

R.A Fisher는 아마 가장 영향력이 있는 anti-Bayesian이라고 할 수 있다. 하지만 Fisher는 빈도주의자라고 부르기에는 적합하지 않다. Fisher가 통계학에 기여한 주요분야, 분산분석, 유의성 검정, 최대우도추정은 빈도주의 측면이 강하지만 Fisherian이라는 또다른 학파가 있음에 주목하자. Fisherian은 최대우도추정에서 중요한 역할을 하며 때로는 베이지안과 빈도주의를 결합한 아이디어를 제공하기도 한다. 이 장에서는 Fisherian의 주요방법에 대해서 알아보자.

#### 4.1 Likelihood and Maximum Likelihood

Fisherian은 우도주의라고도 부를 수 있는데 그 이유는 우도함수에 기반한 추론에서 중요한 역할을 하기 때문이다. 먼저 확률밀도함수의 family  $f_\mu(x)$ 에 대해서 다음과 같이 *log likelihood function*을 정할 수 있다.

$$\ell_x(\mu) = \log(f_\mu(x)),$$

여기서 주목할 점은  $\ell_x(\mu)$ 는  $\mu$ 의 함수라는 점이다. 즉 데이터  $x$ 는 고정되어 있고  $\mu$ 는 다양한 값을 가질 수 있다. 최대우도추정치 (MLE: maximum likelihood estimate)는 다음과 같이 정의된다.

$$\text{MLE: } \hat{\mu} = \underset{\mu \in \Omega}{\operatorname{argmin}} \ell_x(\mu)$$

위의 정의는 최적화의 해가 유일하게 존재할때 성립한다. 가끔씩  $\hat{\mu}$ 가 존재하지 않거나 여러가지 해가 존재할때도 있지만 여기서는 해가 유일하게 존재하는 경우를 가정한다. MLE의 강점은  $\theta = T(\mu)$ 와 같이 원래 모수의 (1-1) 변환으로 이루어진 새로운 모수에 대해서도 plug-in rule을 통해 쉽게 MLE를 구할 수 있다는 점이다.

$$\hat{\theta} = T(\hat{\mu})$$

MLE는 다음과 같은 이유로 최근까지 가장 많이 사용되는 추정방법중 하나이다.

1. MLE 알고리즘은 아주 간편하다. 즉 간단한 최적화 알고리즘으로  $\hat{\mu}$ 을 쉽게 계산할 수 있다. UMVUE의 경우 구하는 방법이 case-by-case라는 걸 고려한다면 쉽게 MLE를 계산할 수 있다는 것은 큰 장점이다.
2. MLE는 빈도주의 장점을 공유한다. 대표본에서 MLE는 (거의) 불편추정치 (unbiased estimate)이며 점근분포의 분산도 다른 추정치와 비교시 가장 작다고 할 수 있다. 표본크기가 작을때에도 MLE는 대부분 좋은 성능을 보여준다.

3. MLE는 베이저안 측면도 가지고 있다. 베이즈 룰을 다음과 같이 표현할 수 있다.

$$g(\mu | x) = c_x g(\mu) e^{\ell_x(\mu)}$$

만약  $g(\mu)$ 가 flat prior라면  $\hat{\mu}$ 는 사후분포  $g(\mu | x)$ 를 최대로 만드는 점이라는 것을 알 수 있다.

그림 4.1은 사구체 여과율 자료에 감마분포와 정규분포를 적합시킨 결과를 보여준다. 각 분포의 모수는 MLE를 이용하여 추정하였다. 구체적으로 다음과 같은 절차를 따랐다. 데이터  $\mathbf{x} = (x_1, x_2, \dots, x_n)$ 가 다음과 같이 확률밀도함수  $f_\mu(x)$ 로 부터 생성되었다고 가정하자. 사구체 투사율의 자료의 경우  $n = 211$ 이다.

$$x_i \stackrel{iid}{\sim} f_\mu(x) \quad \text{for } i = 1, 2, \dots, n,$$

여기서 IID는 "independent and identically distributed" 즉 "서로 독립이고 동일한 분포로부터"를 의미한다. 이 예제에서는  $f_\mu$ 로 정규분포와 감마분포를 고려하였다.

