Applying the importance sampling to regression

Jieun Shin

December 15, 2021

Table of Contents

1. Introduction

1. Introduction

- Objective: Comparison of variance reduction methods for estimating extreme event probability, $\ell = \mathbb{E}[Y > \gamma]$, in simple linear regression.
- Let Y be the response variable of simple linear regression:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \quad \text{where } \epsilon \sim N(0, \sigma^2),$$

  for $i = 1, \ldots, n$.
- Importance sampling using least square estimator.
- Comparison of results when using two suggested distributions in regression.
- Check if variance reduction occur.

2. Methodology

Estimation of rare-event probabilities

- Let the parameter set for target distribution as $\boldsymbol{\theta}^* = (\sigma^{*2}, \beta_0^*, \beta_1^*)$ and proposed distribution as $\boldsymbol{\theta} = (\sigma^2, \beta_0, \beta_1)$.
- We want to estimate of

$$\ell = \mathbb{P}_{\boldsymbol{\theta}^*}(Y \geq \gamma) = \mathbb{E}_{\boldsymbol{\theta}^*}[I\{Y \geq \gamma\}] \tag{1}$$

  for some fixed level γ.
- Y is a random vector with pdf $f(\cdot, \boldsymbol{\theta}^*)$ belonging to parametric family $\{f(\cdot, \boldsymbol{\theta}), \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$, and $\{Y \geq \gamma\}$ is assumed to be a rare event.

Importance sampling estimator

- The importance sampling estimator $\hat{\ell}_{IS}$ is defined as

$$\hat{\ell}_{IS} = \frac{1}{N}\sum_{k=1}^{N} I(Y_k > \gamma) W(Y_k; \boldsymbol{\theta}^*, \boldsymbol{\theta}), \tag{2}$$

  where $W(Y_k; \boldsymbol{\theta}^*, \boldsymbol{\theta}) = f(Y_k; \boldsymbol{\theta}^*)/f(Y_k; \boldsymbol{\theta})$ is the likelihood ratio.
- We consider two methods to choose the *optimal reference parameter* $\boldsymbol{\theta}$.
  1. Cross-entropy method
  2. Least square estimator

3. Choice the optimal reference parameter

3. Choice the optimal reference parameter

Cross-entropy (CE) method

- The optimal CE reference parameter $\hat{\boldsymbol{\theta}}_{CE}$ is defined as

$$\hat{\boldsymbol{\theta}}_{CE} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, \mathbb{E}_{\tilde{\boldsymbol{\theta}}} \left[I(Y_k > \gamma) W(Y_k; \boldsymbol{\theta}^*, \tilde{\boldsymbol{\theta}}) \ln f(Y_k; \boldsymbol{\theta}) \right], \tag{3}$$

  where $\tilde{\boldsymbol{\theta}}$ is an *arbitary tilting parameter*.
- We can estimate $\hat{\boldsymbol{\theta}}_{CE}$ as the solution of the stochastic program

$$\hat{\boldsymbol{\theta}}_{CE} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, \frac{1}{N} \sum_{k=1}^{N} I(Y_k > \gamma) W(Y_k; \boldsymbol{\theta}^*, \tilde{\boldsymbol{\theta}}) \ln f(Y_k; \boldsymbol{\theta}), \tag{4}$$

  where $Y_i, \ldots, Y_k$ is a random sample from $f(Y_k; \tilde{\boldsymbol{\theta}})$.

Cross-entropy (CE) method

1. Find the optimal $\hat{\sigma}^2_{CE}$ as the solution of the stochastic program

$$\hat{\sigma}^2_{CE} = \underset{\sigma^2}{\text{argmax}} \frac{1}{N} \sum_{k=1}^{N} I(\varepsilon_k > \gamma) W(\varepsilon_k; \sigma^{*2}, \tilde{\sigma}^2) \ln f(\varepsilon_k; \sigma^2) \tag{5}$$

$$= \frac{\sum_{k=1}^{N} I(\varepsilon_k > \gamma) W(\varepsilon_k; \sigma^{*2}, \tilde{\sigma}^2) \varepsilon_k^2}{\sum_{k=1}^{N} I(\varepsilon_k > \gamma) W(\varepsilon_k; \sigma^{*2}, \tilde{\sigma}^2)}, \tag{6}$$

where the error $\varepsilon_k \sim N(0, \tilde{\sigma}^2)$.

