LGE - SNU DS 고급 과정

확률통계 및 공통 알고리즘

과제4 -통계적 추론

2022. 01

※필요하다면 다음의 값을 이용하시오

: $z_{0.995} = 2.58$, $z_{0.977} = 2$, $z_{0.95} = 1.645$, $t_{0.975}(8) = 2.306$, $\chi^2_{0.975}(9) = 19.02$

문제 1. 정규분포 $N(\theta, 1^2)$을 따르는 모집단으로부터 2개의 표본 $0, 2$를 뽑았을 때, θ의 최대가능도 추정량을 구하여라.

[풀이].

$$
\begin{aligned}
l(\theta) &= -\frac{1}{2}\{\theta^2 + (\theta-2)^2\} - \log(2\pi) \\
&= -(\theta-1)^2 - 1 - \log(2\pi)
\end{aligned}
$$

위 식은 최고차항의 계수가 음수이므로 $l(\theta)$는 $\theta = 1$ 일 때 최댓값을 가진다.
따라서 $\widehat{\theta}^{MLE} = 1$

문제 2. 확률변수 X가 $U(\theta - 1/2, \theta + 1/2)$의 확률밀도함수를 따른다고 하자. 세 개의 관찰값 $1.2, 1.5, 1.6$ 를 얻었을 때, θ 의 최대가능도 추정량을 구하여라.

[풀이]. $L(\theta) = \{\frac{1}{(\theta+1/2)-(\theta-1/2)}\}^3 = 1, \quad \theta - 1/2 < 1.2, 1.5, 1.6 < \theta + 1/2$
$\Rightarrow \quad 1.1 = 1.6 - 0.5 < \widehat{\theta} < 1.2 + 1/2 = 1.7$ 를 만족하는 모든 $\widehat{\theta}$ 가 최대 가능도 추정량이다.

문제 3. 어느 고등학교의 남학생들의 신장은 근사적으로 평균이 μ이고, 표준편차가 6인 정규분포를 따른다고 한다. 이제 μ에 대하여 알기 위하여 이 고등학교의 남학생들 중 임의로 9명을 뽑아 신장을 측정한 결과 표본 평균은 174.0cm였다. 이를 통해 μ에 대한 99% 신뢰구간을 구하여라.

[풀이].

$$
\begin{aligned}
&\left[\overline{X} - \frac{\sigma}{\sqrt{n}} \times z_{0.995}, \overline{X} + \frac{\sigma}{\sqrt{n}} \times z_{0.995}\right] \\
&= \left[174.0 - \frac{6}{\sqrt{9}} \times 2.58, 174.0 + \frac{6}{\sqrt{9}} \times 2.58\right] \\
&= [168.84, 179.16]
\end{aligned}
$$

문제 4. 동전을 던졌을 때 앞면이 나오는 확률 θ를 알아보고자 한다. θ의 사전분포는 평균이 0.5인 $Beta(10,10)$을 따른다고 하자. 100개의 동전을 던져 그 중 60개가 앞면이 나왔을 때, 앞면이 나오는 확률의 사후분포와 베이즈 추정량(사후평균)을 구하여라.

[풀이].

$$\begin{aligned} f(\theta|X) &= L(\theta|X)\pi(\theta) \\ &= \frac{n!}{X!(n-X)!}\theta^X(1-\theta)^{n-X}\frac{1}{B(10,10)}\theta^{10-1}(1-\theta)^{10-1} \\ &\propto \theta^{10+X-1}(1-\theta)^{10+n-X-1} \end{aligned}$$

이고 $n=100, X=60$이므로 θ의 사후분포는 $Beta(70,50)$이다. 따라서 θ의 베이즈추정량은 $\frac{70}{70+50}=\frac{7}{12}$이다.

문제 5. 연속형 확률변수 X가 어떠한 분포를 따르는지 판단하고자 귀무가설과 대립가설을 아래와 같이 설정하였다.

$$\begin{aligned} H_0 &: X\text{는 } N(10,4)\text{를 따른다.} \\ H_1 &: X\text{는 } N(16,1)\text{를 따른다.} \end{aligned}$$

$X>14$ 일 때 H_0 를 기각하고, 다른 경우에는 H_0를 채택한다고 할 때,

(a) 제 1종의 오류를 범할 확률을 구하여라.

(b) 제 2종의 오류를 범할 확률을 구하여라.

[풀이].

