Generalized Grid Upper Bound

James Yang

April 13, 2022

1 Introduction

The current formulation of the upper bound estimate assumes that the (rectangular) gridding occurs in the (canonical) natural parameter space Ξ. However, it is sometimes more suitable to grid a different space Θ that parametrizes Ξ. For example, an exponential model with the control and treatment arms assumed to be exponentially distributed with hazards λ_c, λ_t, respectively, equipped with the logrank test can be greatly optimized under the parametrization of λ_c, h where $h := \lambda_t/\lambda_c$ is the hazard rate. Moreover, for better scaling, we may want to grid the $(\log(\lambda_c), \log(h))$ space. Such a parametrization defines a mapping from *the grid space* to the natural parameter space. We wish to construct the upper bound estimate under any such parametrization, provided that the mapping is sufficiently smooth.

In the subsequent sections, we will use the notation $\theta \in \Theta \subseteq \mathbb{R}^s$ to denote a point in the grid space and $\eta = \eta(\theta) \in \Xi \subseteq \mathbb{R}^d$ as the canonical natural parameter.

2 Original Upper Bound Estimate

For completion, we give a short overview of the old version of the upper bound estimate.

We begin with a set of multiple hypotheses $H_1, \ldots, H_p$. We define a *configuration* of the multiple hypotheses as an element of $\{0,1\}^p$ where the ith coordinate is 1 if and only if H_i is true. We assume that we have i.i.d. draws of $X^i \sim \mathbb{P}_\eta$ where $\mathbb{P}_\eta$ forms an exponential family. For adaptive trials, we assume that there exists a finite time horizon $\tau_{\max}$ so that $X^i \in \mathbb{R}^{\tau_{\max}}$, though we only observe up to a stopping time τ. We denote T_t as the sufficient statistic of $(X_1, \ldots, X_t)$.

Let $f(\eta) := \mathbb{P}_\eta(X \in A)$ where A is the event of false rejection. Since exponential families are sufficiently smooth, $f(\eta)$ is twice-continuously differentiable. A second-order Taylor expansion gives us

$$f(\eta) = f(\eta_0) + \nabla f(\eta_0)^\top (\eta - \eta_0) + \int_0^1 (1-\alpha)(\eta-\eta_0)^\top \nabla^2 f(\eta_0 + \alpha(\eta - \eta_0))(\eta - \eta_0) d\alpha$$

Note that the derivatives are with respect to η.

Given a bounded set of η values, R, we obtain an upper bound of the true Type I error:

$$\begin{aligned}
\sup_{\eta \in R} f(\eta) &= f(\eta_0) + \sup_{\eta \in R} \left[\nabla f(\eta_0)^\top (\eta - \eta_0) + \int_0^1 (1-\alpha)(\eta-\eta_0)^\top \nabla^2 f(\eta_0 + \alpha(\eta - \eta_0))(\eta - \eta_0) d\alpha \right] \\
&\leq f(\eta_0) + \sup_{v \in R - \eta_0} \left[\nabla f(\eta_0)^\top v + \frac{1}{2} v^\top \sup_{\eta \in R} \mathrm{Var}\left(T_{\tau_{\max}}\right)_\eta v \right]
\end{aligned}$$

where $R - \eta_0 := \{\eta - \eta_0 : \eta \in R\}$.

An obvious estimate for $f(\eta_0)$ is simply

$$\hat{f}(\eta_0) := \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{X^i \in A}$$

and using Clopper-Pearson, for any $\delta_1 \in [0,1]$, we have an exact upper bound $\hat{\delta}_0^u$ for this estimate such that

$$\mathbb{P}_{\eta_0}\left(f(\eta_0) < \hat{f}(\eta_0) + \hat{\delta}_0^u\right) = 1 - \delta_1$$

