Statistics Ⅱ Midterm Homework

name : Sangman Jung **student number** : 2014110374

Professor : Sunmi Lee

Department of Applied Mathematics, Kyung Hee University

PRELIMINARY

❖ 통계적 모형(Statistical Model) : 수학적 모델의 한 종류 (관계와 변수로 이루어지는 어떤 함수 또는 관계)

→ 모델인 함수가 확률함수이고 확률분포를 가지며, 변수가 확률변수 : 통계적 모형

❖ 확률변수(Random Variable) : S가 표본공간일 때 실함수 $X : S \to R$

→ 표본공간 안의 사건(event)들을 실수로 수치화

❖ 확률분포(Probability Distribution) : 확률변수의 분포 / 확률변수가 특정한 값을 가질 확률을 나타내는 함수

* 이산형 RV : 확률질량함수(p.m.f) $f_X(x) = P(X = x)$ s.t. $\forall x, 0 \le f_X(x) \le 1$ and $\sum_x f_X(x) = 1$

* 연속형 RV : 확률밀도함수(p.d.f) $P(a < X < b) = \int_a^b f_X(x)dx \equiv X \sim f_X(x)$

s.t. $\forall x, 0 \le f_X(x) \le 1$ & $\int_{-\infty}^{\infty} f_X(x)dx = 1$

→ Discrete : $E(X) = \Sigma x f_X(x)$, $Var(X) = \Sigma(x - E(X))^2 f_X(x)$

→ Continuous : $E(X) = \int x f_X(x)dx$, $Var(X) = \int (x - E(X))^2 f_X(x)dx$

❖ $i.i.d$ (independent identically distributed) : 각 확률변수는 독립이고, 각각이 동일한 분포를 가짐

❖ 베르누이 분포(Bernoulli Distribution) :

Suppose that X_i is a random variables s.t. $P(X_i = 0) = \theta$, $P(X_i = 1) = 1 - \theta$, $0 \le p \le 1$, and X_i : $i.i.d.$ for $\forall X_i$. Then Probability Distribution of X is Bernoulli Distribution.

* p.m.f. of X_i : $f(x_i|\theta) = \begin{cases} \theta, & x_i = 1 \\ 1 - \theta, & x_i = 0 \end{cases} \Leftrightarrow f(x_i|\theta) = \theta^{x_i}(1-\theta)^{1-x_i}$, $x_i = 0,1$ for $i = 1, \cdots, n$.

* joint p.f. of $\overrightarrow{X} = (X_1, \cdots, X_n)$: $f(\overrightarrow{x}|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$, $\theta \in (0,1)$

* Mean / Variance : $E(X_i) = \theta$ / $Var(X_i) = \theta(1-\theta)$

→ notation) $X_i \sim Ber(\theta)$ / 베르누이 분포는 이항분포의 특수한 경우

❖ 이항 분포(Binomial Distribution) : In Bernoulli Distribution, let $Y = \sum_{i=1}^{n} X_i$. Then

* p.m.f. of Y : $f(y|\theta) = \binom{n}{y}\theta^y(1-\theta)^{1-y}$, $y = 0, \cdots, n$, $\theta > 0$

* Mean / Variance : $E(Y) = n\theta$ / $Var(Y) = n\theta(1-\theta)$

→ notation) $Y \sim Bin(n, \theta)$

❖ 포아송 분포(Poisson Distribution) : 확률변수 X에 대하여 포아송 분포의 확률질량함수는 다음과 같다.

* p.m.f. of X : $P(X=x)=f(x|\theta)=e^{-\theta}\frac{\theta^x}{x!}$, $x=0,1,2,\cdots$, $\theta>0$

* Mean / Variance : $E(X)=\theta$ / $Var(X)=\theta$

→ notation) $X\sim Poisson(\theta)$ / 이항분포의 n이 매우 크고 θ가 매우 작으면 이항분포는 포아송 분포로 수렴한다.

❖ 균등 분포(Uniform Distribution) : 확률변수 X에 대하여 균등분포의 확률밀도함수는 다음과 같다.

* p.d.f. of X : $f(x|\theta_1,\theta_2)=\begin{cases}\frac{1}{\theta_2-\theta_1}, & \theta_1<x<\theta_2 \\ 0 & otherwise\end{cases}$

* Mean / Variance : $E(X)=\frac{\theta_1+\theta_2}{2}$ / $Var(X)=\frac{(\theta_2-\theta_1)^2}{12}$

→ notation) $X\sim U(\theta_1,\theta_2)$

❖ 지수 분포(Exponential Distribution) : 확률변수 X에 대하여 지수분포의 확률밀도함수는 다음과 같다.

* p.d.f. of X : $f(x|\theta)=\theta e^{-\theta x}$, $x>0,\theta>0$

* Mean / Variance : $E(X)=\frac{1}{\theta}$ / $Var(X)=\frac{1}{\theta^2}$

→ notation) $X\sim \mathrm{Exp}(\alpha,\beta)$ / 지수분포는 감마분포의 특수한 경우

❖ 감마 분포(Gamma Distribution) :

In Exponential Distribution, suppose that for each X_i : $i.i.d.$ & $X_i\sim \mathrm{Exp}(\beta)$, $Y=\sum_{i=1}^{n}X_i$. Then Probability Distribution of Y is Gamma Distribution.

* p.d.f. of Y : $f(y|\alpha,\beta)=\frac{\beta^\alpha}{\Gamma(\alpha)}y^{\alpha-1}e^{-\beta y}$, $y>0$, $\alpha,\beta>0$

* Mean / Variance : $E(Y)=\frac{\alpha}{\beta}$ / $Var(Y)=\frac{\alpha}{\beta^2}$

→ notation) $Y\sim Gam(\alpha,\beta)$

❖ 카이제곱 분포(Chi-square Distribution) : In Gamma Distribution,

let $\alpha=\frac{k}{2}$, $\beta=\frac{1}{2}$. then Y has Chi-square(χ^2) Distribution with k degree of freedom. ($Y\sim\chi^2(k)$)

* p.d.f. of Y : $f(y)=\frac{2^{-\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)}y^{\frac{k}{2}-1}e^{-\frac{y}{2}}$, $0<x<\infty$

* Mean / Variance : $E(Y)=k$ / $Var(Y)=2k$

❖ **정규 분포**(Normal Distribution) :

확률변수 X에 대하여 평균 μ와 분산 σ^2를 갖는 정규분포의 확률밀도함수는 다음과 같다.

* p.d.f. of X : $f(x|\theta)=\dfrac{1}{\sqrt{2\pi}\,\sigma}\exp\left[-\dfrac{(x-\mu)^2}{2\sigma^2}\right]$, $-\infty < x < \infty$

이 때, $\mu=0$, $\sigma^2=1$ 이면 이를 **표준정규분포**(Standard Normal Distribution)라 한다.

→ notation) $X \sim N(\mu,\sigma^2)$

❖ **베타 분포**(Beta Distribution) : 확률변수 X에 대하여 베타분포의 확률밀도함수는 다음과 같다.

* p.d.f. of X : $f(x|\alpha,\beta)=\dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$, $0<x<1$

* Mean / Variance : $E(X)=\dfrac{\beta}{\alpha+\beta}$ / $Var(X)=\dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

이 때, $\alpha=\beta=1$ 인 경우 베타분포는 구간 $[0,1]$에서의 균등분포가 된다.

→ notation) $X \sim Be(\alpha,\beta)$

<u>NOTE</u>

* **베이즈 정리**(Bayes' Theorem)

사건 $B_1,\cdots,B_k$ 가 $\Pr(B_j)>0$ for $j=1,\cdots,k$ 을 만족하는 표본공간 S 의 분할(partition)을 형성하고 사건 A가 $\Pr(A)>0$ 을 만족하면, $j=1,\cdots,k$ 에 대하여

$$\Pr(B_i|A)=\frac{\Pr(B_i)\Pr(A|B_i)}{\sum_{j=1}^{k}\Pr(B_j)\Pr(A|B_j)}$$

* **조건부 독립**(Conditional independence)

사건 $A_1,\cdots,A_k$ 가 B 에 대하여 **조건부 독립**(Conditional independent) 일 필요충분조건은 임의의 집합족 $A_{i_1},\cdots,A_{i_j}$, $j=2,3,\cdots,k$ 에 대하여

$$\Pr(A_{i_1}\cap\cdots\cap A_{i_j}|B)=\Pr(A_{i_1}|B)\cdots\Pr(A_{i_j}|B)$$

* **전확률의 법칙**(Law of total probability)

사건 $B_1,\cdots,B_k$ 가 $\Pr(B_j)>0$ for $j=1,\cdots,k$ 을 만족하는 표본공간 S 의 분할(partition)을 형성하면 임의의 사건 $A\subset S$ 에 대하여,

$$\Pr(A)=\sum_{j=1}^{k}\Pr(B_j)\Pr(A|B_j)$$

* **조건부 확률**(Conditional Probability)

The conditional probability of the event A given that the event B has occurred if and only if

$$\Pr(A|B)=\frac{\Pr(A\cap B)}{\Pr(B)} \text{ if } \Pr(B)>0$$

NOTE

* **확률적으로 수렴**(Convergence in Probability)

확률변수들로 이루어진 수열 $Z_1, Z_2, \cdots\cdots$가 확률적으로 b 로 수렴한다.

$\Leftrightarrow\ \forall \epsilon > 0,\ \lim_{n\to\infty} \Pr(|Z_n - b| < \epsilon) = 1 \ \Leftrightarrow\ Z_n \xrightarrow{p} b,$: Z_n converges to b in probability.

* **큰 수의 법칙**(The Law of Large Number)

$X_1, \cdots\cdots, X_n$ 이 $\sigma^2 < \infty$, 모평균을 μ 로 갖는 분포로부터 무작위 표본을 생성한다고 가정하자.

$\overline{X}_n$ 을 표본평균이라고 하면, $\overline{X}_n \xrightarrow{p} \mu.$ 를 만족한다. ($\mu < \infty$, $\sigma^2 = \infty$ 인 경우에도 성립한다.)

NOTE

* Notation) $\overrightarrow{X} = (X_1, \cdots, X_m)$, $\overrightarrow{x} = (x_1, \cdots, x_m)$, $X_1 \mapsto x_i$ … if $S \to R$ 일 때

$$f_m(\overrightarrow{x}\,|\theta) = f(x_1, \cdots, x_m|\theta) = f(x_1|\theta)\cdots f(x_m|\theta)$$

❖ **주변확률밀도함수**(marginal p.d.f.) : 연속형 확률변수 X와 Y에 대하여 주변확률밀도함수는 다음과 같다.

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)dy\ ,\ f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)dx$$

→확률변수가 이산형일 경우 적분(integral)을 각각 x와 y에 대한 합(sum)으로 대체하면 된다.

Definition 7.1.2) Statistical Inference

1. 통계적 모형의 모든, 또는 어떤 부분에 대한 확률적인 진술과정.
2. 표본자료에서 구한 통계량에 의거하여 모집단의 모수에 대한 확률적 의사결정을 하는 과정.
3. 검정통계량의 확률분포에서 검정통계량의 값을 발견할 확률에 의거하여 귀무가설에 대한 거부 여부를 결정하는 과정

Definition 7.1.3) Parameter / Parameter space

* 통계적 추론의 문제에서, random variables of interest(관측 or 가설적 관측 가능한 확률변수) 의 결합 확률분포를 결정하는 특성들의 조합 또는 특성을 그 분포의 **parameter(모수, 매개변수)** 라고 한다.
* parameter θ 또는 vector of parameters $(\theta_1, \cdots, \theta_k)$의 모든 가능한 값들의 집합 Ω을 **parameter space(모수공간)** 라고 한다.