(a)

$$\begin{aligned} (\text{1종 오류의 확률}) &= P(X>14 \mid H_0\text{이 참 }) \\ &= P(\frac{X-10}{2}>2 \mid H_0\text{이 참 }) \\ &= P(Z>2) \\ &= 0.023 \end{aligned}$$

(b)

$$\begin{aligned} (\text{2종 오류의 확률}) &= P(X\le 14 \mid H_1\text{이 참 }) \\ &= P(\frac{X-16}{1}\le -2 \mid H_1\text{이 참 }) \\ &= P(Z\le -2 \mid H_1\text{이 참 }) \\ &= 0.023 \end{aligned}$$

문제 6. 새로운 포장법에 대한 포장시간 단축효과를 평가하고자 한다. 이를 위해 포장시간의 모평균 μ에 대해

$$H_0 : \mu = 1.5 \quad \text{vs} \quad H_1 : \mu < 1.5$$

을 세우고, 시제품 100개를 생산하여 포장시간을 조사한 결과 평균 포장시간이 1.35분 이었다고 한다. 포장시간은 근사적으로 표준편차가 10분인 정규분포를 따른다고 할 때, 유의수준 5%에서 가설검정을 하여라.

[풀이].

(i) 가설설정 : $H_0 : \mu = 1.5 \quad \text{vs} \quad H_1 : \mu < 1.5$
$\sigma = 10$, $n = 100$, $\alpha = 0.05$

(ii) 검정통계량 : $Z = \dfrac{\overline{X} - 1.5}{10/\sqrt{100}} \sim N(0, 1^2)$ under H_0
$z_0 = \dfrac{1.35 - 1.5}{10/\sqrt{100}} = -0.15$

(iii) 기각역 : $P(Z < z|H_0) = 0.05 \quad \Rightarrow \quad z = z_{0.05} = -z_{0.95} = -1.645 \quad \Rightarrow \quad Z < -1.645$

(iv) 기각역을 통한 비교 : $z_0 \not< -1.645$

(v) 결론 : 유의수준 5%에서 귀무가설을 기각할 수 없다.

문제 7. 건물의 소화용으로 사용되는 살수장치의 작동 시작 온도는 근사적으로 평균이 55도인 정규분포를 따르는 것으로 알려져 있다. 제조공정의 이상여부를 판단하기 위하여 생산품 중에서 9개의 표본을 추출하여 작동 시작 온도를 조사한 결과 표본평균이 55.63도, 표본표준편차가 0.9도였다. 공정의 이상여부를 확인하기 위해 적절한 가설을 세우고 유의수준 5%에서 검정하여라.

[풀이].

(i) 가설설정 : $H_0 : \mu = 55 \quad \text{vs} \quad H_1 : \mu \neq 55$
$\sigma = 0.9$, $n = 9$, $\alpha = 0.05$

(ii) 검정통계량 : $T = \dfrac{\overline{X} - 55}{0.9/\sqrt{9}} \sim t(8)$ under H_0
$t_0 = \dfrac{55.63 - 55}{0.9/\sqrt{9}} = 2.1$

(iii) 기각역 : $P(|T| > t|H_0) = 0.05 \quad \Rightarrow \quad t = t_{0.975}(8) = 2.306 \quad \Rightarrow \quad |T| > 2.306$

(iv) 기각역을 통한 비교 : $|t_0| \not> 2.306$

(v) 결론 : 유의수준 5%에서 귀무가설을 기각할 수 없다.

문제 8. 어느 음료 10개의 용량(ml)을 조사한 결과 표본평균이 250, 표본표준편차가 5 이었다. 음료 용량이 정규분포를 따른다고 해도 무방할 때, 표준편차의 참값이 3보다 높다는 뚜렷한 증거가 있는가? 유의수준 2.5%에서 검정을 하여라.

[풀이].

(i) 가설 : $H_0 : \sigma^2 = 9$ **vs** $H_1 : \sigma^2 > 9$
$n = 10$, $\alpha = 0.025$

(ii) 검정통계량 : $V = \frac{(n-1)S^2}{3^2} \sim \chi^2(n-1)$ under H_0
$S = 5$, $n = 10$, $V_0 = 25$

(iii) 기각역 : $P(V > v) = 0.025 \quad \Rightarrow \quad v = \chi^2_{0.975}(9) = 19.02$

(iv) 기각역을 통한 비교 : $V_0 > 19.02$

(v) 결론 : 유의수준 2.5%에서 귀무가설을 기각할 수 있다.