

Aproksimasi Komputasi Bayesian: Suatu Pengantar

Bertha Selviana Djahi

Jurusan Ilmu Komputer, Fakultas Sains dan Teknik, Universitas Nusa Cendana, Kupang

Email: berthadjahi@yahoo.com

Abstrak

Aproksimasi Komputasi Bayesian telah banyak digunakan di berbagai bidang. Metode ini tidak perlu menghitung fungsi kemungkinan sebagaimana diperlukan metode Bayesian lainnya seperti Gibbs. Metode ini sangat sangat berguna dalam teknik inferensi dari parameter-parameter dari model-model yang sangat kompleks. Paper ini akan membahas dan mereview teknik aproksimasi komputasi Bayesian dan memaparkan dua algoritma aproksimasi komputasi Bayesian.

Abstract

Approximation of Computing Bayesian has many applied in various areas. This method not necessarily calculate function of possibility that as required [by] other Bayesian method like Gibbs. This very useful method in inference technique from parameters from a real models complex. This cartridge will study and mereview approximation technique of computing Bayesian and explains two approximation algorithms of computing Bayesian.

Keyword : Approximation of Computing Bayesian, Algorithm, Estimation of Parameter

1. Latar Belakang

Aproksimasi komputasi Bayesian adalah kelompok metode komputasi yang berdasar pada statistika Bayesian. Metode ini pertama kali diperkenalkan pada Tahun 1999 oleh Pritchard dkk. Metode ini kemudian berkembang dengan cepat. Dalam teknik inferensi, fungsi kemungkinan merupakan kunci utama. Namun, ketika sebuah model menjadi sangat kompleks, khususnya model-model yang bersifat probabilistik, perhitungan atau evaluasi dari fungsi kemungkinan dapat tidak mudah dilakukan dan *computationally costly*. Oleh karena itu, untuk mengatasi persoalan tersebut, aproksimasi komputasi Bayesian dapat digunakan. Pendekatan ini tidak mensyaratkan fungsi kemungkinan. Dengan demikian, untuk suatu masalah atau model yang sangat kompleks pendekatan ini lebih tepat digunakan.

Dengan kemajuan perangkat komputer saat ini, pendekatan ini juga berkembang secara pesat. Metode ini telah digunakan dalam berbagai bidang diantaranya psikologi, biologi, kedokteran, dan kesehatan. Review tentang penggunaan metode ini di berbagai bidang dapat dibaca pada artikel Lopes dan Beamont [1]. Pendekatan ini digunakan diantaranya untuk mengestimasi nilai parameter dan pemilihan model. Oleh karena itu, paper ini akan membahas mengenai aproksimasi komputasi Bayesian.

2. Metode Penelitian

Metode yang digunakan dalam paper ini adalah teknik mereview. Yakni, mereview paper-paper yang mengaplikasikan aproksimasi komputasi Bayesian dalam masalah-masalah sains seperti genetika dan sistim dinamika. Kemudian dari paper-paper tersebut diperoleh cara kerja dari metode aproksimasi komputasi Bayesian yang diidentifikasi.

3. Hasil dan Pembahasan

3.1 Hasil Penelitian

3.1.1 Kerangka Bayesian

Dalam aproksimasi komputasi Bayesian, Bayes teori digunakan untuk menggenerasikan distribusi akhir dari parameter. Bayes teori ditulis sebagai berikut, misalkan \mathbf{D} adalah data dan θ adalah parameter, maka peluang dari θ bila terdapat data ditulis:

$$P(\theta|\mathbf{D}) = \frac{P(\mathbf{D}|\theta)P(\theta)}{P(\mathbf{D})}$$

Dimana:

- $P(\theta|\mathbf{D})$ = peluang target
- $P(\mathbf{D}|\theta)$ = peluang kemungkinan
- $P(\theta)$ = peluang awal, dan
- $P(\mathbf{D})$ = peluang prediktif dari data

Peluang awal adalah peluang atau nilai kepercayaan terhadap parameter θ sebelum data tersedia. Nilai ini umumnya dapat dipilih dari distribusi statistik yang tersedia. Ini sering disebut sebagai distribusi akhir yang ingin dicari.

