Online regression analysis for streaming data

By E. CARRARO
Department of Statistical Sciences,
University of Padua,
Via Cesare Battisti 241, 35121 Padova, Italy
enrico.carraro.10@phd.unipd.it

SUMMARY

In the last decades the amount of available data has increased exponentially, and it is more and more common to have data that arrive in batches with similar characteristics. In order to reduce the memory and computational burden of storing all the data and analyze them with canonical offline techniques, many authors proposed new methods to update estimates and perform inference in a streaming scenario. This paper aims to review some of the methods proposed in the literature in order to do inference on the sequentially updating parameters under different assumptions on the covariance structure.

Some key words: Inference; Online Learning; Renewable Estimations; Streaming Data.

1. Introduction

With the advent of the modern computers, clusters and databases the amount of data collected in every field has become huge. It is enough to imagine how many statistics we have in sports and how much we know about every single play, or to think of how electronic devices collect data about our routine and health every single second, in order to catch the relationship between the health of a person and a set of covariates, that can include sex, phenotype and physical characteristics, in addiction to sleep rhythm and physical activity. All the information saved about people are often organized as longitudinal data, where the same information regarding the same people are saved multiple times over time.

With this big amount of data, the classical offline approach can be problematic for different reasons. First of all it requires to have immediate access to all the historical data, that can be too much to be stored over a long time. Secondly it is computationally expensive, as it needs to recompute the model on the whole dataset every time we want to have up to date information about the parameters. In the best cases this results in extended delays, while in the worst cases it is completely unfeasible to obtain any result. The first techniques that have been developed to speed up the computation rely on the distributed-computing paradigm.

When data become available in batches it is more convinient to approach the problem with an online paradigm, which is a natural framework for this particular setting. The typical base assumpion is that the number of observation diverges, which is almost always satisfied. Some model require that also the number of partecipants diverges. In this framework a lot of methods have been developed, but the majority of them take into account only the updating of the estimate of the parameters, without updating the estimate of the variance, giving the idea of the possible true value of the parameter without the possibility to make inference on it to evaluate how significant it is. Most of this methods rely on stochastic gradient descent algorithm and some modifications, or on resampling methods with different properties.

In this paper, I present three methods that update both the estimate and the variability of the estimator under different assumption on the covariance structure. In Section 2 is defined the problem with the relative notation, in Section 3 is presented how to obtain the updating estimator for independent data proposed by Luo & Song (2020), in Section 4 is presented how to deal with dependency between data batches using the estimator proposed by Luo & Song (2023) that relies on state space models and in Section 5 it is presented the estimator proposed by Luo et al. (2023) that can deal with varying dependence and make it easier to evaluate the variability of the estimator.

2. PROBLEM SET-UP

Let $\mathcal{D}_{ij} = \{y_{ij}, X_{ij}\}$ be the data batches sequentially collected at deterministic updating time-points t_j $(j = 1, \ldots, b)$ on the same set of partecipants $i = 1, \ldots, m$, where partecipants are assumed independent one from the other. Now let $y_{ij} \in \mathbb{R}^{n_j}$ be the vector of n_j longitudinal measurements on the same outcome variable in batch j and $X_{ij} = (x_{i,kj}^\top)_{k=1}^{n_j} \in \mathbb{R}^{n_j \times p}$ is the corresponding matrix of p predictors, where $x_{i,kj} = (x_{i,1kj}^\top, \ldots, x_{i,pkj}^\top) \in \mathbb{R}^p$ for $k = 1, \ldots, n_j$ and $j = 1, \ldots, b$. We also have to consider cumulative datasets up to batch b for partecipant i, which are defined as $\mathcal{D}_{ib}^\star = \{\mathcal{D}_{i1}, \ldots, \mathcal{D}_{ib}\}$, and its aggregate response dimension $N_b = \sum_{j=1}^b n_j$. For simplicity it is assumed equal number of repeted measures for N_b for each participant, equal batch size n_j and equal spacing between time-points. In this setting it's easy to see that the marginal generalized linear model with outcome $y_i = (y_{i1}, \ldots, y_{ib})^\top \in \mathbb{R}^{N_b}$ and covariates $X_i = (X_{ij})_{j=1}^b$ is

