Importance Sampling with Applications to Bayesian Logistic Regression

Jung Ádám

Supervisor : Balázs Csanád Csáji

2024

ELTE | FACULTY OF SCIENCE

Overview

- Introduction to Importance Sampling
- Brief introduction to Bayesian Logistic Regression (BLR)
- Present a case study on the usage of importance sampling for the parameters of a BLR model

Overview

- Introduction to Importance Sampling
- Brief introduction to Bayesian Logistic Regression (BLR)
- Present a case study on the usage of importance sampling for the parameters of a BLR model

Overview

- Introduction to Importance Sampling
- Brief introduction to Bayesian Logistic Regression (BLR)
- Present a case study on the usage of importance sampling for the parameters of a BLR model

Importance Sampling

Estimate the expectation $\mu := \mathbb{E}[f(X)]$, using sampling.

Importance Sampling

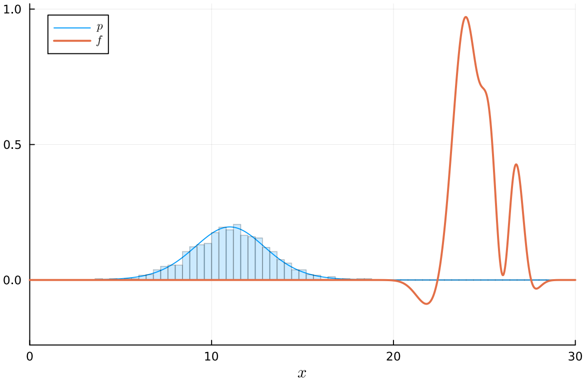
Estimate the expectation $\mu := \mathbb{E}[f(X)]$, using sampling.

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} f(x_i) \qquad x_i \sim p$$

Importance Sampling

Estimate the expectation $\mu := \mathbb{E}[f(X)]$, using sampling.

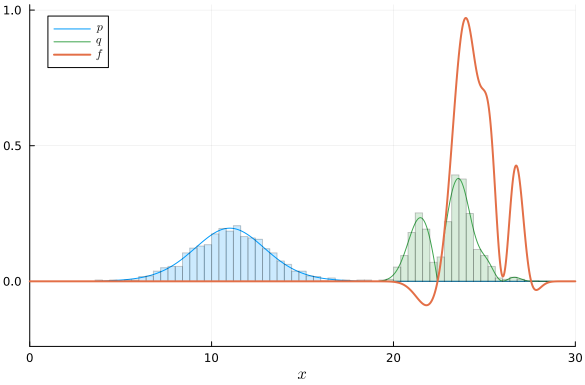
$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} f(x_i) \qquad x_i \sim p$$



Importance Sampling

Estimate the expectation $\mu := \mathbb{E}[f(X)]$, using sampling.

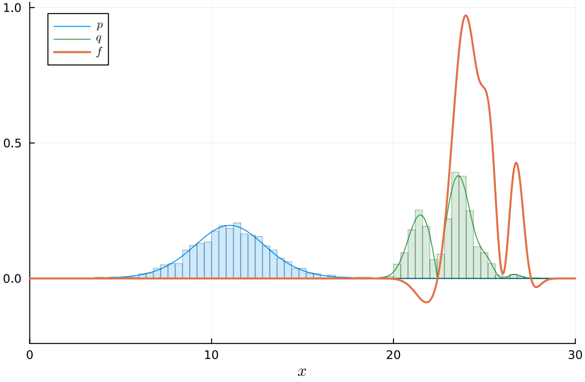
$$\hat{\mu}_q = \frac{1}{n} \sum_{i=1}^{n} f(x_i) \qquad x_i \sim q$$



Importance Sampling

Estimate the expectation $\mu := \mathbb{E}[f(X)]$, using sampling.

$$\hat{\mu}_q = \frac{1}{n}\sum_{i=1}^{n} f(x_i)\frac{p(x_i)}{q(x_i)} \qquad x_i \sim q$$



How to choose q?

