Asymptotics of Maximum Partial Likelihood Estimators in General Semiparametric Multiplicative Hazard Models Under First Order Differentiability

Hanxiang Peng

Department of Mathematical Sciences
Indiana University-Purdue University at Indianapolis

Joint work with Fei Tan, FAMU

November 11, 2009

Outline

Free Knot Spline Cox Models

Consider a parametric Cox model:

$$h(t) = h_0(t)\exp(g_\theta(t, Z(t))), \qquad (1)$$

where g_θ is a “smooth” function. Our candidate of g_θ is a **free knot spline**.

Free Knot Spline Cox Models

Consider a parametric Cox model:

$$h(t) = h_0(t)\exp(g_\theta(t, Z(t))), \tag{1}$$

where g_θ is a "smooth" function. Our candidate of g_θ is a **free knot spline**.

- A **quadratic free-knot polynomial spline with knots in time**:

$$g_\theta(t, z) = (\beta_1 + \beta_2 t + \beta_3 t^2 + \beta_4 (t - \gamma)_+^2) z,$$

where $\theta = [\beta_1, \beta_2, \beta_3, \beta_4, \gamma]^\top$ is the parameter of interest. The knot γ can be a **threshold value** such as a changepoint.

ii A **quadratic free-knot polynomial spline with knots in covariates**:

$$g_\theta(z) = \beta_1 z + \beta_2 z^2 + \beta_3 (z - \kappa)_+^2, \tag{2}$$

where $\theta = [\beta_1, \beta_2, \beta_3, \kappa]^\top$ is the parameter of interest. The knot κ can be a **threshold value** such as a nadir (of BMI), a changepoint, etc.

ii A **quadratic free-knot polynomial spline with knots in covariates**:

$$g_\theta(z) = \beta_1 z + \beta_2 z^2 + \beta_3 (z - \kappa)_+^2, \tag{2}$$

where $\theta = [\beta_1, \beta_2, \beta_3, \kappa]^\top$ is the parameter of interest. The knot κ can be a **threshold value** such as a nadir (of BMI), a changepoint, etc.

iii Applications? B-splines? Natural Splines?

iv g_θ has first continuous derivative,

$$\dot{g}_\theta(z) = [z, z^2, (z-\kappa)^2 1_{\{z>\kappa\}}, -2\beta_3(z-\kappa)1_{\{z>\kappa\}}]^\top,$$

but the 2nd derivative **does not exist at knot** κ.

iv g_θ has first continuous derivative,

$$\dot{g}_\theta(z) = [z, z^2, (z-\kappa)^2 1_{\{z>\kappa\}}, -2\beta_3(z-\kappa)1_{\{z>\kappa\}}]^\top,$$

but the 2nd derivative **does not exist at knot** κ.

v Consistency? Asymptotic Normality? Under continuous first derivative, we have obtained.

iv g_θ has first continuous derivative,

$$\dot{g}_\theta(z) = [z, z^2, (z-\kappa)^2 1_{\{z>\kappa\}}, -2\beta_3(z-\kappa)1_{\{z>\kappa\}}]^\top,$$

but the 2nd derivative **does not exist at knot** κ.

v Consistency? Asymptotic Normality? Under continuous first derivative, we have obtained.

vi Consistency? Asymptotic Normality? **Knots in covariates**:

$$g_\theta(z) = \beta_1 z + \beta_2(z-\kappa)_+, \quad g_\theta(z) = \beta(\mathbf{1}_{z\geq\kappa} - \mathbf{1}_{z<\kappa}).$$

Knots in time:

$$g_\theta(t,z) = (\beta_0 + \beta_1 t + \beta_2(t-\gamma)_+)z, \quad g_\theta(t,z) = \beta(\mathbf{1}_{t\geq\gamma} - \mathbf{1}_{t<\gamma})z,$$

Counting Process Framework

- Consider a semiparametric multiplicative hazard model,

$$h(t) = h_0(t) r_\theta(t, Z(t)), \quad t \geq 0, \tag{3}$$

where r_θ is parametric with $\theta \in \mathbb{R}^p$, h_0 is nonparametric baseline hazard, and $Z(t)$ is a covariate process.

Counting Process Framework

- Consider a semiparametric multiplicative hazard model,

$$h(t) = h_0(t) r_\theta(t, Z(t)), \quad t \geq 0, \tag{3}$$

  where r_θ is parametric with $\theta \in \mathbb{R}^p$, h_0 is nonparametric baseline hazard, and $Z(t)$ is a covariate process.
- n independent individuals continuously monitored over time $t \geq 0$, each corresponds to a process $(N_i(t), Y_i(t), Z_i(t))$, where $N_i(t)$ is a counting proc. recording events (such as deaths), $Y_i(t)$ is the at-risk proc. taking values 1 and 0 depending on whether the individual is under observation, $Z_i(t)$ is a vector covariate proc., all up to time t.

