Ranking and Selection with Covariates

Haihui SHEN

Department of Management Sciences
College of Business
City University of Hong Kong

Joint work with L. Jeff Hong (City University of Hong Kong) and
Xiaowei Zhang (The Hong Kong University of Science and Technology)

Winter Simulation Conference 2017

Contents

Covariates

- Covariates ($\boldsymbol{X}$) are also known as personalized information, side information, auxiliary quantities or contextual variables.
- Given $\boldsymbol{X} = \boldsymbol{x}$, the performance of alternative i is $\mu_i(\boldsymbol{x})$, in many cases.

Covariates

- Covariates ($\boldsymbol{X}$) are also known as personalized information, side information, auxiliary quantities or contextual variables.
- Given $\boldsymbol{X} = \boldsymbol{x}$, the performance of alternative i is $\mu_i(\boldsymbol{x})$, in many cases.
- Examples:
  1. Healthcare: *Personalized medicine*.
  2. Marketing: *Personalized recommendations and promotions*.

Covariates

- Covariates ($\boldsymbol{X}$) are also known as personalized information, side information, auxiliary quantities or contextual variables.
- Given $\boldsymbol{X} = \boldsymbol{x}$, the performance of alternative i is $\mu_i(\boldsymbol{x})$, in many cases.
- Examples:
  1. Healthcare: *Personalized medicine*.
  2. Marketing: *Personalized recommendations and promotions*.
- Covariates allow decisions to be made at individual level.

Value of Covariates

- In traditional ranking and selection (R&S), we may solve

$$\arg\max_{1\leq i\leq k} \mu_i \equiv \mathbb{E}[\mu_i(\boldsymbol{X})],$$

  if we are risk-neutral with respect to the covariates.

- With covariates, we can actually try to solve, given $\boldsymbol{X} = \boldsymbol{x}$,

$$\arg\max_{1\leq i\leq k} \mu_i(\boldsymbol{x}).$$

Value of Covariates

- In traditional ranking and selection (R&S), we may solve

$$\arg\max_{1\le i\le k} \mu_i \equiv \mathbb{E}[\mu_i(\boldsymbol{X})],$$

  if we are risk-neutral with respect to the covariates.

- With covariates, we can actually try to solve, given $\boldsymbol{X} = \boldsymbol{x}$,

$$\arg\max_{1\le i\le k} \mu_i(\boldsymbol{x}).$$

- By Jensen's inequality,

$$\mathbb{E}\left\{\max_{1\le i\le k} \mu_i(\boldsymbol{X})\right\} \ge \max_{1\le i\le k} \mathbb{E}[\mu_i(\boldsymbol{X})].$$

Ranking and Selection with Covariates

- We introduce a new framework of ranking and selection problems in simulation, which is called ranking and selection with covariates (R&S-C):
  - Performance of an alternative depends on some observable random covariates;
  - The best alternative is a function of the covariates;
  - A selection procedure is required to produce a decision rule (i.e., an estimator of the function).

Ranking and Selection with Covariates

- We introduce a new framework of ranking and selection problems in simulation, which is called ranking and selection with covariates (R&S-C):
  - Performance of an alternative depends on some observable random covariates;
  - The best alternative is a function of the covariates;
  - A selection procedure is required to produce a decision rule (i.e., an estimator of the function).
- A decision rule is produced *offline*. But it can be applied *online* to select the best alternative for the subsequent individuals after observing their covariates.

Related Literature

- Traditional R&S:
  - Frequentist approaches: Dudewicz and Dalal (1975), Rinott (1978), Kim and Nelson (2001), Hong (2006), etc.
  - Bayesian approaches: Chen et al. (1997), Chick and Inoue (2001), Frazier et al. (2008), Chick and Frazier (2012), etc.

Related Literature

- Traditional R&S:
  - Frequentist approaches: Dudewicz and Dalal (1975), Rinott (1978), Kim and Nelson (2001), Hong (2006), etc.
  - Bayesian approaches: Chen et al. (1997), Chick and Inoue (2001), Frazier et al. (2008), Chick and Frazier (2012), etc.
- Multi-armed bandit (MAB) with covariates:
  - Parametric bandits: Auer (2002), Rusmevichientong and Tsitsiklis (2010), Goldenshluger and Zeevi (2013), etc.
  - Non-parametric bandits: Rigollet and Zeevi (2010), Perchet and Rigollet (2013), Slivkins (2014), etc.

Related Literature

- Traditional R&S:
  - Frequentist approaches: Dudewicz and Dalal (1975), Rinott (1978), Kim and Nelson (2001), Hong (2006), etc.
  - Bayesian approaches: Chen et al. (1997), Chick and Inoue (2001), Frazier et al. (2008), Chick and Frazier (2012), etc.
- Multi-armed bandit (MAB) with covariates:
  - Parametric bandits: Auer (2002), Rusmevichientong and Tsitsiklis (2010), Goldenshluger and Zeevi (2013), etc.
  - Non-parametric bandits: Rigollet and Zeevi (2010), Perchet and Rigollet (2013), Slivkins (2014), etc.
- R&S with covariates:
  - Not yet defined and studied.
  - Our work serves as an attempt to fill in the gap.

Notations

- There are k alternatives, whose performance depends on the random covariates $\boldsymbol{X}_{\mathsf{c}} = (X_1, \ldots, X_d)^\top$ with support $\Theta_{\mathsf{c}} \subseteq \mathbb{R}^d$.
- Let $\boldsymbol{X} := (1, \boldsymbol{X}_{\mathsf{c}}^\top)^\top$ be the augmented covariates with support $\Theta := \{1\} \times \Theta_{\mathsf{c}}$.
- Let $\boldsymbol{x}$ be the realization of $\boldsymbol{X}$. For each $i = 1, \ldots, k$, let $Y_{i\ell}(\boldsymbol{x})$ denote the ℓth sample (observation) of performance on $\boldsymbol{x}$ from alternative i, $\ell = 1, 2, \ldots$.

