Daeyoung Lim

Grad School

Year of 2008, early

1. 이항시행(Binomial trial) $\text{Bin}(10, \theta)$에서 $X = 9$가 관측되었다 (즉, $n = 10$이고 $x = 9$).

   1. $H_0 : \theta = 0.5$ 대 $H_1 : \theta > 0.5$의 검정을 하고자 한다. 이를 위해 정확한 p-값을 구하라.

   2. θ에 대한 사전분포로 (0,1)에서의 균일분포(uniform distribution)가 상정되었다고 하자. $\theta \leq 0.5$의 사후확률을 정확히 구하라. 즉, $\Pr(\theta \leq 0.5 \,|\, n = 10, x = 9)$은 얼마인가?

Solution:

1. $\Pr(X > 9 \,|\, \theta = 0.5) = 11 \times 0.5^{10}$. 더 이상 설명할 게 없다.

2. 간단하게 Bayesian 추정방법에 대해서 설명을 덧붙인다. Bayesian statistics에서 사후분포(posterior distribution) 모든 것이라 할 수 있다. 모든 추정방법은 사후분포를 구하기 위한 것인데 시작은 이러하다. 우리가 샘플을 관측할 때 지금까지는 $X_1, X_2, \ldots, X_n$이 정규분포, 혹은 지금과 같이 이항분포를 따른다고 가정한다. 하지만 Bayesian은 이 샘플이 어떤 모수에 의존한다고 생각한다. 예를 들어 정규분포에서는 평균 μ나 분산 σ^2이 주어져야만, 혹은 이항분포에서 확률 p가 주어져야만 어떤 분포에 대해서 얘기할 수 있다는 것이다. (이항분포에서 시행횟수 n은 대부분의 경우 모수로 다루지는 않는다.) 즉 우리는 $X_1, \ldots, X_n|\mu, \sigma^2$가 정규분포라고 말하고 있는 것이고 $X_1, \ldots, X_n|p$이 이항분포를 따른다고 표현한다는 것이다. 그리고 목적지인 사후분포는 샘플을 관측했을 때 모수가 따를 분포이다. 예를 들어 $\mu|X_1, \ldots, X_n$나 $p|X_1, \ldots, X_n$이다.

   임의의 모수를 θ라고 할 때 사후분포는 다음과 같이 구해진다.

$$p(\theta|X_1, \ldots, X_n) = \frac{p(X_1, \ldots, X_n|\theta)\, p(\theta)}{\int p(X_1, \ldots, X_n|\theta)\, p(\theta)\; d\theta} \tag{1}$$

   는 많은 경우 (1)의 분모에 있는 적분이 폐쇄형(closed form; 존재하지만 초등함수의 결합으로 표현되지 않음)으로 떨어지지 않아 *Markov Chain Monte Carlo(MCMC)*와 같은 시뮬레이션에 기반을 둔 샘플링 추정법이나 deterministic한 *variational approximation*과 같은 근사 방법을 통해 추정한다.

   현재 문제에서 주어진 것은 간단한 Beta-Binomial 문제로서 사후분포가 정확하게 우리가 아는 분포로 떨어지는 경우이다. 사실 이런 경우가 지수족(Exponential Family)에서만 발생하며, 사전분포와 사후분포가 모수가 다른 같은 분포가 된다. 이를 켤레사전분포(conjugate prior)라 한다. Beta 분포는 Binomial 분포의 켤레사전분포로서 사후분포 역시 Beta 분포가

된다. 즉,

$$p(\theta|x) \propto p(x|\theta)\, p(\theta) \tag{2}$$
$$\propto \theta^x (1-\theta)^{10-x} \cdot 1 \tag{3}$$
$$\sim \mathrm{Be}(x+1, 11-x). \tag{4}$$

따라서 $\Pr(\theta \leq 0.5 \,|\, n=10, x=9)$는 $\mathrm{Be}(10,2)$를 적분하면 된다.

$$\Pr(\theta \leq 0.5|x=9) = \int_0^{0.5} \frac{\Gamma(12)}{\Gamma(10)\Gamma(2)} \theta^9 (1-\theta)\, d\theta \tag{5}$$
$$= 110 \left[\frac{\theta^{10}}{10} - \frac{\theta^{11}}{11}\right]_0^{0.5} \tag{6}$$
$$= \frac{1}{2^{11}}(22-10) \tag{7}$$
$$= \frac{3}{512} \tag{8}$$

2. $X_1, X_2, \ldots, X_n$을 평균이 μ이고 분산이 σ^2인 분포로부터의 임의표본이라고 할 때
   1. $S^2 = \dfrac{1}{n+1}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2$가 σ^2의 일치추정량임을 보이고
   2. 임의표본의 가정에 정규분포의 가정을 추가햇을 때의 위의 S^2의 σ^2에 대한 MSE(Mean Squared Error)을 구하시오.