$$E\left(y_i \mid X_i\right) = \mu_i = \left(\mu_{i,kj}\right)_{k,j=1}^{n_j,b} = \left[h\left\{x_{i,kj}^\top \beta\left(t_j\right)\right\}\right]_{k,j=1}^{n_j,b} \in \mathbb{R}^{N_b}$$

for $i = 1, \ldots, m$, where $\beta(\cdot)$ is the regression coefficient function, assumed to be a smooth, that captures local dynamics, and $h(\cdot)$ is a known link function. Denote with $\beta_j \in \mathbb{R}^p$ the true value of the batch varying coefficient for the batch collected at time t_j. Where not necessary the dependence from the data can be omitted and for simplicity where it is clear it can be used D_b for $\{D_{ib}\}_{i=1}^m$.

3. INDEPENDENT OBSERVATIONS

The method proposed by Luo et al. (2023) aims to find an algorithm to update estimation of a generalised linear model in the case where all observations are assumed independent. Even if this assumption can be quite restrictive, this method can be addressed as a starting point in updating frequentist estimation without the need of historical data, with a very low memory boundary and computational efficiency. The idea behind this method is to find an estimator which is asymptotically equivalent to the maximum likelihood estimator. For this reason the starting point is to obtain the estimator on the first batch of the data solving the score equation $U_1(\{\mathcal{D}_{i1}\}_{i=1}^m; \hat\beta_1) = 0$. When a new generic batch b arrives the maximum likelihood estimator would solve the equation $U_b^\star(\hat\beta_b^\star) = \sum_{j=1}^b U_j(\hat\beta_b^\star) = 0$, that would need to recompute the score equation of the historical batches in the new estimate $\hat\beta_b^\star$. To avoid the need of historical data it is possible to take a first-order Taylor series expansion of each term $U_j(\hat\beta_b^\star)$ for every $j < b$ around the correspondent estimate $\hat\beta_j^\star$. For simplicity we show the case with only two batches $\{\mathcal{D}_{i1}\}_{i=1}^m$ and $\{\mathcal{D}_{i2}\}_{i=1}^m$, where we have

$$U_1\left(\{\mathcal{D}_{i1}\}_{i=1}^m, \hat\beta_1^\star\right) + J_1\left(\{\mathcal{D}_{i1}\}_{i=1}^m, \hat\beta_1^\star\right)\left(\hat\beta_1^\star - \hat\beta_2^\star\right) + U_2\left(\{\mathcal{D}_{i2}\}_{i=2}^m, \hat\beta_2^\star\right) + O_p\left(\left\|\hat\beta_1^\star - \hat\beta_2^\star\right\|^2\right) = 0$$

From which we obtain that the proposed estimator $\tilde{\beta}_2$ satisfies $J_1(\hat{\beta}_1)(\hat{\beta}_1 - \tilde{\beta}_2) + U_2(\tilde{\beta}_2) = 0$, where $\hat{\beta}_1 = \hat{\beta}_1^\star = \tilde{\beta}_1$, and $J_j(\tilde{\beta}_j)$ is the negative Hessian matrix.

Generalising the above case to b batches we obtain that the estimator $\tilde{\beta}_b$ of β is defined as the solution to the incremental equation

$$\sum_{j=1}^{b-1} J_j\left(\tilde{\beta}_j\right)\left(\tilde{\beta}_{b-1} - \tilde{\beta}_b\right) + U_b\left(\tilde{\beta}_b\right) = 0. \tag{1}$$