Definition 7.1.4) Statistic

* 관측 가능한 확률변수들을 $X_1, \cdots, X_n$ 이라 하고, r : n개의 실변수를 갖는 임의의 실함수라고 하자. 그러면, 확률변수 $T = r(X_1, \cdots, X_n)$ 을 통계량(Statistic) 이라고 한다.

* 모집단의 모수를 추정하기 위하여 표본에서 계산한 추정량의 값.
추정량은 통계량을 계산하는 규칙이며, 통계량은 추정량의 계산규칙에 의거하여 표본자료에서 생성된 값.
통계량은 표본에 따라 값이 다른 확률변수로서 확률분포를 가지며, 통계량은 모수에 의존하지 않으나, 통계량의 분포는 모수에 의존한다.

Example 7.2.8) Lifetimes of Fluorescent Lamps

어떤 회사가 자사에서 생산된 형광등에 대한 수명에 관한 분포를 궁금해 한다. 이러한 수명에 관한 확률변수를 X_i로 두고, X_i : $i.i.d.$, $X_i \sim \text{Exp}(\theta)$라 가정하자, 이 때, θ는 알려지지 않았다고 하자. 우리는 이제 이러한 모수 θ를 추정하고자 한다. 이와 같은 확률변수는 사전적인 통계적 경험들을 통하여, θ에 관한 사전분포를 모수가 $\alpha=4$, $\beta=20000$인 다음과 같은 감마분포의 확률밀도함수로 정하였다.

$$\xi(\theta)=\begin{cases}\dfrac{(20000)^4}{3!}\theta^3 e^{-20000\theta} & \theta>0\\ 0 & \theta\le 0\end{cases}$$

> **Definition 7.2.1) 사전 분포(Prior Distribution)**
> 모수 θ를 확률변수로 간주할 때, θ를 제외한 다른 확률변수들을 관측하기 전에 채택한 어떤 통계적 모형의 분포를 사전분포라 한다. 일반적으로 사전분포의 확률함수는 $\xi(\theta)$로 표기한다.

이제 확률변수에 대한 실제 관측 표본 5개를 얻었다고 하자. 즉, $n=5$일 때 시간당 수명에 대한 확률변수는 $X_1=x_1=2911$, $X_2=x_2=3403$, $X_3=x_3=3237$, $X_4=x_4=3509$, $X_5=x_5=3118$로 얻어졌다고 하자. 우리는 이러한 관측값들을 가질 때 모수 θ를 추정하기 위해 업데이트 된(관측된 후의) 분포를 얻어서 모수 θ의 추정을 더 긴밀하게 하고자 한다. 즉, θ에 대한 사후분포를 구하려면 다음의 정의와 정리가 필요하다.

> **Definition 7.2.2) 사후 분포(Posterior Distribution)**
> 관측된 확률변수 $X_1,\cdots,X_n$ 를 갖는 모수 θ를 찾는 통계적 추론 문제를 고려하자. $X_1,\cdots,X_n$ 를 조건부로 갖는 θ에 관한 조건부 분포를 θ에 관한 **사후분포**(Posterior Distribution)라 한다.

> **Theorem 7.2.1)** 확률변수 $X_1,\cdots,X_n$ 가 $f(x|\theta)$ 를 확률함수로 갖는 확률표본을 구성한다고 가정하자.
> 그리고 모수 θ의 값은 알려져 있지 않고, 사전분포가 $\xi(\theta)$라 가정하자. 그러면 사후분포는 다음과 같다.
>
> $$\xi(\theta|\vec{x})=\frac{f(x_1|\theta)\cdots f(x_n|\theta)\xi(\theta)}{g_n(\vec{x})}\quad \text{for } \theta\in\Omega \text{ , where } g_n \text{ : marginal joint p.d.f./p.f. of } X_1,\cdots,X_n.$$

사후분포를 구하기에 앞서, 베이즈 정리로 유도된 위 정리의 식의 우변에 있는 함수들을 구해야한다. 먼저 우도함수(Likelihood function)를 다음과 같이 구하면 된다.

$$f(x\,|\,\theta)=\begin{cases}\theta\, e^{-\theta x} & \text{for } x>0\\ 0 & otherwise\end{cases}\quad (\because X_i \sim \text{Exp}(\theta))$$

> **Definition 7.2.3) Likelihood Function**
> When the joint p.d.f. / joint p.f $f_n(\vec{x}|\theta)$ of the observations in a random sample is regarded as a funtion of θ for given values of $x_1,\cdots,x_n$, it is called the **Likelihood Function(우도함수)**.

그리고 전확률의 법칙을 이용하여 $X_1,\cdots,X_n$에 대한 결합확률밀도함수를 아래와 같이 구하면 된다.

$$f_5(\overrightarrow{x}\,|\theta)=f(x_1,\cdots,x_5|\theta)=f(x_1|\theta)\cdots f(x_5|\theta)=\prod_{i=1}^{n}\theta e^{-\theta x_i}=\theta^n e^{-\theta(x_1+\cdots+x_5)}$$

여기서 $x_1+\cdots+x_5=y$ 로 두고, 정리 7.2.1.의 우변의 분자식을 미리 계산해주면 아래와 같이 된다.

$$f_n(\overrightarrow{x}|\theta)\xi(\theta)=\theta^{n+3}e^{-(y+20000)\theta}\ \text{ for }\ \theta>0$$

이제 우변의 분모인 주변결합확률밀도함수 $g(x)$를 구하면

$$g(\overrightarrow{x})=\int_0^\infty \theta^{n+3}e^{-(y+20000)\theta}d\theta=\frac{\Gamma(n+4)}{(y+20000)^{n+4}}\quad (\because\ \forall\ \alpha,\beta>0,\ \int_0^\infty x^{\alpha-1}e^{-\beta x}dx=\frac{\Gamma(\alpha)}{\beta^\alpha})$$

앞서 구한 우도함수와 사전분포의 곱과 $g(x)$에 실제로 구해진 관측값 $x_1,\cdots,x_5$를 대입해주면 $y=16178$ 이고, 따라서 최종적으로 식을 정리하면 아래와 같은 사후분포를 얻는다.

$$\xi(\theta|\overrightarrow{x})=\frac{f(x_1|\theta)\cdots f(x_5|\theta)\xi(\theta)}{g(\overrightarrow{x})}=\frac{(y+20000)^{n+4}}{\Gamma(n+4)}\theta^{n+3}e^{-(y+20000)\theta}=\frac{(36178)^9}{\Gamma(9)}\theta^8 e^{-(36178)\theta}$$

사후분포의 꼴을 잘 살펴보면, 사후분포는 결국 모수가 9와 36178로 갖는 감마분포임을 알 수 있다.

이제 $X=x_6$으로 새로운 관찰 값을 업데이트 했다고 가정하자, 이 X_6이 3000시간 보다 클 확률을 알고자 한다. 즉, $\Pr(X_6>3000|\overrightarrow{x})$의 값을 구하고자 한다. 이는 다음과 같은 관계를 갖는다.

$$\Pr(X_6>3000|\overrightarrow{x})=\int_{3000}^\infty f(x_6|\overrightarrow{x})dx_6$$

위 식의 우변에서, 피적분함수인 조건부확률함수는 아래의 명제를 이용하여 구한다.

* Sequential Observations and Prediction

$n-1$ 개의 관측값이 주어질 때, X_n에 대한 조건부 확률함수로써, $\xi(\theta|\overrightarrow{x})\propto f(x_n|\theta)\xi(\theta|x_1,\cdots,x_{n-1})$ 인 관계를 가질 때, 비례상수(Proportionality constant) 가 1이면 다음이 성립한다.

$$f(x_n|x_1,\cdots,x_{n-1})=\int f(x_n|\theta)\xi(\theta|x_1,\cdots,x_{n-1})\,d\theta$$

이는 수열적으로 정의되는 n 번째 관측값(observation)을 예측(prediction) 하기 위한 연속분포의 함수이다.

$\xi(\theta|\overrightarrow{x})=2.633\times10^{36}\theta^8e^{-36178\theta}$, $\theta>0$로 구하였으므로, $\theta>0$에 대하여 적분하면 다음을 얻는다.

$$f(x_6|\overrightarrow{x})=\int_0^\infty 2.633\times10^{36}\theta^8e^{-36178\theta}\theta e^{-x_6\theta}d\theta\ =2.633\times10^{36}\frac{\Gamma(10)}{(x_6+36178)^{10}}=\frac{9.555\times10^{41}}{(x_6+36178)^{10}}$$

이제 마지막으로, 구하고자하는 6번째 관측 수명값 X_6가 3000시간 이상일 확률은 아래와 같이 구해진다.

$$\Pr(X_6 > 3000 | \vec{x}) = \int_{3000}^{\infty} \frac{9.555 \times 10^{41}}{(x_6 + 36178)^{10}} \, dx_6 = \frac{9.555 \times 10^{41}}{9 \times 39178^9} = 0.4882$$

민감도 분석을 시도해보자. 마찬가지로 다음 형광등의 수명이 적어도 3000시간 이상일 확률을 알고자 할 때, 우리는 우리가 이전에 행했던 계산에서 사전분포의 선택이 구해지는 확률에 얼마나 영향을 미치는지 볼 수 있다. 우선 본 예제와는 다른 두 번째 사전분포(모수 값을 1, 1000로 갖는 감마분포)를 채택하자. 이 때 사후분포의 모수 값은 6, 17178 로 구해졌다고 하자.

앞서 구했던 사후분포와 마찬가지 방법으로, 새로 바뀐 모수 값에 대한 사후분포와 그 확률을 구하면

$$f(x_6 | \vec{x}) = \frac{1.542 \times 10^{26}}{(x_6 + 17178)^7}, \quad \text{for } \ x_6 > 0 \ ,$$

$$\Pr(X_6 > 3000 | \vec{x}) = \int_{3000}^{\infty} \frac{1.542 \times 10^{26}}{(x_6 + 17178)^7} \, dx_6 = 0.3807 \ .$$

앞서 구한 확률과 다소 차이가 나는 것을 확인할 수 있다. 이처럼 다른 사전분포의 선택은 우리가 하고자 하는 추론에 있어 고려할만한 차이를 만든다. 만약 확률 $\Pr(X_6 > 3000 | \vec{x})$에 대한 두 분포가 더욱 근사하도록 만들고 싶다면, 더 큰 표본을 가져와서 계산할 때 두 사후분포 값이 가깝게 된다.

◀

Example 7.3.3) Glove Use by Nurses.

어떤 연구기관에서 23명의 간호사들에 대해 위생장갑 착용의 중요성에 대한 교육 프로그램 수강 전과 후에 대한 조사를 실시했다. 연구기관은 체액에 접촉할 수도 있는 진료 절차동안 장갑 착용 여부에 대해 기록하였는데, 교육 프로그램 수강 전에는 51개의 진료과정 동안 간호사들을 관찰한 결과, 13명만 장갑을 착용했음을 확인했다.

이제 θ를 간호사들이 교육 프로그램 수강 후 2달 간 장갑을 착용할 확률이라고 하자. 우리는 프로그램 수강 전에 관측된 비율인 13/51 과 θ가 어떻게 비교될 것인지에 초점을 두고자 한다.