Linear Model

Assumption 1 (A1)

For each $i = 1, \ldots, k$ *and* $\ell = 1, 2, \ldots$, *conditioning on* $\boldsymbol{X} = \boldsymbol{x}$,

$$Y_{i\ell}(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{\beta}_i + \epsilon_{i\ell}(\boldsymbol{x}),$$

where $\boldsymbol{\beta}_i = (\beta_{i0}, \beta_{i1}, \ldots, \beta_{id})^\top \in \mathbb{R}^{d+1}$ *is a vector of unknown parameters, and* $\epsilon_{i\ell}(\boldsymbol{x})$ *is random error which satisfies:*

(i) $\epsilon_{i\ell}(\boldsymbol{x}) \sim \mathcal{N}(0, \sigma_i^2(\boldsymbol{x}))$;

(ii) $\epsilon_{i\ell}(\boldsymbol{x})$ *is independent of* $\epsilon_{i'\ell'}(\boldsymbol{x}')$ *for any* $(i, \ell, \boldsymbol{x}) \neq (i', \ell', \boldsymbol{x}')$.

Linear Model

Assumption 1 (A1)

For each $i = 1, \ldots, k$ *and* $\ell = 1, 2, \ldots$, *conditioning on* $\boldsymbol{X} = \boldsymbol{x}$,

$$Y_{i\ell}(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{\beta}_i + \epsilon_{i\ell}(\boldsymbol{x}),$$

where $\boldsymbol{\beta}_i = (\beta_{i0}, \beta_{i1}, \ldots, \beta_{id})^\top \in \mathbb{R}^{d+1}$ *is a vector of unknown parameters, and* $\epsilon_{i\ell}(\boldsymbol{x})$ *is random error which satisfies:*

(i) $\epsilon_{i\ell}(\boldsymbol{x}) \sim \mathcal{N}(0, \sigma_i^2(\boldsymbol{x}))$;

(ii) $\epsilon_{i\ell}(\boldsymbol{x})$ *is independent of* $\epsilon_{i'\ell'}(\boldsymbol{x}')$ *for any* $(i, \ell, \boldsymbol{x}) \neq (i', \ell', \boldsymbol{x}')$.

Remark

- A1 (i) allows the sampling errors to have unequal variances;
- A1 (ii) contains two layers of independence.

Objective

- The objective is to select the alternative with the largest mean performance conditioning on $\boldsymbol{X}$, i.e., to find

$$i^*(\boldsymbol{x}) := \arg\max_{1 \le i \le k} \left\{ \boldsymbol{X}^\top \boldsymbol{\beta}_i \,|\boldsymbol{X} = \boldsymbol{x} \right\}.$$

- Let $\widehat{i^*}(\boldsymbol{x})$ denote the selected alternative based on the decision rule produced by certain procedure.

Objective

- The objective is to select the alternative with the largest mean performance conditioning on $\boldsymbol{X}$, i.e., to find

$$i^*(\boldsymbol{x}) := \arg\max_{1 \le i \le k} \left\{ \boldsymbol{X}^\top \boldsymbol{\beta}_i \,|\boldsymbol{X} = \boldsymbol{x} \right\}.$$

- Let $\widehat{i^*}(\boldsymbol{x})$ denote the selected alternative based on the decision rule produced by certain procedure.

- Indifference-zone (IZ) formulation: Define the event of correct selection (CS) as

$$\left\{ \boldsymbol{X}^\top \boldsymbol{\beta}_{i^*(\boldsymbol{X})} - \boldsymbol{X}^\top \boldsymbol{\beta}_{\widehat{i^*}(\boldsymbol{X})} < \delta \middle| \, \boldsymbol{X} = \boldsymbol{x} \right\},$$

  for a prespecified IZ parameter $\delta > 0$.

Probability of Correct Selection (PCS)

- We first define the *conditional* PCS as

$$\mathrm{PCS}(\boldsymbol{x}) := \mathbb{P}\left\{ \boldsymbol{X}^\top \boldsymbol{\beta}_{i^*(\boldsymbol{X})} - \boldsymbol{X}^\top \boldsymbol{\beta}_{\widehat{i^*}(\boldsymbol{X})} < \delta \middle| \boldsymbol{X} = \boldsymbol{x} \right\},$$

  where the probability is with respect to the distribution of the samples used by the procedure that produces $\widehat{i^*}(\boldsymbol{x})$.

Probability of Correct Selection (PCS)

- We first define the *conditional* PCS as

$$\mathrm{PCS}(\boldsymbol{x}) := \mathbb{P}\left\{\boldsymbol{X}^\top \boldsymbol{\beta}_{i^*(\boldsymbol{X})} - \boldsymbol{X}^\top \boldsymbol{\beta}_{\widehat{i^*}(\boldsymbol{X})} < \delta \middle| \boldsymbol{X} = \boldsymbol{x}\right\},$$

  where the probability is with respect to the distribution of the samples used by the procedure that produces $\widehat{i^*}(\boldsymbol{x})$.

- Forms of *unconditional* PCS:
  - Distribution of $\boldsymbol{X}$ is known: $\mathrm{PCS}_{\mathsf{E}} := \mathbb{E}\left[\mathrm{PCS}(\boldsymbol{X})\right]$.
  - Distribution of $\boldsymbol{X}$ is unknown: $\mathrm{PCS}_{\mathsf{min}} := \min_{\boldsymbol{x}\in\Theta} \mathrm{PCS}(\boldsymbol{x})$.
  - Other forms may also be possible.

Probability of Correct Selection (PCS)

- We first define the *conditional* PCS as

$$\mathrm{PCS}(\boldsymbol{x}) := \mathbb{P}\left\{\boldsymbol{X}^\top \boldsymbol{\beta}_{i^*(\boldsymbol{X})} - \boldsymbol{X}^\top \boldsymbol{\beta}_{\widehat{i^*}(\boldsymbol{X})} < \delta \middle| \boldsymbol{X} = \boldsymbol{x}\right\},$$

  where the probability is with respect to the distribution of the samples used by the procedure that produces $\widehat{i^*}(\boldsymbol{x})$.