Let now $\tilde{J}_b = \sum_{j=1}^{b} J_j(\tilde{\beta}_j)$ denote the aggregated negative Hessian matrix, then the solution of equation (1) can be obtaied via Newton-Rhapson algorithm solving iteratevely

$$\tilde{\beta}_b^{(r+1)} = \tilde{\beta}_b^{(r)} + \left\{\tilde{J}_{b-1} + J_b\left(\mathcal{D}_b; \tilde{\beta}_{b-1}\right)\right\}^{-1} \tilde{U}_b^{(r)}.$$

Moreover a consistent estimator of the dispersion parameter ϕ, that does not enter in the estimation of $\tilde{\beta}_b$, is

$$\tilde{\phi}_b = \frac{N_{b-1} - p}{N_b - p}\tilde{\phi}_{b-1} + \frac{n_b - p}{N_b - p}\hat{\phi}_b$$

where

$$\hat{\phi}_b = \frac{1}{n_b - p}\sum_{i=1}^{m} \frac{\left(y_{ib} - \hat{\mu}_{ib}\right)^2}{v\left(\hat{\mu}_{ib}\right)}$$

and $v(\cdot)$ is the unit variance function. In this way it has been possible to update the estimates of all the parameters storing only the summary statistics $\{\tilde{\beta}_{b-1}, \tilde{J}_{b-1}, \tilde{\phi}_{b-1}\}$, without having access to any other historical data.

Under some mild conditions it is possible to show that the estimator $\tilde{\beta}_b$ is consistent and is approximately distributed as a normal with mean the true value of the parameter β and variance the inverse of the Fisher information, that can be aprroximated with $\tilde{\phi}_b \tilde{J}_b^{-1}$.

4. Dependent Data using State Space Model

There are many cases where the assumption of independence between observation is not respected in the data. for this reason Luo & Song (2023) proposed a new method in order to take into account the dependence intrinsic in the process. In order to do so, they based the new method under a state-space model framework, in which the observed data stream is driven by a latent state process that follows a Markov process. This allow to rely on a real-time Kalman-filter-based regression specifically developed. State-space models are a very flexible class of models for analyzing longitudinal data with large number of repeted observations. The latent process represents temporal or spatial evolving batch-specific effects, which reflect on the observed data as shown in Figure 1. This class of models can focus on estimation and inference both on the latent effect and the fixed effect, which is usually of primary interest. The proposed method is a natural extension of the one proposed by Luo & Song (2020) to take into a account inter-data batch heterogeneity by modeling a batch-specific latent effect that follows a stationary autoregressive process of order 1.

The proposed inference procedure resembles the offline Kalman estimatin equation (Song, 2007), which is itself a generalization of the EM algorithm. In particular the proposed updating inference method consists in tho steps: (i) use the new online Kalman filter in the E-step to update the conditional means of the latent states recursevely; (ii) update population-average fixed effects using only summary statistics of the hystorical data.

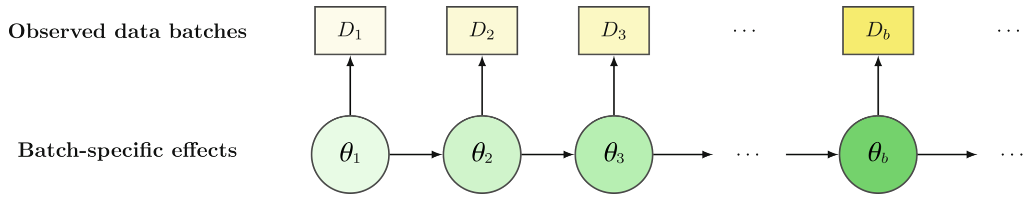

Fig. 1. A structure for a hierarchical dynamic system. Data batches $\{\mathcal{D}_b, b \geq 1\}$ are generated from a state-space model with common fixed effect β and batch-specific latent effects θ_b governed by a Markov process.