Counting Process Framework

Set $g_\theta(t, Z(t)) = \log r_\theta(t, Z(t))$. As usual, assume

- $(N_1, ..., N_n)$ is a multivariate counting proc.
- Each $M_i = N_i - A_i$ is a local martingale w.r.t. a right-cont. filtration, where g_θ a predictable proc., A_i is the compensator $A_i = \int Y_i(s) \exp(g_{\theta_0}(s, Z_i(s)))h_0(s)ds$, and θ_0 is the true parameter value.
- Each Y_i and Z_i is predictable w.r.t. a right-cont. filtration.

Counting Process Framework

Set $g_\theta(t, Z(t)) = \log r_\theta(t, Z(t))$. As usual, assume

- $(N_1, ..., N_n)$ is a multivariate counting proc.
- Each $M_i = N_i - A_i$ is a local martingale w.r.t. a right-cont. filtration, where g_θ a predictable proc., A_i is the compensator $A_i = \int Y_i(s) \exp(g_{\theta_0}(s, Z_i(s))) h_0(s) ds$, and θ_0 is the true parameter value.
- Each Y_i and Z_i is predictable w.r.t. a right-cont. filtration.

The usual partial likelihood function is

$$PL_n(\theta) = \prod_{i=1}^n \left\{ \frac{\exp\left[g_\theta(T_i, Z_i(T_i))\right]}{\sum_{j \in R_i} \exp\left[g_\theta(T_i, Z_j(T_i))\right]} \right\}^{\delta_i},$$

where δ_i is the indicator that the failure of individual i was observed and R_i is the at-risk set at the time of the ith failure.

Assumptions

Let $\mathbb{S}_n$ be the operator defined by

$$\mathbb{S}_n(t, g)[h] = \frac{1}{n}\sum_{i=1}^{n} h(t, Z_i(t))\, Y_i(t) \exp(g(t, Z_i(t))),$$

where g is a function and h is a scalar, vector or matrix function. Formally, let $S_n^{(0)}(\theta, t) = \mathbb{S}_n(t, g_\theta)[\imath]$, where $\imath$ is the identity map; $S_n^{(1)}(\theta, t) = \mathbb{S}_n(t, g_\theta)[\dot{g}_\theta]$; $S_n^{(2)}(\theta, t) = \mathbb{S}_n(t, g_\theta)[\dot{g}_\theta^{\otimes 2}]$.

Assumptions

Let $\mathbb{S}_n$ be the operator defined by

$$\mathbb{S}_n(t,g)[h] = \frac{1}{n}\sum_{i=1}^n h(t,Z_i(t))Y_i(t)\exp(g(t,Z_i(t))),$$

where g is a function and h is a scalar, vector or matrix function. Formally, let $S_n^{(0)}(\theta,t) = \mathbb{S}_n(t,g_\theta)[\imath]$, where $\imath$ is the identity map; $S_n^{(1)}(\theta,t) = \mathbb{S}_n(t,g_\theta)[\dot{g}_\theta]$; $S_n^{(2)}(\theta,t) = \mathbb{S}_n(t,g_\theta)[\dot{g}_\theta^{\otimes 2}]$.
The following assumptions are needed.

(I) There exists time τ such that $\int_0^\tau h_0(x)dx < \infty$.

(II) For any compact neighborhood Θ_0 of θ_0, there exists a scalar $s^{(0)}$, a vector $s^{(1)}$ and a matrix $s^{(2)}$ on $\Theta_0 \times [0,\tau]$ such that for $j = 0,1,2$, $\sup_{t\in[0,\tau],\,\theta\in\Theta_0} \|S_n^{(j)}(\theta,t) - s^{(j)}(\theta,t)\| \xrightarrow{P} 0$,where $\|B\| \equiv \max\{|B_{ij}| : \forall i,j\}$ is a matrix norm.

Assumptions

(III) Define $e \equiv s^{(1)}/s^{(0)}$ and $v \equiv s^{(2)}/s^{(0)} - e^{\otimes 2}$. Then for $\theta \in \Theta_0$ and $t \in [0, \tau]$, $\frac{\partial}{\partial \theta} s^{(0)}(\theta, t) = s^{(1)}(\theta, t)$.

(IV) For $j = 0, 1, 2$, the functions $s^{(j)}(\theta, x)$ are bounded; the function families $s^{(j)}(\cdot, t)$, $t \in [0, \tau]$ are equicontinuous at $\theta = \theta_0$; and $s^{(0)}(\theta, t)$ is bounded away from zero on $\Theta_0 \times [0, \tau]$.

