번스타인-폰 미시스 정리

이경원

Aug 14, 2023

Abstract

이 노트에서는 번스타인-폰 미시스 정리에 대해 간단히 소개한다.

1 기호와 정의

베이즈 추론의 3요소

- 모수 $\theta \in \Theta$ 와 사전분포(prior distribution)

$$\theta \sim \pi(\theta)$$

- 자료 $\mathbf{X}_n = (X_1, \cdots, X_n)^T \in (\mathcal{X}^n, \mathcal{A}^n)$ 와 가능도(likelihood)

$$\mathbf{X}_n \sim p_{n,\theta}(\mathbf{x}_n)$$

- 사후분포(posterior distribution)

$$\pi(\theta|\mathbf{x}_n) = \pi(\theta|\mathbf{X}_n = \mathbf{x}_n)$$

정의 1.1 (통계적 모형(statistical model)). *모수 공간 Θ 에 대해, 표본 공간(sample space) $(\mathcal{X}^n, \mathcal{A}^n)$ 위의 모든 확률 측도의 모임 $\{P_{n,\theta} : \theta \in \Theta\}$ 을 통계적 모형이라 한다.*

참고 1.1. *종종 확률밀도함수는 $\pi(d\theta)$ 와 같이 표기하고*

$$\Pi(\theta \in A) = \int_A \pi(d\theta)$$

와 같이 쓰기도 한다.

베이즈 정리

정리 1.1 (베이즈 정리). *사후분포는 다음과 같이 계산된다*

$$\pi(\theta|\mathbf{x}_n) = \frac{p_{n,\theta}(\mathbf{x}_n)\pi(\theta)}{\int_\Theta p_{n,\theta}(\mathbf{x}_n)\pi(\theta)} \propto p_{n,\theta}(\mathbf{x}_n)\pi(\theta). \tag{1.1}$$

모수적 모형과 비모수적 모형

정의 1.2 (모수적 모형과 비모수적 모형)**.** *모수 공간* Θ*가 유한 차원일 때, 모형* $\{P_{n,\theta} : \theta \in \Theta\}$*를 모수적 모형이라 하고, 무한 차원일 때 비모수적 모형이라 한다.*

점근적 성질 자료의 수(n)이 충분히 커졌을 때 통계적 모형이 갖는 성질들을 점근적 성질(asymptotic property)이라 한다. 대표적으로, 추정량이 참 모수에 충분히 가까워지는 일치성과 어떤 속도로 가까워지는 지를 의미하는 수렴 속도, 어떤 모양으로 가까워지는 지를 의미하는 점근 분포가 있다.

정의 1.3 (일치성(consistency))**.** *모수 공간 위의 적당한 거리* d*에 대해 모수* θ*의 추정량* $\hat{\theta}(\mathbf{X}_n)$*가 다음을 만족할 때, 추정량이 참 모수* θ_0*에 대해 일치성을 갖는다고 한다*

$$d(\hat{\theta}(\mathbf{X}_n), \theta_0) \xrightarrow{P_{n,\theta_0}} 0 \tag{1.2}$$

as $n \to \infty$.

정의 1.4. *사후 일치성(posterior consistency)과 사후 수렴속도(posterior contraction rate) 모수 공간 위의 적당한 거리* d*와 임의의 수열* $M_n \to \infty$, $\epsilon_n \to 0$*에 대해*

$$\Pi(\theta : d(\theta, \theta_0) > M_n \epsilon_n | \mathbf{X}_n) \xrightarrow{P_{n,\theta_0}} 0 \tag{1.3}$$

as $n \to \infty$*를 만족할 때 사후 일치성을 갖는다고 한다. 이때* ϵ_n*을 사후 수렴속도라 한다.*

참고 1.2 (약한 일치성과 강한 일치성)**.** *일치성의 정의에서 확률수렴을 거의 확실한 수렴(almost surely convergence)으로 바꿀 수 있을 때, 강한 일치성을 갖는다고 한다.*