It is necessary to impose a first-order Markov process $\{\theta_b : b \geq 1\}$ to account for cross-batch heterogeneity with the following characteristics:

(A1) given θ_b, y_b is conditionally independent of the other y_js;
(A2) θ_1 is assumed to be a fixed unknown parameter;
(A3) $y_b = X_b\beta + Z_b\theta_b + \epsilon_b$ with ϵ_b are independent observations of $N_{n_b}(0, \phi I)$ and ϕ is the dispersion parameter, β is the vector of fixed effect for the covariates X_b and θ_b is the vector of random effects for the covariates Z_b;
(A4) $\theta_{b+1} = B_b\theta_b + \xi_b$, where B_b is a transition matrix and ξ_b are independent observations from $N_q(0, \delta I_q)$, with ξ_b independent of ϵ_b.

For a stationary auto-regressive process of order 1 we have that B_b is a diagonal matrix with elements ρ_s that is the autocorrelation coefficient between the sth component in θ_{b+1} and θ_b, where $|\rho_s| < 1$ for all $s = 1, \ldots, q$.

In order to obtain an online estimator and inference for the common fixed effect β, it is necessary to write the marginal augmented log-likelihood where the parameter θ_b is treated as a missing data, where $\vec{\theta}_b = (\theta_1^\top, \ldots, \theta_b^\top)^\top$ and $\zeta = (\phi, \rho, \delta)$

$$\ell\left(\beta, \zeta, \mathcal{D}_b^\star, \vec{\theta}_b\right) = \sum_{j=1}^{b} \log P\left(y_j \mid \theta_j, \beta, \zeta\right) + \sum_{j=1}^{b} \log P\left(\theta_{j+1} \mid \theta_j, \zeta\right).$$

To perform maximum likelihood estimation via Expectation Maximization algorithm we need to maximize the expected value of the augmented log-likelihood under the conditional distribution $P(\vec{\theta}_b \mid \mathcal{D}_b^\star, \alpha', \zeta')$, where α' and ζ' are updated parameter values from the previous iteration. Instead of using Monte Carlo techniques to compute the conditional mean $E(\theta_j \mid \mathcal{D}_b^\star, \alpha', \zeta')$, it can be used the best linear unbiased predictior computed via the recursive Kalman formula to speed up the computation. Furthermore the estimator for ζ is obtain via method of moments. In practice the steps of the online estimation are the following:

(1) Initialize the values for β and ζ with $\tilde{\beta}_0$ and $\tilde{\zeta}_0$.
(2) For $b \geq 1$, given the consistent estimates $\tilde{\phi}_{b-1}$, $\tilde{\rho}_{b-1}$ and $\tilde{\xi}_{b-1}$ from the previous iteration, update the estimate $\tilde{\beta_{b-1}}$ to $\tilde{\beta}_b$ by solving the unbiased aggregated Kalman estimating equation

$$\tilde{U}_b(\beta) = \sum_{i,j=1}^{m,b} U_{ij}(\beta) = \sum_{j=1}^{b} X_j^\top\left(y_j - X_j\beta - Z_j m_j\right) = 0$$

where $m_b = E(\theta_b \mid \mathcal{D}_b, \tilde{\beta}_{b-1}, \tilde{\zeta}_{b-1})$ is the conditional mean obtained at the arrival of $\mathcal{D}_b$ using the previous updates for the fixed parameters. The estimate of $\tilde{\beta}_b$ has closed for only for the gaussian linear model

(3) Given $\tilde{\beta}_b$, update ζ_{b-1} to ζ_b by the method of moments, in particular

$$\tilde{\phi}_b = \frac{N_{b-1}}{N_b}\tilde{\phi}_{b-1} + \frac{n_b}{N_b}\hat{\phi}_b$$

where

$$\hat{\phi}_b = \frac{1}{n_b}\left[(y_b - X_b\tilde{\beta}_b - Z_b m_b)^\top (y_b - X_b\tilde{\beta}_b - Z_b m_b) - \sum_{i=1}^{m} P_b(i,i)\right]$$

and $P_j = Z_j C_j Z_j^\top$ where C_j is the conditional variance,

$$\tilde{\delta}_b = \frac{b-2}{b-1}\tilde{\delta}_{b-1} + \frac{1}{b-1}\hat{\delta}_b$$