(V) $\Sigma(\theta_0, \tau) = \int_0^\tau v(\theta_0, x) s^{(0)}(\theta_0, x) h_0(x)\, dx$ is positive definite.

(VI) There exists $\delta > 0$ such that as n tends to infinity,

$$\sup_{1 \le i \le n,\, 0 \le x \le \tau} \|\dot{g}_{\theta_0}(x, Z_i(x))\| Y_i(x) \mathbf{1}_{\{g_{\theta_0}(x, Z_i(x)) > -\delta \|\dot{g}_{\theta_0}(x, Z_i(x))\|\}}$$
$$= o_p(n^{1/2}).$$

Sufficient Conditions

Let T and U be the failure and censoring time and Z be a covariate. Suppose (X_i, δ_i, Z_i) are n i.i.d. observations, where $X_i \equiv \min(T_i, U_i)$, representing the observed time; $\delta_i \equiv 1_{\{T_i \leq U_i\}}$, indicating that the observed time is a death time not a censoring. The counting process $N_i(t) \equiv 1_{\{X_i \leq t, \delta_i = 1\}}$, the at-risk process $Y_i(t) \equiv 1_{\{X_i \geq t\}}$.

Sufficient Conditions

Let T and U be the failure and censoring time and Z be a covariate. Suppose (X_i, δ_i, Z_i) are n i.i.d. observations, where $X_i \equiv \min(T_i, U_i)$, representing the observed time; $\delta_i \equiv 1_{\{T_i \leq U_i\}}$, indicating that the observed time is a death time not a censoring. The counting process $N_i(t) \equiv 1_{\{X_i \leq t, \delta_i = 1\}}$, the at-risk process $Y_i(t) \equiv 1_{\{X_i \geq t\}}$.

The following is sufficient condition for ASSUMPTION.

Proposition Suppose each covariate Z_i is constant in time and takes value in a compact set $\mathbb{Z}$ of $\mathbb{R}^q$; T_i and U_i are conditionally independent given Z_i; and $\mathbb{P}\{Y_i(\tau) > 0\} > 0$ for some $\tau > 0$. Suppose $g_\theta(Z)$ has continuous first derivative for $Z \in \mathbb{Z}, \theta \in \Theta$ with Θ a compact neighborhood of θ_0. Then ASSUMPTION holds with the exception of (V).

Concavity

Theorem 1 Suppose that ASSUMPTION (I)–(V) hold. Assume that there exist a neighborhood Θ_0 of θ_0 and a bounded matrix function $s^{(3)}$ on $\Theta_0^3 \mathbf{x}[0, \tau]$ such that

$$\sup_{t\in[0,\tau],\, \forall\theta_i\in\Theta_0} \|\mathbb{S}_n(t, g_{\theta_1})[\dot{g}_{\theta_2}\dot{g}_{\theta_3}^{\top}] - s^{(3)}(\theta_1, \theta_2, \theta_3, t)\| \xrightarrow{P} 0, \quad (4)$$

and the family of matrix functions $s^{(3)}(\cdot, \cdot, \cdot, x)$, $x \in [0, \tau]$ is equicontinuous at $(\theta_0, \theta_0, \theta_0)$. Then there exists a neighborhood Θ of the true parameter value θ_0 such that the log-partial likelihood function $l_n(\theta) = \log PL_n(\theta)$ is concave in Θ in probability.

Concavity

- For the linear log-relative risk, $g_\theta(t, Z(t)) = \theta^\top Z(t)$, the concavity of $l_n(\theta)$ is a folklore (Andersen and Gill, 1983).

Concavity

- For the linear log-relative risk, $g_\theta(t, Z(t)) = \theta^\top Z(t)$, the concavity of $l_n(\theta)$ is a folklore (Andersen and Gill, 1983).
- If the usual exponential form is relaxed to an arbitrary non-negative twice continuously differentiable function, the global concavity of $l_n(\theta)$ is reduced to the local concavity (Prentice and Self, 1983).

Concavity

- For the linear log-relative risk, $g_\theta(t, Z(t)) = \theta^\top Z(t)$, the concavity of $l_n(\theta)$ is a folklore (Andersen and Gill, 1983).
- If the usual exponential form is relaxed to an arbitrary non-negative twice continuously differentiable function, the global concavity of $l_n(\theta)$ is reduced to the local concavity (Prentice and Self, 1983).
- O'Sullivan (1993) investigated nonparametric estimation in the Cox model,

$$h(t) = h_0(t) \exp[\theta(Z(t)))], \quad t \geq 0,$$

where θ is a nonparametric function. O'Sullivan shows that $l_n(\theta)$ and its limit $l(\theta)$ are concave in θ in some Sobolev space.