먼저  $\mu = (\theta, \sigma)$ 로 주어진 정규분포의 확률밀도함수는 다음과 같다. 여기서  $\theta$ 는 평균이  $\sigma$ 는 표준편차를 나타낸다.

$$f_\mu(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\theta}{\sigma}\right)^2}$$

$\mu = (\lambda, \sigma, \nu)$ 로 이루어진 감마분포의 확률밀도 함수는 다음과 같다. 일반적인 감마분포는  $\lambda = 0$ 으로 2개의 모수로 구성된다.

$$f_\mu(x) = \frac{(x-\lambda)^{\nu-1}}{\sigma^\nu \Gamma(\nu)} e^{-\frac{x-\lambda}{\sigma}} \quad (\text{for } x \geq \lambda, 0 \text{ otherwise}).$$

$x_i$ 가 서로 독립이므로

$$f_\mu(\mathbf{x}) = \prod_{i=1}^n f_\mu(x_i),$$

따라서

$$\ell_{\mathbf{x}}(\mu) = \sum_{i=1}^n \log f_\mu(x_i) = \sum_{i=1}^n \ell_{x_i}(\mu).$$

정규분포의 경우  $\ell_{\mathbf{x}}$ 를 최소화하는 MLE는 다음과 같은 공식을 사용해서 구할 수 있다.

$$(\hat{\theta}, \hat{\sigma}) = (54.3, 13.7) = \left( \bar{x}, \left[ \sum (x_i - \bar{x})^2 / n \right]^{1/2} \right).$$

반면에 감마분포는 수치적 알고리즘으로만 해를 구할 수 있으며 그 값은 다음과 같다.

$$(\hat{\lambda}, \hat{\sigma}, \hat{\nu}) = (21.4, 5.47, 6.0).$$

그림 4.1을 보면 감마분포 MLE가 정규분포 MLE보다 데이터에 적합한 것으로 보이지만 둘다 만족스러운 적합 결과를 보여주지는 않는다.

Fisher가 1920년대에 MLE를 처음 소개했을때 감마분포의 예처럼 계산하는 것이 어렵다는 점에서 비판을 받았

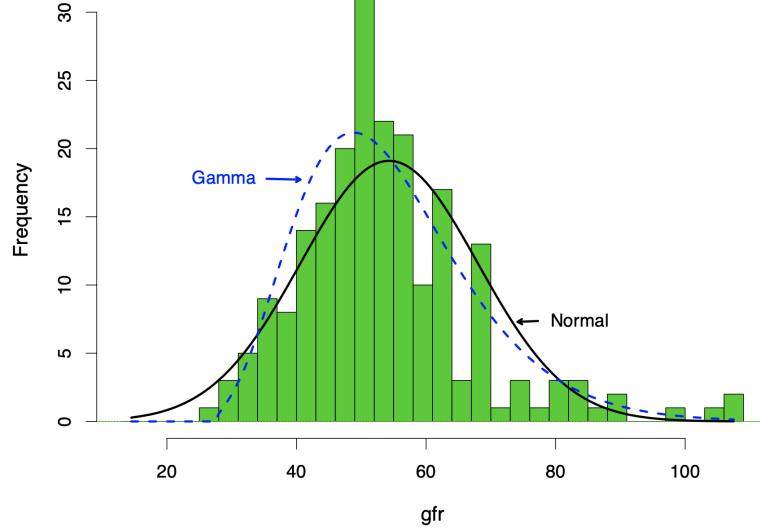


그림 4.1 사구체 여과율 자료와 정규분포 (검은 실선), 감마분포(파란 파선) MLE

다. 지금으로서는 (거의) 문제가 되지 않지만 1920년대에는 간단한 최적화문제도 쉽지 않았다는 점을 고려하자.

MLE는 여전히 광범위하게 사용되는 추정방법이지만 고차원 자료로 대표되는 요즘의 데이터분석에 사용시 주의를 요한다. 3장의 전립선 암 데이터를 생각해보자. 만약 우리가 두 집단 사이의 차이가 가장 큰 유전자를 찾고자 한다면 다음과 같이 관심이 있는 모수를 정할 수 있다.

$$\hat{\theta} = T(\mu) = \max_i \{\mu_i\}$$

만약 여기서 우리가  $T(\hat{\mu}) = \hat{\mu}_{610} = x_{610}$ 을 사용한다면 심각한 과대추정문제를 야기할 수 있다. 따라서 고차원 데이터에서는 "정규화(축소된)"된 MLE를 사용하는 것이 보다 적절해보인다.

## 4.2 Fisher Information and the MLE

사실 MLE를 모수추정에 사용하자고 제안한 처음으로 제안한 사람은 Fisher가 아니다. Fisher는 MLE가 추론에 사용할 경우 여러가지 좋은 성질을 가지고 있다는 것을 증명하였고 그 결과 MLE에 추론에 광범위하게 사용될 수 있도록 이바지했다는 것이 Fisher의 주요 업적이라고 할 수 있다. Fisher의 MLE에 관한 기여중 가장 중요한 업적중 하나인 Fisher information bound에 대해 알아보자.

