

Approximate inference of gene regulatory network models from RNA-Seq time series data

Thomas Thorne*

Division of Brain Sciences, Imperial College London

June 14, 2017

Abstract

Inference of gene regulatory network structures from RNA-Seq data is challenging due to the nature of the data, as measurements take the form of counts of reads mapped to a given gene. Here we present a model for RNA-Seq time series data that applies a negative binomial distribution for the observations, and uses sparse regression with a horseshoe prior to learn a dynamic Bayesian network of interactions between genes. We use a variational inference scheme to learn approximate posterior distributions for the model parameters. The methodology is benchmarked on synthetic data designed to replicate the distribution of real world RNA-Seq data. We compare our method to other sparse regression approaches and information theoretic methods. We demonstrate an application of our method to a publicly available human neuronal stem cell differentiation RNA-Seq time series.

1 Introduction

Methods for the inference of gene regulatory networks from RNA-Seq data are currently not as mature as those developed for microarray datasets. Normalised microarray data possesses the desirable property of being approximately normally distributed so that they are readily amenable to various forms of inference, and in the literature many graphical modelling schemes have been explored that exploit the normality of the data [47, 18, 38, 24, 25, 15, 44, 45, 46].

The data generated by RNA-Seq studies on the other hand present a more challenging inference problem, as the data are no longer approximately normally distributed, and before normalisation take the form of non-negative integers. In the detection of differential expression in RNA-Seq data, negative binomial distributions have been applied [16, 3, 41, 26], providing a good fit to the overdispersion typically seen in the data relative to a Poisson distribution. Following similar graphical modelling approaches as applied in the analysis of microarray data, it is natural to consider Poisson and negative binomially distributed graphical models. Unfortunately in many cases when applying graphical modelling approaches with Poisson distributed observations, only models that represent negative conditional dependencies are available, or inference is significantly

*to whom correspondence should be addressed, t.thorne@imperial.ac.uk

complicated due to lack of conjugacy between distributions[19]. Poisson graphical models have been applied successfully in the analysis of miRNA regulatory interactions [2, 13], but we might expect to improve on these by modelling the overdispersion seen in typical RNA-Seq data sets with a negative binomial model.

One specific case of interest in the analysis of RNA-Seq data is the study of time series in a manner that takes into account the temporal relationships between data points. Previous work in the literature has developed sophisticated models for the inference of networks from microarray time series data [24, 25], but whilst methods have been developed for the analysis of differential behaviour in RNA-Seq time series [1, 21], little attention has been given to the task of learning networks from such data. Here we present a method for the inference of networks from RNA-Seq time series data through the application of a Dynamic Bayesian Network (DBN) approach, that models the RNA-Seq count data as being negative binomially distributed conditional on the expression levels of a set of predictors.

2 Methods

2.1 Dynamic Bayesian Networks

In a DBN framework[23], considering only edges between time points, we can model a sequence of observations using a first order Markov model, where the value of a variable at time t is dependent only on the values of a set of parent variables at time $t - 1$. This can be written as

$$P(Y_t) = P(Y_t | \text{Pa}(Y)_{t-1}) \quad (1)$$

where $\text{Pa}(Y)$ is the set of parents of variable Y in the network. In our case we wish to model the expression level of a gene conditional on a set of parent genes that have some influence on it. To learn the set of parent variables of a given gene, it is possible to perform variable selection in a Markov Chain Monte Carlo framework, proposing to add or remove genes to the parent set in a Metropolis-Hastings sampler. However this can be computationally expensive, and so instead we consider applying a sparse regression approach to learn a set of parents for each gene.

2.2 Sparse negative binomial regression

Given data D consisting of M columns and L rows, with columns corresponding to genes and rows to time points, we seek to learn a parent set for each gene. To do so we can employ a regularised regression approach that enforces sparsity of the regression coefficients, and only take predictors (genes) whose coefficients are significantly larger than zero as parents. To simplify the presentation, below we consider the regression of the counts for a single gene i , $y = D_{2:L,i}$, conditional on the counts of the remaining $W = M - 1$ genes $X = D_{1:(L-1),-i}$. The matrix X is supplemented with a column vector $\mathbf{1}$. Where there are multiple replicates for each time point these can be adjusted appropriately.