Consistency

Theorem 2 Assume that the assumptions in **Theorem 1** hold. Suppose that there exist a neighborhood Θ_0 of θ_0 and scalar functions m_j on $\Theta\mathbf{x}[0,\tau]$ such that

$$\sup_{x\in[0,\tau],\,\theta\in\Theta_0} \|\mathbb{S}_n(x, g_\theta)[g_\theta^j] - m_j(\theta, x)\| \xrightarrow{P} 0, \quad j = 1, 2. \qquad (5)$$

Suppose that differentiation can pass integration, i.e.,

$$\frac{\partial}{\partial\theta} \int_0^\tau s(x, g_{\theta_0})[g_\theta]h_0(x)\,dx\Big|_{\theta=\theta_0} = \frac{\partial}{\partial\theta} \int_0^\tau \log s^{(0)}(\theta, x)s^{(0)}(\theta_0, x)h_0(x)\,dx\Big|_{\theta=\theta_0},$$

Then the MPLE $\hat{\theta}_n$ is consistent, i.e., $\hat{\theta}_n \xrightarrow{p} \theta_0$.

The Score Process

- The score function $U_n(\theta)$ is

$$U_n(\theta) \equiv \frac{\partial l_n(\theta)}{\partial \theta} = \sum_{i=1}^{n} \int_0^\infty \left[\dot{g}_\theta(t, Z_i(t)) - \frac{S_n^{(1)}(\theta, t)}{S_n^{(0)}(\theta, t)} \right] dN_i(t).$$

- Since $dA_i(t) = Y_i(t) \exp(g_{\theta_0}(t, Z_i(t))) h_0(t) dt$, it follows $\sum_{i=1}^n \int_0^\infty \left[\dot{g}_{\theta_0}(t, Z_i(t)) - \frac{S_n^{(1)}(\theta_0,t)}{S_n^{(0)}(\theta_0,t)} \right] dA_i(t) = 0$, so that $U_n(\theta_0) = \sum_{i=1}^n \int_0^\infty \left[\dot{g}_{\theta_0}(t, Z_i(t)) - \frac{S_n^{(1)}(\theta_0,t)}{S_n^{(0)}(\theta_0,t)} \right] dM_i(t)$ is a martingale.
- Let the score process be $U_n(\theta_0, t) = \sum_{i=1}^n \int_0^t \left[\dot{g}_{\theta_0}(x, Z_i(x)) - \frac{S_n^{(1)}(\theta_0,x)}{S_n^{(0)}(\theta_0,x)} \right] dM_i(x).$

Asymptotic Normality of the Score Proc.

Theorem 3 Suppose that ASSUMPTION is satisfied. Then
(a) $n^{-1/2}U_n(\theta_0, t)$ converges in distribution to a Gaussian process, where each component of the Gaussian process has independent increments, the mean of the limiting process is zero and the covariance matrix of the limiting process at time t is

$$\Sigma(\theta_0, t) = \int_0^t v(\theta_0, x) s^{(0)}(\theta_0, x) h_0(x)\, dx.$$

Asymptotic Normality of the Score Proc.

Theorem 3 Suppose that ASSUMPTION is satisfied. Then
(a) $n^{-1/2}U_n(\theta_0, t)$ converges in distribution to a Gaussian process, where each component of the Gaussian process has independent increments, the mean of the limiting process is zero and the covariance matrix of the limiting process at time t is

$$\Sigma(\theta_0, t) = \int_0^t v(\theta_0, x)s^{(0)}(\theta_0, x)h_0(x)\, dx.$$

(b) If $\hat{\theta}_n$ is a consistent estimator of θ_0, then the plug-in estimator $\Sigma(\hat{\theta}_n, t) = \frac{1}{n}\sum_{i=1}^n \int_0^t V_n(\hat{\theta}_n, x)dN_i(x)$ of $\Sigma(\theta_0, t)$ satisfies

$$\sup_{t\in[0,\tau]} \|\frac{1}{n}\sum_{i=1}^n \int_0^t V_n(\hat{\theta}_n, x)\, dN_i(x) - \Sigma(\theta_0, x)\| \xrightarrow{P} 0, \quad n \to \infty.$$

Asymptotic Normality of the MPLE

Theorem 4 Suppose ASSUMPTION holds. Assume

$$\mathbb{E}\left\{\int_0^\tau \|\dot{g}_\theta - \dot{g}_{\theta_0}\|^2(x, Z(x))\, dA(x)\right\} = o(1),\ \theta \to \theta_0. \tag{6}$$