From here, we further assume that R is a convex hull of a finite set of points, $v_1, \ldots, v_M$ so that the supremum is attained at one of the points by convexity. Using Cantelli's inequality, we showed that for any fixed $\delta_2 \in [0,1]$, there exists random $\hat{c}_m$, $m = 1, \ldots, M$, such that

$$\mathbb{P}_{\eta_0}\left(\sup_{v \in R - \eta_0}\left[\nabla f(\eta_0)^\top v + \frac{1}{2} v^\top \sup_{\eta \in R} \mathrm{Var}\left(T_{\tau_{\max}}\right)_\eta v\right] \leq \max_{m=1,\ldots,M} \hat{c}_m\right) \geq 1 - \delta_2$$

In particular, we have

$$\hat{c}_m = \widehat{\nabla f}(\eta_0)^\top v_m + \sqrt{\frac{v_m^\top \mathrm{Var}\left(T_{\tau_{\max}}\right)_{\eta_0} v_m}{N}\left(\frac{1}{\delta_2} - 1\right)} + \frac{1}{2} v_m^\top \sup_{\eta \in R} \mathrm{Var}\left(T_{\tau_{\max}}\right)_\eta v_m$$

where

$$\widehat{\nabla f}(\eta_0) := \frac{1}{N} \sum_{i=1}^{N} (T(X^i) - \nabla A(\eta_0)) \mathbb{1}_{X^i \in A}$$

Combining the two estimates,

$$\begin{aligned}
&\mathbb{P}_{\eta_0}\left(\sup_{\eta \in R} f(\eta) > \hat{f}(\eta_0) + \hat{\delta}_0^u + \max_{m=1,\ldots,M} \hat{c}_m\right) \leq \\
&\quad \mathbb{P}_{\eta_0}\left(\hat{f}(\eta_0) + \hat{\delta}_0^u < f(\eta_0)\right) \\
&\quad + \mathbb{P}_{\eta_0}\left(\max_{m=1,\ldots,M} \hat{c}_m < \sup_{v \in R - \eta_0}\left[\nabla f(\eta_0)^\top v + \frac{1}{2} v^\top \sup_{\eta \in R} \mathrm{Var}\left(T_{\tau_{\max}}\right)_\eta v\right]\right) \\
&\quad \leq \delta_1 + \delta_2
\end{aligned}$$

Given a bounded subset of the natural parameter space $H \subseteq \Xi$ and a finite disjoint covering of H, $\{R_j\}_{j=1}^{M}$, where, without loss of generality, each R_j belongs to exactly one configuration of the multiple hypotheses, we construct the upper bound estimates on each R_j. This gives us a point-wise (in η) guarantee that the true Type I error at η is no larger than the upper bound estimate with probability at least $1 - \delta$ where $\delta := \delta_1 + \delta_2$.

We define a few notations before we conclude this section. For any given $\eta \in H$, if R_0 is a partition where $\eta \in R_0$, and η_0 is a simulation grid-point associated with R_0 (note that η_0 need not be inside R_0), then the upper bound quantity is the sum of the following five quantities:

$$\begin{aligned}
\hat{\delta}_0 &:= \hat{f}(\eta_0) \\
\hat{\delta}_0^u &:= (\text{Clopper-Pearson upper bound with level } \delta_1) - \hat{\delta}_0 \\
\hat{\delta}_1 &:= \widehat{\nabla f}(\eta_0)^\top v_{m^*} \\
\hat{\delta}_1^u &:= \sqrt{\frac{v_{m^*}^\top \mathrm{Var}\left(T_{\tau_{\max}}\right)_{\eta_0} v_{m^*}}{N}\left(\frac{1}{\delta_2} - 1\right)} \\
\hat{\delta}_2^u &:= \frac{1}{2} v_{m^*}^\top \sup_{\eta \in R_0} \mathrm{Var}\left(T_{\tau_{\max}}\right)_\eta v_{m^*}
\end{aligned}$$

where $m^* = \underset{m=1,\ldots,M}{\arg\max}\, \hat{c}_m$.

