Bounds to the normal for group sequential statistics with covariates

Jay Bartroff

USC

Workshop on New Directions in Stein's Method 2015

Goals for this talk

1. Describe bounds on distributional distance to multivariate normal distribution for
$$(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_K)$$
where $\hat{\theta}_k =$ MLE of $\theta \in \mathbb{R}^p$ at kth group sequential analysis, in regression setting
2. Advertise problems in sequential analysis that could (potentially) use Stein's method

Warnings:

1. Work in progress
2. Not hard, but useful

Goals for this talk

1. Describe bounds on distributional distance to multivariate normal distribution for
$$(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_K)$$
where $\hat{\theta}_k =$ MLE of $\theta \in \mathbb{R}^p$ at kth group sequential analysis, in regression setting
2. Advertise problems in sequential analysis that could (potentially) use Stein's method

Warnings:

1. Work in progress
2. Not hard, but useful

Goals for this talk

1. Describe bounds on distributional distance to multivariate normal distribution for
$$(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_K)$$
where $\hat{\theta}_k =$ MLE of $\theta \in \mathbb{R}^p$ at kth group sequential analysis, in regression setting
2. Advertise problems in sequential analysis that could (potentially) use Stein's method

Warnings:

1. Work in progress
2. Not hard, but useful

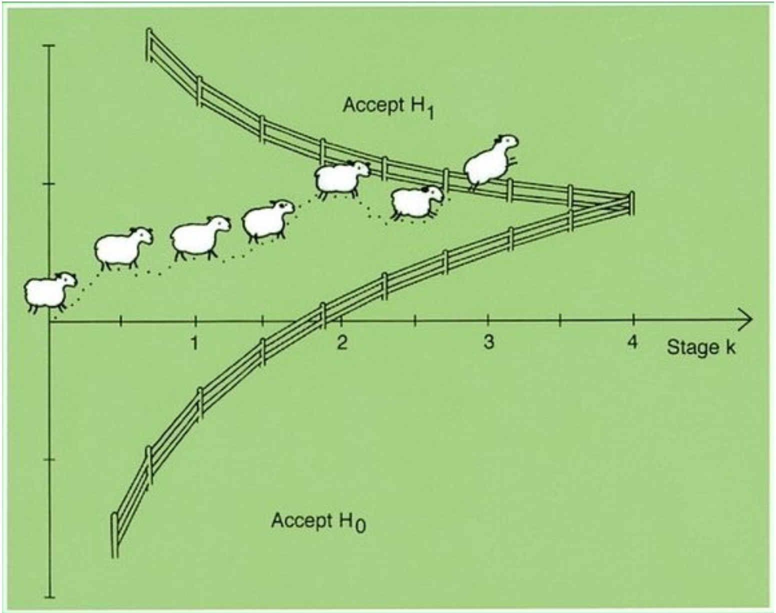
Goals for this talk

1. Describe bounds on distributional distance to multivariate normal distribution for
$$(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_K)$$
where $\hat{\theta}_k =$ MLE of $\theta \in \mathbb{R}^p$ at kth group sequential analysis, in regression setting
2. Advertise problems in sequential analysis that could (potentially) use Stein's method

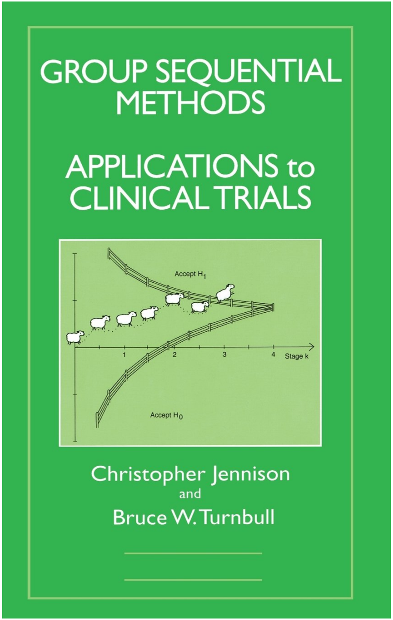
Warnings:

1. Work in progress
2. Not hard, but useful

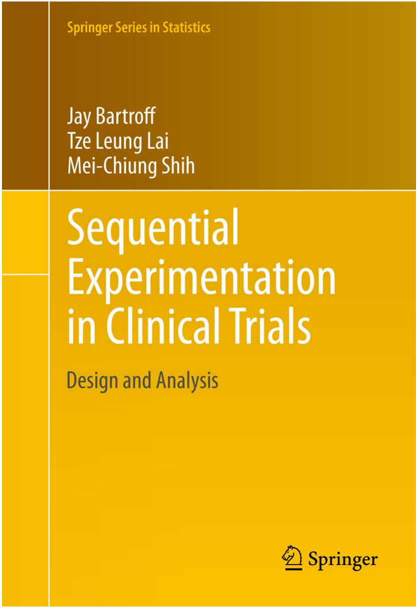
Group sequential analysis



Group sequential analysis



But there are other books on this subject…



Setup

Response $Y_i \in \mathbb{R}$ of ith patient depends on

- known covariate vector x_i
- unknown parameter vector $\theta \in \mathbb{R}^p$

Primary goal: To test a null hypothesis about θ, e.g.,

$$H_0 : \theta = 0$$
$$H_0' : \theta_j \leq 0$$
$$H_0'' : a^T\theta = b, \quad \text{some vector } a \text{, scalar } b$$

Secondary goals: Compute p-values or confidence regions for θ at the end of study

Setup

Response $Y_i \in \mathbb{R}$ of *i*th patient depends on

- known covariate vector x_i
- unknown parameter vector $\theta \in \mathbb{R}^p$

Primary goal: To test a null hypothesis about θ, e.g.,

$$\begin{aligned} H_0 &: \theta = 0 \\ H_0' &: \theta_j \leq 0 \\ H_0'' &: a^T\theta = b, \quad \text{some vector } a\text{, scalar } b \end{aligned}$$