먼저 다음과 같은 모수가 하나인 확률밀도함수의 family를 생각해보자.

$$\mathcal{F} = \{f_{\theta}(x), \theta \in \Omega, x \in \mathcal{X}\},$$

여기서  $\Omega$ 는  $\mathcal{R}^1$ 의 구간을 의미하며  $\mathcal{R}$  자체일 수도 있고 표본공간  $\mathcal{X}$ 는 다차원 공간이 될 수도 있다.

먼저 score function을 다음과 같이 정의하자.

$$\dot{\ell}_x(\theta) = \frac{\partial}{\partial \theta} \log f_\theta(x) = \dot{f}_\theta(x)/f_\theta(x).$$

Score function의 평균이 0임은 쉽게 보일 수 있다.

$$\int_{\mathcal{X}} \dot{\ell}_x(\theta) f_\theta(x) dx = \int_{\mathcal{X}} \dot{f}_\theta(x) dx = \frac{\partial}{\partial \theta} \int_{\mathcal{X}} f_\theta(x) dx = \frac{\partial}{\partial \theta} 1 = 0$$

Fisher information  $\mathcal{I}_\theta$ 는 score function의 분산으로 정의된다.

$$\mathcal{I}_\theta = \int_{\mathcal{X}} (\dot{\ell}_x(\theta))^2 f_\theta(x) dx,$$

따라서

$$\dot{\ell}_x(\theta) \sim (0, \mathcal{I}_\theta)$$

위의 결과를 이용해서 MLE의 점근분포 (asymptotic distribution)가 다음과 같음을 보일 수 있다.

$$\hat{\theta} \sim \mathcal{N}(\theta, 1/\mathcal{I}_\theta)$$

극한분포의 분산이  $1/\mathcal{I}_\theta$ 으로 주어졌다는 점에서  $\mathcal{I}_\theta$ 의 명칭으로 "정보"가 적절하다는 것을 알 수 있다. 즉 정보의 양이 많을 수록 추정량의 정확도가 높아진다고 볼 수 있다.

Log likelihood의 2차 도함수는 다음과 같이 주어진다.

$$\ddot{\ell}_x(\theta) = \frac{\partial^2}{\partial \theta^2} \log f_\theta(x) = \frac{\ddot{f}_\theta(x)}{f_\theta(x)} - \left( \frac{\dot{f}_\theta(x)}{f_\theta(x)} \right)^2$$

2차도함수의 기대값은 다음과 같음을 쉽게 보일 수 있다.

$$\mathbb{E}_\theta(\ddot{\ell}_x(\theta)) = -\mathcal{I}_\theta,$$

$\mathcal{J}_\theta \equiv \text{VAR}_\theta(\ddot{\ell}_x(\theta))$ 로 정의하자. 이 경우는 다음과 같이 표현할 수 있다.

$$-\ddot{\ell}_x(\theta) \sim (\mathcal{I}_\theta, \mathcal{J}_\theta)$$

이제  $\mathbf{x} = (x_1, x_2, \dots, x_n)$ 이고  $x_i \stackrel{iid}{\sim} f_\mu(x)$ ,  $i = 1, \dots, n$ 이라고 하자. 이 경우  $\dot{\ell}_{\mathbf{x}}(\theta)$ 는 다음과 같이 표현할 수 있고

$$\dot{\ell}_{\mathbf{x}}(\theta) = \sum_{i=1}^n \dot{\ell}_{x_i}(\theta),$$

비슷하게

$$-\ddot{\ell}_{\mathbf{x}}(\theta) = -\sum_{i=1}^n \ddot{\ell}_{x_i}(\theta).$$

$x$ 를 기반으로 하는  $\hat{\theta}_{mle}$ 는  $\dot{\ell}_{\mathbf{x}}(\hat{\theta}_{mle}) = 0$ 을 만족하는 해를 이용하여 구할 수 있다. 만약 테일러 급수를 1차항까지 전개하면

$$0 = \dot{\ell}_{\mathbf{x}}(\theta) \doteq \dot{\ell}_{\mathbf{x}}(\theta) + \ddot{\ell}_{\mathbf{x}}(\theta)(\hat{\theta} - \theta),$$

을 얻을 수 있고 이를 다른 방식으로 표현하자면,

$$\hat{\theta} \doteq \theta + \frac{\dot{\ell}_{\mathbf{x}}(\theta)/n}{(-\ddot{\ell}_{\mathbf{x}}(\theta)/n)}.$$

앞의 결과를 토대로 적절한 regularity condition하에서 중심극한정리를 적용하면 다음과 같은 극한분포를 유도할 수 있다.

$$\ell_{\mathbf{x}}(\theta)/n \dot{\sim} \mathcal{N}(0, \mathcal{I}_{\theta}/n),$$

따라서 MLE  $\hat{\theta}$ 의 극한분포는

$$\hat{\theta} \dot{\sim} \mathcal{N}(\theta, 1/(n\mathcal{I}_{\theta})).$$

여기서  $-\ddot{\ell}_{\mathbf{x}}(\theta)/n$ 은 대수의 법칙에 따라  $\mathcal{I}_{\theta}$ 로 수렴한다는 것을 이용하였다.