Solution:

1. 일치추정량의 정의를 이용해서 보이는 방법이 있고 대수의 법칙에 의해 이미 확률수렴함을 알고 있는 통계량을 이용해 보일 수 있다.

   - 먼저 첫번째 정의를 이용한 증명이다. 마코프 부등식(Markov inequality)에 의해 다음의 부등식이 성립한다.

$$\Pr\left(\left|S_n - \sigma^2\right| \geq \epsilon\right) \leq \frac{\mathrm{E}\left(\left|S_n - \sigma^2\right|\right)}{\epsilon} \tag{9}$$

   $\left|S_n - \sigma^2\right|$은 음수일 수도 있고 양수도 있지만 어차피 나중에 무관해진다. 우선 양수라 치고

$$\mathrm{E}(S_n) - \sigma^2 = \frac{n-1}{n+1}\sigma^2 - \sigma^2 \tag{10}$$

이다. $(n-1)^{-1}\sum_{i=1}^{n}\left(X_n-\overline{X}_n\right)^2$가 비편향추정량임을 알고 있기 때문에 이를 이용하면 바로 나온다. 따라서 이를 (9)에 대입하여 극한을 취하면

$$\lim_{n\to\infty}\Pr\left(\left|S_n-\sigma^2\right|\geq\epsilon\right)=\lim_{n\to\infty}\frac{1}{\epsilon}\left(\frac{n-1}{n+1}\sigma^2-\sigma^2\right) \tag{11}$$

$$=0 \tag{12}$$

고로 S^2는 σ^2의 일치추정량이다.

- 다음은 표본적률이 모적률로 확률수렴한다는 대수의 법칙을 이용한 증명이다. 즉

$$\frac{1}{n}\sum_{i=1}^{n}X_i^k\ \xrightarrow{p}\ \mathrm{E}\left(X_1^k\right). \tag{13}$$

S^2를 전개하면

$$S^2=\frac{n}{n+1}\overline{X_n^2}-\frac{n}{n+1}\left(\overline{X}_n\right)^2 \tag{14}$$

이고 $\overline{X_n^2}=n^{-1}\sum_{i=1}^{n}X_i^2$, 2차 표본적률이다. 따라서 이는 다음과 같이 확률수렴한다.

$$S^2\ \xrightarrow{p}\ \mathrm{E}\left(X_1^2\right)-\left(\mathrm{E}\left(X_1\right)\right)^2 \tag{15}$$

1차 표본적률의 제곱이 모적률의 제곱으로 확률수렴한다는 것은 Slutsky 정리의 결과이다. 다시 말해 X와 Y가 각각 μ와 ν로 확률수렴하면 그 곱도 곱으로 확률수렴한다. 따라서

$$S^2\ \xrightarrow{p}\ \left(\sigma^2+\mu^2\right)-\mu^2=\sigma^2 \tag{16}$$

2. MSE는 다음과 같이 분해할 수 있다 (Variance-bias decomposition).

$$\mathrm{MSE}\left(X\right)=\mathrm{Bias}^2\left(X\right)+\mathrm{Var}\left(X\right) \tag{17}$$

이를 이용하여 분산을 계산할 때 정규분포 가정이 없으면 인생이 비참해진다. 왜냐하면 정규성 가정이 있으면 대충 카이제곱 분포를 통해 유도해낼 수 있기 때문이다. 먼저 앞서 $\mathrm{E}\left(S^2\right)$은 구했다. 그러므로 편향의 제곱은 다음과 같다.

$$\mathrm{Bias}^2\left(S^2\right)=\left(\frac{n-1}{n+1}\sigma^2-\sigma^2\right)^2=\frac{4\sigma^4}{\left(n+1\right)^2} \tag{18}$$

그리고 분산을 구하기에 앞서 $\sigma^{-2}\sum_{i=1}^{n}\left(X_i-\overline{X}_n\right)^2\sim\chi^2\left(n-1\right)$임을 이용해보자. 그리고

카이제곱분포와 감마분포와의 관련성을 이용하자.