Assume that there allows a two-term Taylor expansion:

$$\mathbb{E}[n^{-1} l_n(\theta)] = \mathbb{E}[n^{-1} l_n(\theta_0)] - (1/2)(\theta-\theta_0)^\top \Sigma(\theta_0, \tau)(\theta-\theta_0) + o_p(\|\theta-\theta_0\|^2). \tag{7}$$

If (4) holds, then $\hat{\theta}_n$ satisfies the equivalence relation:

$$n^{1/2}(\hat{\theta}_n - \theta_0) = \Sigma^{-1}(\theta_0, \tau) n^{-1/2} U_n(\theta_0, \tau) + o_p(1). \tag{8}$$

Hence, $\hat{\theta}_n$ is asymptotically normal:

$$n^{1/2}(\hat{\theta}_n - \theta_0) \Longrightarrow \mathcal{N}(0, \Sigma^{-1}(\theta_0, \tau)).$$

Asymptotic Normality in Free-knot Quadratic Splines

Theorem 5 Consider the free-knot quadratic spline polynomial model (2). Suppose that the assumptions in **Proposition** are satisfied. Assume the distribution function Q of Z and the integral $\int_{\kappa_0}^{\kappa} z\, dQ(z)$ are continuous at the true knot value κ_0. Then the MPLE $\hat{\theta}_n$ is asymptotic normal with mean θ_0 and covariance $\Sigma(\theta_0, \tau)$, i.e.,

$$n^{1/2}(\hat{\theta}_n - \theta_0) \Longrightarrow \mathcal{N}\left(0, \Sigma^{-1}(\theta_0, \tau)\right).$$

Sketches of Proof of Concavity: Special Case

Suppose $\theta \mapsto g_\theta(t,z)$ has continuous second derivative. Then

$$\frac{\partial^2 \log \mathrm{PL}_n(\theta)}{\partial\theta\partial\theta^\top} = \sum_{i=1}^n \int_0^\tau \left[\ddot{g}_\theta(t, Z_i(t)) + \left(\frac{S_n^{(1)}(\theta,t)}{S_n^{(0)}(\theta,t)}\right)^{\otimes 2} - \frac{S_n^{(2)}(\theta,t)}{S_n^{(0)}(\theta,t)} - \frac{\sum_{j=1}^n \ddot{g}_\theta(t, Z_j(t)) Y_j(t) \exp(g_\theta(t, Z_j(t)))}{S_n^{(0)}(\theta,t)}\right] dN_i(t).$$

Sketches of Proof of Concavity: Special Case

Suppose $\theta \mapsto g_\theta(t, z)$ has continuous second derivative. Then

$$\frac{\partial^2 \log \mathrm{PL}_n(\theta)}{\partial\theta\partial\theta^\top} = \sum_{i=1}^{n} \int_0^\tau \Bigg[\ddot{g}_\theta(t, Z_i(t)) + \left(\frac{S_n^{(1)}(\theta, t)}{S_n^{(0)}(\theta, t)}\right)^{\otimes 2} - \frac{S_n^{(2)}(\theta, t)}{S_n^{(0)}(\theta, t)} - \frac{\sum_{j=1}^{n} \ddot{g}_\theta(t, Z_j(t)) Y_j(t) \exp(g_\theta(t, Z_j(t)))}{S_n^{(0)}(\theta, t)}\Bigg] dN_i(t).$$

In terms of the urn model, the above can be expressed as

$$-\frac{1}{n}\frac{\partial^2 \log \mathrm{PL}_n(\theta)}{\partial\theta\partial\theta^\top} = \frac{1}{n}\sum_{i=1}^{n} \int_0^\tau V_n(\theta, t) dN_i(t)$$

$$-\frac{1}{n}\sum_{i=1}^{n} \int_0^\tau \left\{\ddot{g}_\theta(t, Z_i(t)) - E_{\theta,t}^{I}[\ddot{g}_\theta(t, Z_I(t))]\right\} dN_i(t),$$

Sketches of Proof of Concavity: Special Case

where $E^I_{\theta_1,x}[\ddot{g}_{\theta_2}(t, Z_I(t))] = \sum_{i=1}^n \ddot{g}_{\theta_2}(t, Z_i(t))p_i(\theta_1, x)$, $\theta_1, \theta_2 \in \Theta$ is the expectation calculated under the discrete distribution

$$p_i(\theta_1, x) \equiv \frac{Y_i(x)\exp(g_{\theta_1}(x, Z_i(x)))}{\sum_{j=1}^n Y_j(x)\exp(g_{\theta_1}(x, Z_j(x)))}, \qquad i = 1, \ldots, n. \tag{9}$$