우선, 사전분포의 선택이 θ에 관한 사후분포에 얼마나 민감할지 결정하기 위해 θ에 관한 두 개의 다른 사전분포를 생각하자. 첫 번째 사전분포는 $X \sim U(0,1)$로 하고, 두 번째 사전분포는 $X \sim Be(13,38)$로 정하자. 이 사전분포는 첫 번째 분포보다 훨씬 작은 분산을 갖고 평균을 13/51 로 갖는다. 이 때, 두 번째 사전분포를 택하는 연구자는 교육 프로그램이 주목할 만한 효과를 주지 못할 것이라 굳게 주장한다.

이제 사후분포를 구하기에 앞서 알아두어야 할 것은, 사후분포의 계산이 항상 쉽게 이루어지는 것은 아니라는 점이다. 이러한 사후분포의 계산을 편리하게 하기 위해, 공액분포족에 대한 내용을 알아두어야 한다.

Definition 7.3.1) Conjugate Family

주어진 모형의 특정 모수에 대해 사후분포가 사전분포와 동일한 형태가 되도록 하는 사전분포를 **공액사전분포**(conjugate prior distribution)라 한다. 이러한 공액사전분포들의 모음을 **공액분포족**(Conjugate family) 이라고 말한다.

note) 공액사전분포를 이용하면 사후분포가 적절한(proper)형태로 주어지는 장점을 가진다.

이 때, 적절한 형태란 p.d.f. 의 조건 ($\int_\Omega f(\theta|\vec{x})d\theta = 1$)을 만족하는 것을 일컫는다.

사전분포의 형태에 따라 부적절한 형태로($\int_\Omega f(\theta|\vec{x})d\theta = \infty$) 주어질 수도 있다.

이제 사후분포를 구하기 위해 다음 정리를 알아두자.

Theorem 7.3.1)

$X_1,\cdots,X_n$ 이 알려지지 않은 모수 θ $(0<\theta<1)$를 갖는 베르누이 분포로부터 확률표본을 생성하고, θ의 사전분포는 모수를 $\alpha>0, \beta>0$ 로 갖는 베타분포를 따른다고 가정하자. 그러면 $X_i = x_i,\ i=1,\cdots,n$ 이 주어진 θ의 사후분포는 모수 $\alpha+\sum_{i=1}^n x_i$ 와 $\beta+n-\sum_{i=1}^n x_i$ 를 갖는 베타분포를 따른다.

교육 프로그램 수강 이후 2달간, 56개의 진료동안 간호사들 중 50명이 장갑을 착용했음이 관찰되었다.
그러면 첫 번째 사전분포에 기초한 θ에 관한 사후분포는 정리 7.3.1.에 의하면 모수를 $1+50=51$, $1+6=7$로 갖는 베타분포를 따르게 된다. (균등분포는 모수를 1과 1로 갖는 베타분포) 특히, θ의 사후분포의 평균은 $51/(51+7)=0.88$이고, $\theta > 2\times 13/51$ 이 거의 1인 사후확률을 갖게 된다. 두 번째 사전분포에 기초한 사후분포는 모수를 $13+50=63$, $38+6=44$로 갖는 베타분포를 따르게 된다. 이 때 사후분포의 평균은 0.59이고, $\theta > 2\times 13/51$ 일 사후확률은 0.95이다.
결과적으로, 처음에는 누군가가 프로그램에 회의적이었다고 해도, 교육 프로그램은 매우 효과적이었음을 보여준다. 프로그램을 받기 전보다 받은 후에 간호사들이 적어도 두 번은 장갑을 착용할 비율이 아주 높아졌음을 확인할 수 있다.

◀

❖ *공액사전분포를 이용한 사후분포 구하기*

1. RV : 베르누이 & 사전 : 베타 $\Rightarrow$ 사후 : $\theta \sim Beta(\alpha+\Sigma x_i, \beta+n-\Sigma x_i)$

2. RV : 포아송 & 사전 : 감마 $\Rightarrow$ 사후 : $\theta \sim Gamma(\alpha+\Sigma x_i, \beta+n)$

3. RV : 정규(θ:unknown, σ^2:know) % 사전 : 정규$N(\mu_0, v_0^2)$ $\Rightarrow$ 사후 : $\theta \sim N(\frac{\sigma^2\mu_0+nv_0^2\overline{x_n}}{\sigma^2+nv_0^2}, \frac{\sigma^2 v_0^2}{\sigma^2+nv_0^2})$

4. RV : 지수 & 사전 : 감마 $\Rightarrow$ 사후 : $\theta \sim Gamma(\alpha+n, \beta+\Sigma x_i)$

Example 7.4.7) Chest Measurements of Scottish Solders.

1846년에 5732명의 스코틀랜드 군인들의 가슴둘레 측정(inches)에 대한 약간의 오차가 있는 데이터가 보고되었다. 이 데이터는 1817년 의료 저널에서 먼저 보고되었고, Stigler(1986)에 의해 논의되었다. 이제 우리가 개개인의 가슴둘레 측정값을 모수 θ와 분산이 4로 주어진 정규확률변수의 확률 표본으로써 모형화한다고 가정하자. 가슴둘레 측정값의 평균은 $\overline{x}_n = 39.85$이다. 만약 θ가 모수가 μ_0, v_0^2인 정규 사전분포를 갖는다면, 이는 공액사전분포로써 다음의 평균과 분산을 갖는 정규 사후분포를 갖는다.

$$\mu_1 = \frac{4\mu_0 + 5732 \times v_0^2 \times 39.85}{4 + 5732 \times v_0^2} \ , \ v_1^2 = \frac{4v_0^2}{4 + 5732 v_0^2}$$

우선, 추정량과 추정치에 대한 다음의 정의를 숙지하자.

Definition 7.4.1) 추정량(Estimator) / 추정치(Estimate)

$X_1, \ldots, X_n$이 실직선 Ω의 부분집합에서 취해지는 모수 θ에 의해 정해지는 결합분포의 관측가능한 데이터라 하자. 그러면 θ의 **추정량**(Estimator) 은 실함수 $\delta(X_1, \ldots, X_n)$이고, $X_1 = x_1, \ldots, X_n = x_n$으로 관측될 때, $\delta(x_1, \ldots, x_n)$를 θ의 **추정치**(Estimate) 라 한다.

베이즈 추정량을 구하기 위해서는 손실함수(Loss function)의 개념이 필요하다. 다음의 정의를 숙지하자.

Definition 7.4.2) Loss Function

손실함수(Loss function)는 두 변수 $\theta \in \Omega, a \in R$에 대한 실함수(real-valued function)을 말한다.
이 때, θ는 모수(parameter), a가 estimate인 경우를 일컫는다.

Definition) 기대손실(Expected Loss)

확률벡터 $\overrightarrow{X}$의 관측값 $\overrightarrow{x}$를 관측할 수 있다고 가정하고, $\xi(\theta|\overrightarrow{x})$를 $\theta \in \Omega$의 사후분포함수라고 하자.
각 estimate a에 대하여, 기대손실은 다음과 같다.

$$E[L(\theta, a)] = \int_\Omega L(\theta, a)\xi(\theta)\, d\theta.$$

Definition 7.4.4) Squared Error Loss Function : $L(\theta, a) = (\theta - a)^2$

Corollary 7.4.1) θ가 real-valued parameter, 손실함수가 $L(\theta, a) = (\theta - a)^2$, θ의 사후평균(기대값) $E(\theta|\overrightarrow{X})$이 유한(finite)하면 θ의 Bayes estimator 는 $\delta^*(\overrightarrow{X}) = E(\theta|\overrightarrow{X})$이다.

Definition 7.4.5) Absolute Error Loss Function : $L(\theta, a) = |\theta - a|$

Corollary 7.4.2) θ가 real-valued parameter, 손실함수가 $L(\theta, a) = |\theta - a|$이면 θ의 Bayes estimator $\delta^*(\overrightarrow{X})$는 θ에 관한 사후분포의 중앙값(median)과 같다.

이제, 베이즈 추정량과 추정치에 대한 정의를 살펴보면 아래와 같다.

> Definition 7.4.3) Bayes Estimator / Estimate
> $L(\theta, a)$를 손실함수라고 하자. 그리고 $\overrightarrow{X}$의 각 가능한 $\overrightarrow{x}$에 대하여, $\delta^*(\overrightarrow{x})$를 $E[L(\theta,a)|\overrightarrow{x}]$가 최소가 되게 하는 a의 값이라 하자. 그러면 δ^*를 θ의 Bayes estimator라고 한다.
> 또한 $\overrightarrow{X}=\overrightarrow{x}$로 관측될 때, $\delta^*(\overrightarrow{x})$를 θ의 Bayes estimate라고 한다.
> 이러한 Bayes estimator를 다르게 표현하면, $E[L(\theta,a)|\overrightarrow{x}]=\min_{\forall a}E[L(\theta, a)|\overrightarrow{x}]$ 이다.

제곱오차 또는 절대오차 손실함수를 이용하여 베이즈 추정량을 구하면, 따름정리 7.4.1.에 의하여 베이즈 추정치는 곧 $\delta^*(\overrightarrow{X})=\dfrac{\sigma^2\mu_0+nv_0^2\overline{X}_n}{\sigma^2+nv_0^2}=\mu_1$ 임을 알 수 있다.

(∵ 정규분포에 관련한 추정은 제곱오차와 절대오차 손실함수 중 어떤 것을 사용해도 결국 따름정리에 의해 사후분포의 기댓값으로 구해짐을 주의하자. 여기서 절대오차를 사용했을 때, 사후분포가 정규분포이므로 사후분포의 중앙값이 곧 평균이 됨을 알아두자.)

여기서 μ_0이 굉장히 커지거나 v_0^2이 매우 작아진다고 해도, μ_1이 거의 39.85와 같고 v_1^2이 거의 $4/5732$와 같아진다. 만약 θ에 관한 사전분포함수가 임의의 연속함수로써 $\theta=39.85$의 양의 자리 근처이고 θ가 39.85와 멀지 않을 때 극단적으로 커지지 않는다면 그에 대응하는 사전분포는 평균과 분산이 각각 39.85, 4/5732인 정규분포의 확률밀도함수와 매우 가깝게 된다. 이런 이유로, 그러한 사후분포의 평균과 중앙값은 사전 분포와 상관없이 거의 $\overline{x}_n$에 가깝게 된다.

◀

Example 7.4.8) Lifetimes of Electronic Components.

앞선 관련 예제들에서, θ에 대한 사전분포가 모수 1과 2를 갖는 감마분포이고 사후분포가 모수 4와 8.6을 갖는 감마분포임을 확인했다. 우리는 $\frac{1}{\theta}$ 를 추정하고자 한다. 우선 다음의 정리들을 숙지하자.

> **Theorem 4.7.3)**
> The prediction $d(X)$ that minimizes $\{E[Y-d(X)]\}^2$ is $d(X)=E(Y|X)$

> **Theorem 5.7.3)**
> $$\forall\ \alpha,\beta>0,\ \int_0^\infty x^{\alpha-1}e^{-\beta x}\,dx=\frac{\Gamma(\alpha)}{\beta^\alpha}$$

만약 여기서, 제곱오차 손실함수 $L(\theta,a)=(\frac{1}{\theta}-a)^2$ 를 이용한다면, 정리 4.7.3.에 의하여 베이즈 추정치는 $\frac{1}{\theta}$에 관한 사후분포의 평균(기댓값)이 된다. 즉,

$$\begin{aligned}\delta^*(\vec{x})&=E(\psi|\vec{x})=E\left(\frac{1}{\theta}|\vec{x}\right)\\&=\int_0^\infty\frac{1}{\theta}\xi(\theta|\vec{x})d\theta\\&=\int_0^\infty\frac{1}{\theta}\frac{8.6^4}{6}\theta^3e^{-8.6\theta}d\theta\\&=\frac{8.6^4}{6}\int_0^\infty\theta^2e^{-8.6\theta}d\theta\\&=\frac{8.6^4}{6}\frac{2}{8.6^3}=2.867\end{aligned}$$

여기서 마지막 등식은 정리 5.7.3.으로부터 유도된다. $1/\theta$의 평균은 $1/E(\theta|\vec{x})=8.6/4=2.15$보다 약간 더 높다.