$$S^2 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2}{\sigma^2}\frac{\sigma^2}{n+1} \tag{19}$$

$$\sim \text{Ga}\left(\frac{n-1}{2}, \frac{n+1}{2\sigma^2}\right) \tag{20}$$

$$\text{Var}\left(S^2\right) = \frac{2\left(n-1\right)\sigma^4}{\left(n+1\right)^2} \tag{21}$$

따라서,

$$\text{MSE}\left(S^2\right) = \frac{4\sigma^4}{\left(n+1\right)^2} + \frac{2\left(n-1\right)\sigma^4}{\left(n+1\right)^2} \tag{22}$$

$$= \frac{2\sigma^4}{n+1} \tag{23}$$

3. 확률변수 X가 시행횟수가 n이고 성공확률이 p인 이항분포 $\text{Bin}\left(n,p\right)$를 따른다고 하자.

   1. 귀무가설 $H_0 : p = 1/4$ 대 대립가설 $H_1 : p = 1/2$에 대한 최강력(most powerful) 검정법이 존재한다면 유의수준 $\alpha = 0.05$로 구하고, 구한 검정방법의 검정력을 구하시오.
   2. 문항 1에서 구한 검정법에 대해 n이 큰 경우 근사 임계값을 중심극한정리에 의해 구하시오. 단, 중심극한정리를 정확히 설명하고 이를 근사 임계값을 구하는 데 사용할 수 있는 이유를 쓰시오.
   3. 귀무가설 $H_0 : p = 1/2$ 대 대립가설 $H_1 : p > 1/2$에 대한 균일 최강력 검정방법이 존재한다면 유의수준 $\alpha = 0.05$로 구하시오.
   4. 귀무가설 $H_0 : p = 1/2$ 대 $H_1 : p \neq 1/2$에 대한 균일 최강력 검정 방법이 존재한다면 유의수준 $\alpha = 0.05$로 구하시오.
   5. 귀무가설 $H_0 : p = 1/2$ 대 대립가설 $H_1 : p \neq 1/2$에 대한 유의수준 $\alpha = 0.05$인 일반화 가능도비(likelihood ratio) 검정방법을 구하시오.

      (힌트: $\lambda\left(x\right)$를 가능도비라고 하면 $x \leq n/2$이면 $\lambda\left(x\right)$는 $-\left(2x-n\right)$에 대한 증가함수이며 $\lambda\left(x\right) = \lambda\left(n-x\right)$임을 보인 후, 검정방법을 구할 것.)

Solution:

- 가능도비를 계산하면

$$\frac{L_0}{L_1} = \frac{\left(0.25\right)^x\left(0.25\times 3\right)^{n-x}}{\left(0.5\right)^x} \tag{24}$$

$$= \left(\frac{3}{2}\right)^n\left(\frac{1}{3}\right)^x \leq c \tag{25}$$

이를 정리하면 $x \geq c'$일 때 귀무가설을 기각하는 것이 최강력 검정이 됨을 알 수 있다. 어떤 상수 c'는 유의확률이 α(유의수준)과 같아지는 상수값을 찾으면 된다. 하지만 이 경우 이산확률분포이므로 정확히 $\alpha = 0.05$가 되는 값이 없을 수도 있고, 그러한 경우

$$\sum_{x=c'}^{n} \binom{n}{x} (0.25)^x (0.75)^{n-x} \leq 0.05 \tag{26}$$

인 최소의 양의 정수를 찾으면 된다. 그리고 검정력은 다음과 같다.

$$1 - \beta = 1 - \sum_{x=0}^{c'-1} \binom{n}{x} (0.5)^n \tag{27}$$

- 중심극한정리에서 가장 중요한 가정은 2차 적률이 유한하다는 것이다. 이는 Measure theory의 언어로 확률변수 X가 L_2 space에 속한다는 것이다. 즉, 분산이 유한한 확률변수에서 서로 독립인 임의표본을 뽑은 경우 표본평균이 점근적으로(asymptotically) 정규분포를 따른다는 정리이다. 중심극한정리를 쓸 수 있으려면 다음 두 조건을 만족해야 한다.
  - $\mathrm{E}|X|^2 < \infty$ (확률공간과 같이 유한 측도 공간(finite measure space)를 가정할 경우 $1 \leq p < q \leq \infty$에 대해서 $\mathrm{E}(X^q) < \infty$하면 반드시 $\mathrm{E}(X^p) < \infty$하다. 다시 말해 2차 적률이 유한하다는 말은 1차 적률 또한 유한하다는 뜻이다.)
  - $X_1, X_2, \ldots, X_n$이 서로 독립이다.