- Forms of *unconditional* PCS:
  - Distribution of $\boldsymbol{X}$ is known: $\mathrm{PCS}_{\mathsf{E}} := \mathbb{E}\left[\mathrm{PCS}(\boldsymbol{X})\right]$.
  - Distribution of $\boldsymbol{X}$ is unknown: $\mathrm{PCS}_{\mathsf{min}} := \min_{\boldsymbol{x}\in\Theta} \mathrm{PCS}(\boldsymbol{x})$.
  - Other forms may also be possible.

- Final Goal: to develop some procedures to produce $\widehat{i^*}(\boldsymbol{x})$, which guarantees a particular unconditional PCS is $\geq 1-\alpha$.

Fixed Design

- Choose $m \geq d+1$ design points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m \in \Theta$.
- Assume that alternative i can be sampled at design point $\boldsymbol{x}_j$ as many times as we want, for each $i = 1, \ldots, k$ and $j = 1, \ldots, m$.

Fixed Design

- Choose $m \geq d+1$ design points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m \in \Theta$.
- Assume that alternative i can be sampled at design point $\boldsymbol{x}_j$ as many times as we want, for each $i = 1, \ldots, k$ and $j = 1, \ldots, m$.

Remark

- Fixed design is suitable when a simulation model is available.
- When observations are collected from real experiments, fixed design may sometimes be impossible.

Fixed Design

- Choose $m \geq d+1$ design points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m \in \Theta$.
- Assume that alternative i can be sampled at design point $\boldsymbol{x}_j$ as many times as we want, for each $i = 1, \ldots, k$ and $j = 1, \ldots, m$.

Remark

- Fixed design is suitable when a simulation model is available.
- When observations are collected from real experiments, fixed design may sometimes be impossible.

Assumption 2 (A2)

$\mathcal{X}^\top \mathcal{X}$ is nonsingular, where $\mathcal{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m)^\top \in \mathbb{R}^{m \times (d+1)}$.

Homoscedastic or Heteroscedastic Errors

- Selection procedures are designed separately depending on whether the simulation errors are *homoscedastic* (A3) or *heteroscedastic* (A4).

Assumption 3 (A3)

$\sigma_i^2(\boldsymbol{x}) \equiv \sigma_i^2 < \infty$ *for* $\boldsymbol{x} \in \Theta$ *and* $i = 1, \ldots, k$.

Assumption 4 (A4)

$\sigma_i^2(\boldsymbol{x}) < \infty$ *is a function of* $\boldsymbol{x} \in \Theta$ *and* $i = 1, \ldots, k$.

Homoscedastic or Heteroscedastic Errors

- Selection procedures are designed separately depending on whether the simulation errors are *homoscedastic* (A3) or *heteroscedastic* (A4).

Assumption 3 (A3)

$\sigma_i^2(\boldsymbol{x}) \equiv \sigma_i^2 < \infty$ *for* $\boldsymbol{x} \in \Theta$ *and* $i = 1, \ldots, k$.

Assumption 4 (A4)

$\sigma_i^2(\boldsymbol{x}) < \infty$ *is a function of* $\boldsymbol{x} \in \Theta$ *and* $i = 1, \ldots, k$.

- This analogizes the difference between the *ordinary least squares* method and the *generalized least squares* method in linear regression.

Homoscedastic Errors - Procedure FDHom

Setup. Specify δ, α, m and $\mathcal{X}$. Determine n_0, the first-stage sample size. Calculate the critical constant h (as shown in next slide).

Stage 1. For all $i = 1, \ldots, k$, take n_0 batches of observations on $\mathcal{X}$: $\boldsymbol{Y}_{i\ell} = (Y_{i\ell}(\boldsymbol{x}_1), \ldots, Y_{i\ell}(\boldsymbol{x}_m))^\top, \ell = 1, \ldots, n_0$. Let

$$\widehat{\boldsymbol{\beta}}_i(n_0) = \frac{1}{n_0}(\mathcal{X}^\top\mathcal{X})^{-1}\mathcal{X}^\top \sum_{\ell=1}^{n_0} \boldsymbol{Y}_{i\ell},$$

$$S_i^2 = \frac{1}{n_0 m - 1 - d} \sum_{\ell=1}^{n_0} (\boldsymbol{Y}_{i\ell} - \mathcal{X}\widehat{\boldsymbol{\beta}}_i(n_0))^\top (\boldsymbol{Y}_{i\ell} - \mathcal{X}\widehat{\boldsymbol{\beta}}_i(n_0)).$$

Furthermore, let $N_i = \max\left\{\left\lceil \frac{h^2 S_i^2}{\delta^2} \right\rceil, n_0\right\}$.

Stage 2. For all $i = 1, \ldots, k$, take $N_i - n_0$ batches of observations on $\mathcal{X}$ and denote them as $\boldsymbol{Y}_{i,n_0+1}, \ldots, \boldsymbol{Y}_{iN_i}$. Let $\widehat{\boldsymbol{\beta}}_i = \frac{1}{N_i}(\mathcal{X}^\top\mathcal{X})^{-1}\mathcal{X}^\top \sum_{\ell=1}^{N_i} \boldsymbol{Y}_{i\ell}$.

Selection. Return $\widehat{i^*}(\boldsymbol{x}) = \arg\max_{1\le i\le k}\{\boldsymbol{x}^\top \widehat{\boldsymbol{\beta}}_i\}$ as the decision rule.