Untuk menghitung distribusi akhir dari parameter θ , ada dua syarat yang harus dipenuhi. Pertama, fungsi kemungkinan harus dapat dihitung atau ditentukan. Kedua, distribusi awal harus ditentukan. Penentuan distribusi awal berdasar pada pengalaman atau pengetahuan kita tentang kemungkinan nilai parameter tersebut [2]. Untuk syarat pertama, ketika suatu model semakin kompleks, fungsi kemungkinan tidak mudah dihitung atau ditentukan. Sebagaimana telah diketahui, bahwa aproksimasi komputasi Bayesian merupakan metode yang tidak memerlukan perhitungan fungsi kemungkinan. Perhitungan fungsi kemungkinan diganti oleh simulasi dari model yang menghasilkan suatu data artifisial [3]. Data ini kemudian dibandingkan dengan data real yakni dengan menghitung jarak antara data hasil observasi dan data dari model. Oleh karena itu, metode ini sangat bergantung pada jarak antara data dari model dan data real.

Selanjutnya, fungsi untuk menghitung jarak antara data dari model dan real data menjadi penting. Pemilihan fungsi jarak yang tepat akan memberikan peluang yang besar sehingga distribusi nilai parameter yang dihasilkan lebih baik. Sebaliknya, bila pemilihan fungsi jarak kurang tepat maka distribusi akhir dari suatu parameter menjadi kurang baik. Ada banyak algoritma yang dibuat untuk menentukan distribusi akhir. Pada paper ini akan dibahas algoritma yang paling sederhana dan aplikasinya.

3.1.2 Algoritma Pertama: Algoritma Sederhana Komputasi Bayesian Penolakan Sampel

Algoritma ini merupakan algoritma yang paling sederhana dari aproksimasi komputasi Bayesian. Perhatikan notasi-notasi berikut: θ adalah parameter, π distribusi awal parameter, $\rho(x, x_0)$ adalah jarak antara data observasi dan data yang dihasilkan dari model, dan ε adalah batas toleransi. Algoritma aproksimasi komputasi Bayesian penolakan sampel sebagaimana diusulkan oleh [4] sebagai berikut:

1. Pilih parameter θ dari distribusi awal π
2. Simulasi data seta dari model $f(\cdot|\theta)$ dengan menggunakan nilai θ
3. Hitung jarak antara data dari model dan data hasil observasi $\rho(x, x_0)$
4. Terima θ bila $\rho(x, x_0) < \varepsilon$, jika tidak ulangi langkah satu.

Dari algoritma ini terlihat jelas bahwa parameter yang diusulkan akan diterima bila data yang dihasilkan dari model dengan menggunakan nilai parameter yang diusulkan ketika dibandingkan dengan data observasi atau

real data, lebih kecil dari suatu batas toleransi ε . Bila tidak, akan diulang sampai mendapatkan nilai parameter yang menghasilkan data yang bila dibandingkan dengan data hasil observasi lebih kecil dari batas toleransi. Ini akan diulang sampai jumlah nilai parameter yang diterima mencapai suatu jumlah tertentu.

3.1.3 Algoritma kedua: Algoritma Markov Chain Monte Carlo (MCMC) Aproksimasi Komputasi Bayesian

Algoritma lain yang diusulkan adalah menggunakan konsep Markov Chain Monte Carlo dalam paradigma Bayesian. Algoritma ini berdasarkan yang dipaparkan oleh [5], yakni:

1. Usulkan nilai parameter awal θ
2. Pada waktu t usulkan perpindahan parameter dari parameter θ ke θ' menurut distribusi usulan $q(\theta|\theta')$
3. Simulasi data x dari model menggunakan nilai parameter θ'
4. Jika jarak antara data observasi dan data dari model lebih kecil batasan yang ditetapkan atau ditulis $\rho(x, x_0) < \varepsilon$,
5. Set $\theta_{t+1} = \theta'$ atau terima nilai parameter θ' dengan nilai probabilitas

$$\min \left(1, \frac{\pi(\theta')q(\theta|\theta')}{\pi(\theta)q(\theta'|\theta)} \right)$$

6. Maju ke simulasi berikutnya dan kembali ke langkah 2.

3.2 Pembahasan Hasil Penelitian

Dua algoritma yang telah dipaparkan di atas dapat digunakan untuk mengestimasi distribusi parameter dari model. Selain itu, kesimpulan-kesimpulan statistik seperti median, mean juga dapat ditentukan. Kelemahan dari algoritma yang pertama adalah ketika distribusi awalnya sangat berbeda dengan distribusi akhirnya, maka peluang penerimaan sampelnya sangat rendah. Namun, kelemahan ini dapat diatasi dengan menggunakan algoritma kedua [6], yang dipastikan akan konvergen ke distribusi akhirnya.