  중심극한정리 중 이항분포가 정규근사될 수 있다는 것은 특별히 *DeMoivre-Laplace Theorem*이라 부르고 수학적으로 표현하면 다음과 같다.

$$Y_i \overset{\text{iid}}{\sim} \mathrm{Ber}(p) \implies X \equiv \sum_{i=1}^{n} Y_i \sim \mathrm{Bin}(n, p) \tag{28}$$

  서로 독립인 n개의 베르누이 확률변수의 합이므로 중심극한정리에 따라 그 평균을 정규분포로 근사시킬 수 있다.

  임계값을 구해보면 $\overline{X}_n = n^{-1}X$라 할 때 $\overline{X}_n \sim \mathcal{N}\left(p, n^{-1}p(1-p)\right)$이므로

$$\frac{\sqrt{n}\left(\overline{X}_n - p\right)}{\sqrt{p(1-p)}} \xrightarrow{d} \mathcal{N}(0,1) \tag{29}$$

  이고 1에서 구한 기각역에 따라 다음이 성립하는 c'를 찾으면 된다.

$$\Pr\left(\frac{\sqrt{n}\left(\overline{X}_n - p\right)}{\sqrt{p(1-p)}} \geq \frac{\sqrt{n}\left(c'/n - p\right)}{\sqrt{p(1-p)}}\right) = 0.05 \tag{30}$$

따라서 이를 통해

$$\frac{\sqrt{n}\left(c'/n-p\right)}{\sqrt{p\left(1-p\right)}}=z_{0.05}\implies c'=\sqrt{np\left(1-p\right)}z_{0.05}+np \tag{31}$$

- p_1를 $p_1>1/2$를 만족하는 어떤 모수로 놓고 가능도비를 계산하면

$$\frac{L_0}{L_1}=\left(\frac{1-p_1}{p_1}\right)^x\leq k \tag{32}$$

이 기각역이 되고 $p_1>1/2$이므로 위의 함수는 x에 대하여 단조감소하는 함수가 된다. 따라서 어떤 상수 k보다 작아지려면 x가 충분히 커야 하므로 $c'\geq x$이 기각역이 될 것이다. (가능도비가 단조로우므로(?) 균일 최강력 검정법이 존재한다.) c'값은 1에서와 같이

$$\sum_{x=c'}^{n}\binom{n}{x}0.5^x\leq 0.05$$

를 만족하는 최소의 양의 정수를 찾는다.

- 앞서 보였듯이 대립가설의 모수를 p_1로 놓고 가능도비를 구하면

$$\left(\frac{1-p_1}{p_1}\right)^x\leq k$$

의 형태로 나오게 되는데 $p_1>1/2$일 경우 단조감소, $p_1<1/2$일 경우 단조증가하므로 균일 최강력 검정 방법이 존재하지 않는다.

- 일반화 가능도비 검정법을 쓰기 위해서는 MLE를 알아야 한다. 이항분포에서 MLE는 $\hat{p}=x/n$라는 것을 안다고 치고 가능도비를 구하면

$$\lambda=\left(\frac{1-\hat{p}}{\hat{p}}\right)^x\left(\frac{0.5}{\hat{p}}\right)^n \tag{33}$$

$$=\left(\frac{1-\hat{p}}{\hat{p}}\right)^{n\hat{p}}\left(\frac{0.5}{\hat{p}}\right)^n \tag{34}$$

$$\ln\lambda=n\left(\hat{p}-1\right)\ln\left(1-\hat{p}\right)-n\hat{p}\ln\hat{p}-n\ln 2 \tag{35}$$

$$\frac{d\ln\lambda}{d\hat{p}}=n\ln\left(\frac{1-\hat{p}}{\hat{p}}\right) \tag{36}$$

이고 로그가능도비의 도함수가 0이 되는 지점은 $\hat{p}=1/2$이다. 로그가능도비가 증가하다 $\hat{p}=1/2$를 기점으로 다시 감소하는 형태이므로 $\ln\lambda\leq k$이기 위해서는 $\hat{p}$가 충분히 크거나

충분히 작아야 한다. 즉 기각역이 다음과 같다.

$$\hat{p} \leq c_1 \quad \text{or} \quad \hat{p} \geq c_2 \ (c_1 \leq c_2)$$

따라서

$$\sum_{x=0}^{c_1} \binom{n}{x} 0.5^n \leq 0.025 \tag{37}$$

$$\sum_{x=c_2}^{n} \binom{n}{x} 0.5^n \leq 0.025 \tag{38}$$

를 만족하는 상수값들을 찾으면 된다.