Homoscedastic Errors - Procedure FDHom

- When $\text{PCS}_{\text{E}} \geq 1-\alpha$ is designed, $h = h_{\text{E}}$, which satisfies

$$\mathbb{E}\left\{\int_0^\infty \left[\int_0^\infty \Phi\left(\frac{h_{\text{E}}}{\sqrt{(n_0 m-1-d)\left(\frac{1}{t}+\frac{1}{s}\right)\boldsymbol{X}^\top(\mathcal{X}^\top\mathcal{X})^{-1}\boldsymbol{X}}}\right) f(s)ds\right]^{k-1} f(t)dt\right\} = 1-\alpha,$$

  where $\Phi(\cdot)$ is the standard normal cdf, $f(\cdot)$ is pdf of chi-squared RV with $n_0 m - 1 - d$ degrees of freedom.

- When $\text{PCS}_{\text{min}} \geq 1-\alpha$ is designed, $h = h_{\text{min}}$, which satisfies

$$\min_{\boldsymbol{x}\in\Theta}\left\{\int_0^\infty \left[\int_0^\infty \Phi\left(\frac{h_{\text{min}}}{\sqrt{(n_0 m-1-d)\left(\frac{1}{t}+\frac{1}{s}\right)\boldsymbol{x}^\top(\mathcal{X}^\top\mathcal{X})^{-1}\boldsymbol{x}}}\right) f(s)ds\right]^{k-1} f(t)dt\right\} = 1-\alpha.$$

Homoscedastic Errors - Procedure FDHom

- Procedure FDHom is in a similar form of the classical Rinott's procedure (Rinott 1978) in traditional R&S.
- We have the following statistical validity of Procedure FDHom:

Theorem 1

Under A1 – A3, Procedure FDHom ensures that the unconditional PCS is at least $1-\alpha$*, i.e.,* $PCS_E \geq 1-\alpha$ *or* $PCS_{min} \geq 1-\alpha$.

Homoscedastic Errors - Procedure FDHom

- Procedure FDHom is in a similar form of the classical Rinott's procedure (Rinott 1978) in traditional R&S.
- We have the following statistical validity of Procedure FDHom:

Theorem 1

Under A1 – A3, Procedure FDHom ensures that the unconditional PCS is at least $1-\alpha$*, i.e.,* $PCS_E \geq 1-\alpha$ *or* $PCS_{min} \geq 1-\alpha$.

- The proof of Theorem 1 is based on Lemma 1 (shown in next slide), which is an extension of the result in Stein (1945).

Homoscedastic Errors - Procedure FDHom

Lemma 1

Let $\boldsymbol{Y} = \mathcal{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, *where* $\boldsymbol{\beta} \in \mathbb{R}^d$, $\mathcal{X} \in \mathbb{R}^{m\times d}$, *and* $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathcal{I})$. *Assume that* $\mathcal{X}^\intercal\mathcal{X}$ *is nonsingular. Let* T *be a random variable independent of* $\sum_{\ell=1}^{n} \boldsymbol{Y}_\ell$ *and of* $\{\boldsymbol{Y}_\ell : \ell \geq n+1\}$, *where* $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots$ *are independent samples of* $\boldsymbol{Y}$. *Suppose that* $N \geq n$ *is an integer-valued function of* T *and no other random variables. Let* $\widehat{\boldsymbol{\beta}} = N^{-1}(\mathcal{X}^\top\mathcal{X})^{-1}\mathcal{X}^\top \sum_{\ell=1}^{N} \boldsymbol{Y}_\ell$. *Then, for any* $\boldsymbol{x} \in \mathbb{R}^d$,

(i) $\boldsymbol{x}^\top\widehat{\boldsymbol{\beta}}\big|T \sim \mathcal{N}\left(\boldsymbol{x}^\top\boldsymbol{\beta}, \frac{\sigma^2}{N}\boldsymbol{x}^\top(\mathcal{X}^\top\mathcal{X})^{-1}\boldsymbol{x}\right)$;

(ii) $\dfrac{\sqrt{N}(\boldsymbol{x}^\top\widehat{\boldsymbol{\beta}} - \boldsymbol{x}^\top\boldsymbol{\beta})}{\sigma\sqrt{\boldsymbol{x}^\top(\mathcal{X}^\top\mathcal{X})^{-1}\boldsymbol{x}}}$ *is independent of* T *and has the standard normal distribution.*

Lemma 1 vs Stein Result

- Stein result is a cornerstone for two-stage procedures of R&S:
  - Stage 1: Take n_0 samples for one alternative, i.e., $Y_1, \ldots, Y_{n_0}$, which are i.i.d. $\mathcal{N}(\mu, \sigma^2)$. $\Rightarrow \bar{Y}(n_0)$, S^2.
  - Stage 2: Take $N - n_0$ additional samples, where N depends on $Y_1, \ldots, Y_{n_0}$. $\Rightarrow \bar{Y}(N)$.
  - In general, the distribution of $\bar{Y}(N)|N$ is unknown!
  - But if $N \perp \bar{Y}(n_0)$ (e.g., N is function only of S^2), we have

$$\bar{Y}(N)|N \sim \mathcal{N}(\mu, \sigma^2/N), \quad \frac{\bar{Y}(N) - \mu}{\sigma/\sqrt{N}} \sim \mathcal{N}(0, 1).$$

Lemma 1 vs Stein Result

- Stein result is a cornerstone for two-stage procedures of R&S:
  - Stage 1: Take n_0 samples for one alternative, i.e., $Y_1, \ldots, Y_{n_0}$, which are i.i.d. $\mathcal{N}(\mu, \sigma^2)$. $\Rightarrow \bar{Y}(n_0)$, S^2.
  - Stage 2: Take $N - n_0$ additional samples, where N depends on $Y_1, \ldots, Y_{n_0}$. $\Rightarrow \bar{Y}(N)$.
  - In general, the distribution of $\bar{Y}(N)|N$ is unknown!
  - But if $N \perp \bar{Y}(n_0)$ (e.g., N is function only of S^2), we have

$$\bar{Y}(N)|N \sim \mathcal{N}(\mu, \sigma^2/N), \quad \frac{\bar{Y}(N) - \mu}{\sigma/\sqrt{N}} \sim \mathcal{N}(0, 1).$$

- Lemma 1 extends Stein result to R&S-C setting (i.e., linear regression model) and enables us to analyze the finite-sample property (distribution).