Metode ini dapat digunakan untuk model-model yang bersifat probabilistik seperti model stokastik maupun model lain yang tidak bersifat probabilistik. Salah satu kunci dalam penggunaannya adalah pemilihan fungsi jarak. Pemilihan fungsi jarak ini menjadi krusial. Umumnya, fungsi jarak yang dipakai adalah *sum of square error*. Fungsi jarak lainnya juga dapat digunakan, tergantung dari persoalan yang di teliti dan juga selama memiliki justifikasi ilmiah yang benar. Metode ini telah diaplikasikan dalam berbagai bidang. Dalam bidang genetika [4] dan [7], dalam bidang network diseases [8], ekologi [9], epidemik model [10]. Kemudian berbagai algoritma lainnya telah dikembangkan, yang detailnya dapat ditemukan dalam [11] dan [12].

4. Kesimpulan

Aproksimasi komputasi Bayesian merupakan metode yang sangat bagus dalam mengestimasi distribusi parameter dari model-model yang sangat kompleks khususnya bila fungsi kemungkinan tidak mudah dihitung. Metode ini dapat digunakan untuk model yang bersifat probabilistik ataupun tidak. Tantangannya adalah ketika data observasi tidak berkualitas baik. Misalnya, data hilang atau informasi data yang tidak lengkap. Namun, ini dapat diatasi dengan menggunakan algoritma yang tepat untuk memperlakukan data hilang tersebut. Peluang penelitian dibidang ini sangat luas. Baik itu dari aspek metodologi maupun dari aspek penerapannya. Misalnya, pencarian algoritma yang lebih efisien sehingga waktu yang dibutuhkan untuk mencapai kekonvergenan dan distribusi akhir dari suatu parameter lebih cepat dan tidak memakan memori komputer. Bagaimana ketika data tidak memenuhi distribusi-distribusi standar yang tersedia dan bagaimana cara memperlakukannya.

DAFTAR PUSTAKA

- [1] J.S. Lopes dan M.A. Beaumont. 2009. *ABC: A Useful Bayesian Tool for the Analysis of Population Data*. Infection, Genetics and Evolution.
- [2] B. Turner dan T. Van Zandt. 2012. *A Tutorial on Approximate Bayesian Computation*. Journal of Mathematical Psychology. 56, 69–85.
- [3] B. Turner. 2011. *Likelihood-free Bayesian Modelling*. PhD Thesis. The Ohio State University.
- [4] J. Pritchard, M. T. Seielstad, A. Perez-Lezaun, and M. W. Feldman. 1999. *Population Growth of Human Y Chromosomes: A Study of Y Chromosome Microsatellites*. Molecular Biology Evolutions.
- [5] P. Marjoram, J. Molitor, V. Plagnol, dan S. Tavaré. 2003. *Markov Chain Monte Carlo Without Likelihoods*. Proceeding of National Academy Science, USA. 100: 15324–28.
- [6] T. Toni, D. Welch, N. Strelkowa, A. Ipsen, dan M. P. H. Stumpf. 2009. *Approximate Bayesian Computation Scheme for Parameter Inference and Model Selection in Dynamical Systems*. Journal of the Royal Society Interface. 6,: 187–202. doi:10.1098/rsif.2008.0172.
- [7] M. G. B. Blum dan M. Jakobsson. 2011. *Deep Divergences of Human Gene Trees and Models of Human Origins*. Molecular Biology and Evolution. 28(2): 889-898. doi: 10.1093/molbev/msq265.
- [8] D. M. Walker, D. Allingham, H. W. J. Lee, M. Small. 2010. *Parameter Inference in Small World Network Disease Models with Approximate Bayesian Computational Methods*. Physica A . 389: 540-548.
- [9] M. A. Beaumont. 2010. *Approximate Bayesian Computation in Evolution and Ecology*. The Annual Review of Ecology, Evolution, and Systematics. 41:379–406.
- [10] T. McKinley, A. R. Cook dan R. Deardon. 2009. *Inference in Epidemic Models Without Likelihoods*. The International Journal of Biostatistics: 5(1). doi:10.2202/1557-4679.1171.
- [11] P. D. Moral, A. Doucet-dan A. Jasra. 2011. *An Adaptive Sequential Monte Carlo Method for Approximate*. Statistics Computing. DOI 10.1007/s11222-011-9271-y.
- [12] S. Barthelmé dan N. Chopin. 2013. *Expectation-Propagation for Likelihood-Free Inference*. www.arxiv.org. diakses pada tanggal 10 Mei 2013.