정의 1.5 (점근 분포(asymptotic distribution))**.** *모수* θ*의 추정량* $\hat{\theta}(\mathbf{X}_n)$*가 확률변수* Z*에 대해*

$$P(\hat{\theta}(\mathbf{X}_n) \in A) \to P(Z \in A) \tag{1.4}$$

as $n \to \infty$*가 모든 잴 수 있는* A*에 대해 성립할 때,* Z*의 분포를* $\hat{\theta}(\mathbf{X}_n)$*의 점근 분포라 한다. 특히,* Z*의 분포가 정규 분포일 때는 추정량* $\hat{\theta}(\mathbf{X}_n)$*이 점근 정규성(asymptotic normality)을 갖는다고 한다.*

참고 1.3 (사후 분포의 점근 분포)**.** *사후 분포* $\Pi(\cdot|\mathbf{X}_n)$*은 자료* $\mathbf{X}_n$*에 대한 확률변수이다. 적당한 분포 사이의 괴리도 혹은 거리* β*와 확률 분포* D*에 대해*

$$\beta(\Pi(\cdot|\mathbf{X}_n); D) \xrightarrow{P_{n,\theta_0}} 0 \tag{1.5}$$

가 성립하면 분포 D*를 사후 분포의 점근 분포라 한다. 특히, 사후 분포의 점근 분포가 정규 분포일 때는 번스타인-폰 미시스 정리(Bernstein-von Mises theorem)가 성립한다고 한다.*

참고 1.4. *이번 워크샵의 논문 Castillo and Rousseau [2015]는 비모수적 모형에서 범함수(functional)가 점근 정규성을 갖기 위한 조건들을 규명한 논문이다. 이 발표에서는 간략하게 모수적 모형에서의 점근 점규성에 대해 소개한다.*

2 가능도의 점근 정규성

일치성과 사후 일치성은 추정량, 혹은 사후분포가 참 모수값에 충분히 가까워진다는 것을 알려준다. 점근 정규성은 추정량, 혹은 사후분포가 참 모수에 어떤 형태로 가까워지는가를 알려준다.

다음의 정리는 최대 가능도 추정량의 점근 정규성을 알려준다.

정리 2.1 (최대 가능도 추정량의 점근 정규성)**.** *자료가 서로 독립이고 동일하게 밀도함수가 p_θ 인 모형을 따른다고 하자. 즉, 다음과 같다고 하자*

$$X_1, \cdots, X_n \overset{i.i.d.}{\sim} p_\theta$$

라 하자. 다음의 조건들이 만족될 때,

(R0) *(식별 가능성) $p_{\theta_1} = p_{\theta_2} \Rightarrow \theta_1 = \theta_2$*

(R1) *(공통의 토대) 분포의 토대(support) $\{x : p_\theta(x) > 0\}$ 는 모수 $\theta \in \Theta$ 에 의존하지 않는다.*

(R2) *(열린 모수공간) 모수공간 Θ 는 k 차원 실수 공간 $\mathbb{R}^k$ 에서의 열린 집합이다.*

(R3) *(미분가능한 로그가능도) 모든 관측 자료 $\mathbf{x}_n = (x_1, \cdots, x_n)^T, x_i \in \mathcal{X}$ 에 대하여 로그 가능도함수*

$$l_n(\theta) = \sum_{i=1}^n \log p_\theta(x_i)$$

의 일차 및 이차 편도함수 $\dot{l}_n(\theta)$, $\ddot{l}_n(\theta)$ 가 존재하며 이들은 모두 연속함수이다.

(R4) *(적분 또는 합의 미분의 순서 교환 가능) 자료 $\mathbf{X}_n = (X_1, \cdots, X_n)^T$ 의 함수 $T(\mathbf{X}_n)$ 의 기댓값 $\mathbb{E}_\theta[T(\mathbf{X}_n)]$ 이 존재할 때,*

$$\frac{\partial^r}{\partial\theta^r}\mathbb{E}_\theta[T(\mathbf{X}_n)] = \int T(\mathbf{x}_n)\frac{\partial^r}{\partial\theta^r}p_{n,\theta}(\mathbf{x}_n)d\mathbf{x}_n,$$

(R5) *(정보량의 존재) 모든 모수 $\theta \in \Theta$ 에 대하여 피셔 정보량(Fisher information)*

$$I(\theta) = Var_\theta\left(\frac{\partial}{\partial\theta}\log p_\theta(X_1)\right)$$

가 실수 또는 실수 행렬로 잘 정의될 수 있으며 그 역수 또는 역행렬이 존재한다.