Heteroscedastic Errors - Procedure FDHet

- Procedure FDHet is similar to the Procedure FDHom, except

$$N_{ij} = \max\left\{\left\lceil \frac{h^2 S_{ij}^2}{\delta^2} \right\rceil, n_0\right\}.$$

Theorem 2

Under A1, A2 and A4, Procedure FDHet ensures that the unconditional PCS is at least $1-\alpha$*, i.e.,* $PCS_E \geq 1-\alpha$ *or* $PCS_{min} \geq 1-\alpha$.

- The proof of Theorem 2 shares the same logics as in the proof of Theorem 1 (based on a more general version of Lemma 1).

Least-favorable Configuration (LFC)

- For traditional R&S problem, it is well known that the LFC is the slippage configuration (SC):

$$\mu_1 - \delta = \mu_i, \text{ for } i = 2, 3, \ldots, k.$$

- For R&S-C, the idea of SC can be easily extended to the generalized slippage configuration (GSC):

$$\beta_{10} - \delta = \beta_{i0}, \ \beta_{1j} = \beta_{ij}, \text{ for } j = 1, \ldots, d \text{ and } i = 2, \ldots, k.$$

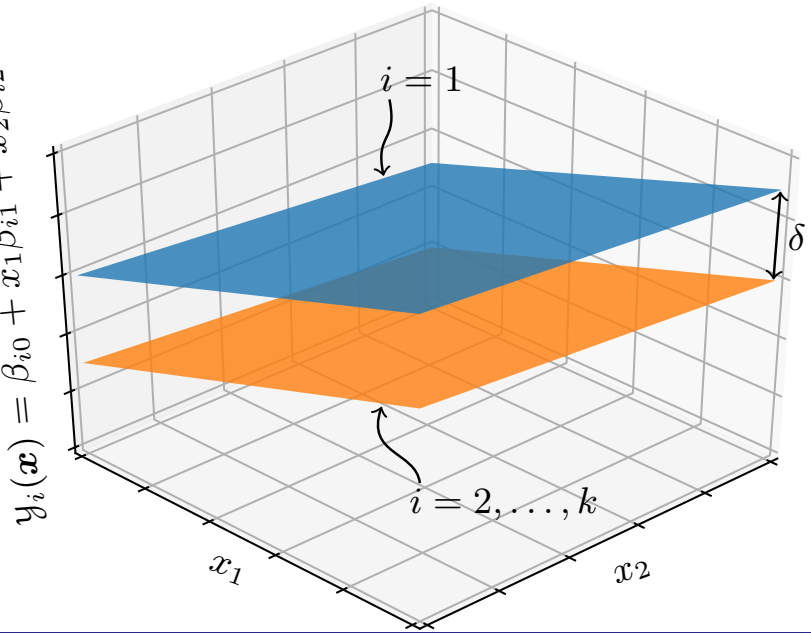
Least-favorable Configuration (LFC)

- For traditional R&S problem, it is well known that the LFC is the slippage configuration (SC):

$$\mu_1 - \delta = \mu_i, \text{ for } i = 2, 3, \ldots, k.$$

- For R&S-C, the idea of SC can be easily extended to the generalized slippage configuration (GSC):

$$\beta_{10} - \delta = \beta_{i0}, \ \beta_{1j} = \beta_{ij}, \text{ for } j = 1, \ldots, d \text{ and } i = 2, \ldots, k.$$



GSC, $d = 2$.

Settings

- Generic setting:
  - Covariates $X_1, \ldots, X_d$, are i.i.d. Unif$[0, 1]$ RVs.
  - Take $m = 2^d$ design points: $\{0, 0.5\} \times \cdots \times \{0, 0.5\}$.
- Benchmark problem (0):
  - $d = 3$ and $k = 5$.
  - Mean configuration: GSC, $\beta_{10} - \delta = \beta_{i0} = 0,\ \beta_{1j} = \beta_{ij} = 1$.
  - Homoscedastic errors: $\sigma_i^2(\boldsymbol{x}) \equiv \sigma_i^2$.
  - Equal variances among alternatives: $\sigma_1 = \cdots = \sigma_k = 10$.

Settings

- Generic setting:
  - Covariates $X_1, \ldots, X_d$, are i.i.d. Unif$[0,1]$ RVs.
  - Take $m = 2^d$ design points: $\{0, 0.5\} \times \cdots \times \{0, 0.5\}$.
- Benchmark problem (0):
  - $d = 3$ and $k = 5$.
  - Mean configuration: GSC, $\beta_{10} - \delta = \beta_{i0} = 0,\ \beta_{1j} = \beta_{ij} = 1$.
  - Homoscedastic errors: $\sigma_i^2(\boldsymbol{x}) \equiv \sigma_i^2$.
  - Equal variances among alternatives: $\sigma_1 = \cdots = \sigma_k = 10$.
- 8 comparing problems:
  (1) Set $k = 2$. (2) Set $k = 8$.
  (3) Mean configuration: Non-GSC, randomly generate all components of $\boldsymbol{\beta}_i$ from Unif$[0,5]$, for $i = 1, \ldots, 5$.
  (4) Increasing variances among alternatives: $\sigma_1 = 5$, $\sigma_2 = 7.5$, $\sigma_3 = 10$, $\sigma_4 = 12.5$, $\sigma_5 = 15$.
  (5) Decreasing variances among alternatives.
  (6) Heteroscedastic errors: $\sigma_i(\boldsymbol{x}) = 10\boldsymbol{x}^\top \boldsymbol{\beta}_i$, for $i = 1, \ldots, 5$.
  (7) Set $d = 1$. (8) Set $d = 5$.

