STAT 701, Eric Slud February 19, 2025

Large-sample Existence of MLEs in Natural Exponential Families

This handout is an expanded detailed proof of Theorem 5.2.2 in Bickel and Doksum. The context is *iid* data $X_i \in \mathbb{R}^d$, $i = 1, \ldots, n$, with full-rank natural exponential family density (or probability mass function)

$$X_i \sim f(x, \theta) = h(x) \exp\{\theta' T_0(x) - A(\theta)\} \quad , \qquad \theta \in \Theta = \mathcal{E} \subset \mathbb{R}^k \tag{1}$$

where the natural parameter space $\mathcal{E}$ is convex and open. Let $\theta_0 \in \mathcal{E}$ denote the 'true' parameter values for the model governing X_i, and let $T(\underline{X}) = \sum_{i=1}^n T_0(X_i) \in \mathbb{R}^k$. Recall that 'full rank' means that 1 and $T_0(X)$ are linearly independent as random variables and implies that $V_0 = \text{Var}_{\theta_0}(T_0(X_1)) = \nabla_\theta^{\otimes 2} A(\theta_0)$ is positive definite. Also recall that under these conditions, $\mu_0 \equiv E_{\theta_0}(T_0(X_1)) = \nabla_\theta A(\theta_0)$ and that $A(\theta)$ is a strictly concave function on $\mathcal{E}$. Finally, recall that the MLE estimator, if it exists, is the unique solution of the GMM (General Method of Moments) estimating equation $T(\underline{X}) = A(\hat{\theta})$.

Bickel and Doksum proved (Theorem 2.3.1) that the MLE exists for the data $\underline{X}$ with $t_0 \equiv T(\underline{X})$ if and only if

$$(*) \qquad \text{for all} \quad c \in \mathbb{R}^k \quad \text{with} \quad \|c\| = 1, \qquad P_{\theta_0}(\, c'\,(T(\underline{X}) - t) > 0) \Big|_{t=t_0} > 0$$

To prove that this event occurs with arbitrarily large probability when n is large, we will prove in terms of independent *iid* P_{θ_0}-distributed samples $\underline{X} = (X_1, \ldots, X_n)$, $\underline{X}^* = (X_1^*, \ldots, X_n^*)$, that

$$\forall\ \epsilon > 0 \ : \ \liminf_{n \to \infty} P_{\theta_0}\Big[\inf_{c:\ \|c\|=1} P_{\theta_0}\Big\{ c'\,(T(\underline{X}) - T(\underline{X}^*)) > 0 \,\Big|\, \underline{X}^* \Big\} > 0 \Big] \geq 1 - \epsilon \tag{2}$$

The main ideas of the proof are the Central Limit Theorem (CLT) and the multivariate *iid* normal density being everywhere positive and decreasing as a function of the norm of its argument. The CLT says that $Z_n \equiv V_0^{-1/2} \sqrt{n}\,(T(\underline{X}) - \mu_0)$ has asymptotic multivariate-normal distribution on $\mathbb{R}^k$ with mean $\underline{0}$ and identity variance-matrix. The CLT says that for fixed large K, which will depend on ϵ in (2), for all sufficiently large n,

$$P_{\theta_0}\Big[\sqrt{n}\, \|V_0^{-1/2}\,(T(\underline{X}^*) - \mu_0)\| \leq K \Big] \geq 1 - \epsilon \tag{3}$$

where $V_0^{-1/2}$ is the inverse of the symmetric square root of V_0, which exists because V_0 is positive definite. Let $B_{n,K}$ denote the event whose probability is lower-bounded in (3).

The CLT applied to $T(\underline{X})$ also implies for all sufficiently large n,

$$\begin{aligned}
& \inf_{c:\, \|c\|=1} P_{\theta_0}\Big\{ \sqrt{n}\, c'\,(T(\underline{X}) - T(\underline{X}^*)) \;>0 \,\Big|\, \underline{X}^*,\; B_{n,k} \Big\} \\
= \;& \inf_{w:\, \|w\|=1} P_{\theta_0}\Big\{ \sqrt{n}\, w'\, V_0^{-1/2}\,(T(\underline{X}) - T(\underline{X}^*)) \;>0 \,\Big|\, \underline{X}^*,\; B_{n,k} \Big\} \\
\geq \;& \inf_{w:\, \|w\|=1} P_{\theta_0}\Big\{ \Big[\sqrt{n}\, w'\, V_0^{-1/2}\,(T(\underline{X}) - T(\underline{X}^*)) \;>0 \Big] \cap \\
& \qquad \Big[\|\sqrt{n}\, V_0^{-1/2}\,(T(\underline{X}) - T(\underline{X}^*))\| \leq 1 \Big] \,\Big|\, \underline{X}^*,\; B_{n,k} \Big\}
\end{aligned}$$

For each unit vector w, the conditional probability in the last line above is the probability that the asymptotically $\mathcal{N}_k(\underline{0},\, I_{k\times k})$ distributed random vector Z_n falls within a half-ball of radius 1 centered at $\sqrt{n}\, V_0^{-1/2}\,(T(\underline{X}^*) - \mu_0)$ which lies within the large ball $B_{K+1}(\underline{0}) \subset \mathbb{R}^k$. But such half-balls all have the same volume $b_k > 0$ not depending on 0, and the limiting multivariate-normal distribution of Z_n has density bounded below everywhere on $B_{K+1}(\underline{0})$. Therefore, we have proved that for all large n

$$\inf_{c:\, \|c\|=1} P_{\theta_0}\Big\{ \sqrt{n}\, c'\,(T(\underline{X}) - T(\underline{X}^*)) \;>0 \,\Big|\, \underline{X}^*,\; B_{n,k} \Big\} \;\geq\; \frac{1}{2}\, b_k\, (2\pi)^{-k/2}\, e^{-(K+1)^2/2}$$

The constant on the right-hand side is depressingly small for large K, but all we had to do is show is that it is positive as long as $\underline{X}^*$ satisfies the event $B_{n,k}$. The ϵ lost in (3) is just the (limiting) upper bound for $P_{\theta_0}(B^c_{n,k})$. So our proof is complete. □