(R6) *(일치성) 자료 $\mathbf{X}_n$ 을 이용한 θ 의 최대가능도 추정량*

$$\hat{\theta}_n^{MLE} = \arg\max_{\theta\in\Theta} p_{n,\theta}(x_i)$$

가 가능도 방정식

$$\dot{l}_n(\theta) = \sum_{i=1}^n \frac{\partial}{\partial\theta}\log p_\theta(X_i) = 0$$

의 단 하나뿐인 근이고 일치성을 갖는다. 즉,

$$\hat{\theta}_n^{MLE} \xrightarrow{P_{n,\theta}} \theta.$$

R7) *(잉여항의 크기) 로그가능도함수의 삼차 편도함수가 연속함수로 존재하며*

$$\sup_{\theta\in\Theta}\left\|\frac{\partial^3}{\partial\theta^3}\log p_\theta(X_1)\right\| \le M(X_1), \quad \mathbb{E}_\theta[M(X_1)] < \infty$$

을 만족하는 $M(X_1)$ 이 존재한다.

참 모수가 θ_0 였을 때, 최대가능도 추정량 $\hat{\theta}_n^{MLE}$ 는 다음의 극한분포를 갖는다.

$$\sqrt{n}(\hat{\theta}_n^{MLE} - \theta_0) \xrightarrow{d} N(0, [I(\theta_0)]^{-1}) \tag{2.1}$$

3 번스타인-폰 미시스 정리

번스타인-폰 미시스 정리(Bernstein-von Mises theorem; BvM theorem)은 사후분포의 정규근사에 대한 정리이다.

Theorem 3.1 (번스타인-폰 미시스 정리 (Lehmann and Casella [1998]의 정리 6.8.2))**.** *자료가 서로 독립이고 동일하게 밀도함수가 p_θ 인 모형을 따른다고 하자. 2.1의 조건에 이어 다음과 같은 조건들을 생각하자*

(R8) *(로그가능도 꼬리 조건) $\delta > 0$ 에 대해,*

$$P_{\theta_0}\left(\sup_{\|\theta-\theta_0\|>\delta}\frac{1}{n}(l_n(\theta) - l_n(\theta_0)) < -\epsilon\right) = 1$$

이 충분히 큰 모든 n 에 대해 성립한다.[1]

(R9) *(양의 사전밀도함수와 연속성) 사전분포의 밀도함수 $\pi(\theta)$ 가 θ_0 에서 연속이고, $\pi(\theta_0 > 0)$.*

(R10) *(사전평균의 존재성)*

$$\int \|\theta\|\pi(\theta)d\theta < \infty.$$

$$t = \sqrt{n}(\theta - T_n), \quad T_n = \theta_0 + \frac{1}{nI(\theta_0)}\dot{l}_n(\theta_0)$$

의 사후분포 $\pi(t|\mathbf{x}_n)$ 에 대해 다음이 성립한다.

(i) 조건 ***(R0)-(R9)*** *하에서 다음이 성립한다*

$$\int\left|\pi^*(t|\mathbf{x}_n) - \sqrt{I(\theta_0)}\phi\left[t\sqrt{I(\theta_0)}\right]\right|dt \xrightarrow{p} 0. \tag{3.1}$$

(ii) 조건 ***(R0)-(R10)*** *하에서 다음이 성립한다*

$$\int(1+|t|)\left|\pi^*(t|\mathbf{x}_n) - \sqrt{I(\theta_0)}\phi\left[t\sqrt{I(\theta_0)}\right]\right|dt \xrightarrow{p} 0. \tag{3.2}$$

증명. 김현연 학생이 증명 예정 □

[1] 이는 모형의 꼬리 쪽의 로그가능도가 충분히 작다는 것을 의미한다.