실제 정규분포의 예를 들어 MLE의 극한분포를 계산해보자. 먼저

$$x_i \stackrel{iid}{\sim} \mathcal{N}(\theta, \sigma^2) \quad \text{for } i = 1, 2, \dots, n,$$

라고 하자. 여기서  $\sigma^2$ 은 알고있다고 가정하자. 이 경우 log likelihood는

$$\ell_{\mathbf{x}}(\theta) = -\frac{1}{2} \sum_{i=1}^n \frac{(x_i - \theta)^2}{\sigma^2} - \frac{n}{2} \log(2\pi\sigma^2).$$

따라서

$$\dot{\ell}_{\mathbf{x}}(\theta) = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \theta) \quad \text{and} \quad -\ddot{\ell}_{\mathbf{x}}(\theta) = \frac{n}{\sigma^2}.$$

위의 결과를 이용하면  $\hat{\theta} = \bar{x}$ 이며  $\mathcal{I}_{\theta} = 1/\sigma^2$ 이므로

$$\hat{\theta} \dot{\sim} \mathcal{N}(\theta, \sigma^2/n)$$

을 알 수 있다.

Fisher는 sufficiency, efficiency, conditionality, ancillarity를 통해 최대우도량이 20세기 추론의 중심으로 자리 잡게 하였다. 최대우도추정량이 왜 좋은 추정량인지 다음에 소개될 *Cramér-Rao bound*를 통해서 알아보자.

먼저  $\tilde{\theta} = t(x)$ 를  $\theta$ 에 대한 임의의 불편추정량 (unbiased estimator)라고 하자. 즉

$$\theta = \mathbb{E}_\theta[t(\mathbf{x})]$$

이 경우

$$\begin{aligned} \int_{\mathcal{X}} t(\mathbf{x}) \dot{\ell}_{\mathbf{x}}(\theta) f_\theta(\mathbf{x}) d\mathbf{x} &= \int_{\mathcal{X}} t(\mathbf{x}) \dot{f}_\theta(\mathbf{x}) d\mathbf{x} \\ &= \frac{\partial}{\partial \theta} \int_{\mathcal{X}} t(\mathbf{x}) f_\theta(\mathbf{x}) d\mathbf{x} = \frac{\partial}{\partial \theta} \theta = 1. \end{aligned}$$

임을 보일 수 있다. 따라서

$$\int (t(\mathbf{x}) - \theta) \dot{\ell}_{\mathbf{x}}(\theta) f_\theta(\mathbf{x}) d\mathbf{x} = 1$$

여기에 Cauchy-Schwarz 부등식을 적용하면

$$\left[ \int (t(\mathbf{x}) - \theta) \dot{\ell}_{\mathbf{x}}(\theta) f_\theta(\mathbf{x}) d\mathbf{x} \right]^2 \leq \left[ \int (t(\mathbf{x}) - \theta)^2 f_\theta(\mathbf{x}) d\mathbf{x} \right] \left[ \int \dot{\ell}_{\mathbf{x}}(\theta)^2 f_\theta(\mathbf{x}) d\mathbf{x} \right]$$

즉

$$1 \leq \text{VAR}_\theta(\tilde{\theta}) \cdot n\mathcal{I}_\theta$$

즉 위의 식을 바탕으로 다음과 같은 *Cramér-Rao bound*를 구할 수 있다.

$$\text{VAR}_\theta(\tilde{\theta}) \geq 1/(n\mathcal{I}_\theta)$$

*Cramér-Rao bound*의 의미는 최대우도추정량의 분산은 불편추정량이 가질 수 있는 분산중 가장 작은 분산보다 작거나 같다는 것을 의미한다. 최대우도추정량의 경우 일반적으로 (거의) 불편추정량에 가깝기 때문에 분산의 크기가 작다는 것은 다른 추정량에 비해서 좋은 추정량이라고 얘기할 수 있다. 참고로 MLE의 bias의 order는  $n^{-1}$ 이고 표준편차의 order는  $n^{-1/2}$ 이다.