where

$$\hat{\delta}_b = \frac{1}{q}\left[\|m_b - B_{b-1}m_{b-1}\|^2 + \sum_{i=1}^{q} F_b(i,i)\right]$$

and $F_j = C_{j+1} + B_j C_j B_j^\top - 2C_b(j+1,j)B_j^\top$.
An estimate of $B = \text{diag}(\rho_1, \ldots, \rho_q)$ is obtained by the moment condition that using the Kalman filter leads to

$$\tilde{B}_b = \left(\sum_{j=1}^{b} m_j^\top m_j\right)^{-1}\left(\sum_{j=1}^{b} m_j^\top m_j\right)$$

for $b \geq 2$, with $\tilde{B}_1 = 0$.

It can be shown that under mild conditions $\tilde{\beta}_b$ is consistent and asymptotically normal distributed with mean the true value of the parameter β_b and covariance matrix that $(\tilde{S}_b \top \tilde{V}_b^{-1}\tilde{S}_b)^{-1}$ where both $\tilde{S}_b$ and $\tilde{V}_b$ are not linear aggregation of inferential quantities but depend of the estimated mean square error. In any case they can be computed without access to historical data, but the inference procedure result more difficult than in the independent case.

5. Dependent Data using Estimating Equations

In the same setting as in Section 4, Luo et al. (2023) proposed a new estimator to make inference on streaming data taking into account possible correlation between observations. The covariance of the outcome is modelled with $\Sigma = \text{cov}(y_i \mid X_i) \propto A_i^{1/2} R(\alpha) A_i^{1/2}$, where $A_i = \text{diag}\{v(\mu_{i,kj})\}_{k,j=1}^{n_j,b}$, $v(\cdot)$ is a known variance function and $R(\alpha)$ is a working correlation matrix that is fully characterized by a correlation parameter α.

Because of the longitudinal nature of the infinite-horizon setting, it has been decided to model the measurements through a first-order autoregressive working correlation structure, which is one of the most widely used correlation models for longitudinal data, since it allows to describe the decay in correlation betwen measurements that occurs over time. In the case were Y_{it} and Y_{is} are two outcomes measured respectively at time-points t and s, with $t \neq s$, the autoregressive structures implies $cor(Y_{it}, Y_{is}) = \alpha^{|t-s|}$ for $\alpha \in (-1, 1)$.

At this point estimation of parameter β is usually carried out starting from the generalized estimating equations of Liang & Zeger (1986) to avoid the use of likelihood and α is estimated via a method of moments approach, so that up to time t_b the estimator of β based on data $\{\mathcal{D}_{ib}^{\star}\}_{i=1}^{m}$ is the solution of $\psi_b^\star(\beta,\alpha;\{\mathcal{D}_{ij}^\star\}_{j=1}^m) = \sum_{i=1}^m D_i^\top \Sigma_i^{-1} W_b(y_i-\mu_i) = 0$, where $D_i = \Delta_\beta \mu_i = (D_{ij})_{j=1}^b \in \mathbb{R}^{N_b\times p}$ and $W_b = \mathrm{diag}\{W_{bj}\}_{j=1}^b \in \mathbb{R}^{N_b\times N_b}$ is a weighting matrix that dynamically adjust the weights assigned to data batches collected at different time points, where in this particular case $W_{bj} = q^{t_b - t_j} I_{n_j}$ for $0<q<1$, which means that data batches further away from batch b receive less weight. This estimator for β is always consistent (even under misspecification) and semiparametrically efficient if $R(\alpha)$ is correctly specified. Unfortunately there exist cases for which the estimator for the correlation parameter α does not exist, but it is possible to use quadratic inference function of Qu et al. (2000) that avoids the estimation of α through an approximation to the inverse of the working correlation matrix by $R^{-1}(\alpha) \approx \sum_{s=1}^S \gamma_s M_s$, where $\gamma_1,\dots,\gamma_S$ are unknown constants. A further approximation that leads to the autoregressive structure of the correlation matrix is to take $S=2$, where $M_1 = I_{N_b}$ and M_2 is a matrix with the only non null values that are the ones on the two main off-diagonals. In this setting $\hat\beta_b^\star = \arg\min_\beta Q_b^\star(\beta)$ with