Secondary goals: Compute *p*-values or confidence regions for θ at the end of study

Setup

Response $Y_i \in \mathbb{R}$ of ith patient depends on

- known covariate vector x_i
- unknown parameter vector $\theta \in \mathbb{R}^p$

Primary goal: To test a null hypothesis about θ, e.g.,

$$\begin{aligned} H_0 &: \theta = 0 \\ H_0' &: \theta_j \le 0 \\ H_0'' &: a^T\theta = b, \quad \text{some vector } a \text{, scalar } b \end{aligned}$$

Secondary goals: Compute p-values or confidence regions for θ at the end of study

Setup: Group sequential analysis

For efficiency, ethical, practical, financial reasons, the standard in trials has become group sequential analysis

A group sequential trial with at most K groups

$$\begin{array}{rl} \text{Group 1:} & Y_1, \ldots, Y_{n_1} \\ \text{Group 2:} & Y_{n_1+1}, \ldots, Y_{n_2} \\ \vdots & \\ \text{Group K:} & Y_{n_{K-1}+1}, \ldots, Y_{n_K} \end{array}$$

Group sequential dominant format for clinical trials since…

Beta-Blocker Heart Attack Trial ("BHAT", *JAMA* 82)

- Randomized trial of propranolol for heart attack survivors
- 3837 patients randomized
- Started June 1978, planned as $\leq$ 4-year study, terminated 8 months early due to observed benefit of propranolol

Setup: Group sequential analysis

For efficiency, ethical, practical, financial reasons, the standard in trials has become group sequential analysis

A group sequential trial with at most K groups

$$
\begin{aligned}
\text{Group 1:} \quad & Y_1, \ldots, Y_{n_1} \\
\text{Group 2:} \quad & Y_{n_1+1}, \ldots, Y_{n_2} \\
& \vdots \\
\text{Group K:} \quad & Y_{n_{K-1}+1}, \ldots, Y_{n_K}
\end{aligned}
$$

Group sequential dominant format for clinical trials since…

Beta-Blocker Heart Attack Trial ("BHAT", *JAMA* 82)

- Randomized trial of propranolol for heart attack survivors
- 3837 patients randomized
- Started June 1978, planned as $\leq$ 4-year study, terminated 8 months early due to observed benefit of propranolol

Setup: Group sequential analysis

For efficiency, ethical, practical, financial reasons, the standard in trials has become group sequential analysis

A group sequential trial with at most K groups

$$
\begin{aligned}
\text{Group 1:} \quad & Y_1, \ldots, Y_{n_1} \\
\text{Group 2:} \quad & Y_{n_1+1}, \ldots, Y_{n_2} \\
\vdots \quad & \\
\text{Group K:} \quad & Y_{n_{K-1}+1}, \ldots, Y_{n_K}
\end{aligned}
$$

Group sequential dominant format for clinical trials since...

Beta-Blocker Heart Attack Trial ("BHAT", *JAMA* 82)

- Randomized trial of propranolol for heart attack survivors
- 3837 patients randomized
- Started June 1978, planned as $\leq$ 4-year study, terminated 8 months early due to observed benefit of propranolol

Setup: Group sequential analysis

Stopping rule related to H_0

- likelihood ratio, t-, F-, χ^2- tests common
- Of the form:

Stop and reject H_0 at stage $\min\{k \leq K : T(Y_1, \ldots, Y_{n_k}) \geq C_k\}$

for some statistic $T(Y_1, \ldots, Y_{n_k})$, often a function of the MLE

$$\hat{\theta}_k = \hat{\theta}_k(Y_1, \ldots, Y_{n_k})$$

The joint distribution of

$$\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_K$$

needed to

- choose critical values C_k
- compute p-value at end of study
- give confidence region for θ at end of study

Setup: Group sequential analysis

Stopping rule related to H_0

- likelihood ratio, t-, F-, χ^2- tests common
- Of the form:

  Stop and reject H_0 at stage $\min\{k \leq K : T(Y_1, \ldots, Y_{n_k}) \geq C_k\}$

  for some statistic $T(Y_1, \ldots, Y_{n_k})$, often a function of the MLE

$$\hat{\theta}_k = \hat{\theta}_k(Y_1, \ldots, Y_{n_k})$$

The joint distribution of

$$\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_K$$

needed to

- choose critical values C_k
- compute p-value at end of study
- give confidence region for θ at end of study

Setup: Group sequential analysis

Stopping rule related to H_0

- likelihood ratio, t-, F-, χ^2- tests common
- Of the form:

  Stop and reject H_0 at stage $\min\{k \leq K : T(Y_1, \ldots, Y_{n_k}) \geq C_k\}$

  for some statistic $T(Y_1, \ldots, Y_{n_k})$, often a function of the MLE

$$\hat{\theta}_k = \hat{\theta}_k(Y_1, \ldots, Y_{n_k})$$

The joint distribution of

$$\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_K$$

needed to

- choose critical values C_k
- compute p-value at end of study
- give confidence region for θ at end of study

Background: Group sequential analysis

Jennison & Turnbull (*JASA* 97)

Asymptotic multivariate normal distribution of

$$(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_K)$$

in a regression setup $Y_i \overset{ind}{\sim} f_i(Y_i, \theta)$, f_i nice

- Asymptotics: $n_k - n_{k-1} \to \infty$ for all k, K fixed
- $E_\infty(\widehat{\theta}_k) = \theta$
- "Independent increments"

$$\mathrm{Cov}_\infty(\hat{\theta}_{k_1}, \hat{\theta}_{k_2}) = \mathrm{Var}_\infty(\hat{\theta}_{k_2}) \quad \text{any} \quad k_1 \leq k_2$$

"Folk Theorem"

- Normal limit widely (over-)used (software packages, etc.) before Jennison & Turnbull paper
- Commonly heard: "Once *n* is 5 or so the normal limit kicks in!"