3 Generalized Upper Bound Estimate

In Section 2, we discussed the old version of the upper bound estimate. Note that we assumed the gridding occured in the canonical natural parameter space. In this section, we extend this framework to allow gridding in a different space Θ where there exists a twice-continuously differentiable mapping $\eta(\cdot) : \Theta \to \Xi$ that maps from the grid space to the natural parameter space, Ξ.

Since we changed the gridding space, we must change the Taylor expansion to be with respect to Θ space. We abuse notation by denoting $f(\theta)$ as $f(\eta(\theta))$ and $f(\eta)$ as in Section 2. Then, for any θ, θ_0,

$$f(\theta) = f(\theta_0) + \nabla f(\theta_0)(\theta - \theta_0) + \int_0^1 (1-\alpha)(\theta - \theta_0)^\top \nabla^2 f(\theta_0 + \alpha(\theta - \theta_0))(\theta - \theta_0) d\alpha$$

Note that all derivatives are with respect to θ.

For now, assume we have a function $U_R(v)$ such that

$$\sup_{\theta \in R} v^\top \nabla^2 f(\theta) v \leq U_R(v)$$

for any v. In Section 3.4, we will propose ways of finding such a U_R. Then,

$$\int_0^1 (1-\alpha) v^\top \nabla^2 f(\theta_0 + \alpha v) v d\alpha \leq \frac{1}{2} U_R(v)$$

where $v = \theta - \theta_0$.

In summary, we have the bound:

$$\begin{aligned}
\sup_{\theta \in R} f(\theta) &= f(\theta_0) + \sup_{v \in R - \theta_0} \left[\nabla f(\theta_0)^\top v + \int_0^1 (1-\alpha) v^\top \nabla^2 f(\theta_0 + \alpha v) v d\alpha \right] \\
&\leq f(\theta_0) + \sup_{v \in R - \theta_0} \left[\nabla f(\theta_0)^\top v + \frac{1}{2} U_R(v) \right]
\end{aligned}$$

3.1 Constant Order Terms: $\hat{\delta}_0, \hat{\delta}_0^u$

The Monte Carlo term $\hat{\delta}_0$ and its corresponding upper bound $\hat{\delta}_0^u$ need no change from reparametrization other than the initial evaluation of $\eta_0 := \eta(\theta_0)$.

3.2 First Order Term: $\hat{\delta}_1$

The first order terms are affected by the η transformation.

$$\begin{aligned}
\nabla f(\theta) := \nabla_\theta P_\theta(A) = \nabla_\theta \int_A \frac{P_\theta}{P_{\theta_0}} dP_{\theta_0} &= \int_A \nabla_\theta \frac{P_\theta}{P_{\theta_0}} dP_{\theta_0} \\
&= \int_A (D_\theta \eta)^\top \nabla_\eta \frac{P_\eta}{P_{\eta_0}} dP_{\eta_0}
\end{aligned}$$

where $\eta_0 = \eta(\theta_0)$. If θ_0 is the point at which we are Taylor expanding, it suffices to compute this gradient at $\theta = \theta_0$. This results in

$$\nabla_\theta P_{\theta_0}(A) = \int_A (D_\theta \eta(\theta_0))^\top (T - \nabla_\eta A(\eta_0)) dP_{\eta_0}$$

Hence, our new gradient Monte Carlo estimate will be

$$\widehat{\nabla} f(\theta_0) := D_\theta \eta(\theta_0)^\top \frac{1}{N} \sum_{i=1}^N (T(X^i) - \nabla_\eta A(\eta_0)) \mathbb{1}_{X^i \in A}$$

Note that the Jacobian of η is known when defining a model and is simulation-independent. Hence, we may save the same gradient estimate as in Section 2 and later multiply by $D_\theta \eta(\theta_0)^\top$.