◀

Example 7.5.6) Sampling from a Normal Distribution with Unknown Mean & Variance.
예제를 살펴보기 전에 최대우도추정량 / 추정치에 대한 정리를 먼저 숙지하자.

> **Definition 7.5.2) 최대우도추정량 / 추정치 (Maximum Likelihood Estimator / Estimate)**
> 관측된 벡터 $\vec{x}$에 대하여, $\delta(\vec{x}) \in \Omega$를 우도함수 $f_n(\vec{x}|\theta)$가 최대가 되게 하는 $\theta \in \Omega$의 값이라고 하자. 그리고 $\hat{\theta} = \delta(\vec{X})$를 이러한 관점으로 정의된 θ의 estimator라고 하자. 그러면 estimator $\hat{\theta}$를 Maximum Likelihood Estimator라 한다. $\vec{X} = \vec{x}$로 관측되었을 때, 그 값인 $\delta(\vec{x})$를 Maximum Likelihood Estimate라 한다.

확률변수 $X_1, \ldots, X_n$이 알려지지 않은 평균 μ와 알려지지 않은 분산 σ^2을 갖는 정규분포로부터 확률표본을 생성한다고 하자. 우도함수 $f_n(\vec{x}|\mu, \sigma^2)$ 는 다음과 같이 구할 수 있다.

$$f_n(\vec{x}|\mu) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right]$$

이 함수는 반드시 $-\infty < \mu < \infty$와 $\sigma^2 > 0$ 의 모든 가능한 값들에 대해 최대가 될 것이다. 직접적으로 이러한 우도함수를 최대화하는 것 대신에, 로그를 취하여 좀 더 쉽게 최대화 하는 방법을 생각하자. 우리는 다음을 얻는다.

$$\begin{aligned} L(\theta) &= \log f_n(\vec{x}|\mu, \sigma^2) \\ &= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2 \end{aligned}$$

이제 다음 세 단계의 과정을 통하여 $L(\theta)$가 최대가 되게 하는 $\theta = (\mu, \sigma^2)$를 찾도록 한다.
(1) 각 고정된 σ^2에 대하여, $L(\theta)$의 우변을 최대화하는 $\hat{\mu}(\sigma^2)$를 찾는다.
(2) $L(\theta')$, $\theta' = (\hat{\mu}(\sigma^2), \sigma^2)$를 최대화하는 $\widehat{\sigma^2}$를 찾는다.
(3) 관측값들의 확률 벡터가 되는 θ의 M.L.E는 $(\hat{\mu}(\sigma^2), \sigma^2)$이 된다.

(1)번의 경우는 예제 7.5.5에서 이미 확인하였으며, $\hat{\mu}(\sigma^2) = \bar{x}_n$ 임을 얻었다. (2)번의 경우에서는, $\theta' = (\bar{x}_n, \sigma^2)$로 두고, 다음을 최대화(maximize)한다.

$$L(\theta) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2$$

이는 σ^2에 관해 미분하여 0으로 두고 σ^2에 대해 풀면 최대화할 수 있다. 그 도함수는 다음과 같이 구한다.

$$\frac{d}{d\sigma^2}L(\theta') = -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2$$

이를 0으로 두고 풀면

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2$$

임을 얻는다.

$L(\theta')$에 대한 2계도함수는 앞서 구한 σ^2에서 음수가 되므로, 우리는 이 σ^2 값이 최대값임을 확인할 수 있다. 그러므로 $\theta=(\mu,\sigma^2)$의 M.L.E 는 다음과 같이 얻을 수 있다.

$$\hat{\theta}=(\hat{\mu},\widehat{\sigma^2})=\left(\overline{X}_n,\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2\right)$$

위의 M.L.E의 첫 번째 좌표는 데이터의 표본평균(sample mean)이라 말한다. 마찬가지로, 두 번째 죄표는 표본분산(sample variance)라고 말한다. 이는 표본분산의 관측값들이 표본의 각 관찰값 $x_1,...,x_n$의 확률이 $\frac{1}{n}$로 배정되는 분포의 분산이 됨을 확인할 수 있다.

◀

Example 7.5.7) Sampling from a Uniform Distribution.

$X_1,...,X_n$이 구간 $[0,\theta]$에서의 균등분포로부터 확률표본을 생성한다고 하자. 이때 모수 $\theta>0$은 알려져있지 않다. 각 관측값의 p.d.f. $f(x|\theta)$는 다음과 같은 형태를 갖는다.

$$f(x|\theta)=\begin{cases}\dfrac{1}{\theta} & \text{for}\, 0\le x\le\theta \\ 0 & otherwise\end{cases}$$

그러면, $X_1,...,X_n$의 joint p.d.f. $f_n(\vec{x}|\theta)$는 다음과 같은 형태를 갖는다.

$$f_n(\vec{x}|\theta)=\begin{cases}\dfrac{1}{\theta^n} & \text{for}\, 0\le x_i\le\theta\ \ (i=1,...,n) \\ 0 & otherwise\end{cases}$$

위의 joint p.d.f.로부터, θ의 M.L.E는 반드시 $\theta\ge x_i$ for $i=1,...,n$인 θ의 값으로써 i중 $1/\theta^n$을 최대화하는 값을 찾는 것임을 알 수 있다. $1/\theta^n$는 θ에 대한 감소함수이므로, estimate는 $\theta\ge x_i$ for $i=1,...,n$를 만족하는 가장 작은 θ값을 갖는다. 이러한 값은 $\theta=\max\{x_1,...,x_n\}$이므로, θ에 대한 M.L.E는 $\hat{\theta}=\max\{X_1,...,X_n\}$ 이 된다.

◀

§. *Additional Examples in M.L.E.*

Example 7.5.1) Lifetimes of Electronic Components.
예제 7.3.11의 데이터인 전자부품의 수명을 포함하는 데이터를 관측한다고 가정하자. 그러면 처음 사전분포와 손실함수 없이 불량률(failure rate) θ를 추정할 수 있는 방법이 있는가?

▲

Example 7.5.2) Lifetimes of Electronic Components.
예제 7.3.11에서, 관측된 데이터는 $X_1 = 3,\ X_2 = 1,5,\ X_3 = 2.1$이었다. 확률변수들은 크기가 3인 확률표본으로써 모수 θ를 갖는 지수분포로부터 모형화 하였다. 여기서 우도함수는 다음과 같다.

$$f_3(\vec{x}|\theta) = \theta^3 \exp(-6.6\theta),\ \theta > 0,\ \vec{x} = (2, 1.5, 2.1).$$

우도함수 $f_3(\vec{x}|\theta)$를 최대가 되게 하는 θ값을 찾는 것은, 로그함수가 증가함수임을 생각하면, $\log f_3(\vec{x}|\theta)$를 최대가 되게 하는 θ값을 찾는 것과 같아진다. 따라서 다음의 식을 최대화하는 θ값을 찾음으로써 M.L.E를 손쉽게 결정할 수 있다.

$$L(\theta) = \log f_3(\vec{x}|\theta) = 3\log(\theta) - 6.6\theta$$

$L(\theta)$의 도함수를 구하여 $\dfrac{dL(\theta)}{d\theta} = 0$ 으로 두면(일계도함수판정), θ에 대한 임계값은 $\theta = 3/6.6 = 0.455$로 구해지고, $\dfrac{d^2L(\theta)}{d\theta^2} = 0$ 으로 두고 θ에 관하여 풀면(이계도함수판정), 방정식을 만족하는 θ는 $\theta < 0$이므로 임계값이 곧 최대값이 된다. 따라서 Maximum Likelihood Estimate는 0.455로 구해진다.

▲

Example 7.5.3) Test for a Disease.
누군가가 길을 걷다가 공공보건소에서 특정 질병에 대한 무료 의학진단을 실시하는 것을 보았다고 하자. 그 의학 진단은 다음과 같은 관점에서 90% 신뢰할 수 있다.

만약 누군가 질병이 있으면, 양성일 진단결과가 나올 확률이 0.9인 반면에, 만약 질병이 없으면 양성일 진단결과가 나올 확률이 0.1 이다. 이러한 의학진단과 같은 예시는 예제 2.3.1에서 다루었다. 우리는 X를 그러한 진단에 대한 결과를 나타낸다고 하고, $X=1$이 양성 결과, $X=0$이 양성이 아닌 결과를 의미한다고 하자. 모수 공간은 $\Omega = \{0.1, 0.9\}$으로, $\theta = 0.1$이면 질병에 걸리지 않았다는 진단 결과를, $\theta = 0.9$는 질병에 걸렸다는 진단 결과를 의미한다고 하자. 이러한 모수공간은 주어진 θ에 대하여 X가 모수 θ를 갖는 베르누이 분포를 따른다고 할 수 있다. 그러므로 우도함수는 다음과 같다.

$$f(x|\theta) = \theta^x (1-\theta)^{1-x}$$

이 때, $x=0$으로 관측된 경우, $f(0|\theta) = \begin{cases} 0.9 & \text{if } \theta = 0.1 \\ 0.1 & \text{if } \theta = 0.9 \end{cases}$ 이고, 여기서 $\theta = 0.1$ 인 경우 우도함수가 최대가 됨을 쉽게 알 수 있다. 만약 $x=1$으로 관측된 경우, $f(1|\theta) = \begin{cases} 0.1 & \text{if } \theta = 0.1 \\ 0.9 & \text{if } \theta = 0.9 \end{cases}$ 이고, 여기서 $\theta = 0.9$인 경우 우도함수가

최대가 됨을 쉽게 알 수 있다. 이런 이유로, 우리는 다음과 같은 M.L.E를 얻는다.

$$\hat{\theta}=\begin{cases} 0.1 & \text{if } \theta=0.1 \\ 0.9 & \text{if } \theta=0.9 \end{cases}$$

▲

Example 7.5.4) Sampling from a Bernoulli Distribution.
확률변수 $X_1,\ldots,X_n$이 알려지지 않은 모수 $\theta(0\le\theta\le 1)$를 갖는 베르누이 분포로부터 확률표본을 형성한다고 하자. 각각의 x_i가 0 또는 1을 갖는 모든 관측값 $x_1,\ldots,x_n$에 대하여, 우도함수는 다음과 같이 구해진다.

$$f_n(\vec{x}|\theta)=\prod_{i=1}^{n}\theta^{x_i}(1-\theta)^{1-x_i}$$

직접 우도함수 $f_n(\vec{x}|\theta)$를 최대화하는 것 대신, 이를 앞선 예제와 마찬가지로 최대화하기 쉽도록 로그를 취해준다.

$$\begin{aligned} L(\theta)=\log f_n(\vec{x}|\theta)&=\sum_{i=1}^{n}\left[x_i\log\theta+(1-x_i)\log(1-\theta)\right] \\ &=\left(\sum_{i=1}^{n}x_i\right)\log\theta+\left(n-\sum_{i=1}^{n}x_i\right)\log(1-\theta). \end{aligned}$$

이제 앞선 예제와 마찬가지로 일계도함수 판정을 수행한다. 만약 $\Sigma_{i=1}^{n}x_i\not\in\{0,n\}$ 이면, 일계도함수가 0이 되는 값은 $\theta=\overline{x}_n$이고, 이는 이계도함수 판전에 의하여 결국 $L(\theta)$과 우도함수를 최대화하는 값이 된다. 만약 $\Sigma_{i=1}^{n}x_i=0$ 이면 모든 θ에 대하여 $L(\theta)$가 감소함수가 되고, 따라서 $L(\theta)$가 $\theta=0$에서 최대가 된다. 마찬가지로 $\Sigma_{i=1}^{n}x_i=n$인 경우도 증가함수가 되므로 $\theta=1$에서 최대가 된다. 즉, $\Sigma_{i=1}^{n}x_i\in\{0,n\}$인 경우 모두 우도함수의 최대값이 $\theta=\overline{x}_n$에서 일어난다. 그러므로 θ의 M.L.E 는 $\hat{\theta}=\overline{X}_n$로 구해진다.