Background: Group sequential analysis

Jennison & Turnbull (*JASA* 97)

Asymptotic multivariate normal distribution of

$$(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_K)$$

in a regression setup $Y_i \overset{ind}{\sim} f_i(Y_i, \theta)$, f_i nice

- Asymptotics: $n_k - n_{k-1} \to \infty$ for all k, K fixed
- $E_\infty(\widehat{\theta}_k) = \theta$
- "Independent increments"

$$\text{Cov}_\infty(\hat{\theta}_{k_1}, \hat{\theta}_{k_2}) = \text{Var}_\infty(\hat{\theta}_{k_2}) \quad \text{any} \quad k_1 \leq k_2$$

"Folk Theorem"

- Normal limit widely (over-)used (software packages, etc.) before Jennison & Turnbull paper
- Commonly heard: "Once *n* is 5 or so the normal limit kicks in!"

Background: Group sequential analysis

Jennison & Turnbull (*JASA* 97)

Asymptotic multivariate normal distribution of

$$(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_K)$$

in a regression setup $Y_i \overset{ind}{\sim} f_i(Y_i, \theta)$, f_i nice

- Asymptotics: $n_k - n_{k-1} \to \infty$ for all k, K fixed
- $E_\infty(\widehat{\theta}_k) = \theta$
- "Independent increments"

$$\mathrm{Cov}_\infty(\hat{\theta}_{k_1}, \hat{\theta}_{k_2}) = \mathrm{Var}_\infty(\hat{\theta}_{k_2}) \quad \text{any} \quad k_1 \le k_2$$

"Folk Theorem"

- Normal limit widely (over-)used (software packages, etc.) before Jennison & Turnbull paper
- Commonly heard: "Once *n* is 5 or so the normal limit kicks in!"

Background: Group sequential analysis

Jennison & Turnbull (*JASA* 97)

Independent increments

$$\mathrm{Cov}_\infty(\hat\theta_{k_1}, \hat\theta_{k_2}) = \mathrm{Var}_\infty(\hat\theta_{k_2}) \quad \text{any} \quad k_1 \le k_2$$

Suppose

$$H_0: a^T\theta = 0, \quad T_k = I_k(a^T\hat\theta_k) \quad \text{where} \quad I_k = [\mathrm{Var}_\infty(a^T\hat\theta_k)]^{-1}.$$

Then

$$\begin{aligned}
\mathrm{Cov}_\infty(T_{k_1}, T_{k_2}) &= I_{k_1} I_{k_2} a^T \mathrm{Cov}_\infty(\hat\theta_{k_1}, \hat\theta_{k_2}) a \\
&= I_{k_1} I_{k_2} a^T \mathrm{Var}_\infty(\hat\theta_{k_2}) a \\
&= I_{k_1} I_{k_2} \mathrm{Var}_\infty(T_{k_2}) \\
&= I_{k_1} = \mathrm{Var}_\infty(T_{k_1})
\end{aligned}$$

$\Rightarrow \mathrm{Cov}_\infty(T_{k_1}, T_{k_2} - T_{k_1}) = 0$

$\Rightarrow T_1, T_2 - T_1, \ldots, T_K - T_{K-1}$ asymptotically independent normals

Background: Group sequential analysis

Jennison & Turnbull (*JASA* 97)

Independent increments

$$\mathrm{Cov}_\infty(\hat\theta_{k_1}, \hat\theta_{k_2}) = \mathrm{Var}_\infty(\hat\theta_{k_2}) \quad \text{any} \quad k_1 \leq k_2$$

Suppose

$$H_0 : a^T\theta = 0, \quad T_k = I_k(a^T\hat\theta_k) \quad \text{where} \quad I_k = [\mathrm{Var}_\infty(a^T\hat\theta_k)]^{-1}.$$

Then

$$\begin{aligned}
\mathrm{Cov}_\infty(T_{k_1}, T_{k_2}) &= I_{k_1} I_{k_2} a^T \mathrm{Cov}_\infty(\hat\theta_{k_1}, \hat\theta_{k_2}) a \\
&= I_{k_1} I_{k_2} a^T \mathrm{Var}_\infty(\hat\theta_{k_2}) a \\
&= I_{k_1} I_{k_2} \mathrm{Var}_\infty(T_{k_2}) \\
&= I_{k_1} = \mathrm{Var}_\infty(T_{k_1})
\end{aligned}$$

$\Rightarrow \mathrm{Cov}_\infty(T_{k_1}, T_{k_2} - T_{k_1}) = 0$

$\Rightarrow T_1, T_2 - T_1, \ldots, T_K - T_{K-1}$ asymptotically independent normals

Background: Group sequential analysis

Jennison & Turnbull (*JASA* 97)

Independent increments

$$\mathrm{Cov}_\infty(\hat\theta_{k_1},\hat\theta_{k_2}) = \mathrm{Var}_\infty(\hat\theta_{k_2}) \quad \text{any} \quad k_1 \le k_2$$

Suppose

$$H_0 : a^T\theta = 0, \quad T_k = I_k(a^T\hat\theta_k) \quad \text{where} \quad I_k = [\mathrm{Var}_\infty(a^T\hat\theta_k)]^{-1}.$$

Then

$$\begin{aligned}
\mathrm{Cov}_\infty(T_{k_1}, T_{k_2}) &= I_{k_1} I_{k_2} a^T \mathrm{Cov}_\infty(\hat\theta_{k_1},\hat\theta_{k_2}) a \\
&= I_{k_1} I_{k_2} a^T \mathrm{Var}_\infty(\hat\theta_{k_2}) a \\
&= I_{k_1} I_{k_2} \mathrm{Var}_\infty(T_{k_2}) \\
&= I_{k_1} = \mathrm{Var}_\infty(T_{k_1})
\end{aligned}$$