4. Let $X_1, \ldots, X_n$ be a random sample from $f(x;\lambda) = \lambda \exp(-\lambda x)$, where $0 < x < \infty$. We want to test $H_0 : \lambda \leq \lambda_0$ versus $H_1 : \lambda > \lambda_0$. Obtain the uniformly most powerful (UMP) test with size α.

Solution: 그만 좀 물어봐라. 가능도 두 개 나눠.

5. 절편이 없는 단순선형회귀모형

$$Y = \beta_1 X_1 + \epsilon$$

을 고려하자. 여기서, Y는 반응변수를, X_1은 설명변수를, 그리고 β_1은 회귀계수를 의미하며 오차항 ϵ은 평균이 0, 분산이 σ^2이라 가정한다.

1. β_1의 최소제곱추정량(least squares estimator)을 구하시오. (10점)
2. 반응변수 Y에 대한 옳은 회귀 모형이 위의 단순선형회귀모형이 아니라

$$Y = \beta_1 X_1 + \beta_2 X_2 + \epsilon$$

이었다고 가정하자. 여기서, X_2는 실수로 포함되지 않은 X_1과 다른 설명변수를 의미하며 β_2는 X_2의 회귀계수를 의미한다. 1에서 구한 최소제곱추정량이 비편향추정량(unbiased estimator)이 되는지 여부를 보이시오. (10점)

3. 2의 회귀모형이 옳은 모형일 때 1에서 구한 β_1의 최소제곱추정량의 분산을 구하시오. (5점)

Solution:

1. $\epsilon_i = Y_i - \beta_1 x_{1i}$이므로

$$\frac{d}{d\beta_1}\sum_{i=1}^{n}\left(Y_i - \beta_1 x_{1i}\right)^2 = -2\sum_{i=1}^{n} x_{1i}\left(Y_i - \beta_1 x_{1i}\right) = 0 \tag{39}$$

$$\widehat{\beta_1} = \frac{\sum_{i=1}^{n} x_{1i}Y_i}{\sum_{i=1}^{n} x_{1i}^2}. \tag{40}$$

2. Omitted variable bias가 나타날 것임이 분명하다. 수학적으로 증명하자면 다음과 같다. 1에서 구한 $\widehat{\beta_1}$에 Y_i가 있다. 여기에 참인 모형을 대입하자.

$$\mathrm{E}\left(\widehat{\beta_1}\right) = \mathrm{E}\left(\frac{\sum_{i=1}^{n} x_{1i}Y_i}{\sum_{i=1}^{n} x_{1i}^2}\right) \tag{41}$$

$$= \frac{1}{\sum_{i=1}^{n} x_{1i}^2}\sum_{i=1}^{n} x_{1i}\mathrm{E}\left(Y_i\right) \tag{42}$$

$$= \frac{1}{\sum_{i=1}^{n} x_{1i}^2}\sum_{i=1}^{n} x_{1i}\left(\beta_1 x_{1i} + \beta_2 x_{2i}\right) \tag{43}$$

$$= \beta_1 + \beta_2\left(\frac{\sum_{i=1}^{n} x_{1i}x_{2i}}{\sum_{i=1}^{n} x_{1i}^2}\right) \tag{44}$$

$$\neq \beta_1 \tag{45}$$

따라서 비편향추정량이 되지 못한다.

3.

$$\mathrm{Var}\left(\widehat{\beta_1}\right) = \frac{1}{\left(\sum_{i=1}^{n} x_{1i}^2\right)^2}\mathrm{Var}\left(x_{11}Y_1 + x_{12}Y_2 + \cdots + x_{1n}Y_n\right) \tag{46}$$

$$= \frac{1}{\left(\sum_{i=1}^{n} x_{1i}^2\right)^2}\left(x_{11}^2\sigma^2 + \cdots + x_{1n}^2\sigma^2\right) \tag{47}$$

$$= \frac{\sum_{i=1}^{n} x_{1i}^2}{\left(\sum_{i=1}^{n} x_{1i}^2\right)^2}\sigma^2 \tag{48}$$

$$= \frac{\sigma^2}{\sum_{i=1}^{n} x_{1i}^2} \tag{49}$$