▲

Example 7.5.5) Sampling from a Normal Distribution with Unknown Mean.
확률변수 $X_1,\ldots,X_n$이 알려지지 않은 평균 μ와 알려진 분산 σ^2을 갖는 정규분포로부터 확률표본을 생성한다고 하자. 모든 관측값 $x_1,\ldots,x_n$에 대하여, 우도함수 $f_n(\vec{x}|\mu)$는 다음과 같이 구해진다.

$$f_n(\vec{x}|\mu)=\frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right]$$

이는 다음의 식을 최소화하는 μ값을 찾으면 우도함수 $f_n(\vec{x}|\mu)$를 최대화할 수 있게 된다.

$$Q(\mu)=\sum_{i=1}^{n}(x_i-\mu)^2=\sum_{i=1}^{n}x_i^2-2\mu\sum_{i=1}^{n}x_i+n\mu^2$$

이제 일계도, 이계도함수 판정을 이용하여 방정식 Q가 최소가 되게 하는 μ을 구하면 $\mu=\overline{x}_n$ 임을 얻고, 따라서 μ에 관한 M.L.E는 $\hat{\mu}=\overline{X}_n$으로 구해진다.

▲

Example 7.6.11) Sampling from an Exponential Distribution.

$X_1, X_2, \ldots$가 모수 θ에 관한 지수분포를 $i.i.d.$하게 갖는다고 가정하자. $T_n = \sum_{i=1}^{n} X_i$ 라 두면, θ의 M.L.E.는 $\hat{\theta}_n = \frac{n}{T_n}$으로 구할 수 있다(7.5절의 연습문제 7 참고).

> *7.5절 연습문제 7번*
>
> $X_1, \cdots, X_n$ 에 대하여 X_i : $i.i.d.$ & $X_i \sim \mathrm{Exp}(\beta)$일 때, β의 최대우도추정량을 구하여라.
>
> solution) X_i에 대한 우도함수는 $f_n(\vec{x}|\beta) = \beta^n e^{-\beta y}$로 구해지고, 이를 최대화하기 위해 양변에 로그를 취해준 후 다음과 같이 정의한 후 최대우도 추정량을 구한다. 즉, $L(\beta) = \log(f_n(\vec{x}|\beta)) = n\log\beta - \beta y$. 이 함수의 일계도함수를 구하면 $L'(\beta) = \frac{n}{\beta} - y = 0 \Rightarrow \beta = \frac{n}{y}$. 이제 이계도함수 판정으로 구한 임계점에서 극대값이 되는지 확인해보면 $L''(\beta) = -\frac{n}{\beta^2} < 0$ since $\beta > 0, n > 0 \Rightarrow \beta$: maximum. $\therefore\ \beta = \frac{n}{y}$.

이는 $1/\hat{\theta}_n$이 유한 분산을 갖는 $i.i.d.$한 확률변수들의 평균이므로, 중심극한정리(the central limit theorem)에 의하면 $1/\hat{\theta}_n$의 분포는 정규분포에 가까워진다. 이 경우에 평균과 분산은, 근사 정규분포의 모수로써 각각 $1/\theta$, $1/(n\theta^2)$이 된다. 앞선 설명에서의 표기로, $V_n(\theta) = \theta^2$가 된다.

다음으로, θ에 관한 사전분포를 모수 α, β를 갖는 감마분포라고 하자. 정리 7.3.4에 의하면 θ의 사후분포는 $\alpha+n, \beta+t_n$인 모수를 갖는 감마분포가 된다. 우리는 이러한 감마분포가 근사적으로 정규분포가 된다는 것을 보임으로써 결론내릴 수 있다. 간단히 하기 위해, α가 정수라고 가정하자. 그러면 θ의 사후분포는 모수 $\beta+t_n$을 갖는 $i.i.d.$한 지수확률변수 $\alpha+n$의 합에 대한 분포와 같아진다. 그러한 합은 근사적으로 평균 $(\alpha+n)/(\beta+t_n)$, 분산 $(\alpha+n)/(\beta+t_n)^2$인 정규분포를 갖는다. 만약 α와 β 모두 작으면, 근사적 평균은 거의 $n/t_n = \hat{\theta}$이고, 근사적 분산은 거의 $n/t_n^2 = \hat{\theta}^2/n = V_n(\hat{\theta})/n$이 된다.

◀

Example 7.6.12) Sampling from an Exponential Distribution.

예제 7.3.14에서, 우리는 280개의 관측값들을 표본으로 하는, 프로이센군 병사들이 말의 발길질로 사망한(연간) 평균 사망횟수 θ의 사후분포를 구하였다. 사후분포는 모수가 196과 280을 갖는 감마분포였음을 상기하자. 예제 7.6.11의 논의를 참고하면, 이러한 감마분포는 근사적으로 모수 280을 갖는 $i.i.d.$인 196개의 지수확률변수들의 합의 분포가 된다. 그러한 합의 분포는 근사적으로 정규분포가 되며 평균을 $196/280$, 분산을 $196/280^2$으로 갖는다.

예제 7.3.14와 같은 데이터를 이용하면, 우리는 280개의 관측값들의 평균인 θ에 관한 M.L.E.를 찾을 수 있다. 중심극한정리에 의하면, 평균 θ를 갖는 280개의 $i.i.d.$한 포아송 확률변수들의 평균에 대한 분포는 근사적으로 평균 θ와 분산 $\theta/280$을 갖는 정규분포에 근사하게 된다. 따라서, 이전의 notation에 의하면 $V_n(\theta)=\theta$으로 갖는다. 그러면 앞선 관측 데이터들의 최대우도추정치(The maximum likelihood estimate)는 사후분포의 평균으로써 $\hat{\theta}=196/280$으로 갖는다. 또한 사후분포의 분산은 $V_n(\hat{\theta})/n=\hat{\theta}/280$이 된다.

◀

note) 다음의 일반적인 두 상황에서는 사후분포와 MLE의 분포가 정규분포로 근사하지 않는 상황이 발생한다.
(1) 표본의 크기가 충분히 크지 않은 경우, (2) 우도함수가 smooth(미분가능한)하지 않은 경우. 다음에 소개할 예제들은 이러한 상황에서의 예제이다.

(1) 표본의 크기가 충분히 크지 않은 경우.

Example 7.6.13) Lifetimes of Electronic Components.

예제 7.3.12에서, 우리는 표본크기가 $n=3$인 모수를 θ로 갖는 지수확률변수들을 갖고 있었다. 그리고 사후분포는 모수를 4와 8.6으로 갖는 감마분포였음을 상기하자. 그러면 M.L.E.는 $\hat{\theta}=3/(X_1+X_2+X_3)$로, 1/모수를 3과 3θ로 갖는 감마확률변수 의 분포를 갖는다. Figure 7.8은 이러한 사후 p.d.f.와 M.L.E. p.d.f.를 보여준다. M.L.E.의 관측값인 θ를, $\theta=3/6.6$으로 가정하자. 그러면 두 p.d.f.의 분포는 비슷할지라도, 조금 다르다. 또한 두 p.d.f.는 서로 비슷하지만, 여전히 사후분포의 표본평균과 분산을 갖는 정규 p.d.f.와 다르다.

◀

(2) 우도함수가 smooth 하지 않은 경우.

Example 7.6.14) Sampling from a Uniform Distribution.

예제 7.5.7에서, 우리는 구간 $[0,\theta]$에서의 균등분포로부터 표본크기가 n인 θ에 대한 M.L.E.를 구하였다. 그러한 M.L.E.는 $\hat{\theta}=\max\{X_1,\dots,X_n\}$으로 구하였다. 우리는 예제 3.9.6.의 결과를 이용하여 정확한 $\hat{\theta}$의 분포를 찾을 수 있다. 여기서 $Y=\hat{\theta}$의 p.d.f.는 다음과 같다.

$$g_n(y|\theta)=n[F(y|\theta)]^{n-1}f(y|\theta),$$

where $f(\bullet\,|\theta)$: p.d.f. of Uniform dist, $F(\bullet\,|\theta)$: c.d.f. of Uniform dist. (on $[0,\theta]$.)

위의 p.d.f.는 정리하면 다음과 같이 쓸 수 있다.

$$g_n(y|\theta)=n\left[\frac{y}{\theta}\right]^{n-1}\frac{1}{\theta}=n\frac{y^{n-1}}{\theta^n}\quad\text{for } 0<y<\theta$$

이러한 p.d.f.는 normal p.d.f.와 조금도 비슷하지 않다. 이는 매우 비대칭적이고 가장 큰 가능한 M.L.E값에서 최대값을 갖는다. 사실, $\hat{\theta}$에 대한 평균과 분산을 구하면 각각 다음과 같다.

$$E(\hat{\theta})=\frac{n}{n+1}\theta\ ,\ \ Var(\hat{\theta})=\frac{n}{(n+1)^2(n+2)}\theta^2$$

이 분산은 우리가 이전에 보던 근사적으로 정규적인 예제에서의 $1/n$ 대신에 $1/n^2$에 점점 가까워진다.
만약 n이 크면, θ에 대한 사후분포는 근사적으로 우도함수와 상수곱의 형태를 갖는 p.d.f.를 갖는다. 그러한 함수에 분모를 θ로 취하여 적분하면, 요구되는 상수는 다음과 같은 θ의 근사 사후 p.d.f.로 유도된다.

$$\xi(\theta|\overrightarrow{x}) \approx \begin{cases} \dfrac{(n-1)\hat{\theta}^{n-1}}{\theta^n} & \text{for } \theta > \hat{\theta} \\ 0 & otherwise \end{cases}$$

이러한 근사 사후분포에서의 평균과 분산은 각각 $(n-1)\hat{\theta}/(n-2)$와 $(n-1)\hat{\theta}^2/\left[(n-2)^2(n-3)\right]$으로 구해진다. 사후평균은 여전히 거의 M.L.E.와 같고, 사후 분산은 M.L.E.의 분산인 $1/n^2$의 비율로 감소한다. 그러나 사후분포는 p.d.f.가 가장 작은 가능한 θ값에 대해 최대값을 가짐으로써 조금도 정규적이지 않다.

◀

Example 8.1.2) Sampling Distribution of the M.L.E. of the Mean of a Normal Distribution.

$X_1, \cdots, X_n$이 평균 μ와 분산 σ^2를 갖는 정규분포로부터 확률 표본을 생성한다고 가정하자. 이전 예제에서 확인했듯이, 표본평균 $\overline{X}_n$이 μ에 대한 최대우도추정량이었다. 더욱이, 이는 따름정리 5.6.2.의 $\overline{X}_n$의 분포가 평균 μ와 분산 σ^2/n을 갖는 정규분포임을 찾을 수 있다.