$\Rightarrow \mathrm{Cov}_\infty(T_{k_1}, T_{k_2} - T_{k_1}) = 0$

$\Rightarrow T_1, T_2 - T_1, \ldots, T_K - T_{K-1}$ asymptotically independent normals

Background: Group sequential analysis

Jennison & Turnbull (*JASA* 97)

Independent increments

$$\text{Cov}_\infty(\hat{\theta}_{k_1}, \hat{\theta}_{k_2}) = \text{Var}_\infty(\hat{\theta}_{k_2}) \quad \text{any} \quad k_1 \leq k_2$$

Suppose

$$H_0 : a^T\theta = 0, \quad T_k = I_k(a^T\hat{\theta}_k) \quad \text{where} \quad I_k = [\text{Var}_\infty(a^T\hat{\theta}_k)]^{-1}.$$

Then

$$\begin{aligned}
\text{Cov}_\infty(T_{k_1}, T_{k_2}) &= I_{k_1} I_{k_2} a^T \text{Cov}_\infty(\hat{\theta}_{k_1}, \hat{\theta}_{k_2}) a \\
&= I_{k_1} I_{k_2} a^T \text{Var}_\infty(\hat{\theta}_{k_2}) a \\
&= I_{k_1} I_{k_2} \text{Var}_\infty(T_{k_2}) \\
&= I_{k_1} = \text{Var}_\infty(T_{k_1})
\end{aligned}$$

$\Rightarrow \text{Cov}_\infty(T_{k_1}, T_{k_2} - T_{k_1}) = 0$

$\Rightarrow T_1, T_2 - T_1, \ldots, T_K - T_{K-1}$ asymptotically independent normals

What I'm Doing

1. Berry-Esseen bound for multivariate normal limit for smooth functions
   - Anastasiou & Reinert 15: Bounds w/ explicit constants for bounded Wasserstein distance for scalar MLE ($K = 1$ analysis)
2. Relax independence assumption: Assume log-likelihood of $\mathcal{Y}_k := (Y_{n_{k-1}+1}, \ldots, Y_{n_k})$ is of the form

$$\sum_{i \in \mathcal{G}_k} \log f_i(Y_i, \theta) + g_k(\mathcal{Y}_k, \theta)$$

   for well-behaved functions f_i, g_k
   - $g_k = 0$ gives Jennison & Turnbull's independent setting
   - Some generalized linear mixed models (GLMMs) with random stage effect U_k take this form
     - U_k = effect due to lab, monitoring board, cohort, etc.
   - Penalized quasi-likelihood (Breslow & Clayton, *JASA* 93)

What I'm Doing

1. Berry-Esseen bound for multivariate normal limit for smooth functions
   - Anastasiou & Reinert 15: Bounds w/ explicit constants for bounded Wasserstein distance for scalar MLE ($K = 1$ analysis)
2. Relax independence assumption: Assume log-likelihood of $\mathcal{Y}_k := (Y_{n_{k-1}+1}, \ldots, Y_{n_k})$ is of the form

$$\sum_{i \in \mathcal{G}_k} \log f_i(Y_i, \theta) + g_k(\mathcal{Y}_k, \theta)$$

   for well-behaved functions f_i, g_k
   - $g_k = 0$ gives Jennison & Turnbull's independent setting
   - Some generalized linear mixed models (GLMMs) with random stage effect U_k take this form
     - $U_k =$ effect due to lab, monitoring board, cohort, etc.
   - Penalized quasi-likelihood (Breslow & Clayton, *JASA* 93)

What I'm Doing

1. Berry-Esseen bound for multivariate normal limit for smooth functions
   - Anastasiou & Reinert 15: Bounds w/ explicit constants for bounded Wasserstein distance for scalar MLE ($K = 1$ analysis)
2. Relax independence assumption: Assume log-likelihood of $\mathcal{Y}_k := (Y_{n_{k-1}+1}, \ldots, Y_{n_k})$ is of the form

$$\sum_{i \in \mathcal{G}_k} \log f_i(Y_i, \theta) + g_k(\mathcal{Y}_k, \theta)$$

   for well-behaved functions f_i, g_k
   - $g_k = 0$ gives Jennison & Turnbull's independent setting
   - Some generalized linear mixed models (GLMMs) with random stage effect U_k take this form
     - $U_k =$ effect due to lab, monitoring board, cohort, etc.
   - Penalized quasi-likelihood (Breslow & Clayton, *JASA* 93)

What I'm Doing

1. Berry-Esseen bound for multivariate normal limit for smooth functions
   - Anastasiou & Reinert 15: Bounds w/ explicit constants for bounded Wasserstein distance for scalar MLE ($K = 1$ analysis)
2. Relax independence assumption: Assume log-likelihood of $\mathcal{Y}_k := (Y_{n_{k-1}+1}, \ldots, Y_{n_k})$ is of the form

$$\sum_{i \in \mathcal{G}_k} \log f_i(Y_i, \theta) + g_k(\mathcal{Y}_k, \theta)$$

   for well-behaved functions f_i, g_k
   - $g_k = 0$ gives Jennison & Turnbull's independent setting
   - Some generalized linear mixed models (GLMMs) with random stage effect U_k take this form
     - U_k = effect due to lab, monitoring board, cohort, etc.
   - Penalized quasi-likelihood (Breslow & Clayton, *JASA* 93)