The counts y_t are then modelled as following a negative binomial distribution with mean $\exp(X\beta)_t$ and dispersion ω , where β is a vector of regression coefficients. The model is then

$$y_t \sim \text{NB}(s_t \exp(X\beta)_t, \omega), \quad (2)$$

where we have applied the NB2 formulation of the negative binomial distribution, and s_t is a scaling factor for each sample to account for sequencing depth. This corresponds to a DBN where

$$p(D_t^i | D_{t-1}^{\text{Pa}(i)}) \sim \text{NB} \left(s_t \exp \left(\sum_{j \in \text{Pa}(i)} \beta_j D_{t-1}^j + \beta_c \right), \omega \right). \quad (3)$$

To enforce sparsity of the β_w we apply a horseshoe prior [6, 7], assuming that $\beta_w \sim \mathcal{N}(0, \zeta_w^2)$, and placing a half-Cauchy prior on the ζ_w^2 ,

$$\beta_w \sim \mathcal{N}(0, \zeta_w^2) \quad (4)$$

$$\zeta_w \sim \text{C}^+(0, \tau). \quad (5)$$

Then as in Carvalho et al. [7] we set a prior on τ that allows the degree of shrinkage to be learnt from the data

$$\tau \sim \text{C}^+(0, \sigma) \quad (6)$$

$$p(\sigma^2) \propto \frac{1}{\sigma^2}. \quad (7)$$

Finally we place a gamma prior on the dispersion parameter ω . This gives a joint probability

$$p(y, X, \theta) = \prod_i p(y_i | X, \beta, \omega) p(\omega | \sigma) \prod_w p(\beta_w | \zeta_w^2) p(\zeta_w^2 | \tau) p(\tau | \sigma) p(\sigma^2) \quad (8)$$

We now apply a variational inference [28, 29, 5, 4, 36] scheme to learn approximate posterior distributions over the model parameters. Unfortunately in this model the optimal distribution \hat{q} for the regression coefficients β_w does not have a tractable solution. However following Luts and Wand [27] we can sidestep this problem by applying non-conjugate variational message passing [22], and we can then derive approximate posterior distributions for each of the model parameters following a straightforward parameter update scheme. The full set of variational updates are given in appendix A.

3 Results

3.1 Synthetic data

We apply our method to the task of inferring directed and undirected networks from 10 simulated gene expression time series. The time series were generated by utilising the GeneNetWeaver [43] software to first generate subnetworks

representative of the structure of the *Saccharomyces cerevisiae* gene regulatory network, and then simulate the dynamics of the networks. Subnetworks of 50 nodes were generated and used to simulate 21 time points with 10 replicates, as in the DREAM4 challenge[39, 31, 32]. Simulated gene expression levels were then transformed by the inverse cumulative distribution function (CDF) F^{-1} of a normal distribution with mean and variance estimated from the data, and subsequently by the CDF of a negative binomial distribution G . The distributions G for each variable were chosen with mean and dispersion parameters randomly sampled from the empirically estimated means and dispersions of each gene from a publicly available RNA-seq count data set from the recount2 database[8] (accession *ERP003613*) consisting of 171 samples from 25 tissues in humans[10]. This was done so as to simulate the observed distributions of RNA-Seq counts in a real world data set.

We compare our approach against the Lasso as implemented in the `lars` R package[17] and the regularised Poisson regression method implemented in the `glmnet` R package[12]. The degree of regularisation was selected using cross validation as implemented in the respective software packages. For the undirected network inference case we also consider networks inferred by the mutual information based approaches implemented in the `minet` R package[35], specifically the CLR[11], ARACNE[33] and MRNET[34] methods.

