 나올 확률이 0.5인 동전을 던져 앞면이 나오면 귀무가설을 기각하고 뒷면이 나오면 귀무가설을 기각하지 않는 검정법이라 할 수 있다.

검정법은 기각역에 의하여 결정되는 것과 같이 기각역을 확장한 검정함수에 의하여 결정된다. 따라서 검정법을 구하는 것은 기각역 또는 검정함수를 구하는 것으로 생각하면 된다. 앞에서 설명한 것과 같이 검정함수는 귀무가설을 기각할 확률을 나타내므로 두 가지 오류를 범할 확률 α와 β는 다음과 같이 나타낼 수 있다.

$$\begin{aligned} \alpha &= E(\phi(\boldsymbol{X})\,|\,H_0) \\ \beta &= E(1-\phi(\boldsymbol{X})\,|\,H_1) \end{aligned}$$

다음은 특정한 조건을 만족하는 검정법에 대한 정의이다.

정의 9.1 귀무가설에 속하는 모든 θ에서 제1종 오류를 범할 확률이 α 이하인 검정법을 **수준 α 검정**(level α test)이라 한다. 즉 $\max_{\theta\in\Theta_0} E(\phi(\boldsymbol{X})\,|\,\theta) \le \alpha$인 검정법 ϕ를 수준 α 검정이라 한다.

정의 9.2 제1종 오류를 범할 확률의 최댓값이 α인 검정법을 **크기 α 검정**(size α test)이라 한다. 즉 $\max_{\theta\in\Theta_0} E(\phi(\boldsymbol{X})\,|\,\theta) = \alpha$인 검정법 ϕ를 크기 α 검정이라 한다.

위 정의에서 $E(\phi(\boldsymbol{X})\,|\,\theta)$는 모수의 값이 θ일 때 $\phi(\boldsymbol{X})$의 기댓값을 나타낸다. 제1종 오류를 범할 확률은 모수 θ에 따라 달라진다. 귀무가설에 속하는 모든 θ에서 제1종 오류를 범할 확률이

α 이하인 검정법을 수준 α 검정이라 하며, 귀무가설에 속하는 모든 θ에서 제1종 오류를 범할 확률의 최댓값이 α인 검정법을 크기 α 검정이라 한다.

앞에서 다룬 예에서 기각역이 $\{0, 1, 2, 3\}$인 검정법은 제1종 오류를 범할 확률이 0.3이었다. 따라서 이 검정법은 크기 0.3 검정법이며 수준 0.3 검정법이라 할 수 있다. 또한 수준 0.4 검정법이라고 할 수도 있다.

9.2 최강력검정

귀무가설과 대립가설이 모두 단순가설인 경우 최적의 검정법을 구하는 방법을 생각하여 보자. 두 가설이 모두 단순 가설일 때 두 가설을 다음과 같이 나타낼 수 있다.

$$H_0:\ \theta = \theta_0 \quad \text{대} \quad H_1:\ \theta = \theta_1$$

앞에서 설명한 것처럼 검정법은 제1종 오류와 제2종 오류를 범할 확률을 기준으로 비교할 수 있다. 가능하면 이 확률을 작게 하는 검정법을 구해야 할 것이다. 그렇지만 일반적으로 한 종류의 오류를 범할 확률을 작게 하는 검정법은 다른 종류의 오류를 범할 확률을 크게 한다. 예를 들어 제1종 오류를 범할 확률을 작게 하기 위해서는 기각역을 작게 해야 하지만 이 경우 제2종 오류를 범할 확률은 커지게 된다. 최적의 검정법은 두 가지 오류를 범할 확률을 동시에 작게 하는 검정법이겠지만 이것은 현실적으로 불가능하므로 최적의 검정법에 대한 다른 기준을 제시하여야 한다.

앞에서 설명한 것처럼 통계적 가설검정은 자료를 통하여 대립가설을 입증하는 과정이라 하였다. 대립가설을 입증하는 한 가지 방법은 대립가설에 반대되는 귀무가설을 가정했을 때 관측한 자료를 얻을 확률이 매우 작다는 것을 보이는 것이다. 이러한 관점에서 보면 제2종 오류를 범할 확률보다는 제1종 오류를 범할 확률을 먼저 고려하여야 한다. 제1종 오류는 귀무가설이 참일 때 귀무가설을 기각하는 경우이다. 즉 입증하려는 가설이 참이 아닌데도 불구하고 이 가설이 옳다고 판단하는 오류이다. 따라서 이러한 오류가 다른 오류보다 더 중요할 수 있다.

구체적인 예를 들어 생각하여 보자. 어느 제약회사에서 새로운 약을 개발하여 시판하려 한다고 하자. 이 제약회사에서는 신약이 기존의 약보다 우수하다는 것을 입증하여야 한다. 따라서 "신약이 기존의 약보다 우수하다"는 가설을 대립가설로 "신약이 기존의 약보다 우수하지 않다"는 것을 귀무가설로 설정하여야 한다. 이때 제1종 오류는 신약이 기존의 약보다 우수하지 않은데도 불구하고 우수하다고 판단하는 오류이고, 제2종 오류는 신약이 기존의 약보다 우수한데도 우수하지 않다고 판단하는 오류이다. 두 가지 오류 중에서 더 치명적인 것은 제1종 오류라고 할 수 있다. 제2종 오류를 범하는 것은 더 좋은 약효를 가지는 신약이 사용되지 못하는 결과를 초래한다. 그러나 기존의 치료제를 계속하여 사용할 것이므로 사회에 끼치는 영향은 그리 크지 않다. 반대로 제1종 오류를 범하는 것은 우수하지 못한 신약을 더 약효가 뛰어나다고 판단을 내리므로 약효가 어느 정도 입증된 기존의 치료제를 버리고 신약을 사용하게 될 것이다. 여기에서 발생하는 부작용은 매우 크다고 할 수 있다. 이처럼 가설검정에서 발생하는 두 가지 오류의 중요도는 다를 수 있다.

가설검정은 대립가설을 입증하는 과정이라는 관점에서 최적의 검정법을 찾기 위한 다음의 기준을 사용하게 된다. 먼저 여러 검정법 중에서 제1종 오류를 범할 확률이 일정한 값 이하가 되는 검정법을 선택하고, 이 중에서 제2종 오류를 범할 확률을 최소로 하는 검정법을 택하게 된다.

다르게 표현하면 제1종 오류를 범할 확률이 일정한 값 이하가 되는 검정법 중에서 검정력을 가장 크게 하는 검정법을 택하는 것이다. 이러한 검정법을 **최강력검정**(most powerful test)이라 한다. 여기서 제1종 오류를 범할 확률이 일정한 값 이하가 되도록 정할 때 이 값을 **유의수준**(significance level)이라 한다. 즉 최강력검정에서는 제1종 오류를 범할 확률과 제2종 오류를 범할 확률을 동시에 고려하는 것이 아니라 먼저 제1종 오류를 범할 확률을 기준으로 일정한 수준에 이른 검정법만으로 한정한다. 그리고 이 중에서 제2종 오류를 범할 확률을 가장 작게 하는 검정법을 선택하는 것이다. 예를 들어 유의수준 0.05에서 최강력검정법을 구하는 것은 먼저 제1종 오류를 범할 확률이 0.05 이하인 검정법을 선택한 후 이 중에서 제2종 오류를 범할 확률이 가장 작은 검정법을 찾는 것이다.

정의 9.3 다음 조건을 만족하는 ϕ를 유의수준 α에서 최강력검정이라 한다.

(i) ϕ는 수준 α 검정이다. 즉 $E(\phi(\boldsymbol{X})\,|\,H_0) \le \alpha$이다.

(ii) 임의의 수준 α 검정 ϕ'에 대하여 ϕ의 검정력이 ϕ'의 검정력보다 크거나 같다. 즉 $E(\phi'(\boldsymbol{X})\,|\,H_0) \le \alpha$인 검정 ϕ'에 대하여 $E(\phi(\boldsymbol{X})\,|\,H_1) \ge E(\phi'(\boldsymbol{X})\,|\,H_1)$이다.

다음 정리는 최강력검정법을 구하는 방법을 보여주고 있다.

정리 9.1 [네이만-피어슨 정리, Neyman-Pearson lemma] 다음과 같은 가설에 대한 가설검정에서

$$H_0:\ \theta = \theta_0 \quad \text{대} \quad H_1:\ \theta = \theta_1$$

다음과 같이 주어지는 검정법 ϕ는 유의수준 α에서 최강력검정법이다.

$$\phi(\boldsymbol{x}) = \begin{cases} 1, & \dfrac{f(\boldsymbol{x}\,|\,\theta_1)}{f(\boldsymbol{x}\,|\,\theta_0)} \ \ge\ k \text{ 일 때} \\ 0, & \dfrac{f(\boldsymbol{x}\,|\,\theta_1)}{f(\boldsymbol{x}\,|\,\theta_0)} \ <\ k \text{ 일 때} \end{cases}$$

단, α는 다음을 만족하는 값이다.

$$E(\phi(\boldsymbol{X})\,|\,H_0) \ = \ P\left(\frac{f(\boldsymbol{X}\,|\,\theta_1)}{f(\boldsymbol{X}\,|\,\theta_0)} \ \ge\ k|H_0\right) = \alpha$$

[증명] 증명은 연속형 확률분포의 경우만을 다루도록 하겠다. 먼저 주어진 조건에 의하여 ϕ는 수준 α 검정이다. ϕ가 최강력검정임을 보이기 위해서는 ϕ'를 임의의 수준 α 검정법이라 할 때 ϕ의 검정력이 ϕ'의 검정력보다 크다는 것을 보여야 한다. 즉 $E(\phi(\boldsymbol{X})\,|\,H_1) \ge E(\phi'(\boldsymbol{X})\,|\,H_1)$

임을 보여야 한다. $0 \le \phi'(\boldsymbol{x}) \le 1$라는 사실과 $\phi(\boldsymbol{x})$에 대한 주어진 조건으로부터 다음을 얻을 수 있다.

$$\Big(\phi(\boldsymbol{x}) - \phi'(\boldsymbol{x})\Big)\Big(f(\boldsymbol{x}\,|\,\theta_1) - k\,f(\boldsymbol{x}\,|\,\theta_0)\Big) \;\ge\; 0$$

따라서 다음이 성립한다.

$$\int_{-\infty}^{\infty} \Big(\phi(\boldsymbol{x}) - \phi'(\boldsymbol{x})\Big)\Big(f(\boldsymbol{x}\,|\,\theta_1) - k\,f(\boldsymbol{x}\,|\,\theta_0)\Big)\,d\boldsymbol{x} \;\ge\; 0$$

위의 부등식을 정리하면 다음을 얻을 수 있다.

$$\int_{-\infty}^{\infty} \phi(\boldsymbol{x})f(\boldsymbol{x}\,|\,\theta_1) - \int_{-\infty}^{\infty} \phi'(\boldsymbol{x})f(\boldsymbol{x}\,|\,\theta_1) \;\ge\; k\Big(\int_{-\infty}^{\infty} \phi(\boldsymbol{x})f(\boldsymbol{x}\,|\,\theta_0) - \int_{-\infty}^{\infty} \phi'(\boldsymbol{x})f(\boldsymbol{x}\,|\,\theta_0)\Big)$$

위 식으로부터 다음이 성립함을 보일 수 있다.

$$\begin{aligned} E(\phi(\boldsymbol{X})\,|\,H_1) - E(\phi'(\boldsymbol{X})\,|\,H_1) \;&\ge\; k(E(\phi(\boldsymbol{X})\,|\,H_0) - E(\phi'(\boldsymbol{X})\,|\,H_0)) \\ &=\; k(\alpha - E(\phi'(\boldsymbol{X})\,|\,H_0)) \end{aligned}$$

한편 ϕ'는 수준 α 검정법이므로 $E(\phi'(\boldsymbol{X})\,|\,H_0) \;\le\; \alpha$이다. 따라서 다음이 성립한다.

$$E(\phi(\boldsymbol{X})\,|\,H_1) \;\ge\; E(\phi'(\boldsymbol{X})\,|\,H_1)$$

■

위 정리가 의미하는 것은 최강력검정법은 확률밀도함수의 비인 $\frac{f(\boldsymbol{x}|\theta_1)}{f(\boldsymbol{x}|\theta_0)}$가 큰 값부터 기각역에 속하여야 한다는 것이다. 그리고 경계 값인 상수 k는 주어진 유의수준에 따라 결정하면 된다. 이를 구체적으로 보이기 위해 다음 예를 살펴보도록 하자.

예 9.1 이산형 확률변수 X의 확률질량함수가 다음과 같을 때 X를 이용하여 다음 가설에 대한 검정을 하려고 한다.

$$H_0: \; \theta = \theta_0 \quad \text{대} \quad H_1: \; \theta = \theta_1$$

	-3	-2	-1	0	1	2	3
$f(x\|\theta_0)$	0.4	0.2	0.1	0.1	0.1	0.05	0.05
$f(x\|\theta_1)$	0.05	0.05	0.1	0.1	0.2	0.2	0.3

(a) 유의수준 0.05에서 최강력검정법을 구하여라.

(b) 유의수준 0.1에서 최강력검정법을 구하여라.

(c) 유의수준 0.075에서 최강력검정법을 구하여라.

[풀이]

(a) 주어진 가설에 대한 최강력검정법을 구하기 위해 $\frac{f(x|\theta_1)}{f(x|\theta_0)}$ 를 구하면 다음과 같다.

	-3	-2	-1	0	1	2	3
$\frac{f(x\mid\theta_1)}{f(x\mid\theta_0)}$	0.125	0.25	1	1	2	4	6

$\frac{f(x\mid\theta_1)}{f(x\mid\theta_0)}$ 의 값을 보면 $x=3$일 때 가장 크고 그 다음은 $x=2$일 때이다. 따라서 최강력검정법의 기각역은 먼저 3을 포함하고 그 다음은 2를 포함하는 순서로 결정되어야 한다. 기각역이 어떤 값까지 포함할 것인지는 주어진 유의수준에 따라 결정된다. 기각역에 제일 먼저 3이 포함되어야 하며, 기각역이 3일 때 제1종 오류를 범할 확률이 0.05이므로 유의수준 0.05에서 최강력검정법의 기각역은 3만 포함 하여야 한다. 따라서 유의수준 0.05에서 최강력검정법은 다음과 같다.

$$\phi(x)=\begin{cases}1, & x=3\text{일 때}\\ 0, & x\neq 3\text{일 때}\end{cases}$$

(b) 같은 방법으로 유의수준 0.1에서 최강력검정법은 다음과 같다.

$$\phi(x)=\begin{cases}1, & x\in\{2,3\}\text{일 때}\\ 0, & x\in\{-3,-2,-1,0,1\}\text{일 때}\end{cases}$$

(c) 먼저 기각역이 3만 포함하면 제1종 오류를 범할 확률이 0.05이므로 주어진 유의수준 0.075보다 작다. 그 다음으로 기각역에 포함될 값은 2이다. 기각역이 3과 2로 이루어지면 제1종 오류를 범할 확률은 0.1이 되어 주어진 유의수준 보다 크게 된다. 따라서 2는 기각역에 완전히 포함될 수는 없고 관측 값이 2일 때 귀무가설을 기각할 확률을 1이 아닌 다른 값으로 정하면 된다. 2일 때 귀무가설을 기각할 확률을 0.5로 정하면 제1종 오류를 범할 확률이 0.075가 된다. 따라서 유의수준 0.075에서 최강력검정법은 다음과 같다.

$$\phi(x)=\begin{cases}1, & x=3\text{일 때}\\ 0.5, & x=2\text{일 때}\\ 0, & x\in\{-3,-2,-1,0,1\}\text{일 때}\end{cases}$$

■

앞의 예와 같이 이산형 확률분포에 대한 검정법은 특정한 관측 값에서 귀무가설을 기각할 확률이 0이나 1이 아닌 값을 가지는 경우가 있다. 이런 검정법을 **랜덤화검정**(randomized test)이라 한다.

예 9.2 $X_1, X_2, \ldots, X_n$ 이 Bernoulli(θ) 로부터의 랜덤샘플일 때 다음 가설에 대해 검정을 하려고 한다.

$$H_0 : \theta = \theta_0 \quad \text{대} \quad H_1 : \theta = \theta_1 \qquad (\text{단}, \theta_1 > \theta_0)$$

이때 최강력검정법의 기각역의 형태를 구하여라. 또 $n = 10$ 이고 $\theta_0 = 0.5$ 일 때 유의수준 $\alpha = 0.5^{10}$, $0.5^{10} + 10(0.5)^{10}$, $0.5^{10} + 5(0.5)^{10}$ 에서 최강력검정법의 기각역을 구하여라.