What I'm Doing

1. Berry-Esseen bound for multivariate normal limit for smooth functions
   - Anastasiou & Reinert 15: Bounds w/ explicit constants for bounded Wasserstein distance for scalar MLE ($K = 1$ analysis)
2. Relax independence assumption: Assume log-likelihood of $\mathcal{Y}_k := (Y_{n_{k-1}+1}, \ldots, Y_{n_k})$ is of the form

$$\sum_{i \in \mathcal{G}_k} \log f_i(Y_i, \theta) + g_k(\mathcal{Y}_k, \theta)$$

   for well-behaved functions f_i, g_k
   - $g_k = 0$ gives Jennison & Turnbull's independent setting
   - Some generalized linear mixed models (GLMMs) with random stage effect U_k take this form
     - ⋆ $U_k =$ effect due to lab, monitoring board, cohort, etc.
   - Penalized quasi-likelihood (Breslow & Clayton, *JASA* 93)

What I'm Doing

1. Berry-Esseen bound for multivariate normal limit for smooth functions
   - Anastasiou & Reinert 15: Bounds w/ explicit constants for bounded Wasserstein distance for scalar MLE ($K = 1$ analysis)
2. Relax independence assumption: Assume log-likelihood of $\mathcal{Y}_k := (Y_{n_{k-1}+1}, \ldots, Y_{n_k})$ is of the form

$$\sum_{i \in \mathcal{G}_k} \log f_i(Y_i, \theta) + g_k(\mathcal{Y}_k, \theta)$$

   for well-behaved functions f_i, g_k
   - $g_k = 0$ gives Jennison & Turnbull's independent setting
   - Some generalized linear mixed models (GLMMs) with random stage effect U_k take this form
     - $U_k =$ effect due to lab, monitoring board, cohort, etc.
   - Penalized quasi-likelihood (Breslow & Clayton, *JASA* 93)

GLMM Example: Poisson regression

Letting $f_\mu = \text{Po}(\mu)$ density,

$$\text{For } Y_i \text{ in } k\text{th stage,} \quad Y_i | U_k \overset{ind}{\sim} f_{\mu_i} \quad \text{where} \quad \mu_i = \exp(\beta^T x_i + U_k)$$
$$\{U_k\} \overset{iid}{\sim} h_\lambda$$
$$\theta = (\beta, \lambda).$$

Then log-likelihood is

$$\log\left(\prod_{k=1}^{K} \int \prod_{i \in \mathcal{G}_k} f_{\mu_i}(Y_i) h_\lambda(U_k) dU_k\right) = \sum_{k=1}^{K}\left(\sum_{i \in \mathcal{G}_k} \log f_{\widetilde{\mu}_i}(Y_i) + g_k(\mathcal{Y}_k, \theta)\right)$$

where $\widetilde{\mu}_i = \exp(\beta^T x_i)$.

GLMM Example: Poisson regression

Letting $f_\mu = \text{Po}(\mu)$ density,

$$\text{For } Y_i \text{ in } k\text{th stage,} \quad Y_i | U_k \overset{ind}{\sim} f_{\mu_i} \quad \text{where} \quad \mu_i = \exp(\beta^T x_i + U_k)$$
$$\{U_k\} \overset{iid}{\sim} h_\lambda$$
$$\theta = (\beta, \lambda).$$

Then log-likelihood is

$$\log\left(\prod_{k=1}^{K} \int \prod_{i \in \mathcal{G}_k} f_{\mu_i}(Y_i) h_\lambda(U_k) dU_k\right) = \sum_{k=1}^{K}\left(\sum_{i \in \mathcal{G}_k} \log f_{\widetilde{\mu}_i}(Y_i) + g_k(\mathcal{Y}_k, \theta)\right)$$

where $\widetilde{\mu}_i = \exp(\beta^T x_i)$.

GLMM Example: Poisson regression

Letting $f_\mu = \text{Po}(\mu)$ density,

$$\text{For } Y_i \text{ in } k\text{th stage,} \quad Y_i | U_k \overset{ind}{\sim} f_{\mu_i} \quad \text{where} \quad \mu_i = \exp(\beta^T x_i + U_k)$$
$$\{U_k\} \overset{iid}{\sim} h_\lambda$$
$$\theta = (\beta, \lambda).$$

Then log-likelihood is

$$\log\left(\prod_{k=1}^{K} \int \prod_{i \in \mathcal{G}_k} f_{\mu_i}(Y_i) h_\lambda(U_k) dU_k\right) = \sum_{k=1}^{K}\left(\sum_{i \in \mathcal{G}_k} \log f_{\widetilde{\mu}_i}(Y_i) + g_k(\mathcal{Y}_k, \theta)\right)$$

where $\widetilde{\mu}_i = \exp(\beta^T x_i)$.

GLMM Example: Poisson regression

Letting $f_\mu = \text{Po}(\mu)$ density,

$$\text{For } Y_i \text{ in } k\text{th stage,} \quad Y_i | U_k \overset{ind}{\sim} f_{\mu_i} \quad \text{where} \quad \mu_i = \exp(\beta^T x_i + U_k)$$
$$\{U_k\} \overset{iid}{\sim} h_\lambda$$
$$\theta = (\beta, \lambda).$$

Then log-likelihood is

$$\log\left(\prod_{k=1}^{K} \int \prod_{i \in \mathcal{G}_k} f_{\mu_i}(Y_i) h_\lambda(U_k) dU_k\right) = \sum_{k=1}^{K}\left(\sum_{i \in \mathcal{G}_k} \log f_{\widetilde{\mu}_i}(Y_i) + g_k(\mathcal{Y}_k, \theta)\right)$$

where $\widetilde{\mu}_i = \exp(\beta^T x_i)$.

Stein's Method for MVN Approximation

Generator approach: Barbour 90, Goetze 91

Size biasing: Goldstein & Rinott 96, Rinott & Rotar 96

Zero biasing: Goldstein & Reinert 05

Exchangeable pair: Chatterjee & Meckes 08, Reinert & Röllin 09

Stein couplings: Fang & Röllin 15

Theorem (Reinert & Röllin 09)

If $W, W' \in \mathbb{R}^q$ exchangeable pair with $EW = 0$, $EWW^T = \Sigma$ PD, and $E(W' - W|W) = \Lambda W + R$ with Λ invertible, then for any 3-times differentiable $h : \mathbb{R}^q \to \mathbb{R}$,

$$|Eh(W) - Eh(\Sigma^{1/2}Z)| \leq \frac{a|h|_2}{4} + \frac{b|h|_3}{12} + c\left(|h|_1 + \frac{q}{2}||\Sigma||^{1/2}|h|_2\right)$$

for certain a, b, c.