### *The Neyman-Scott Paradox*

물론 MLE의 bias<sup>2</sup>과 variance의 order가 위에서 얘기한 것과 다른 경우도 있다. 다음 normal mean 예제를 살펴보자.

$(x_i, y_i)$ 가 서로 독립이고  $N(\mu_i, \sigma^2)$  분포를 따른다고 가정하자. 여기서  $\alpha_i$ 에 대한 MLE는  $\hat{\alpha}_i = (x_i + y_i)/2$ 이고  $\sigma^2$ 의 MLE는  $\hat{\sigma}^2 = \sum_{i=1}^n s_i^2/n$ 이다. 여기서

$$s_i^2 = \frac{1}{2} [(x_i - \hat{\alpha}_i)^2 + (y_i - \hat{\alpha}_i)^2] = \frac{1}{4}(x_i - y_i)^2$$

여기서  $\mathbb{E}(s_i^2) = \sigma^2/2$ 임을 쉽게 보일 수 있다. 따라서  $\hat{\sigma}^2$ 은 inconsistent하다.

### 4.3 Conditional Inference

조건부 추론에 대한 소개를 위해 다음과 같은 예제를 고려해보자. 관측치가 서로 독립이고 평균이  $\theta$ 이고 표준편차가 1인 정규분포에서 생성된다고 가정하자. 즉

$$x_i \stackrel{iid}{\sim} \mathcal{N}(\theta, 1), \quad i = 1, 2, \dots, n$$

이 경우  $\hat{\theta} = \bar{x}$ 이다. 하지만 표본크기를 결정하는데 대한 의견불일치로 결국 동전을 던져서 다음과 같은 방법으로 표본크기를 결정하기로 하였다.

$$N = \begin{cases} 25 & \text{probability } 1/2 \\ 100 & \text{probability } 1/2 \end{cases}$$

$N = 25$ 인 경우로 결정이 되었다고 가정한다면  $\bar{x}$ 의 표준편차는 무엇인가?

만약 여러분의 답이  $\text{VAR}_{\theta}(\bar{X}_N | N = 25) = 1/\sqrt{25} = 0.2$ 라고 한다면 여러분도 Fisher와 마찬가지로 조건부 추론의 신봉자이다. 조건부 추론을 신봉하지 않는다면 표본크기가 동일할 확률로 25 또는 100이 될 가능성을 모두 고려하여

$$\begin{aligned} \text{VAR}_{\theta}(\bar{X}_N) &= \text{VAR}_N [\mathbb{E}_{\theta|N}(\bar{X}_N | N)] + \mathbb{E}_N [\text{VAR}_{\theta|N}(\bar{X}_N | N)] \\ &= \text{VAR}_N(\theta) + \frac{1}{25} \cdot \frac{1}{2} + \frac{1}{100} \cdot \frac{1}{2} = \frac{1}{40} \end{aligned}$$

이라고 대답할 것이다.

조건부 추론은 생각보다 통계학 전반에 광범위하게 이미 사용되고 있다. 예를 들자면 회귀분석에서는 조건부 기대값  $\mathbb{E}(Y|X = x)$ 를 예측하는 것을 고려한다. 물론  $x$ 가 고정된 점이라면 상관이 없지만 많은 경우  $x$ 도 확률 변수로 고려할 수 있다면 이와 같은 조건부 추론이 합당하다.

Fisher는 다음과 같은 2가지 근거를 통하여 조건부 추론을 합리화하였다.

- **More relevant inference:** 위의 표본크기의 예제에서는 조건부 표준편차가 실제로  $\hat{\theta}$ 의 정확성에 보다 관련이 있어보인다. 회귀분석의 경우는 이처럼 명백하지는 않지만 여전히 유효한 논리이다.
- **Simpler inference:** 조건부 추론은 간편하고 해석하기에 편리하다. 예를 들자면 회귀분석에서 예측변수 사이의 상관관계가 추론에 영향을 주지 않는다. 다음예제에서 비슷한 상황을 알아보자.

표 4.1는 45명의 궤양환자를 대상으로 새로운 수술방법 혹은 기존의 수술방법으로 치료한 후 성공여부에 대한 기록을 보여준다. 새 수술방법이 기존의 수술방법보다 진일보 한 것일까? Fisher는 표의 marginal (16, 29, 21, 24)의 값이 주어진 상황에서 조건부 가설검정을 하는 것이 적절하다고 주장했다. 만약 marginal들이 고정이 되었다면 표에서 4개의 cell중 하나의 cell의 값만 결정이 된다면 나머지 3개의 cell값은 자동적으로 결정된다. 예를 들자면 왼쪽 위편에 위치한 cell의 값  $y = 9$ 가 실제로 두 치료방법에 차이가 없다면 관측치로는 너무 큰 값이 아니인지 여부만 검정하면 된다. 이러한 검정방법을 *Fisher's exact test*라고 한다.