Hence, the urn-model variance:

$$\begin{aligned} V_n(\theta, x) &= \frac{S_n^{(2)}(\theta, x)}{S_n^{(0)}(\theta, x)} - \left[\frac{S_n^{(1)}(\theta, x)}{S_n^{(0)}(\theta, x)}\right]^{\otimes 2} \\ &= E^I_{\theta,x}[\dot{g}_\theta(x, Z_I(x))^{\otimes 2}] - \left(E^I_{\theta,x}[\dot{g}_\theta(x, Z_I(x))]\right)^{\otimes 2} \\ &= Var^I_{\theta,x}\left[\dot{g}_\theta(x, Z_I(x))\right] \end{aligned}$$

Sketches of Proof of Concavity: General case

For $t \in [0, \tau]$, let $X_n(\theta, t)$ be defined by

$$\frac{1}{n}\sum_{i=1}^{n}\int_0^t \left[(g_\theta - g_{\theta_0})(x, Z_i(x)) - \log\left(S_n^{(0)}(\theta, x)/S_n^{(0)}(\theta_0, x)\right)\right] dN_i(x).$$

It suffices to show

$$(\theta_1 - \theta_2)^\top \left[\dot{X}_n(\theta_1, t) - \dot{X}_n(\theta_2, t)\right] \leq 0, \quad \theta_1, \theta \in \Theta,$$

as n is sufficiently large and the neighborhood Θ of θ_0 is small.

Sketches of Proof of Asymptotic Normality of the MPLE

Convexity Lemma (Anderson and Gill, 1982; Pollard, 1991) Let $\{\lambda_n(\theta) : \theta \in \Theta\}$ be a seq. of random functions defined on a convex, open subset Θ of $\mathbb{R}^d$. Suppose there exists a seq. $\{C_n\}$ of measurable sets with $P(C_n) \to 1$ such that each $\lambda_n(\theta)$ is convex on C_n for every $\theta \in \Theta$. Suppose $\lambda(\cdot)$ is a real-valued function on Θ for which $\lambda_n(\theta) \xrightarrow{p} \lambda(\theta)$ for each $\theta \in \Theta$. Then

Sketches of Proof of Asymptotic Normality of the MPLE

Convexity Lemma (Anderson and Gill, 1982; Pollard, 1991) Let $\{\lambda_n(\theta) : \theta \in \Theta\}$ be a seq. of random functions defined on a convex, open subset Θ of $\mathbb{R}^d$. Suppose there exists a seq. $\{C_n\}$ of measurable sets with $P(C_n) \to 1$ such that each $\lambda_n(\theta)$ is convex on C_n for every $\theta \in \Theta$. Suppose $\lambda(\cdot)$ is a real-valued function on Θ for which $\lambda_n(\theta) \xrightarrow{p} \lambda(\theta)$ for each $\theta \in \Theta$. Then
(1) For each compact subset K of Θ,

$$\sup_{\theta \in K} |\lambda_n(\theta) - \lambda(\theta)| \xrightarrow{p} 0.$$

The function $\lambda(\cdot)$ is necessarily convex in Θ.

Sketches of Proof of Asymptotic Normality of the MPLE

Convexity Lemma (Anderson and Gill, 1982; Pollard, 1991) Let $\{\lambda_n(\theta) : \theta \in \Theta\}$ be a seq. of random functions defined on a convex, open subset Θ of $\mathbb{R}^d$. Suppose there exists a seq. $\{C_n\}$ of measurable sets with $P(C_n) \to 1$ such that each $\lambda_n(\theta)$ is convex on C_n for every $\theta \in \Theta$. Suppose $\lambda(\cdot)$ is a real-valued function on Θ for which $\lambda_n(\theta) \xrightarrow{p} \lambda(\theta)$ for each $\theta \in \Theta$. Then
(1) For each compact subset K of Θ,

$$\sup_{\theta \in K} |\lambda_n(\theta) - \lambda(\theta)| \xrightarrow{p} 0.$$

The function $\lambda(\cdot)$ is necessarily convex in Θ.
(2) If λ_n has a unique maximum at θ_n and λ has one at θ_0, then $\theta_n \xrightarrow{p} \theta_0$ as $n \to \infty$.