- Minimise the variance of the estimate :

$$\mathbb{E}[(\hat{\mu}_q - \mu)^2] \to \min$$

- Optimum :

$$q^*(x) \propto |f(x)|p(x)$$

  - $f \geq 0 \implies \mathbb{D}^2(\hat{\mu}_{q^*}) = 0$

- In practice : Domain knowledge and educated guessing

How to choose q ?

- Minimise the variance of the estimate :

$$\mathbb{E}[(\hat{\mu}_q - \mu)^2] \to \min$$

- Optimum :

$$q^*(x) \propto |f(x)|p(x)$$

  - $f \geq 0 \implies \mathbb{D}^2(\hat{\mu}_{q^*}) = 0$
- In practice : Domain knowledge and educated guessing

How to choose q?

- Minimise the variance of the estimate :

$$\mathbb{E}[(\hat{\mu}_q - \mu)^2] \to \min$$

- Optimum :

$$q^*(x) \propto |f(x)|p(x)$$

  - $f \geq 0 \implies \mathbb{D}^2(\hat{\mu}_{q^*}) = 0$

- In practice : Domain knowledge and educated guessing

How to choose q ?

- Minimise the variance of the estimate :

$$\mathbb{E}[(\hat{\mu}_q - \mu)^2] \to \min$$

- Optimum :

$$q^*(x) \propto |f(x)|p(x)$$

  - $f \geq 0 \implies \mathbb{D}^2(\hat{\mu}_{q^*}) = 0$
- In practice : Domain knowledge and educated guessing

Bayesian Logistic Regression

$(x_1, y_1), \dots, (x_n, y_n)$ observations, where $y_i \in \{\pm 1\}$, $x_i \in \mathbb{R}^d$

- Model class :

$$\mathbb{P}(Y_i = 1 \mid \theta) = \sigma(\theta^T \Phi(x_i))$$

- Prior distribution : $\pi(\theta)$
- Posterior distribution $\propto$

$$\pi(\theta \mid Y) = \pi(\theta) \prod_{i=1}^{n} \sigma(Y_i \theta^T \Phi(x_i))$$

Bayesian Logistic Regression

$(x_1, y_1), \dots, (x_n, y_n)$ observations, where $y_i \in \{\pm 1\}$, $x_i \in \mathbb{R}^d$

- Model class :

$$\mathbb{P}(Y_i = 1 \mid \theta) = \sigma(\theta^T \Phi(x_i))$$

- Prior distribution : $\pi(\theta)$
- Posterior distribution $\propto$

$$\pi(\theta \mid Y) = \pi(\theta) \prod_{i=1}^{n} \sigma(Y_i \theta^T \Phi(x_i))$$

Bayesian Logistic Regression

$(x_1, y_1), \dots, (x_n, y_n)$ observations, where $y_i \in \{\pm 1\}$, $x_i \in \mathbb{R}^d$

- Model class :

$$\mathbb{P}(Y_i = 1 \mid \theta) = \sigma(\theta^T \Phi(x_i))$$

- Prior distribution : $\pi(\theta)$
- Posterior distribution $\propto$

$$\pi(\theta \mid Y) = \pi(\theta) \prod_{i=1}^{n} \sigma(Y_i \theta^T \Phi(x_i))$$

Bayesian Logistic Regression

$(x_1, y_1), \dots, (x_n, y_n)$ observations, where $y_i \in \{\pm 1\}$, $x_i \in \mathbb{R}^d$

- Model class :

$$\mathbb{P}(Y_i = 1 \mid \theta) = \sigma(\theta^T \Phi(x_i))$$

- Prior distribution : $\pi(\theta)$
- Posterior distribution $\propto$

$$\pi(\theta \mid Y) = \pi(\theta) \prod_{i=1}^{n} \sigma(Y_i \theta^T \Phi(x_i))$$

Importance sampling for BLR

- Approximate $\pi(\theta \mid Y)$ from

$$\log(\pi(\theta \mid Y)) \approx \log(\pi(\theta^* \mid Y)) + \frac{1}{2}(\theta - \theta^*)^T H(\theta - \theta^*)$$

with $\theta \mid Y \sim \mathcal{N}(\theta^*, -H^{-1})$.