표 4.1 45명의 궤양환자를 새로운 수술방법 혹은 기존의 수술방법으로 치료한 후 성공여부 기록. 새 수술방법이 기존의 수술방법보다 훨씬 좋다고 할 수 있나?

	success	failure	
new	9	12	21
old.	7	17	24
	16	29	45

Ancillary statistics (보조통계량)은 관심이 있는 모수에 관한 직접적인 정보가 없는 통계량이지만 조건부 추론에서 중요한 역할을 한다. 예를 들자면 앞의 예제들에서 표본크기  $N$ , 회귀분석에서  $\mathbf{x}$ 의 공분산 행렬, 그리고 표의 marginal들은 ancillary statistic이다. 사실 이런 조건부 확률에서 이런 ancillary statistic을 random이 아니고 주어진 값으로 간주하는 경우 정보의 손실이 있는 것은 사실이다. 하지만 이런 정보의 손실의 앞에서 얘기한 2가지 장점에 비해 크지 않기 때문에 이러한 조건부 추론이 유용하게 사용된다. 이후 9장에서 생존분석을 다룰 때 이러한 조건부 추론의 활용에 대해 보다 자세히 알아보자.

마지막 예제는 최대우도추정량의 정확성을 어떻게 추정하는냐에 관한 문제이다. 우리는 앞에서 최대우도추정량의 극한분포의 분산이  $1/(n\mathcal{I}_\theta)$ 임을 보였다. 여기서  $n\mathcal{I}_\theta$ 는 다음과 같은 2가지 방법으로 추정할 수 있다. 첫번째는 plug-in 방법이다.

$$\hat{\theta} \sim \mathcal{N}(\theta, 1/(n\mathcal{I}_{\hat{\theta}})),$$

Fisher가 제안한 대안은

$$\hat{\theta} \sim \mathcal{N}(\theta, 1/I(\mathbf{x})),$$

으로 여기서  $I(\mathbf{x})$ 는 observed Fisher information으로 다음과 같이 정의된다.

$$I(\mathbf{x}) = -\ddot{\ell}_{\mathbf{x}}(\hat{\theta}) = -\frac{\partial^2}{\partial \theta^2} \ell_{\mathbf{x}}(\theta) \Big|_{\hat{\theta}}$$

$\mathbb{E}(I\mathbf{x}) = n\mathcal{I}_\theta$ 라는 걸 주목한다면 사실 대표본에서 두 방법은 같은 극한분포를 가진다. 하지만 Fisher는 실제  $\hat{\theta}$ 의 정확성을 알고자 할때 observed Fisher information을 사용하는 것이 더 좋은 방법이라고 제안했다. 이를 확인하기 위해 다음과 같은 상황을 가정해보자. 먼저 표본크기가 20개인 데이터를 다음과 같은 코시분포에서 생성해보자.

$$f_\theta(x) = \frac{1}{\pi} \frac{1}{1 + (x - \theta)^2}.$$

이 경우  $\mathcal{I}_\theta = 1/2$ 임을 보일 수 있다.  $\theta = 0$ 인 경우를 가정하고 우리가 위의 과정을 10,000번 반복한다면 표본크기가 20개인 10,000개의 표본을 얻을 수 있다. 각 표본에서  $1/I(\mathbf{x})$ 을 구한 후 이를 기준으로 11개의 구간 (percentage 기준)  $(0 - 5), (5 - 15), \dots, (85 - 95), (95 - 100)$ 나눈 후 각 구간에  $1/I(\mathbf{x})$ 에 대응되는  $\hat{\theta}$ 을 값을 배정하였다. 참고로

$$I(\mathbf{x}) = 2 \sum_{i=1}^n \frac{(1 - (x_i - \theta)^2)}{(1 + (x_i - \theta)^2)^2}.$$