Let $X_n(\theta) = X_n(\theta, \tau)$. For $\alpha \in \mathbb{R}^p$, let $\alpha_n = n^{-1/2}\alpha$ and $D_n(\alpha) = n\big(X_n(\theta_0 + \alpha_n) - \alpha_n^\top \dot{X}_n(\theta_0)\big)$. Then one shows $\mathbb{V}ar(D_n(\alpha)) \to 0$, so that

$$D_n(\alpha) = \mathbb{E}(D_n(\alpha)) + o_p(1).$$

Let $X_n(\theta) = X_n(\theta, \tau)$. For $\alpha \in \mathbb{R}^p$, let $\alpha_n = n^{-1/2}\alpha$ and $D_n(\alpha) = n\big(X_n(\theta_0 + \alpha_n) - \alpha_n^\top \dot{X}_n(\theta_0)\big)$. Then one shows $\mathbb{V}ar(D_n(\alpha)) \to 0$, so that

$$D_n(\alpha) = \mathbb{E}(D_n(\alpha)) + o_p(1).$$

By the two-term Taylor expansion (7), we obtain

$$D_n(\alpha) = -(1/2)\alpha^\top \Sigma(\theta_0, \tau)\alpha + o_p(1), \quad \alpha \in \mathbb{R}^p \qquad (10)$$

Let $X_n(\theta) = X_n(\theta, \tau)$. For $\alpha \in \mathbb{R}^p$, let $\alpha_n = n^{-1/2}\alpha$ and $D_n(\alpha) = n\big(X_n(\theta_0 + \alpha_n) - \alpha_n^\top \dot{X}_n(\theta_0)\big)$. Then one shows $\mathbb{V}ar(D_n(\alpha)) \to 0$, so that

$$D_n(\alpha) = \mathbb{E}(D_n(\alpha)) + o_p(1).$$

By the two-term Taylor expansion (7), we obtain

$$D_n(\alpha) = -(1/2)\alpha^\top \Sigma(\theta_0, \tau)\alpha + o_p(1), \quad \alpha \in \mathbb{R}^p \tag{10}$$

Since $-X_n(\theta)$ is convex in θ in a small neighborhood Θ of θ_0, it follows from **Convexity Lemma** that for any $M > 0$,

$$\sup_{\|\alpha\| \le M} \left| n\left\{ X_n(\theta_0 + n^{-1/2}\alpha) - n^{-1/2}\alpha^\top \dot{X}_n(\theta_0) \right\} + \frac{1}{2}\alpha^\top \Sigma(\theta_0, \tau)\alpha \right| = o_p(1)$$

Let $\tilde{X}_n(\alpha) = X_n(\theta_0 + n^{-1/2}\alpha)$ and $\hat{\alpha}_n = \arg\max_{\alpha \in \mathbb{R}^p} \tilde{X}_n(\alpha)$. Then $\hat{\alpha}_n = n^{1/2}(\hat{\theta}_n - \theta_0)$. For any *random* variable γ_n bounded in probability,

$$\tilde{X}_n(\gamma_n) = \gamma_n^\top n^{-1/2} U_n(\theta_0) - \frac{1}{2}\gamma_n^\top \Sigma(\theta_0, \tau)\gamma_n + o_p(1). \qquad (11)$$

Let $\tilde{X}_n(\alpha) = X_n(\theta_0 + n^{-1/2}\alpha)$ and $\hat{\alpha}_n = \arg\max_{\alpha \in \mathbb{R}^p} \tilde{X}_n(\alpha)$. Then $\hat{\alpha}_n = n^{1/2}(\hat{\theta}_n - \theta_0)$. For any *random* variable γ_n bounded in probability,

$$\tilde{X}_n(\gamma_n) = \gamma_n^\top n^{-1/2} U_n(\theta_0) - \frac{1}{2}\gamma_n^\top \Sigma(\theta_0, \tau)\gamma_n + o_p(1). \tag{11}$$

This shows that $\tilde{X}_n(\gamma_n)$ can be approximated by a quadratic function in γ_n, maximized at $\hat{\gamma}_n = \Sigma^{-1}(\theta_0, \tau) n^{-1/2} U_n(\theta_0)$, with the maximized value approximately $\tilde{X}_n(\hat{\gamma}_n) = \frac{1}{2}\hat{\gamma}_n^\top \Sigma(\theta_0, \tau)\hat{\gamma}_n$.