3.3 Higher Order Upper Bound Terms: $\hat{\delta}_1, \hat{\delta}_1^u, \hat{\delta}_2^u$

Similar to Section 2, once we can show for any $v_m = \theta_m - \theta_0$, where θ_m are the vertices of a convex hull $R \subseteq \Theta$, $m = 1, \ldots, M$, there exists a corresponding random $\hat{c}_m$ such that

$$\mathbb{P}_{\theta_0}\left(\nabla f(\theta_0)^\top v_m + \frac{1}{2}U_R(v_m) \leq \hat{c}_m\right) \geq 1 - \delta_2$$

then we have

$$\mathbb{P}_{\theta_0}\left(\sup_{v \in R - \theta_0}\left[\nabla f(\theta_0)^\top v + \frac{1}{2}U_R(v)\right] \leq \max_{m=1,\ldots,M} \hat{c}_m\right) \geq 1 - \delta_2$$

as soon as we further assume that U_R is convex.

Using Cantelli's inequality with $Y = \widehat{\nabla f}(\theta_0)^\top v_m = \frac{1}{N}\sum_{i=1}^{N} \widehat{\nabla f}(\theta_0)_i^\top v_m$, we only need to provide an upper bound on the variance of $\widehat{\nabla f}(\theta_0)_i^\top v_m$, where $\widehat{\nabla f}(\theta_0)_i := D_\theta \eta(\theta_0)^\top (T(X^i) - \nabla_\eta A(\eta_0))\mathbb{1}_{X^i \in A}$. In that endeavor,

$$\mathrm{Var}\left(\widehat{\nabla f}(\theta)_i^\top v_m\right) = v_m^\top \mathrm{Var}\left(\widehat{\nabla f}(\theta)_i\right) v_m \leq v_m^\top (D_\theta \eta)^\top \mathrm{Var}\left(T_{\tau_{\max}}\right)(D_\theta \eta) v_m$$

The rest of the calculations remain the same.

Hence,

$$\hat{c}_m := \widehat{\nabla f}(\theta_0)^\top v_m + \sqrt{\frac{v_m^\top (D_\theta \eta(\theta_0))^\top \mathrm{Var}\left(T_{\tau_{\max}}\right)_{\eta_0}(D_\theta \eta(\theta_0)) v_m}{N}\left(\frac{1}{\delta_2} - 1\right)} + \frac{1}{2}U_R(v_m)$$

This gives us our new upper bound estimates:

$$\begin{aligned}
\hat{\delta}_{0,\mathrm{new}} &:= \hat{\delta}_0 \\
\hat{\delta}_{0,\mathrm{new}}^u &:= \hat{\delta}_0^u \\
\hat{\delta}_{1,\mathrm{new}} &:= v_{m^*}^\top D_\theta \eta(\theta_0)^\top \hat{\delta}_1 \\
\hat{\delta}_{1,\mathrm{new}}^u &:= \sqrt{\frac{v_{m^*}^\top D\eta(\theta_0)^\top \mathrm{Var}\left(T_{\tau_{\max}}\right)_{\eta_0} D\eta(\theta_0) v_{m^*}}{N}\left(\frac{1}{\delta_2} - 1\right)} \\
\hat{\delta}_{2,\mathrm{new}}^u &:= \frac{1}{2}U_R(v_{m^*})
\end{aligned}$$

where $m^* = \underset{m=1,\ldots,M}{\arg\max}\, \hat{c}_m$.

3.4 Hessian Quadratic Form Bound

As mentioned in Section 3, we will now discuss a way to find the upper bound $U_R(v)$ to $\sup_{\theta \in R} v^\top \nabla^2 f(\theta) v$. In Section 3.3, we made the additional assumption that U_R is convex, so it is crucial this assumption is met.

We will first bound $\nabla^2 f(\theta)$.

$$\nabla^2 f(\theta) = \int_A \nabla^2 P_\theta(x) dx$$

Applying the multivariate chain-rule for the function $\theta \mapsto P_{\eta(\theta)}(x)$, we have that

$$\nabla^2 P_\theta(x) = (D\eta)^\top \nabla^2 P_\eta(x)(D\eta) + \sum_{k=1}^{d} \frac{\partial P_\eta}{\partial \eta_k} \nabla^2 \eta_k$$

[1].