$$Q_b^\star(\beta) = U_b^\star(\beta)^\top \left\{V_b^\star(\beta)\right\}^{-1} U_b^\star(\beta), \tag{2}$$

$$U_b^\star(\beta) = \begin{pmatrix} U_b^\star(\beta)^{(1)} \\ U_b^\star(\beta)^{(2)} \end{pmatrix} = \sum_{i=1}^m \begin{pmatrix} U_{ib}^\star(\beta)^{(1)} \\ U_{ib}^\star(\beta)^{(2)} \end{pmatrix} = \sum_{i=1}^m \begin{pmatrix} D_i^\top A_i^{-1/2} M_1 A_i^{-1/2} W_b(y_i-\mu_i) \\ D_i^\top A_i^{-1/2} M_2 A_i^{-1/2} W_b(y_i-\mu_i) \end{pmatrix} \in \mathbb{R}^{2p}$$

and $V_b^\star(\beta) = \sum_{i=1}^m U_b^\star(\beta) U_b^\star(\beta)^\top$ is the sample covariance matrix of $U_b^\star(\beta)$. It is possible to show that $U_b^\star(\beta)$ can be decomposed into estimating functions for within-batch dependencies through $U_{ij}(\beta)^{(2)}$ and between batch dependencies, trough $U_{i,j,j+1}(\beta)$ and $U_{i,j+1,j}(\beta)$, that require only the last observation of the historical data to be computed. The matrix $S_b^\star(\beta)$, which is the negative gradient of $U_b^\star(\beta)$, can also be decomposed in the same parts as $U_b^\star(\beta)$. The estimator $\hat\beta_b^\star = \arg\min_\beta Q_b^\star(\beta)$ needs to satisfy

$$S_b^\star(\hat\beta_b)^\top \left\{V_b^\star(\hat\beta_b^\star)\right\}^{-1} U_b^\star(\hat\beta_b^\star) = 0.$$

For simplicity let's take just the first updating from $\hat\beta_1$ to $\hat\beta_2^\star$. In order for this updating to be possible it's necessary to plug in the new estimator in the old data batches in order to recompute the score function in the new value of the estimate. Since the aim of this estimator is not to use historical data, a further semplification can be obtain taking the first-order expansion of the pertms $U_1(\hat\beta_2^\star)$ and $S_1(\hat\beta_2^\star)$ about $\hat\beta_1$, to obtain

$$\frac{n_1}{N_2} U_1(\hat\beta_2^\star) = \frac{n_1}{N_2} U_1(\hat\beta_1) + \frac{n_1}{N_2} S_1(\hat\beta_1)(\hat\beta_1 - \hat\beta_2^\star) + O_p\left(\frac{n_1}{N_2}\left\|\hat\beta_1 - \hat\beta_2^\star\right\|^2\right),$$

$$\frac{n_1}{N_2} S_1(\hat\beta_2^\star) = \frac{n_1}{N_2} S_1(\hat\beta_1) + O_p\left(\frac{n_1}{N_2}\left\|\hat\beta_1 - \hat\beta_2^\star\right\|^2\right).$$