Corollary 5.6.2) 확률변수 $X_1, \cdots, X_n$이 평균 μ와 분산 σ^2를 갖는 정규분포로부터 확률 표본을 생성하고 $\overline{X}_n$가 이러한 확률변수들의 표본평균이라 가정하자. 그러면 $\overline{X}_n$는 평균이 μ이고 분산이 σ^2/n인 정규분포이다.

◀

Example 8.1.5) Lifetimes of Electronic Components.

통계학자가 예제 8.1.4의 사후평균 대신에 θ에 대한 최대우도추정량 $\hat{\theta} = 3/T$ 을 채택했다고 가정하자. (이 때, $T = \sum_{i=1}^{3} X_i$ 이었다.) $\Pr(|\hat{\theta} - \theta| < 0.1 | \theta)$를 계산하는 것 대신에, (이전 예제에서 고려한 구하고자하는 확률) 이 학자는 다음을 계산하고자 할 것이다.

$$\Pr\left(\left|\frac{\hat{\theta}}{\theta} - 1\right| < 0.1 | \theta\right)$$

이는 $\hat{\theta}$가 θ의 10% 이내일 확률이다. 이 확률은 최대우도추정량의 표본분포로부터 계산될 수 있고, 또는 $\hat{\theta}/\theta = 3/(\theta T)$이고 θT의 분포가 모수 3과 1을 갖는 감마분포임을 알 수 있다. 이런 이유로, $\hat{\theta}/\theta$는 θ에 의존하지 않는 분포를 갖는다. 따라서 $\Pr(|\hat{\theta} - \theta| < 0.1 | \theta)$가 모든 θ에 대해 같은 수를 가진다.

θT의 누적분포함수는 $G(\cdot | 1)$로 구해지고, 이런 이유로 본래 구하고자 하는 확률은 다음과 같다.

$$\begin{aligned}
\Pr\left(\left|\frac{\hat{\theta}}{\theta} - 1\right| < 0.1 | \theta\right) &= \Pr\left(\left|\frac{3}{\theta T} - 1\right| < 0.1 | \theta\right) \\
&= \Pr\left(0.9 < \frac{3}{\theta T} < 1.1 | \theta\right) \\
&= \Pr(2.73 < \theta T < 3.33 | \theta) \\
&= G(3.33 | 1) - G(2.73 | 1) = 0.134
\end{aligned}$$

결과적으로, 통계학자는 이제 θ가 무엇인지 상관없이 θ의 최대우도추정량 $\hat{\theta}$가 θ의 10% 이내일 확률이 0.134라고 주장할 수 있다.

◀

§. Additional Examples in Sampling Distributions

Example 8.1.1) A Clinical Trial
예제 2.1.4.에서 소개된 Clinical Trial을 다시 생각한다. θ를 모든 가능한 imipramine 복용 환자 중 재발하지 않은(no relapse) 환자의 비율이라 하자. 우리는 θ를 추정하기 위해 imipramine 복용 그룹의 재발(relapse)환자를 제외한 환자들의 관측된 비율을 사용하고자 한다.

데이터를 관측하기에 앞서, 표본추출된 재발하지 않은 환자들의 비율이 모수 θ와 정확히 같진 않은 분포를 가지는 확률변수 T라 하면, 우리는 이러한 T가 높은 확률로 모수 θ와 가까워지는 것을 원한다. 예를 들면 $|T-\theta|<0.1$인 확률의 계산을 시도한다고 하자. 그러면 이 식은 확률변수 T의 분포를 알고 있어야 계산할 수 있다.

Clinical trial에서 우리는 imipramine 복용 그룹의 40명의 환자들의 응답을 모수 θ를 갖는 조건부로 $i.i.d.$인 베르누이 확률변수를 갖도록 모형화 하였다. θ가 주어질 때 $40T$의 조건부 분포는 모수를 40과 θ로 갖는 이항분포이다. T의 분포는 이러한 이항분포로부터 쉽게 유도될 수 있다. 따라서 T는 다음과 같이 유도되는 주어진 θ에 대한 t의 확률함수를 갖는다.

$$f(t|\theta)=\binom{40}{40t}\theta^{40t}(1-\theta)^{40(1-t)} \text{ , for } t=0,\frac{1}{40},\frac{2}{40},\dots,\frac{39}{40},1, \text{ and } f(t|\theta)=0 \text{ otherwise.}$$

◀

note) 예제 8.1.1.에서의 T의 분포를 통계량 T에 대한 표본분포(sampling distribution)이라 한다. 표본분포를 알면 이를 이용하여 데이터를 관측하기 전에 모수 θ에 대해 T가 얼마나 가까이 근접하는가에 대한 예측을 할 수 있다.
또한 이러한 T 의 표본분포는 T를 관찰하는 것으로 모수 θ에 대해 좋은 정보들을 얻을 수 있는데, 예를 들면 두 개의 다른 통계량을 추정량을 사용하여 결정하고자 시도할 때, 그들의 표본분포는 그 둘을 비교하기에 유용하게 사용될 수 있다.

Definition 8.1.1) Sampling Distribution
확률변수 $\overrightarrow{X}=(X_1,\dots,X_n)$이 알려지지 않은 모수 θ와 관련된 분포의 확률 표본을 생성한다고 가정하자.
T를 $\overrightarrow{X}$와 θ에 대한 함수, 즉 $T=r(X_1,\dots,X_n,\theta)$라고 하자. 그러면 θ가 주어졌을 때 T의 분포는 T에 대한 표본분포(Sampling Distribution)이라 한다. T에 대한 평균(기댓값)은 $E_\theta(T)$로 표기한다.

Example8.1.2) Sampling Distribution of the M.L.E. of the Mean of a Normal Distribution.
$X_1,\dots,X_n$을 평균이 μ이고 분산이 σ^2인 정규분포로부터 취해지는 확률 표본을 생성한다고 가정하자. 예제 7.5.5.와 7.5.6.에서, 표본평균 $\overline{X}_n$은 μ의 M.L.E.로 구해졌다. 또한, 따름정리 5.6.2.에서 이러한 표본평균의 분포는 평균이 μ이고 분산이 σ^2/n인 정규분포를 따른다는 것을 확인했다.

◀

§. Purpose of the Sampling Distribution

Example8.1.3) Lifetimes of Electronic Components.
예제 7.1.1.에서의 전기 부품을 판매 회사를 생각하자. 이 회사는 전기 부품의 수명이 $i.i.d.$이고 θ에 조건부인 모수 θ의 지수적 확률변수를 모형화 하였다. 또한 모형은 θ가 1과 2인 감마분포를 따른다고 하였다. 이제 $n=3$인 수명에 대해 관측하였다고 하고, θ의 사후 평균(Posterior mean(expected value))을 추정량으로 사용할 것이라 가정하자.

정리 7.3.4.에 따르면, θ의 사후분포는 모수를 $1+3=4$와 $2+\sum_{i=1}^{3} X_i$를 갖는 감마분포로 구해지고, 따라서 사후평균은 $\hat{\theta}=4/(2+\Sigma_{i=1}^{3} X_i)$로 구해진다.

이 세 개의 수명에 대해 관측하기에 앞서, 회사는 $\hat{\theta}$가 θ에 가까운 값을 가지도록 할 방법을 알고 싶어한다. 예를 들어, $\Pr(\hat{\theta}-\theta<0.1)$을 계산하고 싶어한다. 또한 별개로, 소비자들 또한 $\hat{\theta}$가 θ에 얼마나 가까워지는지에 관심이 있다. 하지만 회사와 소비자 서로 θ에 대한 같은 사전 분포를 선택하지 않길 원한다. 또한 몇몇은 사전분포가 사용되지 않기를 원한다. 우리는 곧 이러한 모든 사람들이 결국 $\hat{\theta}$의 표본분포를 결정하는 것이 유용하다는 것을 확인하게 될 것이다. 그들이 하는 일들은 서로 다르지만, 모두 표본분포를 이용할 수 있다.

◀

note) 위의 예제에서, 3개의 수명을 관측하고 나면, 회사는 θ의 사후분포에 대해 관심을 두게 된다. 그러면 그들은 사후확률 $|\hat{\theta}-\theta|<0.1$을 계산하려고 할 것이다. 그러나 표본을 취하기 전에, $\hat{\theta}$와 θ는 random이고 $\Pr(|\hat{\theta}-\theta|<0.1)$이 $\hat{\theta}$와 θ의 결합분포이므로, 표본분포는 단지 주어진 θ에 대한 $\hat{\theta}$의 조건부 분포가 된다. 이런 이유로, 전확률의 법칙을 적용하면 다음과 같다.

$$\Pr(|\hat{\theta}-\theta|<0.1)=E[\Pr(|\hat{\theta}-\theta|<0.1|\theta)]$$

이 경우에, 회사는 $\Pr(|\hat{\theta}-\theta|<0.1)$를 계산하기 위한 중간 계산으로써 $\hat{\theta}$의 표본분포를 이용할 수 있다.

Example8.1.4) Lifetimes of Electronic Components
앞선 예제 8.1.3.에서, $\hat{\theta}$에 대한 표본분포는 어떤 특정한 이름을 갖는 것은 아니지만, $\hat{\theta}$이 모수를 3과 θ로 갖는 감마분포를 따르는 통계량 $T=\sum_{i=1}^{3} X_i$ 에 대한 단조함수라는 것을 쉽게 알 수 있다. 그래서, 우리는 T에 대한 분포의 누적분포함수 $G(\bullet\,|\theta)$로부터 $\hat{\theta}$의 표본분포에서 누적분포함수 $F(\bullet\,|\theta)$를 계산할 수 있게 된다. 논의는 다음과 같다.

$$\begin{aligned}
\text{For } t>0,\ F(t|\theta) &= \Pr(\hat{\theta}\le t|\theta \\
&= \Pr\left(\frac{4}{2+T}\le t\,|\,\theta\right) \\
&= \Pr\left(T\ge \frac{4}{t}-2\,|\,\theta\right) \\
&= 1-G\left(\frac{4}{t}-2\,|\,\theta\right) \\
\text{For } t\le 0,\ F(t|\theta) &= 0.
\end{aligned}$$

통계 패키지를 이용하여 함수 감마분포의 누적분포함수인 G를 계산한다. 이제 회사는 F를 계산할 수 있으므로,
임의의 θ에 대하여 다음과 같이 구할 수 있다.

$$\Pr(|\hat{\theta}-\theta|<0.1\,|\,\theta)=F(\theta+0.1\,|\,\theta)-F(\theta-0.1\,|\,\theta)$$

◀

Example 8.2.3) Acid Concentration in Cheese.

어떤 연구에서 맛과 치즈의 화학적 합성 사이의 관계에 대한 연구를 진행하였다. 농도가 맛에 영향을 줄 수 있는 화학물질 중 하나는 젖산으로 알려져 있다. 단골고객의 기반을 구축하고자 하는 치즈 제조업체는 고객이 치즈를 구매할 때마다 거의 동일한 맛을 느끼길 원한다. 젖산 같은 화학물질의 농도의 변화는 치즈 맛의 변화를 가져올 수 있다.

우리는 몇개의 치즈덩어리 속의 젖산의 농도를 평균 μ와 분산 σ^2를 갖는 독립적인 정규확률변수로 모형화 하였다고 가정하자. 우린 얼마나 이러한 농도가 평균 μ와 다른지에 관심이 있다. 이제 $X_1,\cdots,X_k$를 k개의 덩어리 속의 농도에 대한 확률변수라 하고, $Z_i=(X_i-\mu)/\sigma$ 라 하자. 그러면 얼마나 많은 k개의 농도가 μ와 다른지에 대한 하나의 척도로 다음을 제시할 수 있다.

$$Y=\frac{1}{k}\sum_{i=1}^{k}|X_i-\mu|^2=\frac{\sigma^2}{k}\sum_{i=1}^{k}Z_i^2$$

u 만큼의 차이, 또는 더 많이 포함된 젖산의 농도 u는 맛의 확연한 차이를 야기하기에 충분하다고 가정하자.
우린 $\Pr(Y\le u^2)$를 계산하길 원한다. 따름정리 8.2.1.에 따르면, $W=kY/\sigma^2$의 분포는 자유도가 k인 χ^2분포를 따른다. 이런 이유로, $\Pr(Y\le u^2)=\Pr(W\le ku^2/\sigma^2)$가 된다.