참고 3.1. *정리에서 (i)는 사후분포가 정규분포에 충분히 가깝다는 것을 의미하고, (ii)는 사후 분포의 점근적 적률에 대한 정보로 사후 평균(Lehmann and Casella [1998] Theorem 6.8.3), 사후 중앙값(Bickel and Doksum [2015] Theorem 5.5.3) 등 베이즈 추정량의 일치성과 점근 정규성을 보장한다.*

참고 3.2. *최대 가능도 추정량의 점근 정규성으로부터 위의* T_n *을* $\hat{\theta}_n^{MLE}$ *로 사용할 수도 있다.*

참고 3.3. *번스타인-폰 미시스 정리는 사후 분포를 정규 분포와 충분히 가까워진다는 것을 알려준다. 즉,*

$$\sqrt{n}(\theta - \hat{\theta}_n^{MLE})|\mathbf{x}_n \approx N(0, I(\theta_0)^{-1}) \tag{3.3}$$

와 같은 근사가 가능하다는 것을 의미하고 이를 바탕으로 근사 신용구간 등을 계산할 수 있다.

4 반더바르트의 번스타인-폰 미시스 정리

3절에서 소개한 번스타인-폰 미시스 정리의 가정들은 사후분포의 테일러 근사에서 필요한 조건들로 다소 강한 조건들이다. 이번에는 결정 규칙의 관점에서 다른 형태의 번스타인-폰 미시스 정리를 소개한다.

정리를 소개하기 앞서 필요한 조건들을 소개한다.

정의 4.1 (제곱평균 미분가능성(differentiable in quadratic mean))**.** *자료가 서로 독립이고 동일하게 밀도 함수가* p_θ *인 모형을 따른다고 하자. 모형* $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ *와 모든 랜덤수열* $h_n = O_{P_{\theta_0}}(1)$ *에 대해*

$$\int \left(p_{\theta_0+h}^{1/2} - p_{\theta_0}^{1/2} - \frac{1}{2} h^T \dot{l}_n(\theta_0) p_{\theta_0}^{1/2} \right)^2 d\mu = o(\|h\|_2^2) \tag{4.1}$$

가 성립하면 모형이 θ_0 *에서 제곱평균 미분가능하다고 한다.*

참고 4.1. *모형이* θ_0 *에서 제곱평균 미분가능하면*

$$\mathbb{E}_{\theta_0}[\dot{l}_{\theta_0}] = 0, \quad \int \dot{l}_n^2(\theta_0) P_{\theta_0}(d\theta) < \infty \tag{4.2}$$

를 만족하고 피셔 정보는 $I(\theta_0) = \mathbb{E}_{\theta_0}[\dot{l}_n(\theta_0)\dot{l}_n(\theta_0)^T]$ *로 정의된다.*

정의 4.2 (국소점근정규성(locally asymptotically normality, LAN))**.** *모형* $\mathcal{P} = \{P_{n,\theta} : \theta \in \Theta\}$ *를 생각하자. 모든* $h_n \to h$ *에 대해 다음을 만족하는 적당한 정칙 행렬* r_n, $I(\theta_0)$, *랜덤 벡터* $\Delta_{n,\theta}$ *가 존재하면 모형이* θ_0 *에서 국소점근정규성을 만족한다고 한다*

$$\begin{aligned} \Delta_{n,\theta_0} &\xrightarrow{d} N(0, I(\theta_0)), \\ \log \frac{dP_{n,\theta_0+r_n^{-1}h_n}}{dP_{n,\theta_0}} &= h^T \Delta_{n,\theta_0} - \frac{1}{2} h^T I(\theta_0) h + o_{P_{n,\theta_0}}(1). \end{aligned} \tag{4.3}$$

국소점근정규성은 모형의 밀도함수가 θ_0 근처에서, 국소적으로 정규분포로 표현될 수 있음을 의미한다. 이 조건은 본 워크샵의 논문Castillo and Rousseau [2015]에서도 등장한다.