[풀이] 귀무가설과 대립가설 하에서 확률질량함수의 비율은 다음과 같다.

$$\begin{aligned} \frac{f(\boldsymbol{x}|\theta_1)}{f(\boldsymbol{x}|\theta_0)} &= \frac{(\theta_1)^{\sum_{i=1}^n x_i}(1-\theta_1)^{n-\sum_{i=1}^n x_i}}{(\theta_0)^{\sum_{i=1}^n x_i}(1-\theta_0)^{n-\sum_{i=1}^n x_i}} \\ &= \left(\frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)}\right)^{\sum_{i=1}^n x_i} \left(\frac{1-\theta_1}{1-\theta_0}\right)^n \end{aligned}$$

$\theta_1 > \theta_0$ 일 때 이 비율은 $\sum_{i=1}^n x_i$ 의 증가함수이다. 따라서 최강력검정법의 기각역의 형태는 다음과 같다.

$$\sum_{i=1}^n X_i \;\geq\; k$$

상수 k 의 값은 주어진 유의수준에 의하여 결정된다. 귀무가설이 참일 때 $\sum_{i=1}^n X_i \sim \mathrm{B}(n, \theta_0)$ 이므로 유의수준 α 에서 최강력검정법은 다음이 성립하도록 상수 k 를 정하면 된다.

$$P\,(\,\mathrm{B}(n, \theta_0) \;\geq\; k\,) \;=\; \alpha$$

$n = 10$ 이고 $\theta_0 = 0.5$ 일 때 다음이 성립하므로

$$P(\,\mathrm{B}(10, 0.5) \;\geq\; 10\,) \;=\; 0.5^{10}$$

$\alpha = 0.5^{10}$ 에서 최강력검정법의 기각역은 $\{10\}$ 이다. 또 $\alpha = 0.5^{10} + 10(0.5)^{10}$ 에서 최강력검정법의 기각역은 $\{9, 10\}$ 이다. 한편 다음이 성립하므로

$$P(\,\mathrm{B}(10, 0.5) \;= 10\,) + 0.5P(\,\mathrm{B}(10, 0.5) \;= 9\,) \;=\; 0.5^{10} + 5(0.5)^{10}$$

$\alpha = 0.5^{10} + 5(0.5)^{10}$ 에서 최강력검정법은 다음과 같다.

$$\phi(\boldsymbol{x}) = \begin{cases} 1, & \sum_{i=1}^{10} x_i = 10 \text{ 일 때} \\ 0.5, & \sum_{i=1}^{10} x_i = 9 \text{ 일 때} \\ 0, & \sum_{i=1}^{10} x_i \in \{0, 1, \ldots, 8\} \text{ 일 때} \end{cases}$$

■

예 9.3 $X_1, X_2, \ldots, X_n$ 이 Poisson(θ) 로부터의 랜덤샘플일 때 다음 가설에 대한 유의수준 α 에서 최강력검정법의 기각역을 구하여라.

$$H_0 : \theta = \theta_0 \quad \text{대} \quad H_1 : \theta = \theta_1 \qquad (\text{단}, \theta_1 > \theta_0)$$

[풀이] 귀무가설과 대립가설 하에서 확률밀도함수의 비율은 다음과 같다.

$$\frac{f(\boldsymbol{x}|\theta_1)}{f(\boldsymbol{x}|\theta_0)} = \frac{\dfrac{1}{x_1!x_2!\cdots x_n!}\theta_1^{\sum_{i=1}^n x_i} e^{-n\theta_1}}{\dfrac{1}{x_1!x_2!\cdots x_n!}\theta_0^{\sum_{i=1}^n x_i} e^{-n\theta_0}} = \left(\frac{\theta_1}{\theta_0}\right)^{\sum_{i=1}^n x_i} e^{-n(\theta_1 - \theta_0)}$$

$\theta_1 > \theta_0$ 일 때 이 비율은 $\sum_{i=1}^n x_i$ 의 증가함수이다. 따라서 최강력검정법의 기각역의 형태는 다음과 같다.

$$\sum_{i=1}^n X_i \geq k$$

귀무가설이 참일 때 $\sum_{i=1}^n X_i \sim \text{Poisson}(n\theta_0)$ 이므로 유의수준 α 에서 최강력검정법은 다음이 성립하도록 상수 k 를 정하면 된다.

$$P\left(\text{Poisson}(n\theta_0) \geq k\right) = \alpha$$

■

예 9.4 $X_1, X_2, \ldots, X_n$ 이 N$(\theta,\, 1)$ 로부터의 랜덤샘플일 때 다음 가설에 대한 유의수준 α 에서 최강력검정법의 기각역을 구하여라.

$$H_0 : \theta = \theta_0 \quad \text{대} \quad H_1 : \theta = \theta_1 \qquad (\text{단}, \theta_1 > \theta_0)$$

[풀이] 귀무가설과 대립가설 하에서 확률밀도함수의 비율은 다음과 같다.

$$\begin{aligned}\frac{f(\boldsymbol{x}\,|\,\theta_1)}{f(\boldsymbol{x}\,|\,\theta_0)} &= \frac{(2\pi)^{-\frac{n}{2}}\exp\left(-\sum_{i=1}^n \frac{(x_i-\theta_1)^2}{2}\right)}{(2\pi)^{-\frac{n}{2}}\exp\left(-\sum_{i=1}^n \frac{(x_i-\theta_0)^2}{2}\right)} \\ &= \exp\left((\theta_1-\theta_0)\sum_{i=1}^n x_i - \frac{n}{2}(\theta_1^2-\theta_0^2)\right)\end{aligned}$$

$\theta_1 > \theta_0$ 일 때 이 비율은 $\sum_{i=1}^n x_i$ 의 증가함수이다. 따라서 최강력검정법의 기각역의 형태는 다음과 같다.

$$\sum_{i=1}^n X_i \;\geq\; k \quad \text{또는} \quad \overline{X} \;\geq\; k$$

따라서 유의수준 α 에서 최강력검정법은 다음이 성립하도록 상수 k 를 정하면 된다.

$$P(\,\overline{X} \geq k\,|\,H_0\,) \;=\; \alpha$$

귀무가설이 참일 때 $\overline{X} \sim \mathrm{N}\left(\theta_0,\, \frac{1}{n}\right)$ 이므로 $P\left(\overline{X} \geq \theta_0 + \frac{Z_\alpha}{\sqrt{n}}\,\middle|\, H_0\right) = \alpha$ 이다. 단, Z_α 는 $P(\mathrm{N}(0,1) \geq Z_\alpha) = \alpha$ 인 상수이다. 따라서 유의수준 α 에서 최강력검정법은 다음과 같다.

$$\phi(\boldsymbol{x}) = \begin{cases} 1, & \sqrt{n}(\overline{X}-\theta_0) \geq Z_\alpha \text{ 일 때} \\ 0, & \sqrt{n}(\overline{X}-\theta_0) < Z_\alpha \text{ 일 때} \end{cases}$$

■

예 9.5 $X_1, X_2, \ldots, X_n$ 이 $\mathrm{Exp}(\theta)$ 로부터의 랜덤샘플일 때 다음 가설에 대한 유의수준 α 에서 최강력검정법의 기각역을 구하여라.

$$H_0:\; \theta = \theta_0 \quad \text{대} \quad H_1:\; \theta = \theta_1 \qquad (\text{단}, \theta_1 > \theta_0)$$

[풀이] 귀무가설과 대립가설 하에서 확률밀도함수의 비율은 다음과 같다.

$$\begin{aligned}\frac{f(\boldsymbol{x}\,|\,\theta_1)}{f(\boldsymbol{x}\,|\,\theta_0)} &= \frac{(\theta_1)^n e^{-\theta_1\sum_{i=1}^n x_i}}{(\theta_0)^n e^{-\theta_0\sum_{i=1}^n x_i}} \\ &= \left(\frac{\theta_1}{\theta_0}\right)^n e^{-(\theta_1-\theta_0)\sum_{i=1}^n x_i}\end{aligned}$$

$\theta_1 > \theta_0$ 일 때 이 비율은 $\sum_{i=1}^n x_i$ 의 감소함수이다. 따라서 최강력검정법의 기각역의 형태는 다음과 같다.

$$\sum_{i=1}^n X_i \le k$$

귀무가설이 참일 때 $\sum_{i=1}^n X_i \sim \text{Gamma}(n, \theta_0)$ 이므로 다음이 성립한다.

$$2\theta_0 \sum_{i=1}^n X_i \sim \text{Gamma}\left(n, \frac{1}{2}\right) = \chi^2_{2n}$$

따라서 유의수준 α 에서 최강력검정법의 기각역은 다음과 같다.

$$2\theta_0 \sum_{i=1}^n X_i \le \chi^2_{2n}(1-\alpha)$$

단, $\chi^2_{2n}(1-\alpha)$ 는 $P(\chi^2_{2n} \ge \chi^2_{2n}(1-\alpha)) = 1-\alpha$ 인 상수이다. ■

9.3 균일최강력검정

균일최강력검정(uniformly most powerful test)은 단순가설에 대한 최강력검정법을 복합가설로 확장한 검정법이다. 유의수준 α 에서 균일최강력검정법은 귀무가설에 속하는 모든 θ 에서 제1종 오류를 범할 확률이 α 이하이며, 임의의 수준 α 검정법과 비교하여 대립가설에 속하는 모든 θ 에서 검정력이 더 큰 검정법을 말한다.

정의 9.4 아래의 귀무가설과 대립가설에 대하여

$$H_0 : \theta \in \Theta_0 \quad \text{대} \quad H_1 : \theta \in \Theta_1$$

다음 조건을 만족하는 ϕ 를 유의수준 α 에서 균일최강력검정이라 한다.

(i) ϕ 는 수준 α 검정이다. 즉 모든 $\theta \in \Theta_0$ 에 대하여 $E(\phi(\boldsymbol{X}) \,|\, \theta) \le \alpha$ 이다.

(ii) ϕ' 이 임의의 수준 α 검정이라 할 때 모든 $\theta \in \Theta_1$ 에서 ϕ 의 검정력이 ϕ' 의 검정력보다 크거나 같다. 즉 모든 $\theta \in \Theta_1$ 에 대하여 $E(\phi(\boldsymbol{X}) \,|\, \theta) \ge E(\phi'(\boldsymbol{X}) \,|\, \theta)$ 이다.

균일최강력검정법은 항상 존재하는 것은 아니며 일반적으로 귀무가설과 대립가설이 $H_0 : \theta \le \theta_0$ 와 $H_1 : \theta > \theta_0$ 처럼 어느 한쪽으로 주어지는 단측검정일 때 최강력검정법을 구할 수 있는 경우가

많다. 예를 들어 $X_1, X_2, \ldots, X_n$ 이 $\mathrm{N}(\theta, 1)$ 로부터의 랜덤샘플일 때

$$H_0 : \theta \leq \theta_0 \quad \text{대} \quad H_1 : \ \theta > \theta_0 \tag{9.1}$$

위의 가설에 대한 유의수준 α 에서 균일최강력검정법의 기각역은 다음과 같음을 보일 수 있다.

$$\phi(\boldsymbol{x}) = \begin{cases} 1, & \sqrt{n}(\overline{X} - \theta_0) \geq Z_\alpha \text{ 일 때} \\ 0, & \sqrt{n}(\overline{X} - \theta_0) < Z_\alpha \text{ 일 때} \end{cases}$$

이를 보이기 위해서는 먼저 ϕ가 수준 α 검정임을 보여야 한다. 귀무가설에 속하는 θ에 대하여 (즉 $\theta \leq \theta_0$ 일 때) 제1종 오류를 범할 확률은 다음과 같이 나타낼 수 있다.

$$\begin{aligned} E(\phi(\boldsymbol{X}) \,|\, \theta) &= P(\sqrt{n}(\overline{X} - \theta_0) \geq Z_\alpha \,|\, \theta) \\ &= P(\sqrt{n}(\overline{X} - \theta) + \sqrt{n}(\theta - \theta_0) \geq Z_\alpha \,|\, \theta) \end{aligned}$$

모수의 참값이 θ일 때 $\sqrt{n}(\overline{X} - \theta) \ \sim \ \mathrm{N}(0,\, 1)$ 이므로 위의 확률은 다음과 같다.

$$\begin{aligned} E(\phi(\boldsymbol{X}) \,|\, \theta) &= P(\,\mathrm{N}(0,1) \geq Z_\alpha - \sqrt{n}(\theta - \theta_0)\,) \\ &\leq P(\,\mathrm{N}(0,1) \geq Z_\alpha\,) \ = \ \alpha \end{aligned}$$

위에서 부등식은 $\sqrt{n}(\theta - \theta_0) < 0$ 이므로 성립하게 된다. 따라서 $\max_{\theta \leq \theta_0} E(\phi(X) \,|\, \theta) \leq \alpha$ 이며 ϕ는 수준 α 검정임을 보일 수 있다.

한편 ϕ' 를 임의의 수준 α 검정법이라 할 때 모든 $\theta > \theta_0$ 에 대하여 다음 부등식이 성립함을 보여야 한다.

$$E(\phi(\boldsymbol{X}) \,|\, \theta) \ \geq \ E(\phi'(\boldsymbol{X}) \,|\, \theta) \tag{9.2}$$

이를 위해 임의의 θ_1(단, $\theta_1 > \theta_0$) 에 대하여 다음과 같은 검정을 생각하여 보자.

$$H_0 : \ \theta = \theta_0 \quad \text{대} \quad H_1 : \ \theta = \theta_1 \tag{9.3}$$

예 9.4로부터 ϕ는 위 가설에 대한 유의수준 α 에서 최강력검정임을 알 수 있다. 또한 ϕ' 는 가설 (9.1) 에 대한 수준 α 검정법이므로 $\max_{\theta \leq \theta_0} E(\phi'(X) \,|\, \theta) \leq \alpha$ 이다. 이는 $E(\phi'(\boldsymbol{X}) \,|\, \theta_0) \leq \alpha$ 를 뜻하므로 ϕ' 는 가설 (9.3) 에 대해서도 수준 α 검정법임을 의미한다. ϕ는 가설 (9.3) 에 대한 최강력검정법이므로 다음 부등식이 성립한다.

$$E[\phi(\boldsymbol{X})|\theta_1] \ \geq \ E[\phi'(\boldsymbol{X})|\theta_1]$$

위의 부등식은 θ_1 이 θ_0 보다 크기만 하면 항상 성립하므로 부등식 (9.2) 가 성립함을 보일 수 있다.

균일최강력검정법이 항상 존재하는 것은 아니다. 예를 들어 $X_1, X_2, \ldots, X_n$ 이 $\mathrm{N}(\theta,\, 1)$ 로부터

의 랜덤샘플일 때 다음 가설에 대한 균일최강력검정법을 구할 수 있는지 알아보도록 하자.

$$H_0: \theta = \theta_0 \quad \text{대} \quad H_1: \theta \neq \theta_0$$

대립가설이 참일 때 모수 θ의 값은 θ_0보다 작을 수도 있고 클 수도 있다. 문제는 θ의 값의 범위에 따라 검정력을 가장 크게 하는 검정법의 형태가 달라진다는 것이다. 먼저 $\theta_1 > \theta_0$일 때 다음 가설에 대한

$$H_0: \theta = \theta_0 \quad \text{대} \quad H_1: \theta = \theta_1$$

최강력검정법은 다음과 같다.

$$\phi(x) = \begin{cases} 1, & \sqrt{n}(\overline{X} - \theta_0) \geq Z_\alpha \text{ 일 때} \\ 0, & \sqrt{n}(\overline{X} - \theta_0) < Z_\alpha \text{ 일 때} \end{cases} \tag{9.4}$$

한편 $\theta_1 < \theta_0$일 때 다음 가설에 대한

$$H_0: \theta = \theta_0 \quad \text{대} \quad H_1: \theta = \theta_1$$

최강력검정법은 다음과 같다.

$$\phi(x) = \begin{cases} 1, & \sqrt{n}(\overline{X} - \theta_0) \leq -Z_\alpha \text{ 일 때} \\ 0, & \sqrt{n}(\overline{X} - \theta_0) > -Z_\alpha \text{ 일 때} \end{cases} \tag{9.5}$$

즉 $\theta > \theta_0$일 때 검정력을 가장 크게 하기 위해서는 검정법의 형태가 (9.4) 이어야 하며 $\theta < \theta_0$일 때 검정력을 가장 크게 하기 위해서는 검정법의 형태가 (9.5) 이어야 한다. 따라서 대립가설에 속하는 모든 θ에 대하여 검정력을 가장 크게 하는 검정법은 존재하지 않는다.

9.4 가능도비 검정

귀무가설과 대립가설이 복합가설인 경우에도 균일최강력검정법을 구할 수 있다면 이를 이용하여 검정을 할 수 있지만 앞에서 설명한 것처럼 균일최강력검정법이 항상 존재하는 것은 아니다. 일반적인 복합가설에 대한 검정법으로 주로 이용되는 것이 **가능도비 검정**(likelihood ratio test) 이다. 가능도비 검정은 제1종 오류를 범할 확률이 일정한 수준 이하인 검정법 중에서 검정력을 가장 크게 하는 검정법을 구하는 기준에 바탕을 둔 검정법은 아니지만 일반적인 가설에 대한 검정법으로 가장 많이 이용되고 있다.

귀무가설과 대립가설이 다음과 같이 단순가설일 때

$$H_0: \theta = \theta_0 \quad \text{대} \quad H_1: \theta = \theta_1$$

최강력검정법의 기각역은 다음과 같이 나타낼 수 있다.

$$\left\{ \boldsymbol{x} : \ \frac{f(\boldsymbol{x}|\theta_0)}{f(\boldsymbol{x}|\theta_1)} \ \le \ k \right\}$$

즉 귀무가설 하에서의 확률밀도함수(확률질량함수)와 대립가설 하에서의 확률밀도함수가 하나로 결정될 때, 기각역은 두 확률밀도함수의 비율에 의하여 결정된다. 그런데 다음과 같은 복합가설에 대해서는 귀무가설과 대립가설 하에서 확률밀도함수가 하나로 결정되지 않는다.

$$H_0 : \ \theta \in \Theta_0 \quad \text{대} \quad H_1 : \ \theta \in \Theta_1$$

하나로 결정되지 않는 확률밀도함수 대신에 각 가설 하에서의 최대가능도를 이용하여 기각역을 결정하는 검정법이 가능도비 검정이다.