The error terms can be ignored if N_2 is big enough and $\|\beta_2 - \beta_1\| = o(1)$. In this setting, generalizing to b batches, it's been proposed an online streaming quadratic inference function estimator $\tilde\beta_2$ of β, which is the solution to the incremental estimating equation

$$\tilde S_b^\top \tilde V_b^{-1} \tilde U_b = 0 \tag{3}$$

where

$$\tilde{U}_{ib} = q^{t_b - t_{b-1}} \tilde{U}_{i,b-1} + q^{t_b - t_{b-1}} \tilde{S}_{i,b-1} \left(\tilde{\beta}_{b-1} - \tilde{\beta}_b\right) + \begin{pmatrix} U_{ib}\left(\tilde{\beta}_b\right)^{(1)} \\ U_{ib}\left(\tilde{\beta}_b\right)^{(2)} + U_{i,b-1,b}\left(\tilde{\beta}_b\right) + q^{t_b - t_{b-1}} U_{i,b,b-1}\left(\tilde{\beta}_b\right) \end{pmatrix}$$

$$\tilde{S}_{ib} = q^{t_b - t_{b-1}} \tilde{S}_{i,b-1} + \begin{pmatrix} S_{ib}\left(\tilde{\beta}_b\right)^{(1)} \\ S_{ib}\left(\tilde{\beta}_b\right)^{(2)} + S_{i,b-1,b}\left(\tilde{\beta}_b\right) + q^{t_b - t_{b-1}} S_{i,b,b-1}\left(\tilde{\beta}_b\right) \end{pmatrix},$$

$$\tilde{U}_b = \sum_{i=1}^{m} \tilde{U}_{ib}, \qquad \tilde{V}_b = \sum_{i=1}^{m} \tilde{U}_{ib} \tilde{U}_{ib}^{\top}, \qquad \tilde{U}_b = \sum_{i=1}^{m} \tilde{S}_{ib}.$$

It is possible to solve iteratively equation (3) via Newton-Rhapson algorithm as follows

$$\tilde{\beta}_b^{(r+1)} = \tilde{\beta}_b^{(r)} + \left\{ \tilde{S}_b^{(r)\top} \left(\tilde{V}_b^{(r)}\right)^{-1} \tilde{S}_b^{(r)} \right\}^{-1} \tilde{S}_b^{(r)\top} \left(\tilde{V}_b^{(r)}\right)^{-1} \tilde{U}_b^{(r)}.$$

The last step to obtain the updated estimates is to select the optimal value of the weighting parameter q. The solution is to find the value within the candidate set C_q that minimize the quadratic inference function constructed with the data in batch b and the last observation in batch $b-1$. The optimal value for q then is

$$q_b^{\text{opt}} = \operatorname*{argmin}_{q \in C_q} U_b\left(\tilde{\beta}_b, q\right)^{\top} \left\{ V_b\left(\tilde{\beta}_b, q\right) \right\}^{-1} U_b\left(\tilde{\beta}_b, q\right).$$

It can be shown that under some assumptions the estimator $\tilde{\beta}_b$ is consistent as $N_b \to \infty$ and that is asymtotically normally distributed with mean β_b and variance $(\tilde{S}_b^{\top} \tilde{V}_b^{-1} \tilde{S}_b)^{-1}$, where it is possible to see that, in contrast with the estimator proposed in the previous section, both $\tilde{S}_b$ and $\tilde{V}_b$ are linear aggregation of inferential quantities computed each time that a new batch arrives.

References

Liang, K.-Y. & Zeger, S. L. (1986). Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika* **73**, 13–22.

Luo, L. & Song, P. X.-K. (2020). Renewable Estimation and Incremental Inference in Generalized Linear Models with Streaming Data Sets. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **82**, 69–97.

Luo, L. & Song, P. X.-K. (2023). Multivariate Online Regression Analysis with Heterogeneous Streaming Data. *Canadian Journal of Statistics* **51**, 111–133.

Luo, L., Wang, J. & Hector, E. C. (2023). Statistical Inference for Streamed Longitudinal Data. *Biometrika* **110**, 841–858.

Qu, A., Lindsay, B. G. & Li, B. (2000). Improving Generalised Estimating Equations Using Quadratic Inference Functions. *Biometrika* **87**, 823–836.

Song, P. X.-K. (2007). *Correlated Data Analysis: Modeling, Analytics, and Applications*. Springer Series in Statistics. New York, NY: Springer.