참고 4.2 (van der Vaart [1998]의 정리 7.7.2)**.** *자료가 서로 독립이고 동일하게 밀도함수가* p_θ *인 모형을 따른다고 하자. 모형이* θ_0 *에서 제곱평균 미분가능하면* $r_n = \sqrt{n}I$ *인 국소점근정규성을 갖는다. 구체적으로는,* $\mathbb{G}_n(T) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} T(X_i)$ *라 했을 때, 모형* $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ *와 모든 랜덤수열* $h_n = O_{P_{\theta_0}}(1)$ *에 대해 다음이 성립한다*

$$\log \prod_{i=1}^{n} \frac{p_{\theta_0 + h_n/\sqrt{n}}(x_i)}{p_{\theta_0}(x_i)} = h_n^T \mathbb{G}_n[\dot{l}_{\theta_0}] - \frac{1}{2} h_n^T I_{\theta_0} h_n + o_{P_{\theta_0}}(1), \tag{4.4}$$

정리 4.1 (번스타인-폰 미시스 정리 (van der Vaart [1998]의 정리 10.1))**.** *자료가 서로 독립이고 동일하게 밀도함수가* p_θ *인 모형을 따른다고 하자. 모형* $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ *에 대해*

1. *모형은* θ_0 *에서 제곱평균 미분가능하다.*
2. *피셔 정보* $I(\theta_0)$ *는 정칙행렬(regular matrix)이다.*
3. *임의의 양수* $\epsilon > 0$ *에 대해 다음을 만족하는 검정열* $\{\phi_n\}$ *이 존재한다*

$$P_{n,\theta_0}\phi_n \to 0, \qquad \sup_{\|\theta - \theta_0\| \geq \epsilon} P_{n,\theta}[1 - \phi_n] \to 0 \tag{4.5}$$

4. *사전분포는* θ_0 *근방에서 르벡 측도에 대해 절대연속(absolutely continuous)이고 양의 밀도함수를 갖는다.*

이 만족되면 사후분포 $\pi^*(t|\mathbf{x}_n)$ *은 정규분포로 전변동 노름(total variation norm)에 관해 수렴한다. 즉, 다음이 성립한다*

$$\sup_A \left| \Pi^*(A|\mathbf{X}_n) - N(A; I(\theta_0)^{-1}\mathbb{G}\dot{l}_n(\theta_0), I(\theta_0)^{-1}) \right| \xrightarrow{P_{n,\theta_0}} 0. \tag{4.6}$$

증명*.* 한재욱 학생이 증명 예정 □

참고 4.3. *세 번째 조건은 임의의 양수* $\epsilon > 0$ *에 대해 두 가설*

$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \|\theta - \theta_0\| \geq \epsilon$$

에 대해 균등일치성을 갖는 검정(uniformly consistent test)의 존재성을 의미한다. 이는 모수 공간에서 참 모수 θ_0 *와 그 주변을 제외한 공간을 구분할 수 있는가에 대한 조건으로 분리 가설(separation hypothesis)이라고도 불리운다. 모형이 식별 가능하고 밀도함수가 연속이고 모수 공간이 옹골집합(compact set)이면 이 조건이 성립함이 알려져 있다 [van der Vaart, 1998].*

참고문헌

P. J. Bickel and K. A. Doksum. *Mathematical Statistics: Basic Ideas and Selected Topics*. Texts in Statistical Science. CRC Press, Boca Raton, 2nd ed edition, 2015. ISBN 978-1-4987-2380-0 978-1-4987-2268-1.

I. Castillo and J. Rousseau. A Bernstein–von Mises theorem for smooth functionals in semiparametric models. *The Annals of Statistics*, 43(6), Dec. 2015. ISSN 0090-5364. doi: 10.1214/15-AOS1336.

P. D. Hoff. *A First Course in Bayesian Statistical Methods*. Springer Texts in Statistics. Springer New York, New York, NY, 2009. ISBN 978-0-387-92299-7 978-0-387-92407-6. doi: 10.1007/978-0-387-92407-6.

E. L. Lehmann and G. Casella. *Theory of Point Estimation*. Springer Texts in Statistics. Springer, New York, 2nd ed edition, 1998. ISBN 978-0-387-98502-2.

A. W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, UK ; New York, NY, USA, 1998. ISBN 978-0-521-49603-2.

김우철. *수리통계학*. 민영사, Mar. 2012. ISBN 978-89-8134-135-0.