Numerical Results

- PCS_{E} is designed to be $\geq 95\%$ (i.e., $\alpha = 0.05$), $\delta = 1, n_0 = 50$.
- 10^4 macro replications are carried out for each procedure.
- 10^5 samples of $\boldsymbol{X}$ are used to calculate APCS_{E} (and APCS_{min}) of $\widehat{i^*}(\boldsymbol{X})$ produced by each procedure.

	Procedure FDHom				Procedure FDHet			
Problem	h_{E}	Sample	APCS_{E}	APCS_{min}	h_{E}	Sample	APCS_{E}	APCS_{min}
(0) Benchmark	3.423	46865	0.9610	0.7439	4.034	65138	0.9801	0.8080
(1) $k = 2$	2.363	8947	0.9501	0.8084	2.781	12380	0.9702	0.8517
(2) $k = 8$	3.822	93542	0.9650	0.7246	4.510	130200	0.9842	0.8052
(3) Non-GSC	3.423	46865	0.9987	0.9410	4.034	65138	0.9994	0.9615
(4) IV	3.423	52698	0.9618	0.7549	4.034	73265	0.9807	0.8147
(5) DV	3.423	52720	0.9614	0.7501	4.034	73246	0.9806	0.8114
(6) Het	3.423	58626	0.9232	0.6336	4.034	81555	**0.9846**	0.8591
(7) $d = 1$	4.612	21288	0.9593	0.7941	4.924	24266	0.9662	0.8223
(8) $d = 5$	2.141	73428	0.9656	0.7446	2.710	117630	0.9895	0.8379

Numerical Results

- PCS_{min} is designed to be $\geq 95\%$ (i.e., $\alpha = 0.05$), $\delta = 1, n_0 = 50$.
- 10^4 macro replications are carried out for each procedure.
- 10^5 samples of $\boldsymbol{X}$ are used to calculate APCS_{min} (and APCS_{E}) of $\widehat{i^*}(\boldsymbol{X})$ produced by each procedure.

	Procedure FDHom				Procedure FDHet			
Problem	h_{min}	Sample	APCS_{E}	APCS_{min}	h_{min}	Sample	APCS_{E}	APCS_{min}
(0) Benchmark	5.927	140540	0.9989	0.9594	6.990	195340	0.9997	0.9825
(1) $k = 2$	4.362	30447	0.9958	0.9466	5.132	42164	0.9987	0.9701
(2) $k = 8$	6.481	268750	0.9993	0.9642	7.651	374720	0.9999	0.9849
(3) Non-GSC	5.927	140540	1.0000	0.9958	6.990	195340	1.0000	0.9981
(4) IV	5.927	158140	0.9989	0.9574	6.990	219870	0.9998	0.9862
(5) DV	5.927	158100	0.9990	0.9617	6.990	219740	0.9998	0.9826
(6) Het	5.927	175700	0.9952	0.8999	6.990	244490	0.9999	**0.9899**
(7) $d = 1$	7.155	51161	0.9954	0.9600	7.648	58493	0.9971	0.9708
(8) $d = 5$	3.792	230220	0.9994	0.9539	4.804	369310	1.0000	0.9907

Background

- Esophageal cancer is the fourth (seventh) leading cause of cancer death among males in China (U.S.).

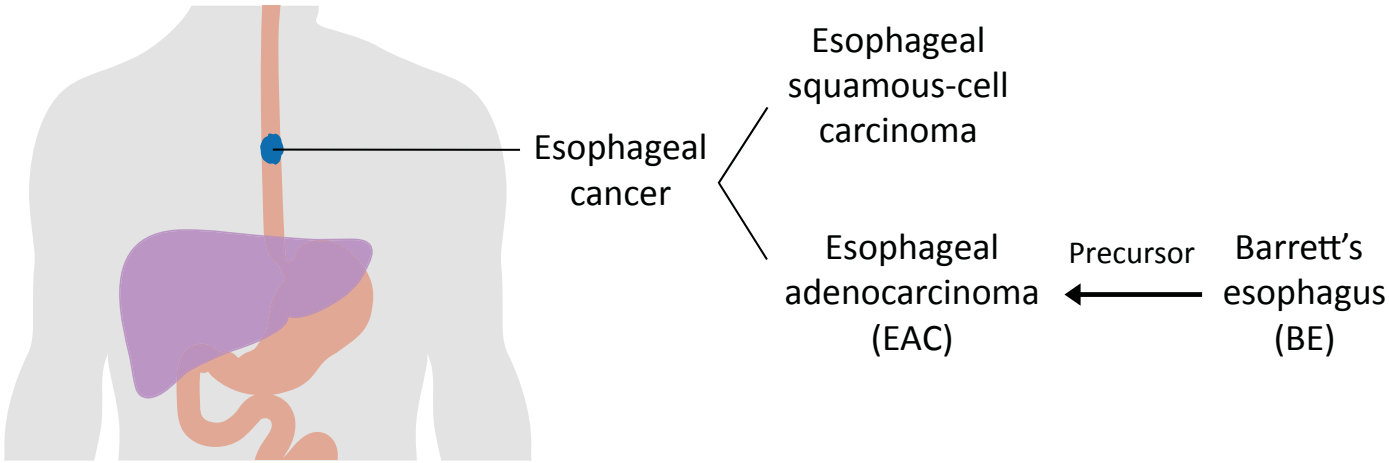


Image Source: Cancer Research UK / Wikimedia Commons.

Background

- Esophageal cancer is the fourth (seventh) leading cause of cancer death among males in China (U.S.).