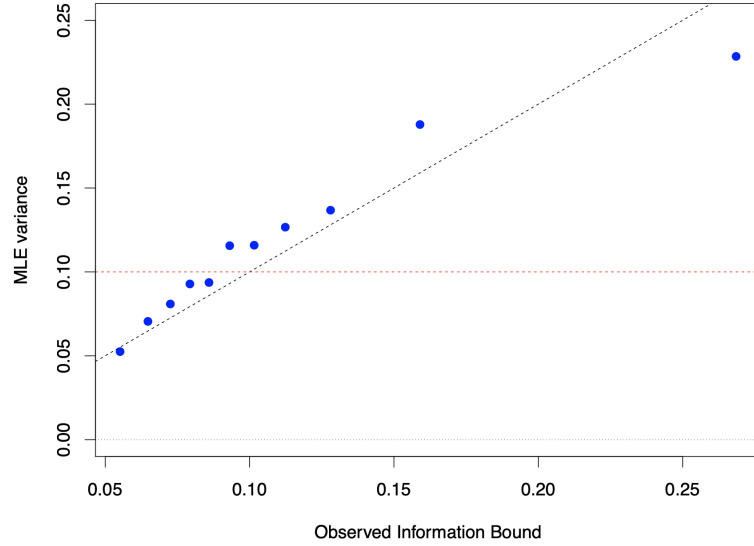


그림 4.2 코시분포에서 생성된 표본( $n = 20$ )을 바탕으로 생성된 MLE의 조건부 분산을 observed information bound  $1/I(\mathbf{x})$ 을 기준으로 제시한 그림. 조건부 분산은  $1/I(\mathbf{x})$  기준으로 11개의 구간 (0-5), (5-15), ..., (85-95), (95-100)으로  $1/I(\mathbf{x})$ 에 대응되는  $\hat{\theta}$ 을 값을 배정한 후 각 구간의  $\hat{\theta}$ 의 표본분산으로 구하였음. 여기서 빨간 수평점선은 unconditional variance  $1/(n\mathcal{I}_\theta)$ 이다.

그림 4.2는 각 구간에 있는  $\hat{\theta}$ 의 값들을 이용해 표본분산을 계산하여 조건부 분산으로 제시하였다. Fisher가 예측한 것처럼 계산된 조건부 분산의 값들이  $1/I(\mathbf{x})$  비슷함을 알 수 있다. 또한 unconditional variance  $1/(n\mathcal{I}_\theta)$ 의 경우는  $1/(20 \cdot 0.5) = 0.10$ 로 고정된 하나의 값을 제시하지만 조건부 분산은 값은 0.05에서 0.20 사이에 다양한 값으로 주어진다.

여기서 observed Fisher information은 보조통계량과 비슷한 역할을 하고 있으며  $n\mathcal{I}(\hat{\theta})$ 보다는  $\hat{\theta}$ 의 정확성에 보다 관련된 정보를 제공하고 계산하는 방법도 간편하다.

또하나 주목할 점이 베이지안과의 밀접한 관련성이다. 만약  $1/I(\mathbf{x})$ 이 클 경우 log-likelihood는 뽀족하고 좁은 봉우리 모양을 가지고 될 것이고 사후분포도 그러한 형태를 띠게 될 것이다.

앞의 결과를 다음과 같이 모수가 여러개인 family로 확장할 수 있다.

$$\hat{\mu} \sim \mathcal{N}_p(\mu, I(x)^{-1})$$

이 경우  $I(\mathbf{x})$ 는 다음과 같은  $p \times p$  행렬로 주어진다.

$$I(\mathbf{x}) = -\ddot{\ell}_{\mathbf{x}}(\mu) = - \left[ \frac{\partial^2}{\partial \mu_i \partial \mu_j} \log f_{\mu}(\mathbf{x}) \right]$$

#### 4.4 Permutation and Randomization

Fisherian의 방법들이 정규분포가정에 너무 의존한다는 비판을 받았을때 Fisher는 정규성 가정의 대안으로 permutation을 제안하였다. 앞에서 언급한 백혈병 환자 예제를 통해서 구체적으로 설명해보자. 47명의 AML(급성

