6.S951 Modern Mathematical Statistics	Fall 2024

Lecture 20 — November 26, 2024

Prof. Stephen Bates

Scribe: Haoyuan Sun

1 Outline

Agenda:

1. Moment estimates
2. Exponential family
3. Asymptotic normality of MLE in exponential family

Last time:

1. Intro to asymptotics
2. Delta method

2 Moment estimator

Setting: i.i.d. random variables $\{X_i\}_{i=1}^n \sim P$ and family of distributions $\{P_\theta \mid \theta \in \Theta\}$ that may not contain P.

Definition 1 (Moment estimator)**.** *Given $f : \mathcal{X} \to \mathbb{R}^k$, a moment estimator $\hat{\theta}$ is the value of θ that solves the equation*

$$\mathbb{E}_\theta f(X) = \frac{1}{n}\sum_{i=1}^{n} f(X_i).$$

Example 2. Given the family of Gaussians $P_\theta = \mathcal{N}(\theta_1, \theta_2^2)$, then $\hat{\theta}_1 = \bar{X}, \hat{\theta}_2 = \sqrt{\overline{X^2} - \bar{X}^2}$ are moment estimators.

Define $e_i = \Theta \to \mathbb{R}^k$ so that $e(\theta) = \mathbb{E}_\theta f(x)$. If e is one-to-one, then we have $\hat{\theta} = e^{-1}(\frac{1}{n}\sum_i f(X))$ (wherever the inverse is well-defined). Furthermore, if e^{-1} is differentiable, then we can employ the delta method to get the asymptotics of $\hat{\theta}$.

Proposition 3. *Let $\mu = \mathbb{E}f(x)$ and $\Sigma = \mathbb{E}(f(x)-\mu)(f(x)-\mu)^\top$. Suppose e^{-1} exists and differentiable at μ, then the moment estimator $\hat{\theta}$ satisfies*

$$\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{d} \mathcal{N}\left(0, (e^{-1})'(\mu)\,\Sigma\,(e^{-1})'(\mu)^\top\right).$$

Proof. Note that $\sqrt{n}(\frac{1}{n}\sum_i X_i - \mu) \xrightarrow{d} \mathcal{N}(0,\Sigma)$. And $\hat{\theta} = e^{-1}(\frac{1}{n}\sum_i f(X_i))$ is differentiable. So, the claim follows immediately from the delta method. □

Remark: This result holds even if P is not contained in the model class.

Q: When is e^{-1} is differentiable at μ?
A: Use inverse function theorem (or implicit function theorem).

Theorem 4 (Inverse function theorem)**.** *Suppose $\Theta \in \mathbb{R}^d$ and $e : \Theta \to \mathbb{R}^k$ is continuously differentiable at θ_0 with a non-singular derivative, then e is locally invertible and e^{-1} is locally continuously differentiable at $e(\theta_0)$.*

Example 5. Consider the family of Beta distributions $P_{(\alpha,\beta)}(x) = c(\alpha,\beta)x^{\alpha-1}(1-x)^{\beta-1}$, then

$$\mathbb{E}X = \frac{\alpha}{\alpha+\beta},$$
$$\mathbb{E}X^2 = \frac{(\alpha+1)\alpha}{(\alpha+\beta+1)(\alpha+\beta)}.$$

Note that $e(\alpha,\beta) = \mathbb{E}_{(\alpha,\beta)}(\bar{X},\overline{X^2})$ is continuously differentiable and has non-singular inverse, so we can apply the previous proposition.

3 Exponential family

We consider a family of distributions (statistical model) with "nice" structures.

Definition 6 (Exponential family)**.** *For $\theta \in \mathbb{R}^m, t : \mathcal{X} \to \mathbb{R}^k$, the exponential family can be parameterized as*

$$P_\theta(x) = c(\theta)h(x)e^{\theta^\top t(x)},$$

where

- *$c(\theta)$ is normalizing constant,*
- *$h(x)$ is the base measure,*
- *$\exp(\theta^\top t(x))$ is the "exponential tilt" that up(down)-weights based on $\theta^\top t(x)$.*

Remark 1: This can be defined over either the Lebesgue or counting measure.

Remark 2: $t(x)$ is a sufficient statistic and the parameter space can be defined as

$$\Theta = \left\{ \theta \mid \int h(x) \exp(\theta^\top t(x)) \, \mathrm{d}x < \infty \right\}.$$

Example 7. Gaussian, binomial, Poisson, and exponential distributions all belong to the exponential family.

Q: Why is the exponential family nice?
A1: They play well with independent samples. Given i.i.d. $\{X_i\}_{i=1}^n \sim P_\theta$, then

$$P_\theta(X_1, \ldots, X_n) = \prod_{i=1}^n P_\theta(X_i) = c(\theta)^n \left(\prod_{i=1}^n h(X_i) \right) \exp\left(\theta^\top \sum_{i=1}^n t(X_i) \right).$$

Therefore, the resulting distribution also belongs to the exponential family with

$$\tilde{c}(\theta) = c(\theta)^n$$
$$\tilde{h}(x_1, \ldots, x_n) = \prod_i h(x_i)$$
$$\tilde{t}(x_1, \ldots, x_n) = \sum_i t(x_i)$$

and a sufficient statistics of the same dimension m.

A2: They have nice smoothness property.

Theorem 8. *The function* $c^{-1} : \theta \mapsto \int h(x) \exp(\theta^\top t(x)) \, \mathrm{d}x$ *is infinitely differentiable with the derivative equal to*

$$\frac{\partial^p c^{-1}}{\partial^{j_1} \theta_1 \cdots \partial^{j_k} \theta_k} = \int h(x) t_1^{j_1}(x) \cdots t_k^{j_k}(x) e^{\theta^\top t(x)} \, \mathrm{d}x,$$

where $j_1 + \cdots + j_k = p$.

Corollary 9. *Let* $\ell_\theta(x) = \log P_\theta(x)$ *and* $\ell'_\theta(x) = t(x) - \mathbb{E}_\theta t(x)$ *be the* *score function. Then, the score has mean 0, i.e.* $\mathbb{E}_\theta \ell'_\theta(x) = 0$. *Therefore, MLE is a moment estimator for the exponential family.*

4 Asymptotics of MLE in exponential family

Theorem 10. *Consider i.i.d. samples* $X_i \sim P$. *Let* $\hat{\theta}$ *be the MLE of the exponential family, then*

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, \mathcal{T}), \quad \mathcal{T} = (e^{-1})'(\theta_0) \, \mathrm{Cov}[t(x)] \, (e^{-1})'(\theta_0),$$

where $\theta_0 = \mathrm{argmax}_\theta \, \mathbb{E}[\ell_\theta(x)]$.

Proof sketch.

1. Note that MLE is a moment estimator,
2. Apply delta method.

□

Remark: This result holds even if P is not in exponential family.