Complexity of crossed random effects

Art Owen

Discussion

Alessandra Salvan and Nicola Sartori
Department of Statistical Sciences
University of Padova

Main points (Ghosh, Hastie & Owen, 2022a, 2022b)

- Big data regression with $N \gg p$ (tall data)
- Random effects
  - Linear model:
    $\rightarrow$ GLS for estimation of β,
    not so straightforward as it seems when random effects are crossed and large N, because of computational complexity
  - GLMM:
    additional difficulty because of high-dimensional integral to get the marginal likelihood
- Standard algorithms $O(N^{3/2})$ complexity

Main points (Ghosh, Hastie & Owen, 2022a, 2022b)

- Big data regression with $N \gg p$ (tall data)
- Random effects
  - Linear model:
    $\rightarrow$ GLS for estimation of β,
    not so straightforward as it seems when random effects are crossed and large N, because of computational complexity
  - GLMM:
    additional difficulty because of high-dimensional integral to get the marginal likelihood
- Standard algorithms $O(N^{3/2})$ complexity

Main points (Ghosh, Hastie & Owen, 2022a, 2022b)

- Big data regression with $N \gg p$ (tall data)
- Random effects
  - Linear model:
    $\rightarrow$ GLS for estimation of β,
    not so straightforward as it seems when random effects are crossed and large N, because of computational complexity
  - GLMM:
    additional difficulty because of high-dimensional integral to get the marginal likelihood
- Standard algorithms $O(N^{3/2})$ complexity

Main points (Ghosh, Hastie & Owen, 2022a, 2022b)

- Big data regression with $N \gg p$ (tall data)
- Random effects
  - Linear model:
    $\rightarrow$ GLS for estimation of β,
    not so straightforward as it seems when random effects are crossed and large N, because of computational complexity
  - GLMM:
    additional difficulty because of high-dimensional integral to get the marginal likelihood
- Standard algorithms $O(N^{3/2})$ complexity

Main points

- The optimization problem is recast as a joint optimization of β and random effects $\to$ backfitting
- The backfitting algorithm has $O(N)$ complexity
- For the linear model, the algorithm requires (consistent) estimates of variance components (Gao & Owen, 2017, method of moments) with same $O(N)$ complexity
- For logistic regression, the algorithm is applied to PQL and iterates also over variance parameters
- Crucial conditions:
  - Sparseness: $N \ll RC$
    Sampling pattern:
    $$Z_{ij} \overset{\text{ind}}{\sim} \text{Bern}(p_{ij})$$
    $$\frac{S}{RC} \leq p_{ij} \leq \Upsilon \frac{S}{RC}$$

Main points

- The optimization problem is recast as a joint optimization of β and random effects $\rightarrow$ backfitting
- The backfitting algorithm has $O(N)$ complexity
- For the linear model, the algorithm requires (consistent) estimates of variance components (Gao & Owen, 2017, method of moments) with same $O(N)$ complexity
- For logistic regression, the algorithm is applied to PQL and iterates also over variance parameters
- Crucial conditions:
  - Sparseness: $N \ll RC$
    Sampling pattern:
$$Z_{ij} \overset{\text{ind}}{\sim} \text{Bern}(p_{ij})$$
$$\frac{S}{RC} \leq p_{ij} \leq \Upsilon \frac{S}{RC}$$

Main points

- The optimization problem is recast as a joint optimization of β and random effects $\rightarrow$ backfitting
- The backfitting algorithm has $O(N)$ complexity
- For the linear model, the algorithm requires (consistent) estimates of variance components (Gao & Owen, 2017, method of moments) with same $O(N)$ complexity
- For logistic regression, the algorithm is applied to PQL and iterates also over variance parameters
- Crucial conditions:
  - Sparseness: $N \ll RC$
    Sampling pattern:
    $$Z_{ij} \overset{\text{ind}}{\sim} \mathrm{Bern}(p_{ij})$$
    $$\frac{S}{RC} \leq p_{ij} \leq \Upsilon \frac{S}{RC}$$

Some questions – backfitting versus MCMC methods

- How crucial are different assumptions about the Z_{ij}?
- Are there configurations where in practice one method fails and the other works?