Let $\tilde{X}_n(\alpha) = X_n(\theta_0 + n^{-1/2}\alpha)$ and $\hat{\alpha}_n = \arg\max_{\alpha \in \mathbb{R}^p} \tilde{X}_n(\alpha)$. Then $\hat{\alpha}_n = n^{1/2}(\hat{\theta}_n - \theta_0)$. For any *random* variable γ_n bounded in probability,

$$\tilde{X}_n(\gamma_n) = \gamma_n^\top n^{-1/2} U_n(\theta_0) - \frac{1}{2}\gamma_n^\top \Sigma(\theta_0, \tau)\gamma_n + o_p(1). \qquad (11)$$

This shows that $\tilde{X}_n(\gamma_n)$ can be approximated by a quadratic function in γ_n, maximized at $\hat{\gamma}_n = \Sigma^{-1}(\theta_0, \tau) n^{-1/2} U_n(\theta_0)$, with the maximized value approximately $\tilde{X}_n(\hat{\gamma}_n) = \frac{1}{2}\hat{\gamma}_n^\top \Sigma(\theta_0, \tau)\hat{\gamma}_n$. By (11),

$$\tilde{X}_n(\gamma_n) = \tilde{X}_n(\hat{\gamma}_n) - \frac{1}{2}(\hat{\gamma}_n - \gamma_n)^\top \Sigma(\theta_0, \tau)(\hat{\gamma}_n - \gamma_n) + o_p(1) \qquad (12)$$

for any γ_n bounded in probability.

Fix $\epsilon > 0$. If $\|\hat{\alpha}_n - \hat{\gamma}_n\| > \epsilon$, then there exists $\hat{\gamma}_n^*$ on the line segment joining $\hat{\alpha}_n$ and $\hat{\gamma}_n$ such that $\hat{\gamma}_n^* - \hat{\gamma}_n = \epsilon \upsilon_n$, where υ_n is a unit vector.

Fix $\epsilon > 0$. If $\|\hat{\alpha}_n - \hat{\gamma}_n\| > \epsilon$, then there exists $\hat{\gamma}_n^*$ on the line segment joining $\hat{\alpha}_n$ and $\hat{\gamma}_n$ such that $\hat{\gamma}_n^* - \hat{\gamma}_n = \epsilon \upsilon_n$, where υ_n is a unit vector. From this equality and the boundedness of $\hat{\gamma}_n$ in probability it follows that $\hat{\gamma}_n^*$ is bounded in probability, Fix $\epsilon > 0$. If $\|\hat{\alpha}_n - \hat{\gamma}_n\| > \epsilon$, then there exists $\hat{\gamma}_n^*$ on the line segment joining $\hat{\alpha}_n$ and $\hat{\gamma}_n$ such that $\hat{\gamma}_n^* - \hat{\gamma}_n = \epsilon \upsilon_n$, where υ_n is a unit vector. From this equality and the boundedness of $\hat{\gamma}_n$ in probability it follows that $\hat{\gamma}_n^*$ is bounded in probability, so that we can substitute $\hat{\gamma}_n^*$ in (12) to obtain

$$\tilde{X}_n(\hat{\gamma}_n^*) = \tilde{X}_n(\hat{\gamma}_n) - \frac{1}{2}\epsilon^2 \upsilon_n^\top \Sigma(\theta_0, \tau)\upsilon_n + o_p(1). \tag{13}$$

Fix $\epsilon > 0$. If $\|\hat{\alpha}_n - \hat{\gamma}_n\| > \epsilon$, then there exists $\hat{\gamma}_n^*$ on the line segment joining $\hat{\alpha}_n$ and $\hat{\gamma}_n$ such that $\hat{\gamma}_n^* - \hat{\gamma}_n = \epsilon \upsilon_n$, where υ_n is a unit vector. From this equality and the boundedness of $\hat{\gamma}_n$ in probability it follows that $\hat{\gamma}_n^*$ is bounded in probability, so that we can substitute $\hat{\gamma}_n^*$ in (12) to obtain

$$\tilde{X}_n(\hat{\gamma}_n^*) = \tilde{X}_n(\hat{\gamma}_n) - \frac{1}{2}\epsilon^2 \upsilon_n^\top \Sigma(\theta_0, \tau)\upsilon_n + o_p(1). \qquad (13)$$

Since $\tilde{X}_n(\cdot)$ is concave and $\tilde{X}_n(\hat{\alpha}_n) \geq \tilde{X}_n(\hat{\gamma}_n)$, it follows that $\tilde{X}_n(\hat{\gamma}_n^*) \geq \tilde{X}_n(\hat{\gamma}_n)$. Accordingly, (13) implies

$$(1/2)\epsilon^2 \upsilon_n^\top \Sigma(\theta_0, \tau)\upsilon_n + o_p(1) \leq 0.$$

Hence $\mathbb{P}(\|\hat{\alpha}_n - \hat{\gamma}_n\| > \epsilon) \leq \mathbb{P}(\frac{1}{2}\epsilon^2 \upsilon_n^\top \Sigma(\theta_0, \tau)\upsilon_n + o_p(1) \leq 0) \to 0$, so that $\hat{\alpha}_n = \hat{\gamma}_n + o_p(1)$ and the desired (8) follows. $\square$

T H A N K S