In figure 1 we show the area under the receiver operating curve (AUC-ROC) and area under the precision recall curve (AUC-PR), as calculated by the `PRROC` R package[14], and Matthews Correlation Coefficient (MCC), for the various methods to be benchmarked in the undirected case. For the MCC, edges were predicted as those where zero was not contained in the 95% credible interval of the corresponding regression coefficients, and for the Lasso and `glmnet` methods, non-zero coefficients were taken as predicted edges. In the undirected case, we ignore the directionality of the edges in the true network generated by the GeneNetWeaver software, and take the strongest prediction in either direction to be the given score for an edge. When benchmarking the methods we found that there was variance in performance depending on the empirical means and dispersions of the negative binomial distributions that were sampled from the RNA-Seq data. Therefore we repeated benchmarking on each network 5 times with resampled negative binomial means and dispersions to take this into account. Running time for our algorithm was under 10 minutes for the 50 node networks considered.

For the undirected case, our method shows an improved performance over the competing methods in terms of the AUC-PR, and also in terms of the MCC. Although the distinction between the approaches is less marked for the AUC-ROC, this is to be expected as the simulated biological network structures have far fewer (< 10%) true positives than true negatives, a situation in which the AUC-ROC does not distinguish performance as well as AUC-PR [9, 14].

As can be seen in figure 2 performance is slightly reduced in the directed case, likely due to the extra complexity of correctly predicting the directionality of edges. However our approach still clearly outperforms the competing methods in terms of the AUC-PR and MCC.

It is clear that, as might be expected, taking into account the distributional properties observed in RNA-Seq data improves on the performance of methods based on assumptions that do not hold for RNA-Seq count data. We also see improvements over the nonparametric mutual information based ap-

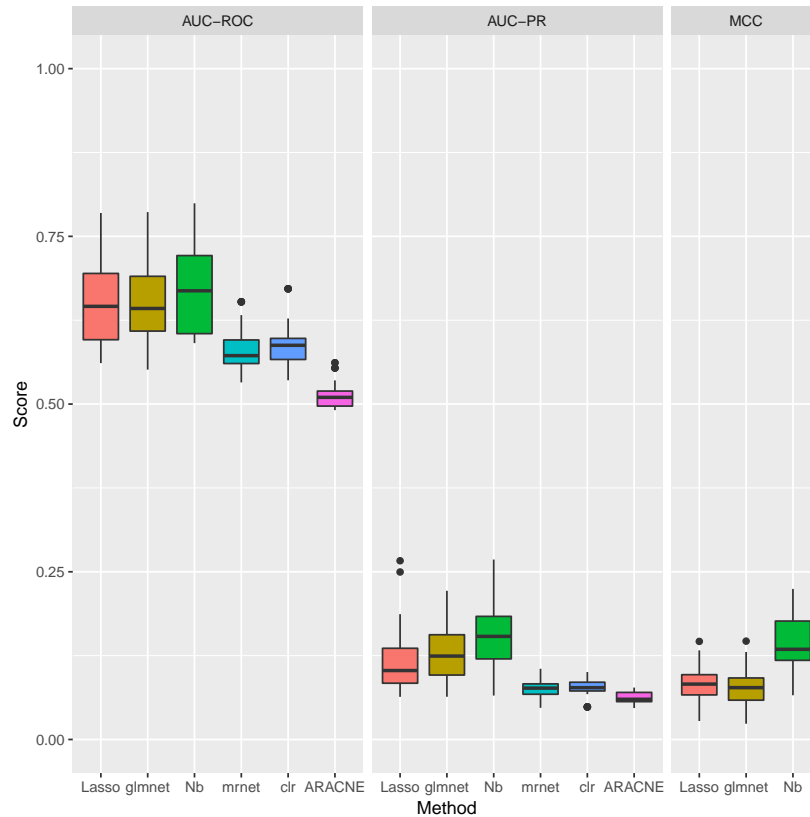


Figure 1: Boxplots of AUC-ROC, AUC-PR, and MCC for our method (Nb) and the competing methods benchmarked when learning undirected networks from synthetic data, for 10 subnetworks sampled from the *S. cerevisiae* gene regulatory network.

proaches when learning undirected networks, perhaps due to our method taking into account the sequential nature of the data, and the constraint of assuming negative binomially distributed observations reducing the complexity of the inference problem.