Some questions – logistic regression and GLMM

- Overall computational complexity larger than $O(N)$: evidence or hints that it could be smaller?
- Accuracy of Laplace approximation with very high-dimensional integrals? (Ogden, 2021; Shun & McCullagh, 1995)
- Is the backfitting algorithm also needed with probit link?
- Could simplifications for probit be used also for logistic regression when approximating the logit link function by normal scale mixtures?
- Other methods?
  - composite likelihood (Bellio & Varin, 2005), with moderate number of subsets of data
  - variational approximations (Menictas, Di Credico, Wand, 2022; Jeon, Rijmen, Rabe-Hesketh, 2017)

Some questions – logistic regression and GLMM

- Overall computational complexity larger than $O(N)$: evidence or hints that it could be smaller?
- Accuracy of Laplace approximation with very high-dimensional integrals? (Ogden, 2021; Shun & McCullagh, 1995)
- Is the backfitting algorithm also needed with probit link?
- Could simplifications for probit be used also for logistic regression when approximating the logit link function by normal scale mixtures?
- Other methods?
  - composite likelihood (Bellio & Varin, 2005), with moderate number of subsets of data
  - variational approximations (Menictas, Di Credico, Wand, 2022; Jeon, Rijmen, Rabe-Hesketh, 2017)

Some questions – logistic regression and GLMM

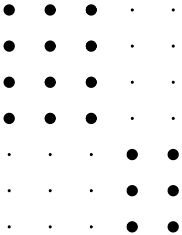
- Overall computational complexity larger than $O(N)$: evidence or hints that it could be smaller?
- Accuracy of Laplace approximation with very high-dimensional integrals? (Ogden, 2021; Shun & McCullagh, 1995)
- Is the backfitting algorithm also needed with probit link?
- Could simplifications for probit be used also for logistic regression when approximating the logit link function by normal scale mixtures?
- Other methods?
  - composite likelihood (Bellio & Varin, 2005), with moderate number of subsets of data
  - variational approximations (Menictas, Di Credico, Wand, 2022; Jeon, Rijmen, Rabe-Hesketh, 2017)

Some questions – logistic regression and GLMM

- Overall computational complexity larger than $O(N)$: evidence or hints that it could be smaller?
- Accuracy of Laplace approximation with very high-dimensional integrals? (Ogden, 2021; Shun & McCullagh, 1995)
- Is the backfitting algorithm also needed with probit link?
- Could simplifications for probit be used also for logistic regression when approximating the logit link function by normal scale mixtures?
- Other methods?
  - composite likelihood (Bellio & Varin, 2005), with moderate number of subsets of data
  - variational approximations (Menictas, Di Credico, Wand, 2022; Jeon, Rijmen, Rabe-Hesketh, 2017)

Some questions – alternative (simpler?) solutions

- In Ghosh, Hastie, Owen (2022a): iterations not $O(1)$, and unbounded mixing time for collapsed Gibbs of Papaspiliopoulos, Roberts, Zanella (2020), when
  there are two disjoint communities of users and two disjoint sets of items and each user in the first community has rated every item in the first item set (and no others) while each user in the second community has rated every item in the second item set (and no others).



Some questions – alternative (simpler?) solutions

- Feasible to work with the two disjoint sets separately and then combine estimates?
- More generally, could ideas from distributed inference (Jordan, 2013) or composite likelihood (with fixed or limited number of factors) be useful?
- In some situation (web applications) data are naturally available as sequential batches: find updating algorithms?

Some references

- Gao, K. & Owen, A. (2017). Efficient moment calculations for variance components in large unbalanced crossed random effects models. Electronic Journal of Statistics, 11, 1235-1296.
- Bellio, R. & Varin, C. (2005). A pairwise likelihood approach to generalized linear models with crossed random effects. Statistical Modelling, 5, 217-227.
- Ghosh, S., Hastie, T. & Owen, A. B. (2022a). Backfitting for large scale crossed random effects regressions. The Annals of Statistics, 50, 560-583.
- Ghosh, S., Hastie, T. & Owen, A. B. (2022b). Scalable logistic regression with crossed random effects. arXiv preprint arXiv:2105.13747, Electronic Journal of Statistics.
- Jeon, M., Rijmen, F. & Rabe-Hesketh, S. (2017). A variational maximization-maximization algorithm for generalized linear mixed models with crossed random effects. Psychometrika, 82, 693-716.
- Jordan, M.I. (2013). On statistics, computation and scalability. Bernoulli, 19, 1378-139.
- Menictas, M., Di Credico, G. & Wand, M. P. (2022). Streamlined variational inference for linear mixed models with crossed random effects. Journal of Computational and Graphical Statistics, to appear.
- Ogden, H. (2021). On the error in Laplace approximations of high-dimensional integrals. Stat, 10(1), e380.
- Papaspiliopoulos, O., Roberts, G. O. & Zanella, G. (2020). Scalable inference for crossed random effects models. Biometrika, 107, 25-40.
- Shun, Z. & McCullagh, P. (1995). Laplace approximation of high dimensional integrals. J. Roy. Stat. Soc. Series B, 57, 749-760.