가능도비 검정의 기각역은 다음과 같이 정해지며

$$\left\{ \boldsymbol{x} : \ \frac{\max_{\theta \in \Theta_0} f(\boldsymbol{x} \,|\, \theta)}{\max_{\theta \in \Theta_1} f(\boldsymbol{x} \,|\, \theta)} \ \le \ k \right\} \tag{9.6}$$

상수 k는 유의수준에 따라 결정된다.

일반적으로 가능도비 검정의 기각역 (9.6) 에서 상수 k는 1보다 작다. k가 1보다 크다면 귀무가설 하에서 최대가능도가 대립가설 하에서 최대가능도보다 큰 경우에도 귀무가설을 기각할 수 있으므로 이러한 검정법은 큰 의미가 없다고 할 수 있다. 상수 k가 1보다 작을 때 가능도비 검정의 기각역은 다음과 같이 나타낼 수 있다.

$$\left\{ \boldsymbol{x} : \ \Lambda \ = \ \frac{\max_{\theta \in \Theta_0} f(\boldsymbol{x} \,|\, \theta)}{\max(\max_{\theta \in \Theta_0} f(\boldsymbol{x} \,|\, \theta), \max_{\theta \in \Theta_1} f(\boldsymbol{x} \,|\, \theta))} \ = \ \frac{\max_{\theta \in \Theta_0} f(\boldsymbol{x} \,|\, \theta)}{\max_{\theta \in \Theta_0 \cup \Theta_1} f(\boldsymbol{x} \,|\, \theta)} \ \le \ k \right\}$$

위 식에서 분자는 귀무가설 하에서 최대가능도이고 분모는 귀무가설과 대립가설을 합한 영역에서 최대가능도이다. 일반적으로 $\Theta_0 \cup \Theta_1 = \Omega$이고, 이 경우에 분모는 모수의 전체 범위에서 구한 최대가능도이다. 귀무가설이 참인 영역에서 구한 θ의 최대가능도추정량을 $\widehat{\theta}_0$라 하고 전체 범위에서 구한 최대가능도추정량을 $\widehat{\theta}$라 할 때, Λ는 다음과 같이 나타낼 수 있다.

$$\Lambda \ = \ \frac{f(\boldsymbol{x} \,|\, \widehat{\theta}_0)}{f(\boldsymbol{x} \,|\, \widehat{\theta}\,)}$$

따라서 가능도비 검정의 기각역은 다음과 같다.

$$\left\{ \boldsymbol{x} : \ \Lambda = \frac{f(\boldsymbol{x} \,|\, \widehat{\theta}_0)}{f(\boldsymbol{x} \,|\, \widehat{\theta}\,)} \ \le \ k \right\}$$

예 9.6 $X_1, X_2, \ldots, X_n$ 이 $\mathrm{N}(\theta, 1)$ 로부터의 랜덤샘플일 때 다음 가설에 대한 가능도비 검정법의

기각역을 구하여라.

$$H_0 : = \theta_0 \quad \text{대} \quad H_1 : \ \theta \neq \theta_0$$

[풀이] 귀무가설이 참일 때 θ의 값은 θ_0로 정해지므로 $\widehat{\theta}_0 = \theta_0$이다. 전체 범위에서 구한 θ의 최대가능도추정량은 $\widehat{\theta} = \overline{X}$이다. 이때 최대가능도의 비율은 다음과 같다.

$$\begin{aligned}
\Lambda = \frac{f(\boldsymbol{x} \,|\, \widehat{\theta}_0)}{f(\boldsymbol{x} \,|\, \widehat{\theta})} &= \frac{(2\pi)^{-\frac{n}{2}} \exp\Big(-\sum_{i=1}^{n} \frac{(x_i-\theta_0)^2}{2}\Big)}{(2\pi)^{-\frac{n}{2}} \exp\Big(-\sum_{i=1}^{n} \frac{(x_i-\overline{x})^2}{2}\Big)} \\
&= \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i-\overline{x})^2 - \frac{n}{2}(\overline{x}-\theta_0)^2 + \frac{1}{2}\sum_{i=1}^{n}(x_i-\overline{x})^2\right) \\
&= \exp\left(-\frac{n}{2}(\overline{x}-\theta_0)^2\right)
\end{aligned}$$

따라서 Λ는 $n(\overline{x}-\theta_0)^2$의 감소함수이고 가능도비 검정의 기각역의 형태는 다음과 같다.

$$n\,(\overline{X}-\theta_0)^2 \ \geq \ k$$

귀무가설이 참일 때 $n(\overline{X}-\theta_0)^2 \sim \chi_1^2$이므로 유의수준 α에서 가능도비 검정법의 기각역은 다음과 같다.

$$n\,(\overline{X}-\theta_0)^2 \ \geq \ \chi_1^2(\alpha)$$

표준정규분포와 카이제곱분포의 관계로부터 위 기각역을 다음과 같이 나타낼 수도 있다.

$$\sqrt{n}\,|\overline{X}-\theta_0| \ \geq \ Z_{\frac{\alpha}{2}}$$

■

지금까지는 모형이 가설검정과 관련이 있는 모수만 포함하는 경우를 다루었다. 다음 예는 가설검정과 관련이 없는 모수를 포함하는 경우를 다루고 있다.

예 9.7 $X_1, X_2, \ldots, X_n$이 $\mathrm{N}(\mu, \sigma^2)$으로부터의 랜덤샘플일 때 다음 가설에 대한 가능도비 검정법의 기각역을 구하여라.

$$H_0 : \ \mu = \mu_0 \quad \text{대} \quad H_1 : \ \mu \neq \mu_0$$

[풀이] 이 예에서 모수 $\theta = (\mu, \sigma^2)$ 는 가설검정과 관련이 있는 모수 μ 와 함께 관련이 없는 모수 σ^2 을 포함하고 있다. 귀무가설이 참일 때 모수의 영역을 나타내는 Θ_0 와 대립가설이 참일 때 모수의 영역을 나타내는 Θ_1 은 다음과 같이 나타낼 수 있다.

$$\Theta_0 \;=\; \{(\mu,\,\sigma^2):\; \mu = \mu_0,\; \sigma^2 > 0\}\,,\quad \Theta_1 \;=\; \{(\mu,\,\sigma^2):\; \mu \neq \mu_0,\; \sigma^2 > 0\}$$

귀무가설이 참일 때 가능도 $L(\mu, \sigma^2)$ 은 다음과 같다.

$$\begin{aligned} L(\mu,\sigma^2) &= L(\mu_0,\sigma^2) \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu_0)^2\right) \end{aligned}$$

이 가능도는 $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_0)^2$ 일 때 최대이므로 $\theta = (\mu, \sigma^2)$ 의 최대가능도추정량 $\widehat{\mu}_0$ 와 $\widehat{\sigma}_0^2$ 은 다음과 같다.

$$\widehat{\mu}_0 = \mu_0\,,\quad \widehat{\sigma}_0^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu_0)^2$$

한편 전체 범위에서 구한 최대가능도추정량 $\widehat{\mu}$ 와 $\widehat{\sigma}^2$ 은 다음과 같으므로

$$\widehat{\mu} = \overline{X}\,,\quad \widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$$

최대가능도의 비율은 다음과 같다.

$$\begin{aligned} \Lambda &= \frac{f(\boldsymbol{x}\,|\,\widehat{\mu}_0,\,\widehat{\sigma}_0^2)}{f(\boldsymbol{x}\,|\,\widehat{\mu},\,\widehat{\sigma}^2)} = \frac{(2\pi\widehat{\sigma}_0^2)^{-\frac{n}{2}}\exp\left(-\frac{1}{2\widehat{\sigma}_0^2}\sum_{i=1}^{n}(x_i-\mu_0)^2\right)}{(2\pi\widehat{\sigma}^2)^{-\frac{n}{2}}\exp\left(-\frac{1}{2\widehat{\sigma}^2}\sum_{i=1}^{n}(x_i-\overline{x}\,)^2\right)} \\ &= \frac{(2\pi\widehat{\sigma}_0^2)^{-\frac{n}{2}}\exp(-\frac{n}{2})}{(2\pi\widehat{\sigma}^2)^{-\frac{n}{2}}\exp(-\frac{n}{2})} \;=\; \left(\frac{\widehat{\sigma}_0^2}{\widehat{\sigma}^2}\right)^{-\frac{n}{2}} \\ &= \left(\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu_0)^2}{\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x}\,)^2}\right)^{-\frac{n}{2}} \;=\; \left(\frac{\sum_{i=1}^{n}(x_i-\overline{x}\,)^2 + n(\overline{x}-\mu_0)^2}{\sum_{i=1}^{n}(x_i-\overline{x}\,)^2}\right)^{-\frac{n}{2}} \\ &= \left(1+\frac{n(\overline{x}-\mu_0)^2}{\sum_{i=1}^{n}(x_i-\overline{x}\,)^2}\right)^{-\frac{n}{2}} \end{aligned}$$

따라서 Λ는 $\dfrac{n(\overline{x}-\mu_0)^2}{\sum_{i=1}^n (x_i-\overline{x})^2}$ 의 감소함수이고 가능도비 검정의 기각역의 형태는 다음과 같다.

$$\frac{n(\overline{X}-\mu_0)^2}{\sum_{i=1}^n (X_i-\overline{X})^2/(n-1)} = \frac{n(\overline{X}-\mu_0)^2}{S^2} \ \geq \ k$$

귀무가설이 참일 때 $\dfrac{n(\overline{X}-\mu_0)^2}{S^2} \sim F_{1,n-1}$ 이므로 유의수준 α에서 가능도비 검정의 기각역은 다음과 같다.

$$\frac{n(\overline{X}-\mu_0)^2}{S^2} \ \geq \ F_{1,\,n-1}(\alpha)$$

t-분포와 F-분포의 관계로부터 위 기각역을 다음과 같이 나타낼 수도 있다.

$$\frac{\sqrt{n}\,|\overline{X}-\mu_0|}{S} \ \geq \ t_{n-1}\left(\frac{\alpha}{2}\right)$$

■

가능도비 검정의 기각역은 $\{\boldsymbol{x} : \Lambda(\boldsymbol{x}) \leq k\}$로 주어진다. 유의수준 α에서 가능도비 검정의 정확한 기각역을 구하기 위해서는 귀무가설이 참일 때 $\Lambda(\boldsymbol{X})$의 확률분포를 알아야 한다. 가능도비 검정이 가지는 장점 중의 하나는 귀무가설이 참일 때 $\Lambda(\boldsymbol{X})$의 점근분포가 알려져 있다는 것이다. 따라서 귀무가설 하에서 $\Lambda(\boldsymbol{X})$의 정확한 확률분포를 알지 못하더라도 $\Lambda(\boldsymbol{X})$의 점근분포를 이용하여 가능도비 검정의 근사적인 기각역을 구할 수 있다.

정리 9.2 적절한 조건 하에서 $-2\log\Lambda(\boldsymbol{X})$는 귀무가설이 참일 때 점근적으로 카이제곱분포를 따른다. 즉 자료의 수가 충분히 크면 근사적으로 다음이 성립한다.

$$-2\log\Lambda(\boldsymbol{X}) \ \sim \ \chi^2_d$$

단, 카이제곱분포의 자유도 d는 $(\Theta_0\cup\Theta_1)$에서 모수의 차원과 Θ_0에서 모수의 차원의 차이이다.

위 정리에서 자유도는 추정하는 전체 모수의 수와 귀무가설이 참일 때 추정하는 모수의 수의 차이라 할 수도 있다. 정리 9.2에 의하여 유의수준 α에서 가능도비 검정의 점근적인 기각역은 다음과 같다.

$$\{\,\boldsymbol{x} :\ -2\log\Lambda(\boldsymbol{x}) \ \geq \ \chi^2_d(\alpha)\,\}$$

여기서 위의 정리를 증명하지는 않겠지만 예를 통해서 앞의 정리가 성립하는지 살펴보도록

하자. $X_1, X_2, \ldots, X_n$ 이 $\mathrm{N}(\theta, 1)$ 로부터의 랜덤샘플일 때 다음 가설에 대한

$$H_0 : \ \theta = \theta_0 \quad \text{대} \quad H_1 : \ \theta \neq \theta_0$$

예 9.6의 가능도비 검정으로부터 $-2\log\Lambda$ 는 다음과 같다.

$$-2\log\Lambda \ = \ -2\log\left(\exp\left[-\frac{n}{2}(\overline{X}-\theta_0)^2\right]\right) \ = \ n(\overline{X}-\theta_0)^2$$

귀무가설이 참일 때 $\sqrt{n}(\overline{X}-\theta_0) \ \sim \ \mathrm{N}(0,1)$ 이므로 $-2\log\Lambda \ \sim \ \chi_1^2$ 임을 쉽게 확인 할 수 있다. 여기서 $(\Theta_0 \cup \Theta_1)$ 에서 모수의 차원은 1이고 Θ_0 는 한 점 $\{\theta_0\}$ 이므로 차원은 0이다. 따라서 차원의 차이는 1이 된다. 이 예는 $-2\log\Lambda(\boldsymbol{X})$ 의 정확한 확률분포가 카이제곱분포인 경우이다.

예 9.8 $X_1, X_2, \ldots, X_n$ 이 $\mathrm{N}(\mu, \sigma^2)$ 으로부터의 랜덤샘플일 때 다음 가설에 대한 가능도비 검정에서 $-2\log\Lambda(\boldsymbol{X})$ 는 점근적으로 자유도 1인 카이제곱분포임을 보여라.

$$H_0 : \mu = \mu_0 \quad \text{대} \quad H_1 : \ \mu \neq \mu_0$$

[풀이] 예 9.7에서 구한 Λ 로부터 $-2\log\Lambda$ 는 다음과 같다.

$$\begin{aligned} -2\log\Lambda \ &= -2\log\left(1+\frac{n(\overline{X}-\mu_0)^2}{\sum_{i=1}^n (X_i-\overline{X})^2}\right)^{-\frac{n}{2}} \\ &= n\log\left(1+\frac{n(\overline{X}-\mu_0)^2}{\sum_{i=1}^n (X_i-\overline{X})^2}\right) \end{aligned}$$

테일러 전개를 이용하면 $|x| \approx 0$ 일 때 $\log(1+x) \approx x$ 이므로 다음이 성립한다.

$$-2\log(\Lambda) \ \approx \ \frac{n(\overline{X}-\mu_0)^2}{\dfrac{1}{n}\displaystyle\sum_{i=1}^n (X_i-\overline{X})^2} \tag{9.7}$$

한편 다음 결과로부터

$$\begin{aligned} E\left(\frac{1}{n}\sum_{i=1}^n (X_i-\overline{X})^2\right) \quad &= \quad \frac{(n-1)\sigma^2}{n} \\ Var\left(\frac{1}{n}\sum_{i=1}^n (X_i-\overline{X})^2\right) \quad &= \quad \frac{2(n-1)(\sigma^2)^2}{n^2} \end{aligned}$$

σ^2 의 추정량 $\frac{1}{n}\sum_{i=1}^n (X_i-\overline{X})^2$ 의 편향과 분산 모두 0으로 수렴함을 알 수 있다. 따라서

다음이 성립한다.

$$\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2 \ \xrightarrow{p}\ \sigma^2$$

이를 식 (9.7) 에 대입하면 다음과 같으므로

$$-2\log\Lambda \ \approx\ \frac{n(\overline{X}-\mu_0)^2}{\sigma^2} \sim \chi_1^2$$

이 예에서도 정리 9.2가 성립함을 보일 수 있다. 여기서 $(\Theta_0\cup\Theta_1)$ 에서 모수의 차원은 2이고 Θ_0 에서 모수의 차원은 1이다. 따라서 차원의 차이는 1이 된다. ■

분할표로 주어지는 자료에 대한 카이제곱검정도 가능도비 검정의 한 예이다. 분할표의 가장 간단한 형태인 m 개의 범주에 속하는 빈도수 자료에 대한 적합도검정 (goodness-of-fit test)을 살펴보도록 하자. 각 범주에 속하는 빈도수 자료는 다항분포를 따른다고 볼 수 있다. $\boldsymbol{X}=(X_1,X_2,\ldots,X_m)$ 이 시행횟수는 n 이고 각 범주에 속할 확률이 $(p_1,p_2,\ldots,p_m)$ 인 다항분포를 따를 때 $\boldsymbol{X}$ 의 확률질량함수는 다음과 같다.

$$f(x_1,x_2,\ldots,x_m\,|\,p_1,p_2,\ldots,p_m) \ =\ \frac{n!}{x_1!x_2!\cdots x_m!}\,p_1^{x_1}p_2^{x_2}\cdots p_m^{x_m}$$

이때 각 범주에 속할 확률에 대한 다음의 가설에 대하여 가능도비 검정을 구하여 보자.

$$H_0:\ (p_1,p_2,\ldots,p_m)\in\Theta_0 \quad \text{대} \quad H_1:(p_1,p_2,\ldots,p_m)\in\Theta_1$$

Θ_0 영역에서 p_i 의 최대가능도추정량을 $\widehat{p}_{i0}$ 라 하고 $\Theta_0\cup\Theta_1$ 영역에서 p_i 의 최대가능도추정량을 $\widehat{p}_i$ 라 하자. $\Theta_0\cup\Theta_1$ 이 모수의 전체 범위일 때 $\widehat{p}_i$ 는 각 범주의 관측 비율이다.

$$\widehat{p}_i=\frac{x_i}{n}$$

이때 Λ 와 $-2\log\Lambda$ 는 다음과 같다.