Corollary 8.2.1) 확률변수 $X_1,\cdots,X_m$이 $i.i.d.$인 표준정규분포를 따른다고 하면, 이들의 제곱합 $X_1^2+\cdots+X_m^2$ 은 자유도가 m인 χ^2-분포를 따른다.

예를 들면, $\sigma^2=0.09$이고, 우리는 $k=10$개의 치즈 덩어리에만 관심이 있다고 가정하자. 또한, 이러한 관심에서 $u=0.3$은 중요한 차이라고 가정하자. 그러면 다음과 같이 쓸 수 있다.

$$\Pr(Y\le 0.3^2)=\Pr\left(W\le\frac{10\times 0.09}{0.09}\right)=\Pr(W\le 10).$$

자유도가 10인 카이제곱분포표를 이용하면, 우리는 10이 0.5와 0.6 분계선 사이에 위치함을 볼 수 있다.
사실, 위에서 제시된 확률은 컴퓨터 소프트웨어를 이용하여 찾으면 0.56으로 계산된다. 따라서 젖산 농도와 10개의 치즈덩어리의 평균농도 사이의 평균제곱차가 원하는 양보다 많을 확률은 44%가 된다. 만약 이 확률이 너무 크다면, 제조업테는 젖산 농도의 변화가 줄어들게 하기 위한 노력을 지속하길 바랄 것이다.

◀

Exercise 13) Prove that the distribution of $\widehat{\sigma^2}$ in Example 8.2.1 and 8.2.2 is the gamma distribution with parameters $n/2$ and $n/(2\sigma^2)$.

solution)

> *Example 8.2.1. M.L.E. of the Variance of a Normal Distribution*
> $X_1,\cdots,X_n$이 알려진 평균 μ와 알려지지 않은 분산 σ^2를 갖는 정규분포로부터 확률 표본을 생성한다고 가정하자. 이러한 분산 σ^2의 최대우도추정량은 7.5절 연습문제 6번에서 다음과 같이 찾을 수 있다.
> $$\widehat{\sigma_0^2}=\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$$
> 이 최대우도추정량의 분포와 $\widehat{\sigma_0^2}/\sigma^2$의 분포는 몇몇 통계적 문제에 대해서 유용하다.

> *Example 8.2.2. M.L.E. of the Variance of a Normal Distribution*
> 예제 8.2.1에서, 확률변수 $Z_i=(X_i-\mu)/\sigma$ for $i=1,\cdots,n$이 표준정규분포로부터 확률 표본을 생성한다고 가정하자. 이는 따름정리 8.2.1로부터 $\sum_{i=1}^{n}Z_i$의 분포가 n의 자유도를 갖는 χ^2-분포를 가지게 됨을 알 수 있다. 이는 예제 8.2.1에서 $\sum_{i=1}^{n}Z_i$가 $n\widehat{\sigma_0^2}/\sigma^2$와 정확하게 같은 형태임을 쉽게 확인할 수 있다. 따라서 $n\widehat{\sigma_0^2}/\sigma^2$는 마찬가지로 n의 자유도를 갖는 χ^2-분포를 갖게 된다. 여기서 또한 $\widehat{\sigma_0^2}$ 는 그 자체로 모수 $n/2$와 $n/(2\sigma^2)$를 갖는 감마분포임을 확인할 수 있다. (연습문제 13번)

예제에서 이미 $W=n\widehat{\sigma^2}/\sigma^2$ 가 n의 자유도를 갖는 카이제곱분포임을 확인했고, 또한 카이제곱분포는 감마분포의 특수한 경우이므로(모수를 $\alpha=m/2$, $\beta=1/2$로 갖는), 이 경우에는 모수를 $n/2$와 $1/2$로 갖는 감마분포가 된다. 식을 대수적으로 조작해주면, $\widehat{\sigma^2}=(\sigma^2/n)W$ 임을 얻고, 결국 $\widehat{\sigma^2}$의 분포는 모수가 $n/2$와 $n/(2\sigma^2)$인 감마분포를 따르게 된다.

◀

Example 8.3.2) Rain from Seeded Clouds.

> Theorem 8.3.1)
> 확률변수 $X_1,\cdots,X_m$이 평균 μ와 분산 σ^2를 갖는 정규분포로부터 확률 표본을 생성한다고 가정하자. 그러면 표본평균 $\overline{X}_n$과 표본분산 $(1/n)\Sigma_{i=1}^{n}(X_i-\overline{X}_n)^2$는 독립적인 확률변수로, $\overline{X}_n$는 평균이 μ, 분산이 σ^2/n인 정규분포를 가지며, $\Sigma_{i=1}^{n}(X_i-\overline{X}_n)^2/\sigma^2$는 $n-1$ 자유도를 갖는 χ^2-분포를 갖는다.

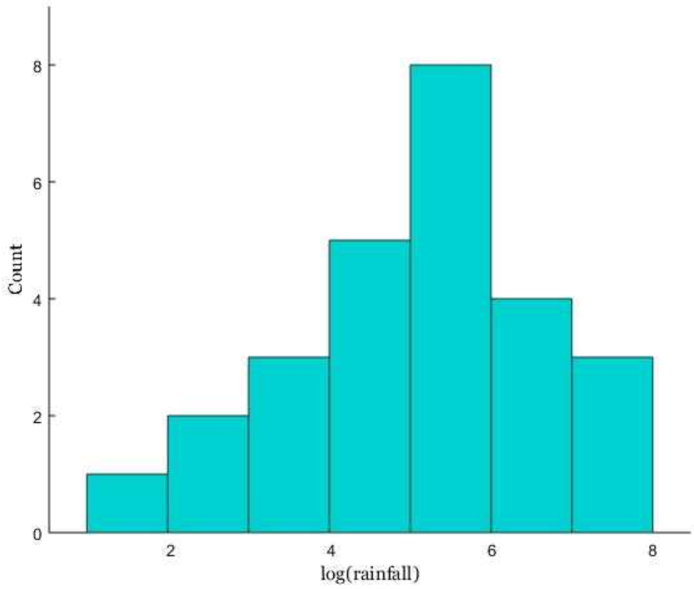
Figure 8.3은 구름 낀 날로부터 강우량에 대한 로그 히스토그램을 보여준다. 이러한 로그값 $X_1,\cdots,X_{26}$가 $i.i.d.$인 정규확률변수로써 모수를 평균 μ와 분산 σ^2로 갖는다고 가정하자. 만약 우리가 구름 낀 날과 강우량 사이에 얼만큼의 변화가 있는지에 관심이 있다면, 우리는 표본분산 $\widehat{\sigma^2}=\sum_{i=1}^{26}(X_i-\overline{X}_n)^2/26$ 를 계산할 수 있다.

$U=26\widehat{\sigma^2}/\sigma^2$의 분포는 25 자유도를 갖는 카이제곱분포를 따른다. 우리는 이 분포를 사용하여 얼마나 $\widehat{\sigma^2}$가 σ^2를 과추정 또는 저추정하게 되는지 알아볼 수 있다. 예를 들면, χ^2 분포표에서 카이제곱분포의 자유도 25인 0.25분위수는 19.94이고, 따라서 $\Pr(U\le 19.94)=0.25$ 이다. 이는 다음과 같이 할 수 있다.

$$0.25=\Pr\left(\frac{\widehat{\sigma^2}}{\sigma^2}\le\frac{19.94}{26}\right)=\Pr(\widehat{\sigma^2}\le 0.77\sigma^2)$$

이는, $\widehat{\sigma^2}$가 σ^2를 23%이상 저추정할 확률이 0.25라는 것을 의미한다. 관측 값 $\widehat{\sigma^2}$는 이 예제에서 2.460 으로 구하였다. 그러면 바로 위에 계산된 확률은 σ^2에서 2.460이 얼마나 먼지에 대해 말할 수 있는 게 없다. 위 확률(데이터를 관측하기 전애)은 우리에게 $\widehat{\sigma^2}$이 σ^2보다 적어도 23% 아래의 값일 확률만 말해줄 수 있다.

Figure 8.3. Histogram of log-rainfalls from seeded clouds.



Example 8.4.3) Rain from Seeded Clouds.

Example 8.4.1. Rain from Seeded Clouds.
예제 8.3.2의 강우량에 대한 문제를 생각하자. 강우량에 관련한 관측치들의 표본평균이 평균 μ로부터 얼마나 먼지에 대하여 관심이 있다고 가정하자. 우리는 $n^{1/2}(\overline{X}_n-\mu)/\sigma$가 표준정규분포를 따름을 알고 있다. 하지만 여기서 σ는 알지 못한다. 만약 우리가 σ대신 M.L.E. 또는 비슷한 방법으로 구한 추정량 $\hat{\sigma}$로 대체한다면, $n^{1/2}(\overline{X}_n-\mu)/\hat{\sigma}$의 분포는 어떨 것이며 μ에 대한 추론을 하기 위한 이 확률변수의 사용을 어떻게 할 수 있을 것인가?

위 예제에 대한 답을 주기 위하여, 다음 정의와 정리를 숙지하도록 한다.

Definition 8.4.1) t Distribution
(student) t 분포는 다음 확률변수의 분포로 정의된다.

$$X=\frac{Z}{\sqrt{V/v}}$$

이 때, Z : 표준정규분포, V : 자유도가 v인 카이제곱분포이다.

* p.d.f. of X : $\dfrac{\Gamma\left(\frac{v+1}{2}\right)}{(v\pi)^{1/2}\Gamma\left(\frac{v}{2}\right)}\left(1+\dfrac{x^2}{v}\right)^{-(v+1)/2}$ for $-\infty<x<\infty$

* Mean / Variance : $E(X)=0$ if $v>0$ / $Var(X)=\dfrac{v}{v-2}$ $(v>2)$, $Var(X)=\infty$ $(1<v\le 2)$.

Theorem 8.4.2)
$X_1,\cdots,X_n$가 평균이 μ이고 분산이 σ^2인 정규분포로부터 확률 표본을 생성한다고 가정하자. 이 때, $\overline{X_n}$은 표본평균으로, σ'는 다음과 같이 정의한다.

$$\sigma'=\left[\frac{\sum_{i=1}^{n}(X_i-\overline{X_n})^2}{n-1}\right]^{1/2}$$

그러면, $n^{1/2}(\overline{X_n}-\mu)/\sigma'$는 자유도가 $n-1$인 $t-$분포를 갖는다.

Example 8.4.2. Rain from Seeded Clouds.
예제 8.4.1로 돌아가보자. 우린 이미 $Z=n^{1/2}(\overline{X}_n-\mu)/\sigma$가 표준정규분포임을 확인했다. 또한 정리 8.3.1에 따르면, $\overline{X}_n$은 $Y=n\widehat{\sigma^2}/\sigma^2$와 독립적이고, 그러므로 Z도 마찬가지로 독립적이며 이는 $n-1$ 자유도를 갖는 카이제곱분포를 따른다. 그러므로 $Z/(Y/[n-1])^{1/2}$는 $n-1$ 자유도인 t 분포를 따르게 된다.