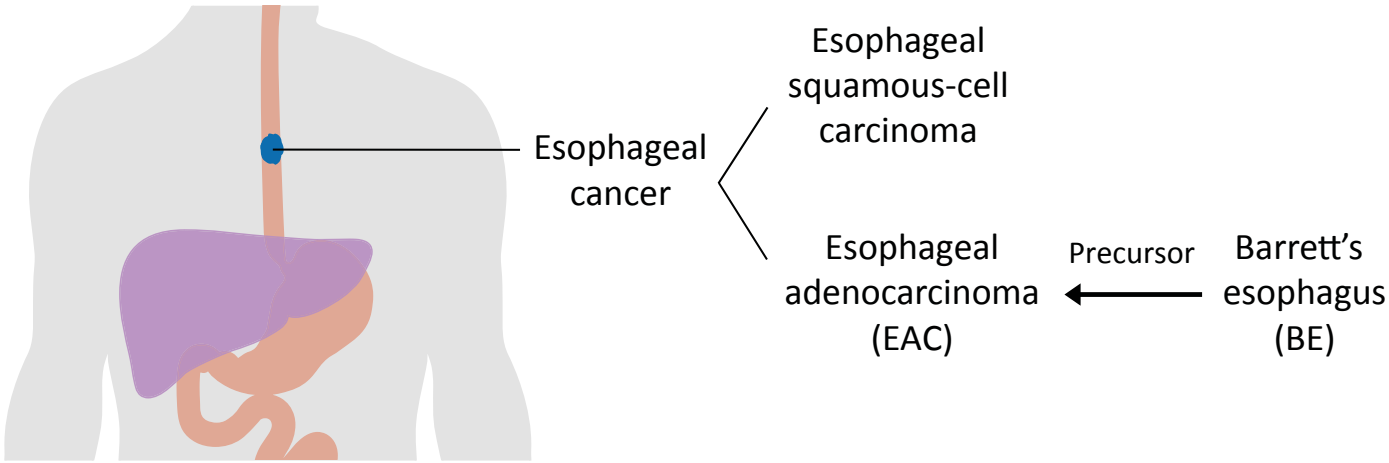


Image Source: Cancer Research UK / Wikimedia Commons.

- EAC is one sub-type of esophageal cancer, and its incidence has increased by 500% over the past 40 years (Bollschweiler et al. 2001, Hur et al. 2013).
- BE is a precursor to EAC, and its management is important and attracts many attentions.

Best Treatment Regimens

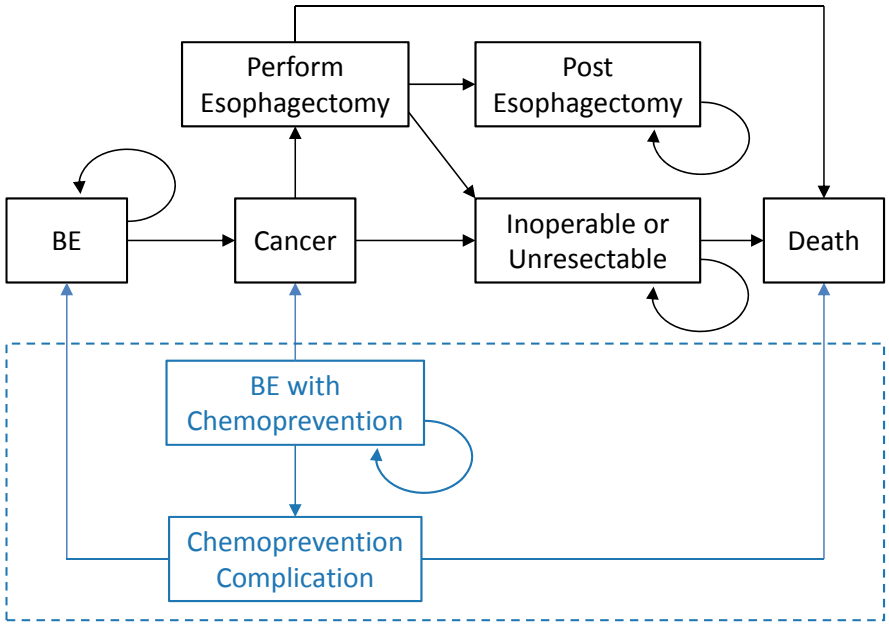
- Consider 3 treatment regimens (i.e., alternatives) for BE; all regimens include standard endoscopic surveillance:
  (1) No drug;
  (2) Aspirin chemoprevention;
  (3) Statin chemoprevention.

- Consider some individual characteristics (i.e., covariates):
  X_1 – Age;
  X_2 – Risk (i.e., the annual progression rate of BE to EAC);
  X_3 – Effect of aspirin (i.e., progression reduction effect);
  X_4 – Effect of statin.

- The best decision of treatment regimen for BE is patient-specific (depending on individual covariates).

Simulation Model

- A Markov simulation model was developed by Hur et al. (2004) and Choi et al. (2014) to study the effectiveness of aspirin and statin chemoprevention against EAC.



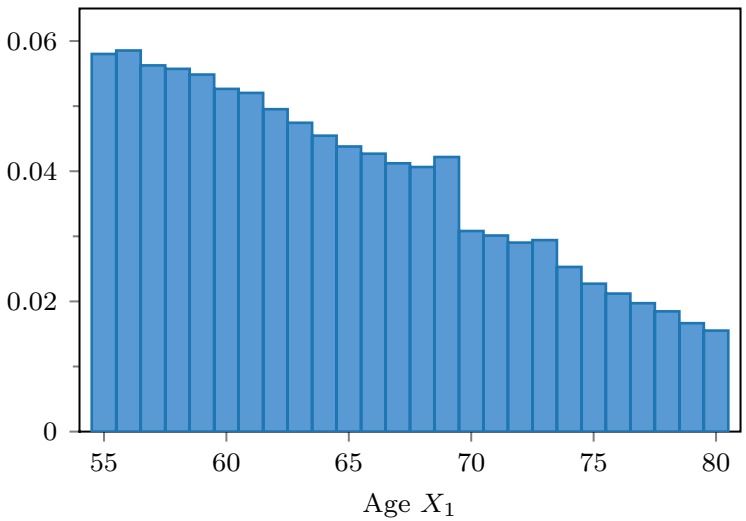
- A male with BE goes through various health state until death.
- The person in each state can die from age-related all-cause mortality.
- The time length between state transition is one month.
- Detailed structure inside dotted box depends on drug.
- Parameters are well calibrated.

