6회차 과제 - 기초 통계

Date	@2024년 7월 30일 오후 11:59
Tag	과제

1. **Frequentist 방법과 Bayesian 방법론의 차이점을 10줄 이내로 정리하여 작성해주세요.**

   Frequentist 방법과 Bayesian 방법은 확률을 해석하는 관점에 따라 구분됩니다. 먼저 Frequentist는 확률을 상대 빈도의 극한값으로 봅니다. 즉, 무한히 많은 시행을 할 때, 특정 사건이 발생하는 비율을 확률로 정의합니다. 따라서 Frequentist 방법에서는 상대 빈도를 통해 미지의 상수인 모수를 추정합니다. 반면 Bayesian은 확률을 어떠한 사건이 발생할 것임에 대한 주관적인 믿음을 수치로 나타낸 것으로 봅니다. 이러한 주관적인 믿음은 확률변수로서의 모수에 반영됩니다. 이렇듯 Frequentist와 Bayesian은 확률을 해석하는 관점에 따라 구분되며, 이에 따라 Frequentist는 모수를 미지의 상수, 즉 하나의 정해진 값으로 보고, Bayesian은 모수를 확률변수, 즉 정해지지 않은 값으로 봅니다.

2. **베이지안은 사전분포 g(θ)를 완벽하게 안다고 가정한다는 문제점이 존재합니다. 이를 어떻게 해결할 수 있는지 10줄 이내로 정리하여 작성해주세요.**

   Bayesian에서 사후분포는 사전분포 g(θ)와 sampling distribution으로부터 얻은 정보에 의존합니다. 이 때, 어떠한 이유로 사전분포에 대한 정보를 가지고 있지 않은 경우 non-informative prior distribution을 이용하여 sampling distribution으로부터 얻은 정보만을 이용하여 사후분포를 결정할 수 있습니다.

   non-informative prior distribution은 사용하는 사전분포의 pdf를 적분한 값에 따라 proper prior distribution과 improper prior distribution으로 구분됩니다. 먼저 proper prior distribution은 그 밀도함수를 적분한 값이 양수로 수렴합니다. 이러한 사전분포의 예시로 $\mathrm{Uniform}(0, 1)$ 등이 있습니다. 반면 improper prior distribution의 경우 사전분포의 밀도함수를 적분했을 때 특정한 값으로 수렴하지 않습니다.

3. **분포수렴은 누적분포함수(cumulative density function)를 통해 정의됩니다. 확률밀도함수(probability density function)를 이용하여 분포수렴을 정의할 수 없는 이유를 반례를 통해 서술해주세요.**

   분포 수렴은 다음과 같이 정의합니다.

   Let the distribution function $F_n(y)$ of Y_n depend on $n, n = 1, 2, \cdots$.

If $F_n(y)$ is a distribution function and if $\lim_{n\to\infty} F_n(y) = F(y)$ for every point y at which $F(y)$ is continuous, the sequence $Y_1, Y_2, \cdots$ converges in distribution to a random variable with distribution function $F(y)$.

위 정의에서 분포 수렴을 보이기 위해서 cdf의 극한값을 이용하는데, 이 과정에서 cdf 대신 pdf를 이용하면 극한값이 0이 되기에 pdf의 극한값으로 확률변수의 distribution function을 결정할 수 없다는 문제점이 있어 pdf가 아닌 cdf를 이용하여 분포 수렴을 정의합니다. 아래 반례를 통해 이를 확인할 수 있습니다.

$$f(x) = \begin{cases} 1 & \text{if } x = 2 + \frac{1}{n} \\ 0 & \text{otherwise} \end{cases}$$

When $x = 3, f_1(3) = 1, f_2(3) = 0, f_3(3) = 0, \cdots \lim_{n\to\infty} f_n(3) = 0.$

When $x = 1, f_1(1) = 0, f_2(1) = 0, f_3(1) = 0, \cdots \lim_{n\to\infty} f_n(1) = 0.$

When $x = 2, f_1(2) = 1, f_2(2) = 0, f_3(2) = 0, \cdots \lim_{n\to\infty} f_n(2) = 0$

Therefore, $\lim_{n\to\infty} f_n(x) = 0$ for all x, but 0 cannot be a pdf.

4. **MGF(moment generating function)를 이용해서 CLT를 증명해주세요.**

Let $X_1, \cdots X_n$ be a random sample from a distribution which has mean μ and variance σ^2 ($\sigma^2 < \infty$). Then

$Y_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$ has limiting distribution $N(0,1)$.

We assume the existence of mgf $M(t) = E(e^{tX}), -h < t < h$.

$$m(t) = E(e^{t(X-\mu)}) = e^{-t\mu} M(t) \text{ is mgf of } X - \mu$$

$$m(0) = 1$$

$$m'(t) = E[(X-\mu)e^{t(X-\mu)}], \quad m'(0) = E[X-\mu] = 0$$

$$m''(t) = E[(X-\mu)^2 e^{t(X-\mu)}], \quad m''(0) = E[(X-\mu)^2] = \sigma^2$$

$$m(t) = m(0) + m'(0)t + \tfrac{1}{2}m''(\xi)t^2 \text{ for } 0 < \xi < t \text{ by Taylor's formula}$$
$$= 1 + \tfrac{1}{2}m''(\xi)t^2$$
$$= 1 + \tfrac{\sigma^2 t^2}{2} + \tfrac{1}{2}[m''(\xi) - \sigma^2]t^2 \cdots (1)$$

$$M(t;n) = E[\exp(t\left(\sum \tfrac{X_i - n\mu}{\sigma\sqrt{n}}\right))] : \text{mgf of } Y_n$$
$$= \left\{E\left[\exp(t\tfrac{X-\mu}{\sigma/\sqrt{n}})\right]\right\}^n := E\left[\exp(t\tfrac{X_1-\mu}{\sigma/\sqrt{n}})\cdots\exp(t\tfrac{X_n-\mu}{\sigma/\sqrt{n}})\right]$$

$= [m(\frac{t}{\sigma/\sqrt{n}})]^n = \left[1 + \frac{t^2}{2n} + \frac{m''(\xi)-\sigma^2}{2n\sigma^2}t^2\right]^n, \quad -h < \frac{t}{\sigma\sqrt{n}} < h$ by replacing t with $\frac{t}{\sigma\sqrt{n}}$ in equation (1).

Now, $0 < \xi < \frac{t}{\sigma\sqrt{n}}$

As $n \to \infty, \xi \to 0,\ \lim_{n\to\infty}[m''(\xi) - \sigma^2] = 0$

$\because m''(t) = \sigma^2 \quad m''(t)$ is continuous at $t = 0$

$\lim_{n\to\infty} M(t;n) = \lim_{n\to\infty}[1 + \frac{t^2}{2n} + \frac{[m''(\xi)-\sigma^2]}{2n\sigma^2}t^2]^n = \exp(\frac{1}{2}t^2)$

$\therefore Y_n$ has a limiting standard normal distribution.

$\bar{X}_n$ follows approximately $N(\mu, \frac{\sigma^2}{n})$