$$\begin{aligned}
\Lambda \ &=\ \frac{\max\limits_{(p_1,p_2,\ldots,p_m)\in\Theta_0} f(x_1,x_2,\ldots,x_m\,|\,p_1,p_2,\ldots,p_m)}{\max\limits_{(p_1,p_2,\ldots,p_m)\in\Theta_0\cup\Theta_1} f(x_1,x_2,\ldots,x_m\,|\,p_1,p_2,\ldots,p_m)}\\
&= \frac{\dfrac{n!}{x_1!x_2!\cdots x_m!}\,\widehat{p}_{10}^{x_1}\widehat{p}_{20}^{x_2}\cdots\widehat{p}_{m0}^{x_m}}{\dfrac{n!}{x_1!x_2!\cdots x_m!}\,\widehat{p}_{1}^{x_1}\widehat{p}_{2}^{x_2}\cdots\widehat{p}_{m}^{x_m}}\\
&= \left(\frac{\widehat{p}_{10}}{\widehat{p}_1}\right)^{x_1}\left(\frac{\widehat{p}_{20}}{\widehat{p}_2}\right)^{x_2}\cdots\left(\frac{\widehat{p}_{m0}}{\widehat{p}_m}\right)^{x_m},
\end{aligned}$$

$$\begin{aligned} -2\log\Lambda &= -2\sum_{i=1}^{m} x_i \log\left(\frac{\widehat{p}_{i0}}{\widehat{p}_i}\right) \;=\; -2\sum_{i=1}^{m} x_i \log\left(\frac{n\widehat{p}_{i0}}{x_i}\right) \\ &= 2\sum_{i=1}^{m} x_i \log\left(\frac{x_i}{n\widehat{p}_{i0}}\right) \end{aligned}$$

각 범주에서 관측도수 X_i를 O_i라 하고 귀무가설 하에서 기대빈도수인 $n\widehat{p}_{i0}$를 E_i라 할 때 $-2\log\Lambda$는 다음과 같이 나타낼 수 있다.

$$-2\log\Lambda \;=\; 2\sum_{i=1}^{m} O_i \log\left(\frac{O_i}{E_i}\right) \tag{9.8}$$

위 식을 다르게 표현할 수도 있다. 함수 $f(x) = x\log\left(\frac{x}{x_0}\right)$를 $x = x_0$에서 테일러 전개하면 다음과 같다.

$$\begin{aligned} f(x) &\approx f(x_0) + (x-x_0)f'(x_0) + \frac{1}{2}(x-x_0)^2 f''(x_0) \\ &= 0 + (x-x_0)\cdot 1 + \frac{1}{2}(x-x_0)^2 \cdot \frac{1}{x_0} \end{aligned}$$

따라서 다음이 성립한다.

$$O_i \log\left(\frac{O_i}{E_i}\right) \;\approx\; (O_i - E_i) + \frac{1}{2}\frac{(O_i-E_i)^2}{E_i}$$

한편 $\sum_{i=1}^{m} O_i = \sum_{i=1}^{m} E_i = n$이므로 식 (9.8)은 다음과 같이 나타낼 수 있다.

$$\begin{aligned} -2\log\Lambda &= 2\sum_{i=1}^{m} O_i \log\left(\frac{O_i}{E_i}\right) \\ &\approx 2\sum_{i=1}^{m}(O_i - E_i) + \sum_{i=1}^{m}\frac{(O_i-E_i)^2}{E_i} \\ &= \sum_{i=1}^{n}\frac{(O_i-E_i)^2}{E_i} \end{aligned} \tag{9.9}$$

식 (9.8)의 값은 분할표로 주어지는 자료에 대한 적합도검정에 이용되는 G^2 통계량이며, 식 (9.9)의 값은 X^2 통계량이다.

독립인 두 모집단을 비교하는데 자주 이용되는 것이 t 검정(t-test)이다. t 검정도 가능도비 검정에 의하여 유도될 수 있다. $\boldsymbol{X} = (X_1, X_2, \ldots, X_m)$은 $\mathrm{N}(\mu_X, \sigma^2)$로부터의 랜덤샘플이고 $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)$은 $\mathrm{N}(\mu_Y, \sigma^2)$로부터의 랜덤샘플이며 서로 독립이라 하자. 두 모집단의 평균에 대한 다음 가설에 대한 가능도비 검정을 구하여 보자.

$$H_0 : \mu_X = \mu_Y \quad \text{대} \quad H_1 : \; \mu_X \neq \mu_Y$$

이때 모수는 $\theta = (\mu_X, \mu_Y, \sigma^2)$ 이고 $(\boldsymbol{X}, \boldsymbol{Y})$ 의 확률밀도함수는 다음과 같다.

$$
\begin{aligned}
f(\boldsymbol{x}, \boldsymbol{y} \,|\, \mu_X, \mu_Y, \sigma^2) &= f(\boldsymbol{x} \,|\, \mu_X, \sigma^2) f(\boldsymbol{y} \,|\, \mu_Y, \sigma^2) \\
&= (2\pi\sigma^2)^{-\frac{m+n}{2}} \exp\left[-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{m}(x_i - \mu_X)^2 + \sum_{j=1}^{n}(y_j - \mu_Y)^2\right)\right]
\end{aligned}
$$

가능도 $L(\mu_X,\, \mu_Y,\, \sigma^2) = f(\boldsymbol{x}, \boldsymbol{y} \,|\, \mu_X, \mu_Y, \sigma^2)$ 를 이용하여 전체 영역에서 구한 $(\mu_X, \mu_Y,\, \sigma^2)$ 의 최대가능도추정량은 다음과 같다.

$$
\begin{aligned}
\widehat{\mu}_X &= \overline{X}\,, \quad \widehat{\mu}_Y = \overline{Y} \\
\widehat{\sigma}^2 &= \frac{1}{m+n}\left(\sum_{i=1}^{m}(X_i - \widehat{\mu}_X)^2 + \sum_{j=1}^{n}(Y_j - \widehat{\mu}_Y)^2\right)
\end{aligned}
$$

귀무가설이 참일 때 두 모집단의 평균을 $\mu_X = \mu_Y = \mu$ 로 나타내면 가능도 $L(\mu,\, \sigma^2)$ 은 다음과 같다.

$$
L(\mu,\, \sigma^2) = (2\pi\sigma^2)^{-\frac{m+n}{2}} \exp\left[-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{m}(x_i - \mu)^2 + \sum_{j=1}^{n}(y_j - \mu)^2\right)\right]
$$

이를 이용하여 $\theta = (\mu, \sigma^2)$ 의 최대가능도추정량 $\widehat{\mu}_0$ 와 $\widehat{\sigma}_0^2$ 을 구하면 다음과 같다.

$$
\begin{aligned}
\widehat{\mu}_0 &= \frac{1}{m+n}\left(\sum_{i=1}^{m} X_i + \sum_{j=1}^{n} Y_j\right) = \frac{m\overline{X} + n\overline{Y}}{m+n} \\
\widehat{\sigma}_0^2 &= \frac{1}{m+n}\left(\sum_{i=1}^{m}(X_i - \widehat{\mu}_0)^2 + \sum_{j=1}^{n}(Y_j - \widehat{\mu}_0)^2\right)
\end{aligned}
$$

따라서 가능도비 검정의 Λ 에서 분자는 다음과 같으며

$$
\begin{aligned}
&\max_{\theta \in \Theta_0} f(\boldsymbol{x}, \boldsymbol{y} \,|\, \theta) \\
&\quad = \left(2\pi\widehat{\sigma}_0^2\right)^{-\frac{m+n}{2}} \exp\left[-\frac{1}{2\widehat{\sigma}_0^2}\left(\sum_{i=1}^{m}(x_i - \widehat{\mu}_0)^2 + \sum_{j=1}^{n}(y_j - \widehat{\mu}_0)^2\right)\right] \\
&\quad = \left(2\pi\widehat{\sigma}_0^2\right)^{-\frac{m+n}{2}} \exp\left(-\frac{m+n}{2}\right)
\end{aligned}
$$

분모는 다음과 같다.

$$\begin{aligned}
&\max_{\theta\in\,\Theta_0\cup\Theta_1} f(\boldsymbol{x},\boldsymbol{y}\,|\,\theta)\\
&\quad=\ \left(2\pi\widehat{\sigma}^2\right)^{-\frac{m+n}{2}}\exp\left[-\frac{1}{2\widehat{\sigma}^2}\left(\sum_{i=1}^{m}(x_i-\widehat{\mu}_X)^2+\sum_{j=1}^{n}(y_j-\widehat{\mu}_Y)^2\right)\right]\\
&\quad=\ \left(2\pi\widehat{\sigma}^2\right)^{-\frac{m+n}{2}}\exp\left(-\frac{m+n}{2}\right)
\end{aligned}$$

이때 Λ는 다음과 같으므로

$$\begin{aligned}
\Lambda\ &=\ \frac{\max_{\theta\in\,\Theta_0} f(\boldsymbol{x},\boldsymbol{y}\,|\,\theta)}{\max_{\theta\in\,\Theta_0\cup\Theta_1} f(\boldsymbol{x},\boldsymbol{y}\,|\,\theta)}\ =\ \left(\frac{\widehat{\sigma}_0^2}{\widehat{\sigma}^2}\right)^{-\frac{m+n}{2}}\\
&=\ \left(\frac{\sum_{i=1}^{m}(x_i-\widehat{\mu}_0)^2+\sum_{j=1}^{n}(y_j-\widehat{\mu}_0)^2}{\sum_{i=1}^{m}(x_i-\widehat{\mu}_X)^2+\sum_{j=1}^{n}(y_j-\widehat{\mu}_Y)^2}\right)^{-\frac{m+n}{2}}\\
&=\ \left(\frac{\sum_{i=1}^{m}(x_i-\overline{x}\,)^2+\sum_{j=1}^{n}(y_j-\overline{y}\,)^2+m(\overline{x}-\widehat{\mu}_0)^2+n(\overline{y}-\widehat{\mu}_0)^2}{\sum_{i=1}^{m}(x_i-\overline{x})^2+\sum_{j=1}^{n}(y_j-\overline{y})^2}\right)^{-\frac{m+n}{2}}\\
&=\ \left(1+\frac{mn^2(\overline{x}-\overline{y})^2/(m+n)^2+m^2n(\overline{x}-\overline{y})^2/(m+n)^2}{\sum_{i=1}^{m}(x_i-\overline{x})^2+\sum_{j=1}^{n}(y_j-\overline{y})^2}\right)^{-\frac{m+n}{2}}\\
&=\ \left(1+\frac{\frac{mn}{m+n}(\overline{x}-\overline{y})^2}{\sum_{i=1}^{m}(x_i-\overline{x})^2+\sum_{j=1}^{n}(y_j-\overline{y})^2}\right)^{-\frac{m+n}{2}}
\end{aligned}$$

기각역의 형태는 다음과 같다.

$$\frac{\frac{mn}{m+n}(\overline{X}-\overline{Y})^2}{S_p^2}\ \geq\ k$$

위 식에서 S_p^2은 다음과 같다.

$$S_p^2\ =\ \frac{1}{m+n-2}\left(\sum_{i=1}^{m}(X_i-\overline{X})^2+\sum_{j=1}^{n}(Y_j-\overline{Y})^2\right)$$

귀무가설이 참일 때 다음이 성립하므로

$$\frac{\frac{mn}{m+n}(\overline{X}-\overline{Y})^2}{S_p^2}\ =\ \frac{\left(\frac{(\overline{X}-\overline{Y})}{\sqrt{(1/m+1/n)\sigma^2}}\right)^2}{S_p^2/\sigma^2}\ =\ \frac{\mathrm{N}(0,1)^2}{\chi^2_{m+n-2}/(m+n-2)}\ \sim\ F_{1,\,m+n-2}$$

유의수준 α 에서 가능도비 검정법의 기각역은 다음과 같다.

$$\frac{\frac{mn}{m+n}(\overline{X} - \overline{Y})^2}{S_p^2} \;\geq\; F_{1,\, m+n-2}(\alpha)$$

t-분포와 F-분포의 관계로부터 위 기각역을 다음과 같이 나타낼 수도 있다.

$$\frac{\sqrt{\frac{mn}{m+n}}\left|\overline{X} - \overline{Y}\right|}{S_p} \;\geq\; t_{m+n-2}\left(\frac{\alpha}{2}\right)$$

연습문제

1. 이산형 확률변수 X는 다음과 같은 확률분포를 가지고 있다.

	0	1	2	3	4	5
$f(x\|\theta_0)$	0.15	0.10	0.05	0.05	0.25	0.40
$f(x\|\theta_1)$	0.10	0.15	0.30	0.25	0.15	0.05

X를 이용하여 $H_0:\ \theta = \theta_0$ 대 $H_1:\ \theta = \theta_1$을 검정하려 한다.

a) 유의수준 $\alpha = 0.1$에서 최강력검정법의 기각역을 구하시오.
b) 위 검정법의 검정력을 구하여라.
c) 유의수준 $\alpha = 0.15$에서 최강력검정법의 기각역을 구하시오.

2. 이산형 확률변수 X는 다음과 같은 확률분포를 가지고 있다.

	-3	-2	-1	0	1	2	3
$f(x\|\theta_0)$	0.3	0.15	0.05	0.05	0.05	0.1	0.3
$f(x\|\theta_1)$	0.1	0.1	0.25	0.3	0.15	0.05	0.05
$f(x\|\theta_2)$	0.05	0.05	0.15	0.1	0.35	0.15	0.15

X를 이용하여 $H_0:\ \theta = \theta_0$ 대 $H_1:\ \theta = \theta_1$를 검정하려 한다.

a) 기각역이 $R = \{-1, 0, 1\}$인 검정법의 제1종 오류를 범할 확률과 제2종 오류를 범할 확률을 구하여라.
b) 기각역이 $R_1 = \{-2, 2\}$인 검정법과 기각역이 $R_2 = \{0, 2\}$인 검정법 중에서 어느 검정법이 더 좋은가? 그 이유는?
c) 유의수준 $\alpha = 0.1$에서 최강력검정법의 기각역을 구하시오.
d) $H_0:\ \theta = \theta_0$ 대 $H_1:\ \theta \in \{\theta_1, \theta_2\}$에 대하여 유의수준 $\alpha = 0.1$에서 가능도비 검정법을 구하시오.

3. $X_1, X_2, \ldots, X_n$이 확률밀도함수 $f(x|\theta) = \theta^2 x e^{-\theta x},\ x > 0$인 Gamma$(2, \theta)$로부터의 랜덤샘플이다. $H_0:\ \theta = \theta_0$ 대 $H_1:\ \theta = \theta_1$ (단, $\theta_1 > \theta_0$)에 대한 유의수준 α에서 최강력검정법의 기각역을 카이제곱분포를 이용하여 정확히 나타내어라.

4. $X_1, X_2, \ldots, X_n$이 N$(\theta, 1)$로부터의 랜덤샘플일 때 $H_0:\ \theta = \theta_0$ 대 $H_1:\ \theta = \theta_1$ (단, $\theta_1 > \theta_0$)에 대한 유의수준 α에서 최강력검정법의 검정력이 $1 - \beta$가 되기 위한 표본의 크기 n을 구하여라.

5. $X \sim \mathrm{N}(\theta, 1^2)$, $Y \sim \mathrm{N}(\theta, 2^2)$ 이며 서로 독립이다. X 와 Y 를 이용하여 다음 가설에 대한 유의수준 α 에서 최강력검정법을 구하여라.

$$H_0:\ \theta = 0 \quad \text{대} \quad H_1:\ \theta = \theta_1 \quad (\text{단}, \theta_1 > 0)$$

6. $X \sim \mathrm{Uniform}(0, \theta)$ 일 때 $H_0:\ \theta = \theta_0$ 대 $H_1:\ \theta = \theta_1$ (단, $\theta_1 > \theta_0$) 에 대한 유의수준 α 에서 최강력검정법을 구하여라.

7. $X_1, X_2, \ldots, X_n$ 이 Bernoulli(θ) 로부터의 랜덤샘플일 때 다음 가설에 대한

$$H_0:\ \theta \le \theta_0 \quad \text{대} \quad H_1:\ \theta > \theta_0 \qquad (\text{단}, \theta_1 > \theta_0)$$

유의수준 α 에서 균일최강력검정법의 기각역이 다음과 같음을 보여라.

$$\sum_{i=1}^{n} X_i \ \ge \ k$$

단, k 는 $P(\, \mathrm{B}(n, \theta_0) \ge k\,) = \alpha$ 를 만족하는 상수이다.

8. $X_1, X_2, \ldots, X_n$ 이 Poisson(θ) 로부터의 랜덤샘플일 때 다음 가설에 대한

$$H_0:\ \theta \le \theta_0 \quad \text{대} \quad H_1:\ \theta > \theta_0 \qquad (\text{단}, \theta_1 > \theta_0)$$

유의수준 α 에서 균일최강력검정법의 기각역이 다음과 같음을 보여라.

$$\sum_{i=1}^{n} X_i \ \ge \ k$$

단, k 는 $P(\, \mathrm{Poisson}(n\theta_0) \ge k\,) = \alpha$ 를 만족하는 상수이다.

9. $X_1, X_2, \ldots, X_n$ 이 Exp(θ) 로부터의 랜덤샘플일 때 다음 가설에 대한

$$H_0:\ \theta \le \theta_0 \quad \text{대} \quad H_1:\ \theta > \theta_0 \qquad (\text{단}, \theta_1 > \theta_0)$$

유의수준 α 에서 균일최강력검정법의 기각역이 다음과 같음을 보여라.

$$2\theta_0 \sum_{i=1}^{n} X_i \ \le \ \chi^2_{2n}(1-\alpha)$$

10. $X_1, X_2, \ldots, X_n$ 이 $\mathrm{N}(0, \sigma^2)$ 로부터의 랜덤샘플이다. $H_0:\ \sigma^2 = \sigma_0^2$ 대 $H_1:\ \sigma^2 = \sigma_1^2$ (단, $\sigma_1^2 > \sigma_0^2$) 을 검정하려 한다.

   a) 유의수준 α 에서 최강력검정법을 구하여라.
   b) (a) 에서 구한 최강력검정법의 검정력을 구하여라.

c) (a)에서 구한 최강력검정법이 $H_0: \ \sigma^2 \le \sigma_0^2$ 대 $H_1: \ \sigma^2 > \sigma_0^2$에 대한 균일최강력검정법임을 보여라.

d) $X_1, X_2, \ldots, X_n$이 $\mathrm{N}(1, \sigma^2)$로부터의 랜덤샘플일 때 최강력검정법을 구하여라.

