STATS300B – Lecture 8

Julia Palacios
Scribed by Michael Howes

01/27/22

Contents

1 Fisher information

Suppose we have a model $\mathcal{P} = \{P_\theta : \theta \in \Omega\}$ and $\Omega \subseteq \mathbb{R}^d$. Define $l_\theta(x) = \log(p_\theta(x))$ where p_θ is the density of P_θ. The *Fisher information of* θ is defined to be the matrix,

$$I(\theta) = \mathbb{E}[\nabla l_\theta(X) \nabla l_\theta(X)^T] = \mathrm{Var}_\theta(\nabla l_\theta(X)).$$

If integration and differentiation can be exchanged, then

$$I(\theta) = -\mathbb{E}[\nabla^2 l_\theta(X)].$$

Under appropriate assumptions about the regularity of the map $\theta \mapsto l_\theta$, then for all θ in the interior of Ω,

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{d} \mathsf{N}_k(0, I(\theta)^{-1}),$$

where $\widehat{\theta}_n$ is the MLE from a sample $X_1, \ldots, X_n \overset{\text{iid}}{\sim} P_\theta$. We can combine this result with the delta method. If $g : \Omega \to \mathbb{R}$ is differentiable at θ, then

$$\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \xrightarrow{d} N(0, \nabla g(\theta)^T I(\theta)^{-1} \nabla g(\theta)).$$

And there is also a multivariate version, if $g : \Omega \to \mathbb{R}^p$, then

$$\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \xrightarrow{d} \mathsf{N}_p\left(0, (Dg(\theta))^T I(\theta)^{-1} (Dg(\theta))\right).$$