Stein's Method for MVN Approximation

Generator approach: Barbour 90, Goetze 91

Size biasing: Goldstein & Rinott 96, Rinott & Rotar 96

Zero biasing: Goldstein & Reinert 05

Exchangeable pair: Chatterjee & Meckes 08, Reinert & Röllin 09

Stein couplings: Fang & Röllin 15

Theorem (Reinert & Röllin 09)

If $W, W' \in \mathbb{R}^q$ exchangeable pair with $EW = 0$, $EWW^T = \Sigma$ PD, and $E(W' - W|W) = \Lambda W + R$ with Λ invertible, then for any 3-times differentiable $h : \mathbb{R}^q \to \mathbb{R}$,

$$|Eh(W) - Eh(\Sigma^{1/2}Z)| \leq \frac{a|h|_2}{4} + \frac{b|h|_3}{12} + c\left(|h|_1 + \frac{q}{2}||\Sigma||^{1/2}|h|_2\right)$$

for certain a, b, c.

Why Stein?

Dependent Case: No characteristic function based results (that I know of)

Independent Case: There are characteristic function based methods to handle sums of independent but non-identically distributed vectors

- Ulyanov 79, 87, 86
- Fujikoshi et al. 10

but no explicit constants (that I know of)

Why Stein?

Dependent Case: No characteristic function based results (that I know of)

Independent Case: There are characteristic function based methods to handle sums of independent but non-identically distributed vectors

- Ulyanov 79, 87, 86
- Fujikoshi et al. 10

but no explicit constants (that I know of)

Why Stein?

Dependent Case: No characteristic function based results (that I know of)

Independent Case: There are characteristic function based methods to handle sums of independent but non-identically distributed vectors

- Ulyanov 79, 87, 86
- Fujikoshi et al. 10

but no explicit constants (that I know of)

Bounds to normal for $\hat{\theta}^K := (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_K)$

Approach: Apply Reinert & Röllin 09 result with $W =$ score function increments to get smooth function bounds to normal.

Result

In the group sequential setup above, if the Y_i are independent or follow GLMMs with the log-likelihood of the kth group data $\mathcal{Y}_k = (Y_{n_{k-1}+1}, \ldots, Y_{n_k})$ of the form

$$\sum_{i \in \mathcal{G}_k} \log f_i(Y_i, \theta) + g_k(\mathcal{Y}_k, \theta),$$

then under regularity conditions on the f_i and g_k there are a, b, c, d s.t.

$$|Eh(J^{-1/2}(\hat{\theta}^K - \theta^K)) - Eh(Z)| \leq \frac{aK^2||J^{-1/2}||^2|h|_2}{4} + \frac{bK^3||J^{-1/2}||^3|h|_3}{12} + cK||J^{-1/2}||\left(|h|_1 + \frac{pK^2}{2}||\Sigma||^{1/2}||J^{-1/2}|||h|_2\right) + d.$$

Bounds to normal for $\hat{\theta}^K := (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_K)$

Approach: Apply Reinert & Röllin 09 result with $W =$ score function increments to get smooth function bounds to normal.

Result

In the group sequential setup above, if the Y_i are independent or follow GLMMs with the log-likelihood of the kth group data $\mathcal{Y}_k = (Y_{n_{k-1}+1}, \ldots, Y_{n_k})$ of the form

$$\sum_{i \in \mathcal{G}_k} \log f_i(Y_i, \theta) + g_k(\mathcal{Y}_k, \theta),$$

then under regularity conditions on the f_i and g_k there are a, b, c, d s.t.

$$|Eh(J^{-1/2}(\hat{\theta}^K - \theta^K)) - Eh(Z)| \leq \frac{aK^2||J^{-1/2}||^2|h|_2}{4} + \frac{bK^3||J^{-1/2}||^3|h|_3}{12} + cK||J^{-1/2}|| \left(|h|_1 + \frac{pK^2}{2}||\Sigma||^{1/2}||J^{-1/2}|||h|_2 \right) + d.$$

Comments on result

- a, b, c terms directly from Reinert & Röllin 09 bound
- c term $\propto$ Var(R) in

$$E(W' - W|W) = \Lambda W + R,$$

  vanishes in independent case
- d term is from Taylor Series remainders
- Rate $O(1/\sqrt{n_K})$ under regularity conditions and

$$\frac{n_k - n_{k-1}}{n_K} \to \gamma_k \in (0, 1)$$

Sketch of argument

Independent Case

Score statistic

$$S_i(\theta) = \frac{\partial}{\partial\theta} \log f_i(Y_i, \theta) \in \mathbb{R}^p, \quad W = \left(\sum_{i \in \mathcal{G}_1} S_i(\theta), \ldots, \sum_{i \in \mathcal{G}_K} S_i(\theta) \right) \in \mathbb{R}^q,$$

where $q = pK$.

Fisher Information

$$J_i(\theta) = -E\left(\frac{\partial}{\partial\theta} S_i(\theta)^T \right) \in \mathbb{R}^{p \times p}$$

$$J(\theta_1, \ldots, \theta_K) = \operatorname{diag}\left(\sum_{i=1}^{n_1} J_i(\theta_1), \ldots, \sum_{i=1}^{n_K} J_i(\theta_K) \right) \in \mathbb{R}^{q \times q}$$

$$\Sigma := \operatorname{Var}(W) = \operatorname{diag}\left(\sum_{i \in \mathcal{G}_1} J_i(\theta), \ldots, \sum_{i \in \mathcal{G}_K} J_i(\theta) \right) \in \mathbb{R}^{q \times q}$$

Sketch of argument

Independent Case

Score statistic

$$S_i(\theta) = \frac{\partial}{\partial\theta} \log f_i(Y_i, \theta) \in \mathbb{R}^p, \quad W = \left(\sum_{i \in \mathcal{G}_1} S_i(\theta), \ldots, \sum_{i \in \mathcal{G}_K} S_i(\theta) \right) \in \mathbb{R}^q,$$

where $q = pK$.