11. $X_1, X_2, \ldots, X_m$은 $\mathrm{N}(\mu_X, 1)$로부터의 랜덤샘플이고 $Y_1, Y_2, \ldots, Y_n$은 $\mathrm{N}(\mu_Y, 1)$로부터의 랜덤샘플이며 서로 독립이다. $H_0: \ \mu_X = \mu_Y$ 대 $H_1: \ \mu_X \neq \mu_Y$에 대한 유의수준 α에서 가능도비 검정법의 기각역이 다음과 같음을 보여라.

$$|\overline{X} - \overline{Y}| \ > Z_{\frac{\alpha}{2}} \sqrt{\frac{1}{m} + \frac{1}{n}}$$

12. $X_k \sim \mathrm{N}(k\theta, 1), \ k = 1, 2, \ldots, n$이며 서로 독립일 때 다음 가설을 검정하고자 한다.

$$H_0: \ \theta = 0 \quad \text{대} \quad H_1: \ \theta \neq 0$$

a) 가능도비 검정을 위한 Λ를 구하여라.

b) 유의수준 α에서 가능도비 검정법의 기각역을 구하여라.

13. $X_1, X_2, \ldots, X_n$이 아래와 같은 확률밀도함수를 가지는 지수족으로부터의 랜덤샘플일 때 다음 물음에 답하여라.

$$f(x|\theta) \ = \ \exp\left(\theta T(x) + d(\theta) + S(x)\right) I_A(x)$$

a) $H_0: \ \theta = \theta_0$ 대 $H_1: \ \theta = \theta_1$ (단, $\theta_1 > \theta_0$)에 대하여 다음과 같은 검정함수를 가지는 검정법이

$$\phi(\boldsymbol{X}) \ = \ \begin{cases} 1, & \sum_{i=1}^n T(X_i) \ge k \text{ 일 때} \\ 0, & \sum_{i=1}^n T(X_i) < k \text{ 일 때} \end{cases}$$

유의수준 α에서 최강력검정법임을 보여라. 단, k는 $P(\sum_{i=1}^n T(X_i) \ge k \,|\, \theta_0) = \alpha$를 만족하는 상수이다.

b) $P(\sum_{i=1}^n T(X_i) \ge k \,|\, \theta)$는 θ의 증가함수임을 보여라.

c) (a)에서 주어진 검정법이 $H_0: \ \theta \le \theta_0$ 대 $H_1: \ \theta > \theta_0$에 대한 균일최강력검정법임을 보여라.

14. $X_i \sim \mathrm{B}(n_i, p_i), \ i = 1, 2, \ldots, m$이며 서로 독립이다. 다음 가설에 대한 유의수준 α에서 가능도비 검정법을 구하여라.

$$H_0: \ p_1 = p_2 = \cdots = p_n \quad \text{대} \quad H_1: \ \text{not } H_0$$

15. 다음을 일원배치 모형이라 한다. 단, $\epsilon_{ij} \sim \mathrm{N}(0, \sigma^2)$ 이며 서로 독립이다.

$$X_{ij} \;=\; \mu_i + \epsilon_{ij} \quad i = 1, 2, \ldots m,\; j = 1, 2, \ldots, n$$

이때 다음 가설 에 대한 가능도비 검정법을 구하여라.

$$H_0 :\; \mu_1 = \mu_2 = \cdots = \mu_m \quad \text{대} \quad H_1 :\; \text{not } H_0$$

16. $X_1, X_2, \ldots, X_n$ 이 확률밀도함수(확률질량함수) $f(x|\theta)$ 를 가지는 랜덤샘플일 때, 다음 가설에 대한 최강력검정법은 θ 의 충분통계량에 의하여 결정됨을 설명하여라.

$$H_0 :\; \theta = \theta_0 \quad \text{대} \quad H_1 :\; \theta = \theta_1$$

제 10 장

통계적 결정이론과 베이즈 추론

10.1 개요

겨울의류를 생산하는 회사에서 얼마만큼의 제품을 생산해야 하는지 결정을 하려 한다. 겨울의류의 판매량은 일반적으로 온도에 따라 영향을 받으며 날씨가 추울수록 많은 제품을 판매할 수 있다. 문제는 겨울이 되기 전에 미리 제품을 생산해야 하기 때문에 합리적인 판단에 의하여 제품 생산량을 결정하여야 한다. 이러한 문제를 해결하기 위해 간단한 설정 하에서 생각하여 보자.

춥지 않은 겨울인 경우 겨울의류 판매량은 10벌, 추운 겨울인 경우 판매량은 20벌이라 하자. 이 회사는 제품 생산량을 10벌로 할 것인지 아니면 20벌로 할 것인지 결정하려 한다. 겨울의 날씨를 미리 알 수 있다면 거기에 맞추어 제품을 생산하면 되지만 제품을 생산해야 할 시점에서 다가올 겨울의 날씨를 알 수는 없다. 이 회사에서 결정할 수 있는 것은 제품의 생산량에 관한 것이다. 겨울의 날씨를 두 가지로 나누어 춥지 않은 겨울은 θ_1로 나타내고 추운 겨울은 θ_2로 나타내도록 하자. 또 제품 생산량을 10벌로 하는 행동을 a_1으로 나타내고 제품 생산량을 20벌로 하는 행동을 a_2로 나타내도록 하자.

이 회사에서는 이익을 최대화하는 의사결정을 내리려 할 것이다. 통계학에서는 일반적으로 이익보다는 손실의 관점에서 살펴보게 된다. 따라서 이 회사의 목표는 손실을 최소화하는 의사결정을 내리려 한다고 볼 수 있다. 그런데 손실은 회사의 의지로 결정할 수 있는 생산량과 회사에서 제어할 수 없는 날씨에 따라 달라진다. 한편 제품 하나의 생산원가는 2만원이고 판매가는 3만원이라 하자. 또한 날씨가 추울 경우는 생산량에 상관없이 2만원의 난방 비용이 들어간다고 하자. 예를 들어 날씨는 춥지 않고(θ_1) 생산량은 20벌(a_2)인 경우에 제품의 총 생산 비용은 40만원이고 총 판매액은 30만원이므로 10만원의 손실을 입게 된다. 한편 날씨는 춥고(θ_2) 생산량은 10벌(a_1)인 경우에 제품의 총 생산 비용은 20만원이고 총 판매액은 30만원이므로 10만원의 이익을 얻는다. 2만원의 난방 비용을 고려하면 총 이익은 8만원이고 손실의 관점에서 본다면 손실은 -8만원이 된다. 날씨와 생산량에 따른 이 회사의 손실은 다음 표와 같이 나타낼 수 있다.

	a_1	a_2
θ_1	-10	10
θ_2	-8	-18

이 회사에서 행동을 a_1으로 할 것인지 아니면 a_2로 할 것인지 결정을 하려 한다. 가능한 손실을 최소화하도록 결정을 내려야 하지만 날씨를 미리 알지 못하기 때문에 어떤 손실을 입을지 정확히 알 수는 없다. 예를 들어 a_2의 행동을 취했을 때 날씨가 춥지 않다면 손실은 10만원이지만 날씨가 춥다면 손실은 -18만원이다. 이러한 상황 하에서 최적의 의사결정을 위해 사용되는 일반적인 방법은 다음의 두 가지이다.

- 손실의 최댓값을 최소화하는 행동을 취한다.
- 손실의 기댓값을 최소화하는 행동을 취한다.

첫 번째 방법을 사용하여 최적의 의사결정 방법을 구하여 보자. a_1의 행동을 취했을 때 날씨의 변화에 따른 최대 손실은 -8만원이고 a_2의 행동을 취했을 때 최대 손실은 10만원이므로 최대 손실을 작게 하는 행동은 a_1이다. 따라서 첫 번째 방법을 기준으로 의사결정을 한다면 생산량을 10벌로 하는 것이 타당하다고 할 수 있다. 두 번째 방법을 사용하기 위해서는 날씨의 변화에 따른 손실의 기댓값을 계산하여야 한다. 이를 위해서는 날씨가 추울 확률을 알아야 한다는 전제 조건이 있다. 장기예보나 다른 정보를 통하여 날씨가 추울 확률이 0.7이라는 것을 알고 있다고 하자. 각 행동을 취했을 때 손실의 기댓값은 다음과 같으므로

$$\begin{aligned} a_1\text{의 행동을 취했을 때 손실의 기댓값} &: 0.3 \times (-10) + 0.7 \times (-8) = -8.6 \\ a_2\text{의 행동을 취했을 때 손실의 기댓값} &: 0.3 \times 10 + 0.7 \times (-18) = -9.6 \end{aligned}$$

손실의 기댓값을 작게 하는 행동은 a_2이다. 따라서 두 번째 방법을 기준으로 의사결정을 한다면 생산량을 20벌로 하는 것이 타당하다고 할 수 있다.

생산하는 제품의 수에 대한 단순한 행동 대신에 다음과 같은 의사결정 방법 d_p를 사용한다고 하자.

$$d_p = \begin{cases} a_1, & p\text{의 확률로} \\ a_2, & 1-p\text{의 확률로} \end{cases}$$

즉 d_p는 p의 확률로 a_1의 행동을 취하고 $1-p$의 확률로 a_2의 행동을 취하게 된다. 앞에서 설명한 두 가지 방법을 기준으로 p의 값이 얼마일 때 최적의 의사결정을 내릴 수 있는지 알아보자. 의사결정 방법 d_p을 사용했을 때 손실은 다음과 같다.

$$\begin{aligned} \text{날씨가 춥지 않을 때}(\theta_1)\text{ 손실} &: (-10) \times p + 10 \times (1-p) = 10 - 20p \\ \text{날씨가 추울 때}(\theta_2)\text{ 손실} &: (-8) \times p + (-18) \times (1-p) = -18 + 10p \end{aligned}$$

날씨의 변화에 따른 손실의 최댓값은 $\max(10-20p, -18+10p)$이다. 이 값은 $p = \frac{14}{15}$일 때 최소가 된다. 따라서 손실의 최댓값을 최소화하려면 $\frac{14}{15}$의 확률로 생산량을 10벌로 하고 $\frac{1}{15}$의 확률로 생산량을 20벌로 하여야 한다. 한편 θ_1일 때 손실은 $10-20p$이고, θ_2일 때 손실은 $-18+10p$이므로 날씨의 변화에 따른 손실의 기댓값은 $(10-20p) \times 0.3 + (-18+10p) \times 0.7 = p - 9.6$이다. 따라서 손실의 기댓값을 최소로 하는 p는 0이다. 따라서 손실의 기댓값을 최소화하려면 생산량을 20벌로 하여야 한다.

10.2 통계적 결정이론

통계적 추론의 주요 목표는 자료를 이용하여 모집단을 규명하는데 있다. 자료 $\boldsymbol{x}$를 이용하여 모집단에 대하여 추론하는 규칙(rule)을 $d(\boldsymbol{x})$라 하자. 예를 들어 점추정에서 $d(\boldsymbol{x})$는 모수 θ를 한 값으로 추측하는 추정량이라 할 수 있으며 가설검정에서 $d(\boldsymbol{x})$는 가설을 기각할 것인지 아닐지를 나타내는 것으로 볼 수 있다. 모집단에 대한 추론에서 발생하는 손실은 자료로부터 추론한 결과 $d(\boldsymbol{x})$와 모수 θ에 따라 달라질 것이다. 자료 $\boldsymbol{x}$를 이용한 추론의 결과가 $d(\boldsymbol{x})$이고 모수가 θ일 때 발생하는 손실을 $l(\theta, d(\boldsymbol{x}))$로 나타내며 이를 **손실함수**(loss function)라 한다. 일반적으로 모수에 대한 추론의 성격에 따라 손실함수는 달라질 수 있다. 모수에 대한 점추정에서는 추정값 $d(\boldsymbol{x})$이 θ로부터 멀리 떨어질수록 손실이 크다고 할 수 있다. 따라서 점추정에서 주로 사용하는 손실함수는 다음과 같다.

$$l(\theta,\, d(\boldsymbol{x})) \;=\; (d(\boldsymbol{x}) - \theta)^2$$

이를 제곱오차손실(squared error loss)이라 한다. 한편 다음과 같은 가설검정에 대해서는

$$H_0:\ \theta \in \Theta_0 \quad \text{대} \quad H_1:\ \theta \in \Theta_1$$

추론의 결과가 맞거나 틀리거나 둘 중의 하나이므로 다음과 같은 손실함수를 사용할 수 있다.

	$d(\boldsymbol{x})$에 의하여 H_0를 기각하지 않을 때	$d(\boldsymbol{x})$에 의하여 H_0를 기각할 때
$\theta \in \Theta_0$	0	1
$\theta \in \Theta_1$	1	0

이를 $0-1$ 손실(zero-one loss)이라 한다.

확률변수 $\boldsymbol{X}$에 대한 손실함수의 기댓값을 **위험함수**(risk function)라 하며 $R(\theta, d)$로 나타낸다. 즉 위험함수는 다음과 같다.

$$R(\theta,\, d) \;=\; E[l(\theta,\, d(\boldsymbol{X}))]$$

확률밀도함수가 $f(\boldsymbol{x}\,|\,\theta)$ 인 연속형 확률변수인 경우 위험함수는 다음과 같이 나타낼 수 있다.

$$R(\theta,\, d) \;=\; \int_{-\infty}^{\infty} l(\theta,\, d(\boldsymbol{x}))\, f(\boldsymbol{x}\,|\,\theta)\; d\boldsymbol{x}$$

예를 들어 점추정에서 손실함수가 제곱오차손실일 때 위험함수는 다음과 같으므로

$$R(\theta,\, d) \;=\; E(d(\boldsymbol{X}) - \theta)^2$$

위험함수는 추정량 d 의 평균제곱오차이다. 또 가설검정에서 $0-1$ 손실에 대한 위험함수는 다음과 같다.

$$R(\theta,\, d) \;=\; \begin{cases} P(H_0 \text{ 를 기각}|\theta) = \alpha, & \theta \in \Theta_0 \text{ 일 때} \\ P(H_0 \text{ 를 기각하지 않음}|\theta) = \beta, & \theta \in \Theta_1 \text{ 일 때} \end{cases}$$

앞에서 설명한 것처럼 최적의 추론 규칙 $d(\boldsymbol{X})$ 를 구하는 방법으로는 크게 두 가지가 있다. 첫 번째 방법은 θ 가 변함에 따라 달라지는 위험함수의 최댓값을 가장 작게 하는 추론 규칙을 구하는 것이다. 두 번째 방법은 θ 에 대한 확률분포를 가정하고 위험함수의 기댓값을 최소로 하는 추론 규칙을 구하는 것이다.

정의 10.1 위험함수의 최댓값 $\max_\theta R(\theta,\, d)$ 를 최소로 하는 추론 규칙 $d^M(\boldsymbol{X})$ 를 **미니맥스 규칙**(minimax rule)이라 한다. 즉 미니맥스 규칙은 다음과 같이 정의할 수 있다.

$$d^M = \underset{d}{\operatorname{argmin}}(\max_{\theta} R(\theta,\, d))$$

점추정에서는 미니맥스 규칙을 미니맥스 추정량이라 한다.

최적의 추론 규칙 $d(\boldsymbol{X})$ 를 구하는 두 번째 방법을 사용하기 위해서는 모수 θ 에 대한 확률분포를 가정하여야 한다. 이러한 가정은 모수는 알지는 못하지만 고정된 상수라는 지금까지의 관점에서 벗어난 것이라 할 수 있다. 모수를 고정된 값이 아닌 확률분포를 가지는 랜덤으로 간주하는 통계학을 **베이즈 통계학**(**Bayesian Statistics**)이라 한다.

모수 θ 에 대한 확률분포를 가정했을 때 위험함수에 대한 기댓값을 구할 수 있고 이 값을 **베이즈 위험**(Bayes risk)이라 하고 $B(d)$ 로 나타낸다. 즉 베이즈 위험은 다음과 같다.

$$\begin{aligned} B(d) &= E_\Theta[R(\Theta, d)] \\ &= E_\Theta[E_{\boldsymbol{X}|\Theta}(l(\Theta, d(\boldsymbol{X})))] \end{aligned}$$

위에서 Θ 는 랜덤인 모수를 나타낸다. 또 $E_\Theta(\cdot)$ 는 Θ 에 대한 기댓값을 나타내고 $E_{\boldsymbol{X}|\Theta}(\cdot)$ 는 모수의 값이 주어졌을 때 $\boldsymbol{X}$ 에 대한 조건부기댓값을 나타내다. 베이즈 통계학에서는 $\boldsymbol{X}$ 의 확률밀도함수 $f(\boldsymbol{x}|\theta)$ 를 모수의 값이 θ 로 주어졌을 때 $\boldsymbol{X}$ 의 조건부분포에 대한 확률밀도함수로 해석한다. Θ 의

확률밀도함수가 $\pi(\theta)$ 일 때 베이즈 위험은 다음과 같이 나타낼 수 있다.

$$
\begin{aligned}
B(d) &= \int_{-\infty}^{\infty} R(\theta, d)\pi(\theta)\ d\theta \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} l(\theta, d(\boldsymbol{x}))f(\boldsymbol{x}|\theta)\pi(\theta)\ d\boldsymbol{x}\, d\theta
\end{aligned}
$$

정의 10.2 베이즈 위험 $B(d)$ 를 최소로 하는 추론 규칙 $d^B(\boldsymbol{X})$ 를 **베이즈 규칙**(Bayes rule)이라 한다. 즉 베이즈 규칙은 다음과 같이 정의할 수 있다.

$$
d^B = \operatorname*{argmin}_d B(d)
$$

점추정에서는 베이즈 규칙을 베이즈 추정량이라 한다.