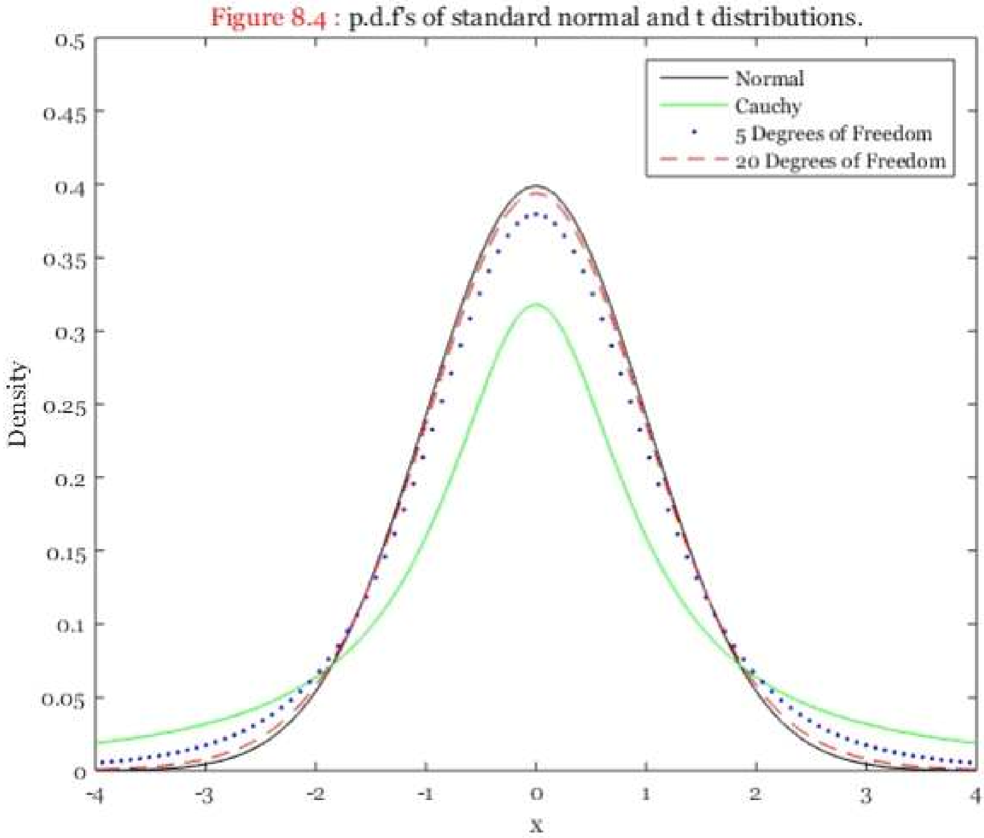
예제 8.4.2로 돌아가보자. 관측값들인 $X_1, \cdots, X_n$ (log-rainfalls) 가 정규분포를 따르고, 각각이 독립적이며 $U = n^{1/2}(\overline{X}_n - \mu)/\sigma'$이 $n-1$ 자유도인 t 분포를 따른다. $n=26$일 때, t 분포표를 참조하면 자유도가 25일 때 0.9 분계선에 위치하는 값은 1.316이다. 따라서 $\Pr(U \le 1.316) = 0.9$이다. 이는 곧 다음을 유도할 수 있다.

$$\Pr(\overline{X}_n \le \mu + 0.2581\sigma') = 0.9 \qquad (\because\ 1.316/(26)^{1/2} = 0.2581)$$

즉, $\overline{X}_n$이 μ 위로 $0.2581\sigma'$ 더 크지 않을 확률이 0.9임을 의미한다. 물론, σ'는 $\overline{X}_n$과 마찬가지로 확률변수이고, 이 결과는 우리가 원하는 것에는 크게 도움이 될 만한 정보는 아니다. 8.5절과 8.6절에서 우리는 알려지지 않은 평균 μ에 대한 일반적인 추론을 하기 위해 t 분포를 어떻게 쓸 것인지를 알아보게 될 것이다.

◀

MATLAB CODE IN SECTION 8.4.




```
clear all
clc

sample_size = 6000;
rv=randn(sample_size,1); % Observations

[y,x]=hist(rv,100); % Using Histogram to find p.d.f.
% y : size of the samples,
% x : intervals (each size of the interval are equal)

figure();
plot(x,pdf('norm',x,0,1),'k-'); % Standard Normal Distribution
hold on
plot(x,pdf('T',x,1),'g-'); % t - Distribution degree 1 (Cauchy)
plot(x,pdf('T',x,5),'b.'); % t - Distribution degree 5
plot(x,pdf('T',x,20),'r--'); % t - Distribution degree 20
xlim([-4 4]);
ylim([0 0.5]);
xlabel('x');
ylabel('Density');
legend('Normal','Cauchy','5 Degrees of Freedom','20 Degrees of Freedom');
title(['\color{red}Figure 8.4 :' '\color{black} pdfs of standard normal and t
distributions.']);
```

Example 8.5.4) Acid Concentration in Cheese.

이 예제에서 다룰 내용들을 이해하기 위해, 다음 정리와 정리들을 숙지하도록 한다.

> Definition 8.5.1) Confidence Interval.
>
> 확률변수 벡터 $\overrightarrow{X}=(X_1,\cdots,X_n)$가 모수 또는 모수벡터 θ에 의존하는 분포로부터 표본을 생성한다고 하자. 그리고 $g(\theta)$를 θ에 대한 실함수라고 하고, $A \le B$ 가 두 개의 통계량(statistic)으로써 모든 θ에 대하여 다음 성질을 갖는다고 하자.
>
> $$\Pr(A < g(\theta) < B) \ge \gamma$$
>
> 그러면, 확률구간 (A, B)를 $g(\theta)$에 대한 계수 γ의 **신뢰구간** 또는 $g(\theta)$에 대한 $100\gamma\%$ **신뢰구간**이라 한다.
> 만약 부등식이 모든 θ에 대하여 등식이 된다면, 신뢰구간이 **정확**(exact)하다고 한다.

> Theorem 8.5.1) Confidence Interval for the Mean of a Normal Distribution.
>
> $X_1,\cdots,X_n$가 평균이 μ이고 분산이 σ^2인 정규분포로부터 확률 표본을 생성한다고 가정하자. $0<\gamma<1$인 각 γ에 대하여, 다음 끝점(endpoint)을 갖는 구간 (A, B)은 정확한 μ에 대한 계수 γ의 신뢰구간이다.
>
> $$A=\overline{X}_n - T_{n-1}^{-1}\left(\frac{1+\gamma}{2}\right)\frac{\sigma'}{n^{1/2}}$$
>
> $$B=\overline{X}_n - T_{n-1}^{-1}\left(\frac{1+\gamma}{2}\right)\frac{\sigma'}{n^{1/2}}$$
>
> 이 때, T_{n-1}은 자유도 $n-1$인 t 분포의 누적분포함수(c.d.f)이다.

앞선 예제 8.2.3에서, 우리는 치즈로부터 10개의 젖산 측정값의 확률 표본에 대해 논의했다. 우리는 알려지지 않은 평균 젖산 농도 μ에 대한 90% 신뢰구간을 계산하길 원한다. 우선 90% 신뢰구간을 구해야하므로, 신뢰수준 $\gamma=0.9$이고, $n=10$인 사실을 알고 있다. 이제 다음 식을 이용하도록 하자.

$$\Pr\left(\overline{X_n}-\frac{c\sigma'}{n^{1/2}}<\mu<\overline{X_n}+\frac{c\sigma'}{n^{1/2}}\right)=\gamma$$

이 때, 우리는 c를 구해야하므로, 다음 식을 이용하여 c를 구한다.

$$c=T_{n-1}^{-1}([1+\gamma]/2) \Leftrightarrow c \equiv (1+\gamma)/2 \text{ quantile of the } t \text{ distribution with } n-1 \text{ degrees of freedom.}$$

앞서 구한 값들을 이에 대입하면 $(1+\gamma)/2=0.95$로 구해진다. 이는 분포표를 참고하면 자유도 9인 t분포의 0.95 분계선에 위치하므로 $c=1.833$이 된다. 이제 맨 처음의 확률 식을 계산하여 보면 90% 신뢰구간은 $\overline{X}_n \pm 1.833\sigma'/(10)^{1/2}$ 로 구해진다. 이제 10개의 젖산 농도에 대한 관측값이 다음과 같이 구해졌다고 가정해보자.

0.86, 1.53, 1.57, 1.81, 1.09, 1.29, 1.78, 1.29, 1.58

그러면 이들 10개의 값에 대한 표본평균 $\overline{X}_n=1.379$로 구해지고, $\sigma'=0.3277$로 구해진다. 이제 모든 값을 구하였으므로 최종적인 90% 신뢰구간은 $1.379 \pm 1.833 \times 0.3277/(10)^{1/2} \Rightarrow (1.189, 1.569)$이다.

◀

Exercise 7) In the June 1986 issue of *Consumer Reports*, some data on the calorie content of beef hot dogs is given. Here are the numbers of calories in 20 different hot dog brands :

186, 181, 176, 149, 184, 190, 158, 139, 175, 148,
152, 111, 141, 153, 190, 157, 131, 149, 135, 132.

Assume that these numbers are the observed values from a random sample of twenty independent normal random variables with mean μ and variance σ^2, both unknown.

Find a 90% confidence interval for the mean number of calories μ.

solution)

모수인 μ와 σ^2 모두 알려져 있지 않고, 확률변수는 정규분포를 따르므로, t-분포를 이용하여 신뢰구간을 구하고자 한다. 우선 신뢰수준 $\gamma = 0.9$이고, 표본 크기는 $n = 20$으로 주어졌다. 주어진 20개의 서로 다른 핫도그 브랜드의 칼로리 수에 대한 표본 평균은 $\overline{X}_{20} = 156.85$ 로 구할 수 있다. 이제 다음을 참고하여 신뢰구간을 구해보자.

$$\Pr\left(\overline{X}_n - \frac{c\sigma'}{n^{1/2}} < \mu < \overline{X}_n + \frac{c\sigma'}{n^{1/2}}\right) = \gamma \text{ (모수 } \mu\text{가 양 끝점 구간 안에 들어갈 확률 = 신뢰수준)}$$

우리가 구해야할 것은 c와 σ'이다. $1+\gamma/2 = (1+0.9)/2 = 0.95$이므로, 자유도 19인 t-분포의 분포표의 0.95 분계선에 위치한 값을 찾으면 $c = 1.729$임을 확인할 수 있다. 이제 표본표준편차 σ'을 구하자. σ'는 값을 적절히 잘 대입하면 아래와 같이 구해질 수 있다.

$$\sigma' = \left[\frac{\sum_{i=1}^{20}(X_i - \overline{X_{20}})^2}{20-1}\right]^{1/2} = \sqrt{\frac{[(186-156.85)^2 + \cdots + (132-156.85)^2]}{19}} = \sqrt{512.6605} = 22.64$$

이제 모든 값을 구하였으므로, μ에 대한 90% 신뢰구간은 다음과 같이 구해진다.

$$\mu \in \left(\overline{X}_n - \frac{c\sigma'}{n^{1/2}}, \overline{X}_n + \frac{c\sigma'}{n^{1/2}}\right) = \left(156.85 - \frac{1.729\times 22.64}{20^{1/2}}, 156.85 + \frac{1.729\times 22.64}{20^{1/2}}\right) = (156.85 - 8.753, 156.85 + 8.753).$$

$$\therefore\ (148.097, 165.603)$$

◀