Fisher Information

$$J_i(\theta) = -E\left(\frac{\partial}{\partial\theta} S_i(\theta)^T \right) \in \mathbb{R}^{p \times p}$$

$$J(\theta_1, \ldots, \theta_K) = \mathrm{diag}\left(\sum_{i=1}^{n_1} J_i(\theta_1), \ldots, \sum_{i=1}^{n_K} J_i(\theta_K) \right) \in \mathbb{R}^{q \times q}$$

$$\Sigma := \mathrm{Var}(W) = \mathrm{diag}\left(\sum_{i \in \mathcal{G}_1} J_i(\theta), \ldots, \sum_{i \in \mathcal{G}_K} J_i(\theta) \right) \in \mathbb{R}^{q \times q}$$

Sketch of argument

Independent Case

Score statistic

$$S_i(\theta) = \frac{\partial}{\partial\theta} \log f_i(Y_i, \theta) \in \mathbb{R}^p, \quad W = \left(\sum_{i \in \mathcal{G}_1} S_i(\theta), \ldots, \sum_{i \in \mathcal{G}_K} S_i(\theta) \right) \in \mathbb{R}^q,$$

where $q = pK$.

Fisher Information

$$J_i(\theta) = -E\left(\frac{\partial}{\partial\theta} S_i(\theta)^T \right) \in \mathbb{R}^{p \times p}$$

$$J(\theta_1, \ldots, \theta_K) = \mathrm{diag}\left(\sum_{i=1}^{n_1} J_i(\theta_1), \ldots, \sum_{i=1}^{n_K} J_i(\theta_K) \right) \in \mathbb{R}^{q \times q}$$

$$\Sigma := \mathrm{Var}(W) = \mathrm{diag}\left(\sum_{i \in \mathcal{G}_1} J_i(\theta), \ldots, \sum_{i \in \mathcal{G}_K} J_i(\theta) \right) \in \mathbb{R}^{q \times q}$$

Sketch of argument: Exchangeable pair

Independent Case

1. Choose $i^* \in \{1, \ldots, n_K\}$ uniformly, independent of $Y_1, \ldots, Y_{n_K}$
2. Replace Y_{i^*} by independent copy Y'_{i^*} (keeping x_{i^*}), call result W'

$\Rightarrow W, W'$ exchangeable

$\Rightarrow W, W'$ satisfy linearity condition

$$E(W' - W | W) = -n_K^{-1} W$$

which is easy to check on each sub-p-vector

Sketch of argument: Exchangeable pair

Independent Case

1. Choose $i^* \in \{1, \ldots, n_K\}$ uniformly, independent of $Y_1, \ldots, Y_{n_K}$
2. Replace Y_{i*} by independent copy Y'_{i*} (keeping x_{i*}), call result W'

$\Rightarrow W, W'$ exchangeable

$\Rightarrow W, W'$ satisfy linearity condition

$$E(W' - W|W) = -n_K^{-1} W$$

which is easy to check on each sub-p-vector

Sketch of argument: Exchangeable pair

Independent Case

1. Choose $i^* \in \{1, \ldots, n_K\}$ uniformly, independent of $Y_1, \ldots, Y_{n_K}$
2. Replace Y_{i^*} by independent copy Y'_{i^*} (keeping x_{i^*}), call result W'

$\Rightarrow W, W'$ exchangeable

$\Rightarrow W, W'$ satisfy linearity condition

$$E(W' - W|W) = -n_K^{-1} W$$

which is easy to check on each sub-p-vector

Sketch of argument: Exchangeable pair

Independent Case

1. Choose $i^* \in \{1, \ldots, n_K\}$ uniformly, independent of $Y_1, \ldots, Y_{n_K}$
2. Replace Y_{i^*} by independent copy Y'_{i^*} (keeping x_{i^*}), call result W'

$\Rightarrow W, W'$ exchangeable

$\Rightarrow W, W'$ satisfy linearity condition

$$E(W' - W|W) = -n_K^{-1} W$$

which is easy to check on each sub-p-vector

Sketch of argument: Relating $\hat{\theta}^K$ to W

Independent Case

By standard Taylor series,

$$\hat{\theta}^K - \theta^K = J(\theta^{*K})^{-1} S, \quad \text{where} \quad S = \left(\sum_{i=1}^{n_1} S_i(\theta_1), \ldots, \sum_{i=1}^{n_K} S_i(\theta_K) \right) \in \mathbb{R}^q$$

and $\theta^{*K} \in \mathbb{R}^q$ on line segment connecting $\hat{\theta}^K, \theta^K$.

Then

$$|Eh(J^{1/2}(\hat{\theta}^K - \theta^K)) - Eh(Z)| \leq$$
$$|Eh(J^{-1/2}S) - Eh(Z)| + |Eh(J^{1/2}J(\theta^{*K})^{-1}S) - Eh(J^{-1/2}S)|$$

Sketch of argument: Relating $\hat{\theta}^K$ to W

Independent Case

By standard Taylor series,

$$\hat{\theta}^K - \theta^K = J(\theta^{*K})^{-1} S, \quad \text{where} \quad S = \left(\sum_{i=1}^{n_1} S_i(\theta_1), \ldots, \sum_{i=1}^{n_K} S_i(\theta_K) \right) \in \mathbb{R}^q$$

and $\theta^{*K} \in \mathbb{R}^q$ on line segment connecting $\hat{\theta}^K, \theta^K$.