예 10.1 모수는 $\theta_1, \theta_2, \theta_3$ 의 세 가지 값이 가능하고 세 가지 행동 a_1, a_2, a_3 에 따른 손실함수가 다음과 같다고 한다.

	a_1	a_2	a_3
θ_1	2	0	4
θ_2	1	4	0
θ_3	5	6	2

위의 세 가지 행동 중에서 어느 행동을 할 것인지는 관측 값 X 를 이용하여 결정하며 관측 값은 x_1 과 x_2 의 두 가지 값만 가능하다고 하자. 여기서는 아래의 세 가지 규칙 d_1, d_2, d_3 를 고려하여 보자.

	d_1	d_2	d_3
x_1	a_1	a_2	a_2
x_2	a_1	a_3	a_1

예를 들어 규칙 d_2 는 관측 값이 x_1 이면 a_2 의 행동을 하고 x_2 이면 a_3 의 행동을 하는 것이다. 한편 각 모수의 값에서 X 의 확률분포는 다음과 같다고 한다.

	x_1	x_2
θ_1	0.4	0.6
θ_2	0.5	0.5
θ_3	0.3	0.7

(a) 각 규칙 d_1, d_2, d_3 에 대한 위험함수를 구하여라.

(b) d_1, d_2, d_3 중에서 위험함수의 최댓값을 가장 작게 하는 규칙을 구하여라.

(c) $\pi(\theta_1) = \pi(\theta_2) = \pi(\theta_3) = \frac{1}{3}$ 라 할 때 d_1, d_2, d_3 중에서 베이즈 위험을 가장 작게 하는 규칙을 구하여라.

[풀이]

(a) 모수가 θ_1 일 때 각 규칙에 대한 위험함수는 다음과 같다.

$$\begin{aligned}
R(\theta_1,\, d_1) &= E(l(\theta_1,\, d_1(X))) = l(\theta_1, a_1)f(x_1|\theta_1) + l(\theta_1, a_1)f(x_2|\theta_1) \\
&= 2 \times 0.4 + 2 \times 0.6 = 2 \\
R(\theta_1,\, d_2) &= E(l(\theta_1,\, d_2(X))) = l(\theta_1, a_2)f(x_1|\theta_1) + l(\theta_1, a_3)f(x_2|\theta_1) \\
&= 0 \times 0.4 + 4 \times 0.6 = 2.4 \\
R(\theta_1,\, d_3) &= E(l(\theta_1,\, d_3(X))) = l(\theta_1, a_2)f(x_1|\theta_1) + l(\theta_1, a_1)f(x_2|\theta_1) \\
&= 0 \times 0.4 + 2 \times 0.6 = 1.2
\end{aligned}$$

같은 방법으로 모수가 θ_2 와 θ_3 일 때 각 규칙에 대한 위험함수를 구하면 다음과 같다.

$$R(\theta_2,\, d_1) = 1,\; R(\theta_2,\, d_2) = 2,\; R(\theta_2,\, d_3) = 2.5$$
$$R(\theta_3,\, d_1) = 5,\; R(\theta_3,\, d_2) = 3.2,\; R(\theta_3,\, d_3) = 5.3$$

(b) (a)의 결과를 이용하여 각 규칙에 대한 최대 위험함수를 구하면 다음과 같다.

$$\begin{aligned}
\max_\theta R(\theta, d_1) &= \max(R(\theta_1, d_1), R(\theta_2, d_1), R(\theta_3, d_1)) = 5 \\
\max_\theta R(\theta, d_2) &= \max(R(\theta_1, d_2), R(\theta_2, d_2), R(\theta_3, d_2)) = 3.2 \\
\max_\theta R(\theta, d_3) &= \max(R(\theta_1, d_3), R(\theta_2, d_3), R(\theta_3, d_3)) = 5.3
\end{aligned}$$

따라서 d_1, d_2, d_3 중에서 위험함수의 최댓값을 가장 작게 하는 규칙은 d_2 이다.

(c) $\pi(\theta_1) = \pi(\theta_2) = \pi(\theta_3) = \frac{1}{3}$ 일 때 베이즈 위험은 다음과 같다.

$$\begin{aligned} B(d_1) &= R(\theta_1, d_1)\pi(\theta_1) + R(\theta_2, d_1)\pi(\theta_2) + R(\theta_3, d_1)\pi(\theta_3) = \frac{8}{3} \\ B(d_2) &= R(\theta_1, d_2)\pi(\theta_1) + R(\theta_2, d_2)\pi(\theta_2) + R(\theta_3, d_2)\pi(\theta_3) = \frac{7.6}{3} \\ B(d_3) &= R(\theta_1, d_3)\pi(\theta_1) + R(\theta_2, d_3)\pi(\theta_2) + R(\theta_3, d_3)\pi(\theta_3) = 3 \end{aligned}$$

따라서 d_1, d_2, d_3 중에서 베이즈 위험을 가장 작게 하는 규칙은 d_2 이다.

■

10.3 베이즈 추정량

이 절에서는 모수를 랜덤으로 간주하는 베이즈 분석방법을 이용하여 모수를 추정하는 방법을 알아보도록 하자. 베이즈 추정량은 베이즈 위험을 최소로 하는 추정량이며, 베이즈 위험을 구하기 위해서는 모수에 대한 확률분포가 주어져야 한다. 모수에 대한 확률분포를 자료를 관측하기 이전의 모수에 대한 정보를 나타낸다는 의미에서 **사전분포**(prior distribution)라 한다. 모수의 사전분포에 대한 확률밀도함수 또는 확률질량함수를 $\pi(\theta)$ 라 하자. 앞에서 설명한 것처럼 베이즈 분석방법에서는 $\boldsymbol{X}$ 의 확률밀도함수 $f(\boldsymbol{x}|\theta)$ 를 모수의 값이 θ 로 주어졌을 때 $\boldsymbol{X}$ 의 조건부분포에 대한 확률밀도함수로 해석한다.

모수를 나타내는 확률변수 Θ 의 사전분포와 $\boldsymbol{X}$ 의 확률분포를 이용하여 $(\boldsymbol{X}, \Theta)$ 의 결합확률밀도함수를 다음과 같이 구할 수 있다.

$$f(\boldsymbol{x},\, \theta) = f(\boldsymbol{x}|\theta)\pi(\theta)$$

$\boldsymbol{X}$ 의 주변확률밀도함수를 $f_{\boldsymbol{X}}(\boldsymbol{x})$ 라하고 $\boldsymbol{X} = \boldsymbol{x}$ 로 주어졌을 때 Θ 의 조건부분포의 확률밀도함수를 $h(\theta|\boldsymbol{x})$ 라 하자. 이때 $f_{\boldsymbol{X}}(\boldsymbol{x})$ 와 $h(\theta|\boldsymbol{x})$ 는 다음과 같이 구할 수 있다.

$$\begin{aligned} f_{\boldsymbol{X}}(\boldsymbol{x}) &= \int_{-\infty}^{\infty} f(\boldsymbol{x}, \theta)\, d\theta = \int_{-\infty}^{\infty} f(\boldsymbol{x}|\theta)\pi(\theta)\, d\theta \\ h(\theta|\boldsymbol{x}) &= \frac{f(\boldsymbol{x}, \theta)}{f_{\boldsymbol{X}}(\boldsymbol{x})} = \frac{f(\boldsymbol{x}|\theta)\pi(\theta)}{\int_{-\infty}^{\infty} f(\boldsymbol{x}|\theta)\pi(\theta)\, d\theta} \end{aligned}$$

$\boldsymbol{X} = \boldsymbol{x}$ 로 주어졌을 때 Θ 의 조건부분포를 자료를 관측한 후의 모수에 대한 정보를 나타낸다는 의미에서 **사후분포**(posterior distribution)라 한다.

추정량 $d(\boldsymbol{X})$ 의 베이즈 위험은 다음과 같이 나타낼 수 있다.

$$\begin{aligned} B(d) &= \int_{-\infty}^{\infty} R(\theta, d)\pi(\theta)\, d\theta \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} l(\theta, d(\boldsymbol{x}))f(\boldsymbol{x}|\theta)\pi(\theta)\, d\boldsymbol{x}d\theta \\ &= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} l(\theta, d(\boldsymbol{x}))h(\theta|\boldsymbol{x})\, d\theta\right) f_{\boldsymbol{X}}(\boldsymbol{x})\, d\boldsymbol{x} \end{aligned}$$

손실함수가 제곱오차손실인 경우 베이즈 위험은 다음과 같다.

$$B(d) = \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}(d(\boldsymbol{x})-\theta)^2 h(\theta|\boldsymbol{x})\, d\theta\right) f_{\boldsymbol{X}}(\boldsymbol{x})\, d\boldsymbol{x} \tag{10.1}$$

베이즈 위험을 최소로 하기 위해서는 각 고정된 $\boldsymbol{x}$ 에서 다음 값을 최소로 하면 된다.

$$\int_{-\infty}^{\infty}(d(\boldsymbol{x})-\theta)^2 h(\theta|\boldsymbol{x})\, d\theta \tag{10.2}$$

일반적으로 a 가 상수일 때 $E(X-a)^2$ 는 $a = E(X)$ 일 때 최소가 되므로 식 (10.2) 를 최소로 하는 $d(\boldsymbol{x})$ 는 $\int_{-\infty}^{\infty}\theta h(\theta|\boldsymbol{x})\, d\theta$ 이다. 따라서 베이즈 위험 (10.1) 을 최소로 하는 베이즈 추정량 $\widehat{\theta}^B$ 는 다음과 같다.

$$\widehat{\theta}^B = E(\Theta|\boldsymbol{x}) = \int_{-\infty}^{\infty}\theta\, h(\theta|\boldsymbol{x})\, d\theta$$

즉 제곱오차손실에 대한 베이즈 추정량은 Θ 의 사후분포의 기댓값이다.

예 10.2 모수가 θ 일 때 X 의 확률분포는 Bernoulli(θ) 이며 모수 Θ 의 사전분포는 Uniform(0, 1) 이라 한다. 이때 제곱오차손실에 대한 베이즈 추정량을 구하여라.

[풀이] $X\,|\,\theta \sim \text{Bernoulli}(\theta)$ 이고 $\Theta \sim \text{Uniform}(0,1)$ 이므로 $f(x|\theta)$ 와 $\pi(\theta)$ 는 다음과 같다.

$$\begin{aligned} f(x|\theta) &= \theta^x(1-\theta)^{1-x}, \quad x = 0 \text{ 또는 } 1 \\ \pi(\theta) &= 1, \quad 0 < \theta < 1 \end{aligned}$$

따라서 X 의 주변분포에 대한 확률질량함수는 다음과 같고

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x|\theta)\pi(\theta)\, d\theta = \int_0^1 \theta^x(1-\theta)^{1-x}\, d\theta \\ &= \begin{cases} \int_0^1 \theta\, d\theta = \frac{1}{2}, & x = 1 \text{ 일 때} \\ \int_0^1 (1-\theta)\, d\theta = \frac{1}{2}, & x = 0 \text{ 일 때} \end{cases} \end{aligned}$$

Θ의 사후분포에 대한 확률밀도함수는 다음과 같다.

$$h(\theta|x) \;=\; \frac{f(x|\theta)\pi(\theta)}{f_X(x)} \;=\; \frac{\theta^x(1-\theta)^{1-x}}{1/2} \;=\; 2\,\theta^x(1-\theta)^{1-x}$$

제곱오차손실에 대한 베이즈 추정량은 사후분포의 기댓값이므로 다음과 같다.

$$\begin{aligned}\widehat{\theta}^B &= \int_0^1 \theta\, 2\theta^x(1-\theta)^{1-x}\, d\theta \\ &= \begin{cases} \int_0^1 2\theta^2\, d\theta \;=\; \frac{2}{3}, & x=1\text{일 때} \\ \int_0^1 2\theta(1-\theta)\, d\theta \;=\; \frac{1}{3}, & x=0\text{일 때}\end{cases}\end{aligned}$$

예를 들어 동전을 던져 앞면이 나올 확률을 θ라 하고 X를 동전을 한 번 던져 나온 앞면의 수라 할 때 주어진 사전분포에 대한 θ의 베이즈 추정량은 앞면이 나오면 $\frac{2}{3}$이고, 뒷면이 나오면 $\frac{1}{3}$이다. ■

베이즈 추정량을 구할 때 핵심은 Θ의 사후분포를 구하는 것이다. Θ의 사후분포에 대한 확률밀도함수 $h(\theta|\boldsymbol{x})$는 다음과 같이 구할 수 있다.

$$h(\theta|\boldsymbol{x}) \;=\; \frac{f(\boldsymbol{x}|\theta)\pi(\theta)}{f_{\boldsymbol{X}}(\boldsymbol{x})} \;=\; \frac{f(\boldsymbol{x}|\theta)\pi(\theta)}{\int_{-\infty}^{\infty} f(\boldsymbol{x}|\theta)\pi(\theta)\, d\theta} \tag{10.3}$$

사후분포를 구하는 위 식에서 분모는 $\boldsymbol{x}$만의 함수로서 θ와 관련이 없다. 즉 Θ의 사후분포를 구할 때 식 (10.3)의 분모는 상수 역할만을 하므로 사후분포는 분자에 의하여 결정된다고 할 수 있다. 즉 사후분포의 확률밀도함수는 적적한 상수 c에 대하여 다음이 성립한다.

$$h(\theta|\boldsymbol{x}) \;=\; c\, f(\boldsymbol{x}|\theta)\pi(\theta)$$

혹은 사후분포가 $f(\boldsymbol{x}|\theta)$와 $\pi(\theta)$의 곱에 비례한다는 의미에서 다음과 같이 나타내기도 한다.

$$h(\theta|\boldsymbol{x}) \;\propto\; f(\boldsymbol{x}|\theta)\pi(\theta)$$

예 10.2에서 $\pi(\theta)=1,\ 0<\theta<1$이고 $f(x|\theta)=\theta^x(1-\theta)^{1-x}$이므로, Θ의 사후분포는 다음 식에 비례한다.

$$h(\theta|x) \;\propto\; f(x|\theta)\pi(\theta) \;=\; \theta^x(1-\theta)^{1-x}$$

여기서 주의할 사항은 θ가 변수 역할을 하며 x는 상수라는 것이다. 이러한 형태의 확률분포는 베타분포이므로 Θ의 사후분포는 $\text{Beta}(x+1, 2-x)$이다.

예 10.3 모수가 θ일 때 X의 확률분포는 $\text{B}(n,\,\theta)$이며 모수 Θ의 사전분포는 $\text{Uniform}(0,1)$일 때

Θ의 사후분포를 구하여라.

[풀이] Θ의 사후분포는 다음 식에 비례한다.

$$h(\theta|x) \quad \propto \; f(x|\theta)\pi(\theta) \; = \; \binom{n}{x}\theta^x(1-\theta)^{n-x}$$

위에서 x는 상수이므로 사후분포의 확률밀도함수는 적절한 상수 c에 대하여 다음이 성립한다.

$$h(\theta|x) \; = \; c\,\theta^x(1-\theta)^{n-x}$$

이러한 형태의 확률분포는 베타분포이므로 Θ의 사후분포는 $\mathrm{Beta}(x+1, n-x+1)$ 이다. ■

예 10.4 모수가 θ일 때 X의 확률분포는 $\mathrm{B}(n,\,\theta)$ 이며 모수 Θ의 사전분포는 $\mathrm{Beta}(\alpha, \beta)$라 한다.

(a) Θ의 사후분포를 구하여라.

(b) 제곱오차손실에 대한 베이즈 추정량을 구하여라.

[풀이]

(a) Θ의 사후분포는 다음 식에 비례한다.

$$\begin{aligned} h(\theta|x) \quad &\propto \; f(x|\theta)\pi(\theta) \; = \; \binom{n}{x}\theta^x(1-\theta)^{n-x}\,\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} \\ &\propto \; \theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1} \end{aligned}$$

따라서 Θ의 사후분포는 $\mathrm{Beta}(x+\alpha, n-x+\beta)$ 이다.

(b) 제곱오차손실을 사용할 때 베이즈 추정량은 사후분포의 기댓값이다. $\mathrm{Beta}(\alpha, \beta)$의 기댓값은 $\dfrac{\alpha}{\alpha+\beta}$ 이므로 베이즈 추정량은 다음과 같다.

$$\widehat{\theta}^B \; = \; \frac{x+\alpha}{n+\alpha+\beta}$$

참고로 이 값은 다음과 같이 나타낼 수 있다.

$$\widehat{\theta}^B \; = \; \frac{n}{n+\alpha+\beta}\left(\frac{x}{n}\right) + \frac{\alpha+\beta}{n+\alpha+\beta}\left(\frac{\alpha}{\alpha+\beta}\right)$$

즉 베이즈 추정량은 $\frac{x}{n}$ 와 $\dfrac{\alpha}{\alpha+\beta}$ 의 가중평균이라 할 수 있다. 여기서 $\frac{x}{n}$ 는 관측 값을 이용하여 구할 수 있는 θ 의 추정량이고 $\dfrac{\alpha}{\alpha+\beta}$ 는 사전분포를 이용하여 구할 수 있는 θ 의 추정량이다.