- Use $t(\nu, \theta^*, -H^{-1})$ as q for importance sampling

Importance sampling for BLR

- Approximate $\pi(\theta \mid Y)$ from

$$\log(\pi(\theta \mid Y)) \approx \log(\pi(\theta^* \mid Y)) + \frac{1}{2}(\theta - \theta^*)^T H(\theta - \theta^*)$$

with $\theta \mid Y \sim \mathcal{N}(\theta^*, -H^{-1})$.

- Use $t(\nu, \theta^*, -H^{-1})$ as q for importance sampling

Importance sampling for BLR

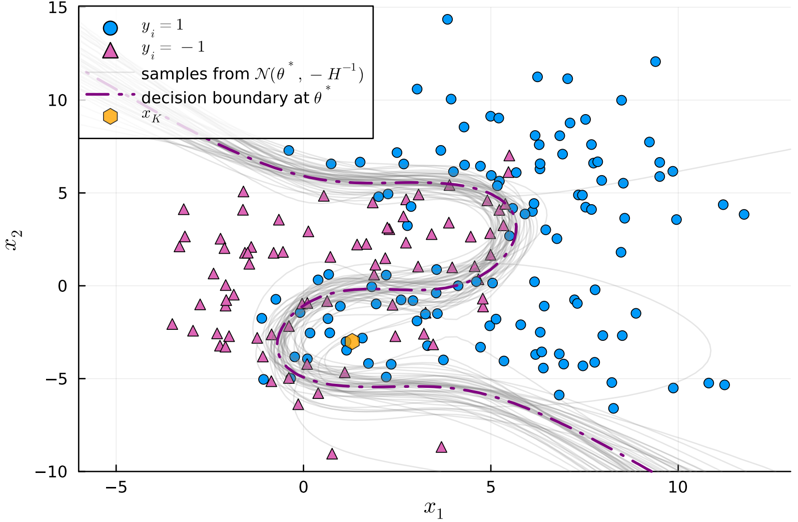
- Approximate $\pi(\theta \mid Y)$ from

$$\log(\pi(\theta \mid Y)) \approx \log(\pi(\theta^* \mid Y)) + \frac{1}{2}(\theta - \theta^*)^T H(\theta - \theta^*)$$

  with $\theta \mid Y \sim \mathcal{N}(\theta^*, -H^{-1})$.
- Use $t(\nu, \theta^*, -H^{-1})$ as q for importance sampling

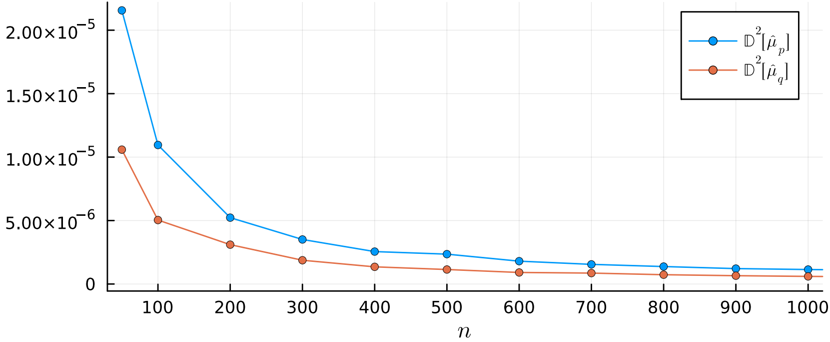
Case study

Let $x_K \in \mathbb{R}^2$ be a fixed point, and let $f(\theta) = \mathbb{I}(\sigma(\theta^T \Phi(x_K)) < 1/2)$.



Variance of the estimates

- Sampling from q did improved on the variances of the estimates
- The improvement is slightly more significant at low sample sizes



Empirical variances computed from 3000 repeated estimations at each sample size n.

Thank you for listening