Then

$$|Eh(J^{1/2}(\hat{\theta}^K - \theta^K)) - Eh(Z)| \leq$$
$$|Eh(J^{-1/2}S) - Eh(Z)| + |Eh(J^{1/2}J(\theta^{*K})^{-1}S) - Eh(J^{-1/2}S)|$$

Sketch of argument: Relating $\hat{\theta}^K$ to W
Independent Case

Using $S = AW$ where

$$A = \begin{bmatrix} 1_p & 0_p & \cdots & 0_p \\ 1_p & 1_p & \cdots & 0_p \\ \vdots & \vdots & \ddots & \vdots \\ 1_p & 1_p & \cdots & 1_p \end{bmatrix}, \qquad 1_p, 0_p \in \mathbb{R}^{p\times p} \quad \text{identity and 0 matrices,}$$

1st term is

$$|Eh(J^{-1/2}S) - Eh(Z)| = |E\widetilde{h}(W) - E\widetilde{h}(\Sigma^{1/2}Z)|$$

where $\widetilde{h}(w) = h(J^{-1/2}Aw)$, then apply Reinert-Röllin and simplify.

2nd term is bounded by Taylor series arguments.

Sketch of argument: Relating $\hat{\theta}^K$ to W

Independent Case

Using $S = AW$ where

$$A = \begin{bmatrix} 1_p & 0_p & \cdots & 0_p \\ 1_p & 1_p & \cdots & 0_p \\ \vdots & \vdots & \ddots & \vdots \\ 1_p & 1_p & \cdots & 1_p \end{bmatrix}, \qquad 1_p, 0_p \in \mathbb{R}^{p \times p} \quad \text{identity and 0 matrices,}$$

1st term is

$$|Eh(J^{-1/2}S) - Eh(Z)| = |E\widetilde{h}(W) - E\widetilde{h}(\Sigma^{1/2}Z)|$$

where $\widetilde{h}(w) = h(J^{-1/2}Aw)$, then apply Reinert-Röllin and simplify.

2nd term is bounded by Taylor series arguments.

Sketch of argument: Exchangeable pair

GLMM Case

1. Choose $i^* \in \{1, \ldots, n_K\}$ uniformly, independent of $Y_1, \ldots, Y_{n_K}$
2. If i^* in kth group, replace Y_{i^*} by independent copy Y'_{i^*} with mean

$$\varphi(\beta^T x_{i^*} + U_k), \quad \varphi^{-1} = \text{ link function}$$

   (same covariates x_{i^*}, group effect U_k), call result W'

$\Rightarrow W, W'$ exchangeable

$\Rightarrow W, W'$ satisfy linearity condition

$$E(W' - W|W) = -n_K^{-1} W + R$$

where $R = R(g_1, \ldots, g_K)$

Sketch of argument: Exchangeable pair
GLMM Case

1. Choose $i^* \in \{1, \ldots, n_K\}$ uniformly, independent of $Y_1, \ldots, Y_{n_K}$
2. If i^* in kth group, replace Y_{i^*} by independent copy Y'_{i^*} with mean

$$\varphi(\beta^T x_{i^*} + U_k), \quad \varphi^{-1} = \text{ link function}$$

   (same covariates x_{i^*}, group effect U_k), call result W'

$\Rightarrow W, W'$ exchangeable

$\Rightarrow W, W'$ satisfy linearity condition

$$E(W' - W | W) = -n_K^{-1} W + R$$

where $R = R(g_1, \ldots, g_K)$

Sketch of argument: Exchangeable pair

GLMM Case

1. Choose $i^* \in \{1, \ldots, n_K\}$ uniformly, independent of $Y_1, \ldots, Y_{n_K}$
2. If i^* in kth group, replace Y_{i^*} by independent copy Y'_{i^*} with mean

$$\varphi(\beta^T x_{i^*} + U_k), \quad \varphi^{-1} = \text{ link function}$$

(same covariates x_{i^*}, group effect U_k), call result W'

$\Rightarrow W, W'$ exchangeable

$\Rightarrow W, W'$ satisfy linearity condition

$$E(W' - W|W) = -n_K^{-1} W + R$$

where $R = R(g_1, \ldots, g_K)$

Sketch of argument: Exchangeable pair

GLMM Case

1. Choose $i^* \in \{1, \ldots, n_K\}$ uniformly, independent of $Y_1, \ldots, Y_{n_K}$
2. If i^* in kth group, replace Y_{i^*} by independent copy Y'_{i^*} with mean

$$\varphi(\beta^T x_{i^*} + U_k), \quad \varphi^{-1} = \text{ link function}$$

(same covariates x_{i^*}, group effect U_k), call result W'

$\Rightarrow W, W'$ exchangeable

$\Rightarrow W, W'$ satisfy linearity condition

$$E(W' - W | W) = -n_K^{-1} W + R$$

where $R = R(g_1, \ldots, g_K)$

Sketch of argument: Exchangeable pair

GLMM Case

1. Choose $i^* \in \{1, \ldots, n_K\}$ uniformly, independent of $Y_1, \ldots, Y_{n_K}$
2. If i^* in kth group, replace Y_{i^*} by independent copy Y'_{i^*} with mean

$$\varphi(\beta^T x_{i^*} + U_k), \quad \varphi^{-1} = \text{ link function}$$

(same covariates x_{i^*}, group effect U_k), call result W'

$\Rightarrow W, W'$ exchangeable

$\Rightarrow W, W'$ satisfy linearity condition

$$E(W' - W|W) = -n_K^{-1} W + R$$

where $R = R(g_1, \ldots, g_K)$

Other Sequential Problems

- Dose finding problems
- Distribution of stopped sequential test statistic
- Overshoot over the boundary
- Changepoint problems


THANK YOU FOR LISTENING
TROLL.ME