■

예 10.5 모수가 θ 일 때 X 의 확률분포는 $\mathrm{N}(\theta, \sigma^2)$ 이며 모수 Θ 의 사전분포는 $\mathrm{N}(\theta_0, \sigma_0^2)$ 라 한다. 단, $\theta_0, \sigma^2, \sigma_0^2$ 은 모두 상수이다.

(a) Θ 의 사후분포를 구하여라.

(b) 제곱오차손실에 대한 베이즈 추정량을 구하여라.

[풀이]

(a) $X\,|\,\theta \ \sim \mathrm{N}(\theta, \sigma^2)$ 이고 $\Theta \ \sim \mathrm{N}(\theta_0, \sigma_0^2)$ 이므로 $f(x|\theta)$ 와 $\pi(\theta)$ 는 다음과 같다.

$$\begin{aligned} f(x|\theta) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\theta)^2}{2\sigma^2}\right] \\ \pi(\theta) &= \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left[-\frac{(\theta-\theta_0)^2}{2\sigma_0^2}\right] \end{aligned}$$

이때 $h(\theta|x)$ 는 다음과 같으므로

$$\begin{aligned} h(\theta|x) \ &\propto f(x|\theta)\pi(\theta) \ = \ \frac{1}{2\pi\sqrt{\sigma^2\sigma_0^2}} \exp\left[-\frac{(x-\theta)^2}{2\sigma^2} - \frac{(\theta-\theta_0)^2}{2\sigma_0^2}\right] \\ &\propto \exp\left[-\frac{1}{2}\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_0^2}\right)\theta^2 + \left(\frac{x}{\sigma^2}+\frac{\theta_0}{\sigma_0^2}\right)\theta - \frac{1}{2}\left(\frac{x^2}{\sigma^2}+\frac{\theta_0^2}{\sigma_0^2}\right)\right] \\ &= \exp\left[-\frac{1}{2}\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_0^2}\right)\left(\theta - \frac{\frac{x}{\sigma^2}+\frac{\theta_0}{\sigma_0^2}}{\frac{1}{\sigma^2}+\frac{1}{\sigma_0^2}}\right)^2 + \frac{1}{2}\frac{\left(\frac{x}{\sigma^2}+\frac{\theta_0}{\sigma_0^2}\right)^2}{\frac{1}{\sigma^2}+\frac{1}{\sigma_0^2}} - \frac{1}{2}\left(\frac{x^2}{\sigma^2}+\frac{\theta_0^2}{\sigma_0^2}\right)\right] \\ &\propto \exp\left[-\frac{1}{2}\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_0^2}\right)\left(\theta - \frac{\frac{x}{\sigma^2}+\frac{\theta_0}{\sigma_0^2}}{\frac{1}{\sigma^2}+\frac{1}{\sigma_0^2}}\right)^2\right] \end{aligned}$$

Θ 의 사후분포는 다음과 같다.

$$N\left(\frac{\frac{x}{\sigma^2}+\frac{\theta_0}{\sigma_0^2}}{\frac{1}{\sigma^2}+\frac{1}{\sigma_0^2}},\ \frac{1}{\frac{1}{\sigma^2}+\frac{1}{\sigma_0^2}}\right)$$

(b) 제곱오차손실을 사용할 때 베이즈 추정량은 사후분포의 기댓값이므로 다음과 같다.

$$\widehat{\theta}^B = \frac{\frac{x}{\sigma^2} + \frac{\theta_0}{\sigma_0^2}}{\frac{1}{\sigma^2} + \frac{1}{\sigma_0^2}}$$

이 값은 관측 값을 이용한 θ의 추정량인 x와 사전분포의 기댓값인 θ_0의 가중평균이라 할 수 있다. 가중값은 두 추정값의 정확도를 결정하는 분산에 의하여 결정된다. 관측값의 분산 σ^2와 사전분포의 분산 σ_0^2의 역수를 가중값으로 한 평균이 베이즈 추정량이 된다.

■

베이즈 분석방법을 사용하기 위해서는 모수를 고정된 값이 아닌 랜덤이라는 가정을 받아들여야 하며 모수에 대한 사전분포를 알아야 한다는 가정이 필요하다. 또한 사전분포가 특별한 형태를 가져야만 사후분포를 손쉽게 구할 수 있다. 예를 들어 X가 정규분포를 따를 때 모수의 사전분포가 정규분포인 경우 사후분포 역시 정규분포를 따른다는 것을 쉽게 확인할 수 있지만 사전분포가 정규분포가 아닌 다른 확률분포라면 사후분포를 구하는 것이 쉽지 않다. 최근에는 특별한 형태의 사전분포를 가정하지 않더라도 컴퓨터를 이용하여 사후분포를 구할 수 있게 되어 통계학의 여러 분야에서 베이즈 분석방법의 활용도가 증가하고 있다.

연습문제

1. $X_1, X_2, \ldots, X_n$ 이 확률질량함수 $f(x|\theta) = \frac{\theta^x e^{-\theta}}{x!},\ x = 0, 1, 2, \ldots$ 인 포아송분포로부터의 랜덤샘플이며 Θ의 사전분포는 확률밀도함수 $\pi(\theta) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\lambda\theta}$ 인 감마분포라 한다.

   a) Θ의 사후분포를 구하여라.

   b) 제곱오차손실에 대한 Θ의 베이즈 추정량을 구하여라.

   c) 위에서 구한 베이즈 추정량은 최대가능도추정량과 사전분포의 기댓값의 가중평균이 됨을 설명하여라.

2. $X_1, X_2, \ldots, X_n$ 이 확률밀도함수 $f(x|\theta) = \theta e^{-\theta x},\ x > 0$ 인 지수분포로부터의 랜덤샘플이며 Θ의 사전분포는 $\text{Gamma}(\alpha, \lambda)$ 라 한다.

   a) Θ의 사후분포를 구하여라.

   b) 제곱오차손실에 대한 Θ의 베이즈 추정량을 구하여라.

3. $X_1, X_2, \ldots, X_n$ 이 확률밀도함수 $f(x|\theta) = \frac{\theta^3}{2} x^2 e^{-\theta x},\ x > 0$ 인 $\text{Gamma}(3, \theta)$ 로부터의 랜덤샘플이며 Θ의 사전분포의 확률밀도함수는 $\pi(\theta) = \lambda e^{-\lambda\theta},\ \theta > 0$ 이다.

   a) Θ의 사후분포를 구하여라.

   b) 제곱오차손실에 대한 Θ의 베이즈 추정량을 구하여라.

4. $H_0 : \theta = \theta_0$ 대 $H_1 : \theta = \theta_1$ 에 대한 가설검정에서 다음과 같은 세 가지의 행동, $a_1 = \{H_0$ 를 기각하지 않음$\}$, $a_2 = \{$판단을 유보$\}$, $a_3 = \{H_0$ 를 기각$\}$ 을 한다고 한다. 이때 잘못된 판단을 하면 손실은 2, 올바른 판단을 하면 손실은 0, 판단을 내리지 못하면 손실은 1이라고 한다. 관측 값 X 의 확률분포는 다음과 같다.

	x_1	x_2	x_3
$f(x\|\theta_0)$	0.6	0.3	0.1
$f(x\|\theta_1)$	0.2	0.2	0.6

이 관측 값에 근거한 네 가지 규칙 d_1, d_2, d_3, d_4 는 다음과 같다.

	x_1	x_2	x_3
d_1	a_1	a_1	a_3
d_2	a_1	a_2	a_3
d_3	a_1	a_3	a_3
d_4	a_2	a_2	a_2

Θ의 사전분포는 $\pi(\theta_0) = 0.6,\ \pi(\theta_1) = 0.4$일 때 다음 물음에 답하여라.

a) 각 Θ의 값과 세 가지 행동에 대한 손실함수를 구하여라.

b) 각 규칙의 위험함수를 구하고 d_1, d_2, d_3, d_4 중에서 위험함수의 최댓값을 가장 작게 하는 규칙을 구하여라.

c) 각 규칙의 베이즈 위험을 구하고 d_1, d_2, d_3, d_4 중에서 베이즈 위험을 가장 작게 하는 규칙을 구하여라.

5. X는 $\mathrm{N}(0, 1^2)$ 또는 $\mathrm{N}(0, 2^2)$을 따르는 확률변수이다. 두 확률분포 중에서 어느 확률분포를 따르는지 결정하려고 한다. X의 확률분포가 $\mathrm{N}(0, 1^2)$인데 $\mathrm{N}(0, 2^2)$을 따른다고 잘못 판단했을 때 발생하는 손실은 1, X의 확률분포가 $\mathrm{N}(0, 2^2)$인데 $\mathrm{N}(0, 1^2)$을 따른다고 잘못 판단했을 때 발생하는 손실은 2, 올바르게 판단했을 때 발생하는 손실은 0이라 한다. X를 이용하여 판단하는 방법으로 아래의 규칙을 사용할 때 다음 물음에 답하여라.

$$d_c(X) = \begin{cases} |X| \le c \text{일 때 } X \text{는 } \mathrm{N}(0, 1^2) \text{을 따른다고 판단} \\ |X| > c \text{일 때 } X \text{는 } \mathrm{N}(0, 2^2) \text{을 따른다고 판단} \end{cases}$$

a) $X \sim \mathrm{N}(0, 1^2)$일 때 d_c의 위험함수를 구하여라.

b) $X \sim \mathrm{N}(0, 2^2)$일 때 d_c의 위험함수를 구하여라.

c) $X \sim \mathrm{N}(0, 1^2)$일 사전확률이 0.6, $X \sim \mathrm{N}(0, 2^2)$일 사전확률이 0.4일 때 d_c의 베이즈 위험을 구하여라.

6. 앞면이 나올 확률이 θ인 동전을 이용하여 다음과 같은 게임을 한다고 하자. 먼저 동전을 한 번 던진 후 다음 번 동전의 결과를 예측하는 게임이다. 이때 예측이 맞으면 3을 얻고 예측이 틀리면 2를 잃게 된다. 예측을 하지 않으면 잃는 것도 얻는 것도 없다. 이 게임의 방법으로 다음의 세 가지 전략을 생각하여 보자. d_1: 첫 번째 동전의 결과와 같게 예측한다, d_2: 첫 번째 동전의 결과와 반대로 예측한다, d_3: 예측을 하지 않는다.

a) 세 가지 전략의 위험함수를 구하여라.

b) 세 가지 전략 중에서 위험함수의 최댓값을 가장 작게 하는 전략을 구하여라.

c) Θ의 사전분포가 $\mathrm{Uniform}(0, 1)$일 때 세 가지 전략 중에서 베이즈 위험을 가장 작게 하는 전략을 구하여라.

7. Θ는 $\theta_1, \theta_2, \theta_3, \theta_4$ 중의 하나이며 사전분포는 다음과 같다.

$$\pi(\theta_1) = 0.2,\ \pi(\theta_2) = 0.2,\ \pi(\theta_3) = 0.3,\ \pi(\theta_4) = 0.3$$

이산형 확률변수 X는 x_1, x_2, x_3 중 하나의 값을 가지며 각 θ에서 다음과 같은 확률분포를 가진다.

	θ_1	θ_2	θ_3	θ_4
x_1	0.25	0.40	0.30	0.50
x_2	0.50	0.40	0.20	0.30
x_3	0.25	0.20	0.50	0.20

a) $X = x_2$를 관측하였을 때 Θ의 사후분포를 구하여라.

b) 첫 번째 관측 값 이후에 그와 독립인 또 다른 관측 값이 x_3일 때 Θ의 사후분포를 구하여라.

8. 손실함수가 $l(\theta, d(\boldsymbol{x})) = \frac{(\theta - d(\boldsymbol{x}))^2}{\theta^2}$ 일 때 베이즈 위험을 최소로 하는 베이즈 추정량은 다음과 같음을 보여라.

$$\widehat{\theta}^B = \frac{E\left(\frac{1}{\Theta}|\boldsymbol{x}\right)}{E\left(\frac{1}{\Theta^2}|\boldsymbol{x}\right)}$$

9. $X \sim \mathrm{N}(\theta, 1)$이며 Θ의 사전분포는 $\mathrm{N}(0, \sigma_0^2)$이다. θ의 추정량 $d_c(X) = cX$의 제곱오차손실에 대한 위험함수와 베이즈 위험을 구하고 $d_c(X) = cX$ 중에서 베이즈 위험을 최소로 하는 상수 c를 구하여라.

해답

1장

1. a) $\frac{9!}{3!2!} = 30240$
   b) ${}_{10}\mathrm{P}_6 = 151200$
   c) $\binom{13}{3} = 286$
   d) $4! \cdot 4! + 4! \cdot 4! = 1152$
   e) $9! \cdot 2! - 8! \cdot 2! \cdot 2! = 564480$
   f) $13 \cdot 12 \cdot \binom{4}{3} \cdot \binom{4}{2} = 3744$
2. a) $\frac{5}{18}$
   b) $\frac{3}{10}$
3. a) $\frac{142}{225}$
   b) $\frac{36}{71}$
4. a) $\frac{117}{200}$
   b) $\frac{72}{117}$
5. $p_1p_2 + p_1(1-p_2)(1-p_1^3) + (1-p_1)p_2(1-p_2^3)$
6.
7. a) 참
   b) 참
   c) 참
   d) 거짓

2장

1. 꺼낸 공이 흰색인 경우를 B, 흰색인 경우를 W 라고 나타낼 때

   a) $S = \{WW, WB, BW, BB\}$

   b) $P(X=-2) = \frac{1}{10},\ P(X=0) = \frac{6}{10},\ P(X=2) = \frac{3}{10}$

2.

x	1	2	3	4	5	6
$p_X(x)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

3. 상자 A에서 꺼낸 검정색 공의 수를 X, 상자 B에서 꺼낸 공이 흰색인 사건을 E라 할 때,

   a) $P(E) = \frac{1}{\lambda}(1-e^{-\lambda})$

   b) $P(X=k|E) = \frac{\lambda^{k+1}e^{-\lambda}}{(k+1)!(1-e^{-\lambda})}, \quad k=0,1,2,\ldots$

4. a) 4

   b) $f_Y(y) = 4e^{-4y}, \quad y>0$

5.

y	0	1	4	9
$p_Y(y)$	0.1	0.5	0.1	0.3

6. $f_Y(y) = \begin{cases} \frac{1}{3}y^{-\frac{1}{2}}, & 0 \le y \le 1 \\ \frac{1}{6}y^{-\frac{1}{2}}, & 1 < y \le 4 \end{cases}$

7. $f_Y(y) = \frac{1}{2}e^{-\sqrt{y}}, \quad y>0$

8. a) $f_Y(y) = \frac{2}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}, \quad y>0$

   b) $f_Y(y) = \frac{3}{16}y^{\frac{1}{2}}, \quad 0<y<4$

9. a) $f_Y(y) = e^{y}, \quad y<0$

   b) $f_Y(y) = \frac{3}{\sqrt{y}}\left(y+\frac{1}{4}\right), \quad 0 \le y < \frac{1}{4}$

10. Gamma $\left(\alpha, \frac{\lambda}{c}\right)$

11. $P(X=2k) = p^{k-1}(1-p)^{k-1}[p^2+(1-p)^2],\ P(X=2k+1) = p^k(1-p)^k$

12. $P(X=k) = F(k) - F(k-1)$

13. $P(X=k) = e^{-\lambda k}(1-e^{-\lambda}), \quad k=0,1,2,\ldots$

3장

1.

		y			
		1	2	3	4
x	1	$\frac{1}{64}$	$\frac{6}{64}$	$\frac{12}{64}$	$\frac{18}{64}$
	2	0	$\frac{1}{64}$	$\frac{6}{64}$	$\frac{12}{64}$
	3	0	0	$\frac{1}{64}$	$\frac{6}{64}$
	4	0	0	0	$\frac{1}{64}$

2. a)

x	0	1	2
$p_X(x)$	$\frac{2}{9}$	$\frac{1}{3}$	$\frac{4}{9}$

b)

y	10	20	30
$p_Y(y)$	$\frac{5}{18}$	$\frac{1}{3}$	$\frac{7}{18}$

c)

x	0	1	2
$p_{X\|Y}(x\|10)$	$\frac{2}{5}$	$\frac{3}{5}$	0

3. a) $f_X(x) = 2(1-x), \quad 0 < x < 1$

   b) $f_Y(y) = 2y, \quad 0 < y < 1$

   c) 독립이 아니다.

4. a) $f_X(x) = e^{-x}, \quad x > 0$

   b) $f_Y(y) = \frac{2}{(1+2y)^2}, \quad y > 0$

   c) 독립이 아니다.

5. $Z = X + Y$ 라 할 때, $f_Z(z) = \frac{\lambda_1\lambda_2}{\lambda_1-\lambda_2}\left(e^{-\lambda_2 z} - e^{-\lambda_1 z}\right), \quad z > 0$

6. $\mathrm{B}\left(n, \frac{\lambda_1}{\lambda_1+\lambda_2}\right)$

7. a) $P(X=0) = e^{-\lambda p}$

   b) Poisson(λp)

8. a) $f_X(x) = \lambda e^{-\lambda x}, \quad x > 0$

   b) $f_{Y|X}(y|x) = \frac{\lambda}{2} e^{-\frac{\lambda}{2}(y-x)}, \quad y > x$

9. $f_Z(z) = \int_{-\infty}^{\infty} f(x, x-z)dx$

10.