Cross-entropy (CE) method

2. Find the optimal $\hat{\beta}_{0CE}$ as the solution of the stochastic program:

$$\hat{\beta}_{0CE} = \underset{\beta_0}{\operatorname{argmax}} \frac{1}{N} \sum_{k=1}^{N} I(Y_k > \gamma) W(Y_k; \beta_0^*, \tilde{\beta}_0) \ln f(Y_k; \beta_0) \tag{7}$$

$$= \frac{\sum_{k=1}^{N} I(Y_k > \gamma) W(Y_k; \beta_0^*, \tilde{\beta}_0)(Y_k - \tilde{\beta}_1 X_k)}{\sum_{k=1}^{N} I(Y_k > \gamma) W(Y_k; \beta_0^*, \tilde{\beta}_0)}, \tag{8}$$

where the response variable $Y_k \sim N(\tilde{\beta}_0 + \tilde{\beta}_1 X, \tilde{\sigma}^2)$.

Cross-entropy (CE) method

3. Find the optimal $\hat{\beta}_{1CE}$ as the solution of the stochastic program:

$$\hat{\beta}_{1CE} = \underset{\beta_1}{\operatorname{argmax}} \frac{1}{N} \sum_{k=1}^{N} I(Y_k > \gamma) W(Y_k; \beta_1^*, \tilde{\beta}_1) \ln f(Y_k; \beta_1) \quad (9)$$

$$= \frac{\sum_{k=1}^{N} I(Y_k > \gamma) W(Y_k; \beta_1^*, \tilde{\beta}_1) X_k (Y_k - \tilde{\beta}_0)}{\sum_{k=1}^{N} I(Y_k > \gamma) W(Y_k; \beta_1^*, \tilde{\beta}_1) X_k^2}, \quad (10)$$

where the response variable $Y_k \sim N(\tilde{\beta}_0 + \tilde{\beta}_1 X, \tilde{\sigma}^2)$.

Least square estimator

We choose the optimal reference parameter as least square estimator as follows:

$$\hat{\sigma}^2_{LS} = \underset{\sigma^2}{\operatorname{argmin}} \sum_{k=1}^{N} \varepsilon_k^2, \tag{11}$$

$$\hat{\beta}_{0LS} = \underset{\beta_0}{\operatorname{argmin}} \sum_{k=1}^{N} (Y_k - \tilde{\beta}_0 - \tilde{\beta}_1 X_k)^2, \tag{12}$$

$$\hat{\beta}_{1LS} = \underset{\beta_1}{\operatorname{argmin}} \sum_{k=1}^{N} (Y_k - \tilde{\beta}_0 - \tilde{\beta}_1 X_k)^2, \tag{13}$$

where the response variable $\varepsilon_k \sim N(0, \tilde{\sigma}^2),\ Y_k \sim N(\tilde{\beta}_0 + \tilde{\beta}_1 X, \tilde{\sigma}^2)$.

4. Simulation

Simulation setting

- Set the parameter of target distribution as
  $\sigma^{*2} \in \{\sqrt{0.4}, \sqrt{12}\}, \beta_0^* = 5, \beta_1^* = 3.$
  and $X_k \sim U(0,1)$ and $\varepsilon_k \sim N(0, \sigma^{*2})$ then
  $Y_k = \beta_0^* + \beta_1^* X_k + \varepsilon_k, k = 1, \dots, N.$
- Set the large value γ is 99.95% quantile value of Y.
- Sample size: $N = 10^6$.

4. Simulation

Table 1: $\sigma^{*2} = \sqrt{0.4}, \tilde{\sigma}^2 = 1, \tilde{\beta}_0 = 7, \tilde{\beta}_1 = 7$

Method	$\hat{\sigma}^2$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\ell}$	Var($\hat{\ell}$)	relative error
CMC	-	-	-	0.000500	0.0000223551	0.0447102
IS-CE	0.40025	3.048124	6.999594	0.000464	0.0000032378	0.0069722
IS-LS	0.99992	6.999022	7.000913	0.000349	0.0000972918	0.2782808

Table 2: Table2. $\sigma = \sqrt{12}, \tilde{\sigma}^2 = 1, \tilde{\beta}_0 = 10, \tilde{\beta}_1 = 10$

Method	$\hat{\sigma}^2$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\ell}$	Var($\hat{\ell}$)	relative error
CMC	-	-	-	0.000500	0.0000223551	0.0447102
IS-CE	7.86831	10.14025	9.936445	0.000520	0.0000020892	0.0040148
IS-LS	0.99992	9.999022	10.00091	0.000440	0.0000105115	0.0238859