It is easy to see that

$$-\mathrm{Var}\left(T_{\tau_{\max}}\right)_\eta \preceq \int_A \nabla^2 P_\eta(x)dx \preceq \mathrm{Var}\left(T_{\tau_{\max}}\right)_\eta$$

Note that if $S \preceq T$ for any square matrices S, T, then we must have that for any matrix A, $A^\top SA \preceq A^\top TA$. This is because $S \preceq T$ if and only if $T - S$ is positive semi-definite, and $A^\top(T-S)A$ is clearly positive semi-definite as well. Rearranging, we have our claim. Hence,

$$-(D\eta)^\top \mathrm{Var}\left(T_{\tau_{\max}}\right)_\eta (D\eta) \preceq (D\eta)^\top \int_A \nabla^2 P_\eta(x)dx(D\eta) \preceq (D\eta)^\top \mathrm{Var}\left(T_{\tau_{\max}}\right)_\eta (D\eta)$$

This gives us the first bound:

$$v^\top \nabla^2 f(\theta) v \leq v^\top (D\eta)^\top \mathrm{Var}\left(T_{\tau_{\max}}\right)_\eta (D\eta)v + \sum_{k=1}^d v^\top \nabla^2 \eta_k v \int_A (T(x) - \nabla A(\eta))_k P_\eta(x)dx \tag{1}$$

We next bound the second term in Eq. 1.

$$\begin{aligned}
\int_A \left|(T(x) - \nabla A(\eta))_k\right| P_\eta(x)dx &\leq \int \left|(T(x) - \nabla A(\eta))_k\right| P_\eta(x)dx \\
&\leq \left(\int \left|(T(x) - \nabla A(\eta))_k\right|^2 P_\eta(x)dx\right)^{1/2} \\
&= \sqrt{\mathrm{Var}\left(T_k\right)_\eta}
\end{aligned}$$

Combining with Eq. 1,

$$\begin{aligned}
\sup_{\theta \in R} v^\top \nabla^2 f(\theta) v &\leq \sup_{\theta \in R}\left[v^\top (D\eta(\theta))^\top \mathrm{Var}\left(T_{\tau_{\max}}\right)_{\eta(\theta)} (D\eta(\theta))v\right] + \sum_{k=1}^d \sup_{\theta \in R}\left[\left|v^\top \nabla^2 \eta_k(\theta) v\right| \sqrt{\mathrm{Var}\left(T_k\right)_{\eta(\theta)}}\right] \\
&\leq \sup_{\theta \in R}\left[v^\top (D\eta(\theta))^\top \mathrm{Var}\left(T_{\tau_{\max}}\right)_{\eta(\theta)} (D\eta(\theta))v\right] + ||v||^2 \sum_{k=1}^d \sup_{\theta \in R}\left[||\nabla^2 \eta_k(\theta)||_{op} \sqrt{\mathrm{Var}\left(T_k\right)_{\eta(\theta)}}\right] \\
&=: U_1(v) + U_2(v) =: U_R(v)
\end{aligned}$$

where U_i are defined to be the respective terms. Note that U_R is convex. Also note that $U_2(v) = 0$ for any linear transformation η. In particular, for the identity transformation, it simplifies to

$$U_R(v) = U_1(v) = \sup_{\theta \in R} v^\top \mathrm{Var}\left(T_{\tau_{\max}}\right)_{\eta(\theta)} v$$

which can be further bounded above by the usual formula

$$v^\top \sup_{\theta \in R} \mathrm{Var}\left(T_{\tau_{\max}}\right)_{\eta(\theta)} v$$

where the sup is element-wise. Note that in general, we can make $U_R(v)$ even more conservative by taking further upper bounds to make the computations more tractable. The only constraint is that the resulting bound must be convex.

References

[1] Maciej Skorski. Chain rules for hessian and higher derivatives made easy by tensor calculus. *CoRR*, abs/1911.13292, 2019.