11. $P(X < 0) = \sum_{n=0}^{\infty} \Phi(-n)\frac{\lambda^n e^{-\lambda}}{n!}$

12. a) $f_{S,T}(s,t) = \lambda^2 s e^{-\lambda s}, \quad s > 0,\ 0 < t < 1$
    b) $\text{Uniform}(0, 1)$
13. a) $f_{S,T}(s,t) = \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} s^{\alpha+\beta-1} t^{\alpha-1}(1-t)^{\beta-1} e^{-\lambda s}, \quad s > 0,\ 0 < t < 1$
    b) $\text{Beta}(\alpha, \beta)$
14.

4장

1. $\frac{1}{\lambda^2}(1 - e^{-\lambda} - \lambda e^{-\lambda})$
2. $\frac{\alpha(\alpha+1)(\alpha+2)}{\lambda^3}$
3. $\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+3)\Gamma(\beta+2)}{\Gamma(\alpha+\beta+5)}$
4. a) $\frac{11}{9}$
   b) $E(X|Y=10) = \frac{3}{5},\ E(X|Y=20) = \frac{4}{3},\ E(X|Y=30) = \frac{11}{7}$
   c)
   d) $Var(X|Y=10) = \frac{6}{25}$
5. a) $E(Y|X=0) = \frac{4}{3}, Var(Y|X=0) = \frac{5}{9},\ E(Y|X=1) = \frac{7}{4}, Var(Y|X=0) = \frac{11}{16}$
   b) $\frac{3}{2}$
   c) $\frac{1}{24}$
   d) $\frac{73}{120}$
6. a) $\frac{3}{\lambda}$
   b) $x + \frac{2}{\lambda}$
   c) $\frac{3}{\lambda}$
   d) $\frac{4}{\lambda^2}$
7. a) $E(X|Y) = \frac{Y}{2},\ Var(X|Y) = \frac{Y^2}{12}$
   b) 1
   c) $\frac{1}{2}$
   d) $\frac{1}{2}$
8.
9. $E(U) = \frac{1}{2\lambda},\ Var(U) = \frac{5}{12\lambda^2}$
10. $E(X) = \lambda p,\ Var(X) = \lambda p$

11. $Cov(X,Y) = \frac{1}{12}$

12. $E(Y) \approx 29,\ Var(Y) \approx 400$

13. $E(Y) \approx 4,\ Var(Y) \approx 4$

14. $\sigma_X^2\sigma_Y^2 + \mu_X^2\sigma_Y^2 + \mu_Y^2\sigma_X^2$

15. $\sigma_X^2\sigma_Y^2\sigma_Z^2 + \mu_X^2\sigma_Y^2\sigma_Z^2 + \mu_Y^2\sigma_X^2\sigma_Z^2 + \mu_Z^2\sigma_X^2\sigma_Y^2 + \mu_X^2\mu_Y^2\sigma_Z^2 + \mu_X^2\mu_Z^2\sigma_Y^2 + \mu_Y^2\mu_Z^2\sigma_X^2$

16. $e^{\lambda(\mu-1)}$

17. $E(X) = \frac{7}{4},\ Var(X) = \frac{3}{16}$

18. $M_Y(t) = \left(\frac{\lambda}{\lambda - ct}\right)^\alpha,\ Y \sim$ Gamma $\left(\alpha, \frac{\lambda}{c}\right)$

19. a) $M_{X_1}(t) = (p_0 + p_1e^t + p_2e^{2t})$
    b) $M_Y(t) = (p_0 + p_1e^t + p_2e^{2t})^2$
    c) $2(p_1 + 2p_2)^2 + 2(p_1 + 4p_2)$

5장

1. a) χ_3^2
   b) t_2
   c) $F_{2,1}$

2. a) $\beta = \alpha^2$
   b) $\gamma = \frac{1}{\alpha^2}$

3. $F_{2,4}$

4. a) $\chi_{m(n-1)}^2$
   b) χ_{m-1}^2

5. a) Gamma$(n\alpha, n\lambda)$
   b) $2n\lambda$

6. a) $f_{X_{(i)}|X_{(j)}}(x|y) = \frac{(j-1)!}{(i-1)!(j-i-1)!}\left(\frac{x}{y}\right)^{i-1}\left(1-\frac{x}{y}\right)^{j-i-1}\frac{1}{y}, \quad 0 < x < y$
   b) $R = X_{(i)} - X_{(i-2)}$ 라 할 때, $f_R(r) = n(n-1)r(1-r)^{n-2}, \quad 0 < r < 1$

7. $f_{X_{(i)},X_{(j)}}(x,y) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!}(1-e^{-\lambda x})^{i-1}(e^{-\lambda x} - e^{-\lambda y})^{j-i-i}(e^{-\lambda y})^{n-j}\lambda^2 e^{-\lambda(x+y)}$, $0 < x < y$

8. a) $P(X_{(i)} \le k) = \sum_{l=i}^n \binom{n}{l}[1-(1-p)^k]^l(1-p)^{k(n-l)}$

b) $P(X_{(i)} = k) = \sum_{l=i}^{n} \binom{n}{l} [1-(1-p)^k]^l (1-p)^{k(n-l)} - \sum_{l=i}^{n} \binom{n}{l} [1-(1-p)^{k-1}]^l$ $(1-p)^{(k-1)(n-l)}$

9.

10. a) $\mathrm{N}\left(c, \frac{\sigma^2}{\sum_{k=1}^{n} k^2}\right)$

b) $\frac{(\sum_{k=1}^{n} kX_k)^2}{\sigma^2 \sum_{k=1}^{n} k^2} \sim \chi_1^2,\ \frac{\sum_{k=1}^{n} X_k^2}{\sigma^2} \sim \chi_n^2$

c)

d) χ_{n-1}^2

6장

1. a) $\mathrm{Exp}\left(\frac{1}{\theta}\right)$

b) $\mathrm{Exp}\left(\frac{1}{\theta}\right)$

2.

3. $P(\mathrm{N}(0,1) > -1)$

4. a) $a = \pm \frac{\lambda}{\sqrt{n}},\ b = \mp\sqrt{n}$

b) $\mathrm{Gamma}(n, \lambda)$는 $\mathrm{Gamma}(1, \lambda)$의 합으로 나타낼 수 있으므로 중심극한정리에 의하여 표준정규분포에 분포수렴한다.

5. a) $M_{S^2}(t) = \left(1 - 2\frac{t\sigma^2}{n-1}\right)^{-\frac{n-1}{2}}$

b) σ^2

7장

1. $(\sum_{i=1}^{n} X_i,\ \sum_{i=1}^{n} \log X_i)$

2. $(X_{(1)},\ \sum_{i=1}^{n} X_i)$

3. a) $X_1 \cdot X_2 \cdots X_n$ (또는 $\sum_{i=1}^{n} \log X_i$)

b) 완비통계량이다.

4. a)

b) 완비통계량이 아니다.

5. $\sum_{i=1}^{n} X_i^2$

6. $\sum_{i=1}^{n} (X_i - 1)^2$

7.

8.

9. a) $\sum_{i=1}^{n} |X_i|$

   b) 완비통계량이 아니다.

8장

1. $\widehat{\theta} = \frac{\overline{X}}{2}$

2. a) $X = -3$일 때 $\widehat{\theta} = \theta_1$, $X = -2$일 때 $\widehat{\theta} = \theta_1$, $X = -1$일 때 $\widehat{\theta} = \theta_2$, $X = 0$일 때 $\widehat{\theta} = \theta_2$, $X = 1$일 때 $\widehat{\theta} = \theta_3$, $X = 2$일 때 $\widehat{\theta} = \theta_3$, $X = 3$일 때 $\widehat{\theta} = \theta_1$,

   b) $\widehat{\theta} = \theta_3$

3. a) $\widehat{\mu} = \overline{X} - \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}$, $\widehat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}$

   b) $\widehat{\mu} = X_{(1)}$, $\widehat{\sigma} = \frac{1}{n}\sum_{i=1}^{n}(X_i - X_{(1)})$

4. a) $Bias(T_1(\boldsymbol{X})) = 0$, $Bias(T_2(\boldsymbol{X})) = -\frac{1}{n+1}\theta$

   b) $Var(T_1(\boldsymbol{X})) = \frac{\theta^2}{n(n+2)}$, $Var(T_2(\boldsymbol{X})) = \frac{n\theta^2}{(n+1)^2(n+2)}$

   c) $MSE(T_1(\boldsymbol{X})) = \frac{\theta^2}{n(n+2)}$, $MSE(T_2(\boldsymbol{X})) = \frac{2\theta^2}{(n+1)(n+2)}$

   d) 평균제곱오차가 더 작은 $T_1(\boldsymbol{X})$이 더 좋은 추정량이라 할 수 있다.

5. a)

   b) $a = \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}$

6. a) $a + 2b = 1$

   b) $a^2\theta + 2b^2\theta + \theta^2(a + 2b - 1)^2$

   c) $\widehat{\theta} = \frac{1}{3}X + \frac{1}{3}Y$

7. a)

   b) $c = \frac{a^2 d}{\frac{1}{n}+a^2 d}$, 단, $d = 1 - \frac{2}{n-1}\left(\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\right)^2$

   c) $MSE(T_{c_1,c_2}(\boldsymbol{X})) = \theta^2\left[(c_1 + c_2 - 1)^2 + \frac{c_1^2}{n} + c_2^2 a^2\left(1 - \frac{2}{n-1}\left(\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\right)^2\right)\right]$

8. a) $MSE(T_c(\boldsymbol{X})) = \theta^2\left[(c - 1)^2 + \frac{c^2}{n}\right]$

   b) $c = \frac{n}{n+1}$

9. a) $\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$

   b) $\widehat{(\sigma^2)^2} = \left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right)^2$

c) $I(\sigma^2) = \frac{1}{2(\sigma^2)^2}$

10. a) $\widehat{\theta} = \overline{X}$
    b) $\widehat{\theta} = \overline{X}$
    c) $Var(\widehat{\theta}) = \frac{\theta^2}{n}$
    d) $\frac{1}{nI(\theta)} = \frac{\theta^2}{n}$

11. a) $\widehat{\theta} = \frac{1}{2n}\sum_{i=1}^{n} X_i$
    b) $I(\theta) = \frac{2}{\theta^2}$
    c) $\frac{1}{nI(\theta)} = \frac{\theta^2}{2n}$

12. a) $\widehat{\theta} = \frac{n}{\sum_{i=1}^{n} X_i^2}$
    b) $\mathrm{N}(0, \theta^2)$
    c) $\left(\widehat{\theta} - 1.96\frac{\widehat{\theta}}{\sqrt{n}},\ \widehat{\theta} + 1.96\frac{\widehat{\theta}}{\sqrt{n}}\right)$

13. $\dfrac{\begin{pmatrix} n-4 \\ \sum_{i=1}^{n} X_i - 4 \end{pmatrix}}{\begin{pmatrix} n \\ \sum_{i=1}^{n} X_i \end{pmatrix}}$

14. a) $\sum_{i=1}^{n} X_i$
    b)
    c) $\dfrac{\begin{pmatrix} (n-1)m \\ \sum_{i=1}^{n} X_i - m \end{pmatrix}}{\begin{pmatrix} nm \\ \sum_{i=1}^{n} X_i \end{pmatrix}}$

15. a)
    b) $\frac{\sum_{i=1}^{n} X_i}{n} \left(\frac{n-1}{n}\right)^{\sum_{i=1}^{n} X_i}$

16. a) $\frac{1}{n}\sum_{i=1}^{n} X_i^2$
    b) $\frac{1}{n(n+2)}\left(\sum_{i=1}^{n} X_i^2\right)^2$

17. a) $L(\theta) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \exp\left(-\frac{\sum_{k=1}^{n}(x_k - k\theta)^2}{2}\right)$
    b) $\widehat{\theta} = \frac{\sum_{k=1}^{n} kX_k}{\sum_{k=1}^{n} k^2}$

18.

9장

1. a) $\{2,3\}$

   b) 0.55

   c) $\phi(x) = \begin{cases} 1, & x \in \{2,3\}\text{일 때} \\ \frac{1}{2}, & x = 1\text{일 때} \\ 0, & x \in \{0,4,5\}\text{일 때} \end{cases}$

2. a) $\alpha = 0.15,\ \beta = 0.3$

   b) 제1종 오류를 범할 확률과 제2종 오류를 범할 확률이 더 작은 R_2가 더 좋다.

   c) $\{-1,0\}$

   d) $\{0,1\}$

3. $2\theta_0 \sum_{i=1}^n X_i \ \le \ \chi^2_{4n}(1-\alpha)$

4. $n = \left(\frac{Z_\alpha + Z_\beta}{\theta_1 - \theta_0}\right)^2$

5. $\sqrt{\frac{4}{5}}\left(X + \frac{Y}{4}\right) \ > \ Z_\alpha$

6. 한 가지 방법은 $\phi(x) = \begin{cases} \alpha, & 0 \le x \le \theta_0 \\ 1, & x > \theta_0 \end{cases}$

7.

8.

9.

10. a) $\frac{1}{\sigma_0^2}\sum_{i=1}^n X_i^2 \ > \ \chi^2_n(\alpha)$

    b) $P\left(\chi^2_n \ > \ \frac{\sigma_0^2}{\sigma_1^2}\chi^2_n(\alpha)\right)$

    c)

    d) $\frac{1}{\sigma_0^2}\sum_{i=1}^n (X_i - 1)^2 \ > \ \chi^2_n(\alpha)$

11.

12. a) $\Lambda = \exp\left(-\frac{1}{2}\frac{(\sum_{k=1}^n kX_k)^2}{\sum_{k=1}^n k^2}\right)$

    b) $\frac{(\sum_{k=1}^n kX_k)^2}{\sum_{k=1}^n k^2} \ > \ \chi^2_1(\alpha)$

13.

14. $\sum_{i=1}^m \left(X_i \log\frac{\widehat{p}_i}{\widehat{p}} + (n_i - X_i)\log\frac{1-\widehat{p}_i}{1-\widehat{p}}\right) \ > \chi^2_{n-1}(\alpha)$, 단, $\widehat{p}_i = \frac{X_i}{n_i}$, $\widehat{p} = \frac{\sum_{i=1}^m X_i}{\sum_{i=1}^m n_i}$

15. $\frac{n\sum_{i=1}^m (\overline{X}_i - \overline{X})^2/(m-1)}{\sum_{i=1}^m \sum_{j=1}^n (X_{ij} - \overline{X}_i)^2/m(n-1)} \ > \ F_{m-1,m(n-1)}(\alpha)$, 단, $\overline{X}_i = \frac{1}{n}\sum_{j=1}^n X_{ij}$, $\overline{X} = \frac{1}{mn}\sum_{i=1}^m \sum_{j=1}^n X_{ij}$

16.

10장

1. a) $\text{Gamma}(\alpha + \sum_{i=1}^{n} x_i,\ n + \lambda)$
   b) $\widehat{\theta}^B = \frac{\alpha + \sum_{i=1}^{n} x_i}{n + \lambda}$
   c)
2. a) $\text{Gamma}(n + \alpha,\ \lambda + \sum_{i=1}^{n} x_i)$
   b) $\widehat{\theta}^B = \frac{n + \alpha}{\lambda + \sum_{i=1}^{n} x_i}$
3. a) $\text{Gamma}(3n + 1,\ \lambda + \sum_{i=1}^{n} x_i)$
   b) $\widehat{\theta}^B = \frac{3n + 1}{\lambda + \sum_{i=1}^{n} x_i}$
4. a)

	a_1	a_2	a_3
θ_0	0	1	2
θ_1	2	1	0

   b) $R(\theta_0, d_1) = 0.2, R(\theta_1, d_1) = 0.8,\ R(\theta_0, d_2) = 0.5, R(\theta_1, d_2) = 0.6,\ R(\theta_0, d_3) = 0.8, R(\theta_1, d_3) = 0.4,\ R(\theta_0, d_4) = 1.0, R(\theta_1, d_4) = 1.0$
   위험함수의 최댓값을 가장 작게하는 규칙은 d_2 이다.
   c) $B(d_1) = 0.44,\ B(d_2) = 0.54,\ B(d_3) = 0.64,\ B(d_4) = 1.0$
   베이즈 위험을 가장 작게하는 규칙은 d_1 이다.
5. a) $P(|\text{N}(0,1)| > c)$
   b) $2P\left(|\text{N}(0,1)| \le \frac{c}{2}\right)$
   c) $B(d_c) = 0.6 \times P(|\text{N}(0,1)| > c) + 0.8 \times P\left(|\text{N}(0,1)| \le \frac{c}{2}\right)$
6. a) $R(\theta, d_1) = -10\theta^2 + 10\theta - 3,\ R(\theta, d_2) = 10\theta^2 - 10\theta + 2,\ R(\theta, d_3) = 0$
   b) d_1
   c) d_1
7. a) $P(\Theta = \theta_1 | X = x_2) = \frac{10}{33},\ P(\Theta = \theta_2 | X = x_2) = \frac{8}{33},\ P(\Theta = \theta_3 | X = x_2) = \frac{6}{33},\ P(\Theta = \theta_4 | X = x_2) = \frac{9}{33}$
   b) $P(\Theta = \theta_1 | X_1 = x_2, X_2 = x_3) = \frac{25}{89},\ P(\Theta = \theta_2 | X_1 = x_2, X_2 = x_3) = \frac{16}{89},\ P(\Theta = \theta_3 | X_1 = x_2, X_2 = x_3) = \frac{30}{89},\ P(\Theta = \theta_4 | X_1 = x_2, X_2 = x_3) = \frac{18}{89}$
8.
9. $c = \frac{\sigma_0^2}{\sigma_0^2 + 1}$

찾아보기