Advanced Statistical Computing Proejct

Hierarchical-block conditioning approximations for high-dimensional multivariate normal probabilities

Hyunseok Yang, Kyeongwon Lee, Songhyun Kim

Department of Statistics, Seoul National University

December 7, 2019

1 Introduction

The computation of multivariate normal probability appears various fields. For instance, the inferences based on the central limit theorem, which holds when the sample size is large enough, is widely used in the social sciences and engineering as well as in the natural sciences. Recently, the dimensionality of data and models has been grown significantly, and in this respect, so does a need for the methodology to efficiently calculate high-dimensional multivariate normal probability.

Cao, Genton, Keyes, and Turkiyyah (2019) proposes new approaches to approximate high-dimensional multivariate normal probability

$$\Phi_n(\mathbf{a}, \mathbf{b}; 0, \Sigma) = \int_a^b \frac{1}{\sqrt{(2\pi)^n|\Sigma|}} \exp\left(-\frac{1}{2}\mathbf{x}^T\Sigma^{-1}\mathbf{x}\right) d\mathbf{x}, \tag{1}$$

using the hierarchical matrix $\mathcal{H}$ (Hackbusch, 2015) for the covariance matrix Σ. The methods are based on two state-of-arts methods, among others, are the bivariate conditioning method (Trinh & Genz, 2015) and the hierarchical Quasi-Monte Carlo method (Genton, Keyes, & Turkiyyah, 2018). Specifically, Cao et al. (2019) generalize the bivariate conditioning method to a d-dimension and combine it with the hierarchical representation of the covariance matrix.

2 Multidimensional Conditioning Approximations

2.1 d-Dimensional Conditioning Approximation

2.2 CMVN

2.3 RCMVN

3 Hierarchical-Block Conditioning Approximations

In this section, we suggest methods to solve the n-dimensional MVN problem with the hierarchical covariance matrix using the d-dimensional conditioning method with that of the Monte Carlo-based method for solving the m-dimensional MVN problems presented by the diagonal blocks.

Let $\phi_m(\mathbf{x}; \mathbf{\Sigma})$ be a pdf of the m-dimensional normal distribution $N(\mathbf{0}, \mathbf{\Sigma})$ and $(\mathbf{B}, \mathbf{U}\mathbf{V}^T)$ be the hierarchical Cholesky decompostion of the covariance matrix Σ. Then, we can express (1) as

$$\Phi_n(\mathbf{a}, \mathbf{b}; \mathbf{0}, \mathbf{\Sigma}) = \int_{\mathbf{a}'_1}^{\mathbf{b}'_1} \phi_m(\mathbf{x}_1; \mathbf{B}_1\mathbf{B}_1^T) \cdots \int_{\mathbf{a}'_r}^{\mathbf{b}'_r} \phi_r(\mathbf{x}_r; \mathbf{B}_r\mathbf{B}_r^T) d\mathbf{x}_r \cdots d\mathbf{x}_1. \tag{2}$$

Where $\mathbf{a}'$, $\mathbf{b}'$, $i = 1, \cdots, r$, are the corresponding segments of the updated $\mathbf{a}$ and $\mathbf{b}$. Speciffically, we compute n-dimensional MVN problem using hierarchical structure as algorithm 1.

Algorithm 1 Hierarchical-block conditioning algorithm

procedure HMVN(a, b, Σ, d)
  $\mathbf{x} \leftarrow \mathbf{0}$ and $P \leftarrow 1$
  $[\mathbf{B}, \mathbf{UV}] \leftarrow$ `choldecomp_hmatrix`(Σ)
  for $i = 1 : r$ **do**
    $j \leftarrow (i-1)m$
    if $i > 1$ **then**
      $o_r \leftarrow$ row offset of $\mathbf{U}_{i-1}\mathbf{V}_{i-1}^T$
      $o_c \leftarrow$ column offset of $\mathbf{U}_{i-1}\mathbf{V}_{i-1}^T$
      $l \leftarrow \dim(\mathbf{U}_{i-1}\mathbf{V}_{i-1}^T)$
      $\mathbf{g} \leftarrow \mathbf{U}_{i-1}\mathbf{V}_{i-1}^T\mathbf{x}[o_c+1 : o_c+l]$
      $\mathbf{a}[o_r+1 : o_r+l] = \mathbf{a}[o_r+1 : o_r+l] - \mathbf{g}$
      $\mathbf{b}[o_r+1 : o_r+l] = \mathbf{a}[o_r+1 : o_r+l] - \mathbf{g}$
    end if
    $\mathbf{a}_i \leftarrow \mathbf{a}[j+1 : j+m]$
    $\mathbf{b}_i \leftarrow \mathbf{b}[j+1 : j+m]$
    $P = P * \Phi_m(\mathbf{a}_i, \mathbf{b}_i; \mathbf{0}, \mathbf{B}_i\mathbf{B}_i^T)$
    $\mathbf{x}[j+1 : j+m] \leftarrow \mathbf{B}_i^{-1}\mathbb{E}[\mathbf{X}_i]$
  end for
end procedure

Note the probabilities $\Phi_m(\mathbf{a}_i, \mathbf{b}_i; \mathbf{0}, \mathbf{B}_i\mathbf{B}_i^T)$ can be computed using d-dimensional conditioning algorithm (`HCMVN`) or with d-dimensional conditioning algorithm with univariate reordering (`HRCMVN`). These methods are more effective and easily parallelizable than the classical methods.

4 Block Reordering

5 Results

5.1 Data

5.2 Multivariate Normal Probabilities

To implement `*MVN` functions, we need to calculate n-dimensional normal probability (1),

$$\Phi_n(a, b; 0, \Sigma) = \int_a^b \frac{1}{\sqrt{(2\pi)^n|\Sigma|}} \exp\left(-\frac{1}{2}\mathbf{x}^T\Sigma^{-1}\mathbf{x}\right) d\mathbf{x},$$

numerically. We implement `mvn`, the function that calculate multivariate normal probabilities using Richtmyer Quasi-Monte Carlo(QMC) method proposed by Genz and Bretz (2009).

It is well-known that QMC methods is more effective than classical Monte Carlo(MC) method. All the multivariate normal distribution probabilities required in the next algorithms are calculated using the `mvn` function.

6 Conclusion

References

Cao, J., Genton, M. G., Keyes, D. E., & Turkiyyah, G. M. (2019). Hierarchical-block conditioning approximations for high-dimensional multivariate normal probabilities. *Statistics and Computing*, *29*(3), 585–598.

Genton, M. G., Keyes, D. E., & Turkiyyah, G. (2018). Hierarchical decompositions for the computation of high-dimensional multivariate normal probabilities. *Journal of Computational and Graphical Statistics*, *27*(2), 268–277.

Genz, A., & Bretz, F. (2009). *Computation of multivariate normal and t probabilities* (Vol. 195). Springer Science & Business Media.

Hackbusch, W. (2015). *Hierarchical matrices: algorithms and analysis* (Vol. 49). Springer.

Trinh, G., & Genz, A. (2015). Bivariate conditioning approximations for multivariate normal probabilities. *Statistics and Computing*, *25*(5), 989–996.