- Output $Y_{i\ell}(\boldsymbol{X})$: Quality-adjusted life years (QALYs) after the starting age under treatment regimen i conditioning on $\boldsymbol{X}$.

Distribution of Covariates

- In this study we just assume that $X_1, \ldots, X_4$ are independent with distributions as listed in following table.

Covariates	Distributions	Support	Mean
X_1	Discrete (Figure below)	$\{55, \ldots, 80\}$	64.78
X_2	Unif $(0, 0.1)$	$[0, 0.1]$	0.05
X_3	Triangular $(0, 0.59, 1)$	$[0, 1]$	0.53
X_4	Triangular $(0, 0.62, 1)$	$[0, 1]$	0.54



Probability mass function of X_1 (truncated). *Data Source:* U.S. 2016 population data, U.S. Census Bureau.

Value of Personalized Medicine

- Decision of treatment regimen:
  - (i) In traditional way, the best treatment is the one that works best for the average of population ($i^\dagger \equiv 3$).
  - (ii) In personalized way, the best treatment is $i^*(\boldsymbol{x})$.

    To get its estimate, $\widehat{i^*}(\boldsymbol{x})$, apply Procedure FDHet with $\text{PCS}_\text{E} \geq 95\%, \delta = 0.2, n_0 = 100$.

Value of Personalized Medicine

- Decision of treatment regimen:
  - (i) In traditional way, the best treatment is the one that works best for the average of population ($i^\dagger \equiv 3$).
  - (ii) In personalized way, the best treatment is $i^*(\boldsymbol{x})$.

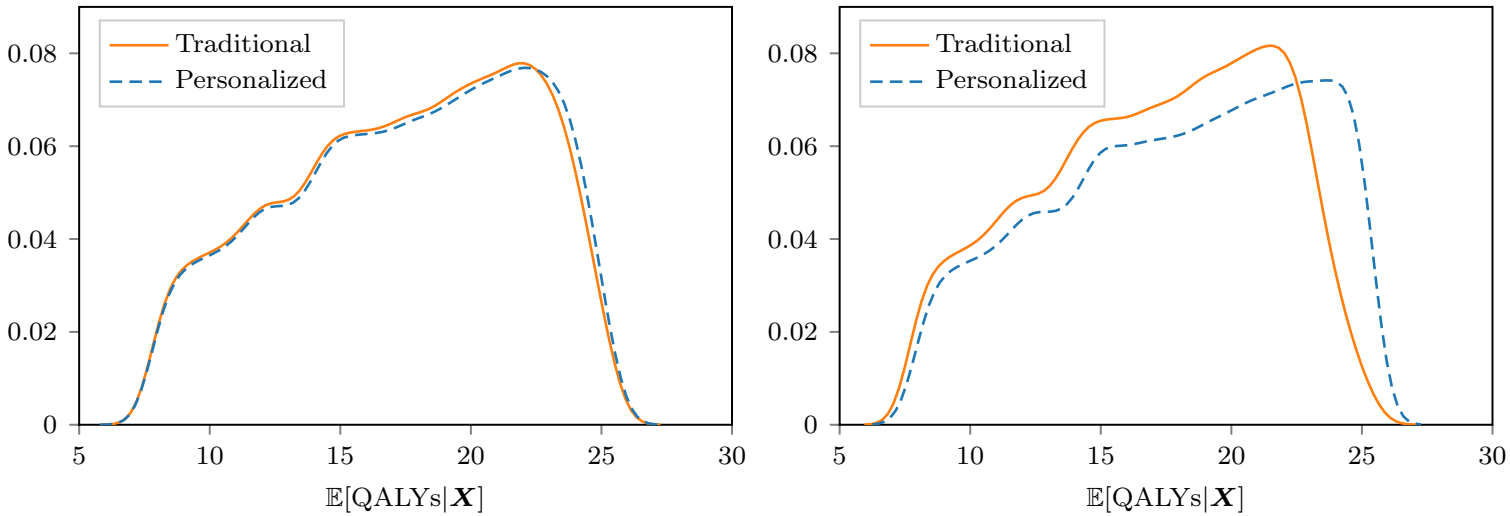
    To get its estimate, $\widehat{i^*}(\boldsymbol{x})$, apply Procedure FDHet with $\mathrm{PCS}_{\mathsf{E}} \geq 95\%, \delta = 0.2, n_0 = 100$.

- In order to evaluate the $\mathrm{APCS}_{\mathsf{E}}$, we run very long simulation to get the "true" surfaces of the expected QALYs $\mathbb{E}[Y_{i\ell}(\boldsymbol{x})]$ for $\boldsymbol{x} \in \Theta$ and $i = 1, 2, 3$.

- We find that
  - (i) Traditional way: $\mathrm{APCS}_{\mathsf{E}} \approx 78.0\%$.
  - (ii) Personalized way: $\mathrm{APCS}_{\mathsf{E}} \approx 99.7\%$.

Value of Personalized Medicine

- Distributions of expected QALYs under the two ways.
  - (a) For the *entire population* considered.
  - (b) For a more *specific group* of patients, i.e., patients with $\boldsymbol{X} = (X_1, X_2, 0.9, 0.2)^\top$.



(a) (b)

  - (c) For a *specific individual* with $\boldsymbol{X} = (55, 0.1, 0.9, 0.2)^\top$, expected QALYs increases by 2.43 years when personalized medicine is performed.

Conclusions

- Personalized decisions lead us to consider Ranking and Selection with Covariates.
- We use a linear model to capture the relationship between the response and the covariates. It is the simplest yet most useful parametric model in practice.
- There are many directions that R&S-C may be studied, e.g., non-parametric models, Bayesian formulations, random designs and sequential procedures.

Thank You!

Haihui SHEN
Dec. 5, 2017