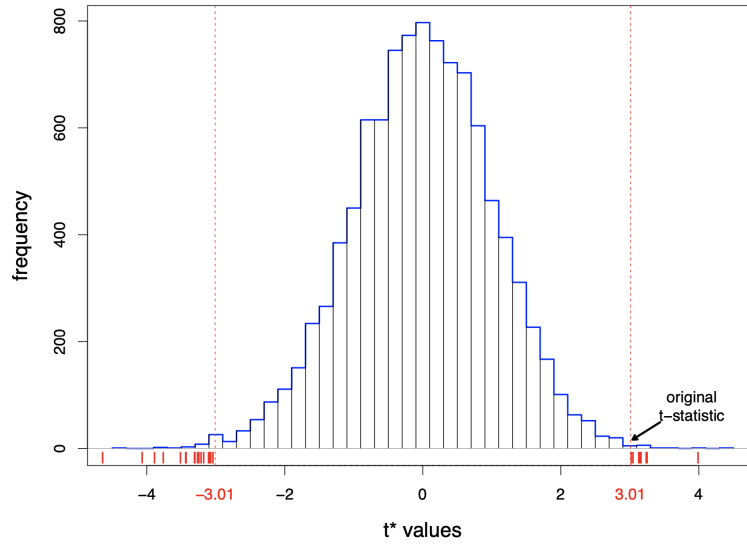


그림 4.3 백혈병 예제에서 유전자 136에 대한 10,000번의 permutation을 한  $t$  통계량의 값. 여기서 빨간색 체크표시는 관측된  $t$  통계량 값 3.13보다 permutation을 통해 계산된  $t$ -통계량의 절대값이 3.13보다 큰 26개의 경우를 나타낸다.

골수성 백혈병)환자와 25명의 ALL(급성 림프구성 백혈병)환자의 차이를 특정 유전자를 통해 알아볼 수 있는지 여부를 검사하기 위해 그룹별 136번 유전자의 활성도를  $t$  통계량을 이용하여 비교하였다.  $t$  통계량의 값은 3.01 이었고 정규성(과 두 집단의 유전자 활성도의 분산이 같다)가정을 하고 귀무가설하에서  $t$  통계량은 자유도 70인  $t$ 분포를 따른다. 이 경우 양측검정을 시행할 경우  $p$ 값은 0.0036이다.

만약 정규성 가정을 하지 않는다면 (그렇지만 여전히 분산은 같다는 가정하에서) 어떤 방법으로 두 그룹의 평균을 비교할 수 있을까? Fisher는 72개 데이터의 permutation을 제안하였다. 만약 72개의 데이터를 임의로 크기가 47과 25인 2개의 집단으로 임의로 나눈다고 가정하자. 이 경우 모든 가능한 permutation의 숫자는  $75!/(47!25!) = 3.146323 \times 10^{20}$ 이다! 실제로는 이중 일부의 permutation만 임의로 선정하여 실행한다. 이렇게 실행된 permutation의 개수를  $B$ 라고 하고 각 permutation에서 생성된 결과를 바탕으로  $t$  검정통계량  $t_i^*$ 을 계산할 수 있다. 이렇게 계산된  $t_i^*$ 를 이용하여 다음과 같이 양측검정의  $p$ 값을 계산한다.

$$p\text{-value} = \#\{|t_i^*| \geq |t|\}/B$$

그림 4.3은 136번 유전자 자료를 이용한  $B = 10,000$ 개의  $t_i^*$ 값으로 이루어진 히스토그램을 보여주고 있다. 총 26개의  $t_i^*$ 의 절대값이 3.01이 보다 큰 것으로 관측되었다. 즉  $p$ 값은 0.0026으로 정규성 가정하에서의  $p$ 값 0.0036과 크게 차이나지는 않는다.

그렇다면 우리는 왜 이렇게 계산된  $p$ 값을 신뢰할 수 있을까? Fisher는 다음과 같은 2가지 점을 주목했다.

- 백혈병 예제에서 72개의 136번 유전자 발현데이터가 모두 같은 확률분포에서 생성되었다고 가정하자. 즉

$$x_i \stackrel{iid}{\sim} f_\mu(x) \text{ for } i = 1, 2, \dots, n.$$

여기서 우리는  $f$ 에 대한 특정확률분포 (즉 정규분포)를 가정하지 않는다. Permutation test에서 우리의 유일한 가정은 exchangeable이다. 즉 두 집단사이에 임의로 자료를 배정하는 것이 어떤 차이도 만들지 않는다는 것이다. Permutation test의 결과로 계산된  $p$ 값이 작은 경우 우리는 두 집단이 나눈 기준이 랜덤이 아니라 귀무가설이 맞지 않기때문에 일어난 결과라고 생각하는 것이 합리적이다.

- Fisher의 무작위 임상시험 (RCT: randomized clinical trial)은 의학시험에 통계적추론을 위한 표준으로 자리잡고 있다. 이런 randomization은 permutation의 결론을 강화해주는 역할을 한다. 물론 위의 백혈병 예제에서 실질적으로 임의로 한명을 특정질병그룹에 배정하는 것은 불가능하다. 하지만 우리가 AML 그룹에서 유전자 발현도가 전반적으로 크다는 사실은 알고 있지만 두 그룹의 차이가 꼭 백혈병의 종류에 기인한 것인지는 알기 어렵다. 즉 AML 그룹의 환자들의 나이가 많거나 또는 체중이 전반적으로 많이 나가는 것이 원인일 수도 있다. 하지만 randomization을 통해서 이런 요인에 의한 차이의 거의 없어짐을 알 수 있다.