4 Neural progenitor cell differentiation

To illustrate an application of our model to a real world RNA-Seq data set, we consider a publicly available RNA-Seq time course data set available from the recount2 database[8], accession *SRP041159*. The data consist of RNA-Seq counts from neuronal stem cells for 3 replicates over 7 time points after the induction of neuronal differentiation[42]. To select a subset of genes to analyse we performed a differential expression test between time points using the DESeq2 R package[26], and selected the 25 genes with the largest median fold-change between time points that were also differentially expressed between all time points. We chose to consider only an undirected network, as our benchmarking

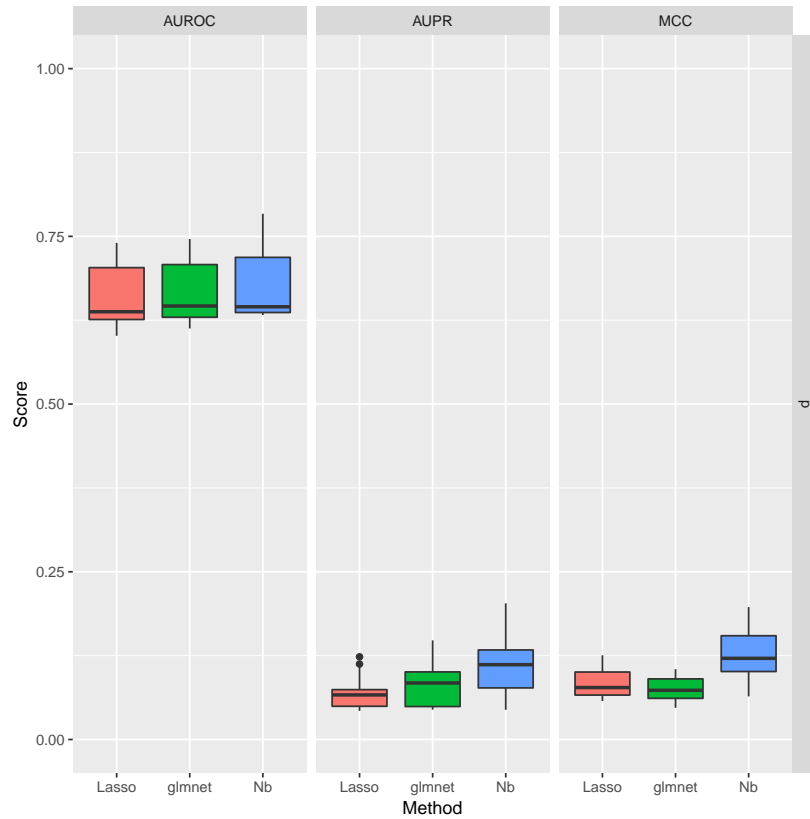


Figure 2: Boxplots of AUC-ROC, AUC-PR, and MCC for our method (Nb) and the methods benchmarked when learning directed networks from synthetic data, for 10 subnetworks sampled from the *S. cerevisiae* gene regulatory network.

shows that such networks are likely more reliable.

Applying our method produced the network shown in figure 3, where it can be seen that there is a strongly interconnected network with four genes (MCUR1, PARP12, COL17A1, CDON) acting as hubs, suggesting these genes may be important in neuronal differentiation. Of these genes, CDON has been shown to promote neuronal differentiation through the activation of p38MAPK pathway[48, 37] and inhibition of Wnt signalling[20], whilst MCUR1 is known to bind to MCU[30], that in turn has been shown to influence neuronal differentiation[40].

5 Conclusions

We have developed a fast and robust methodology for the inference of gene regulatory networks from RNA-Seq data that specifically models the observed count data as being negative binomially distributed. Our approach outperforms other sparse regression and mutual information based methods in learning directed and undirected networks from time series data.

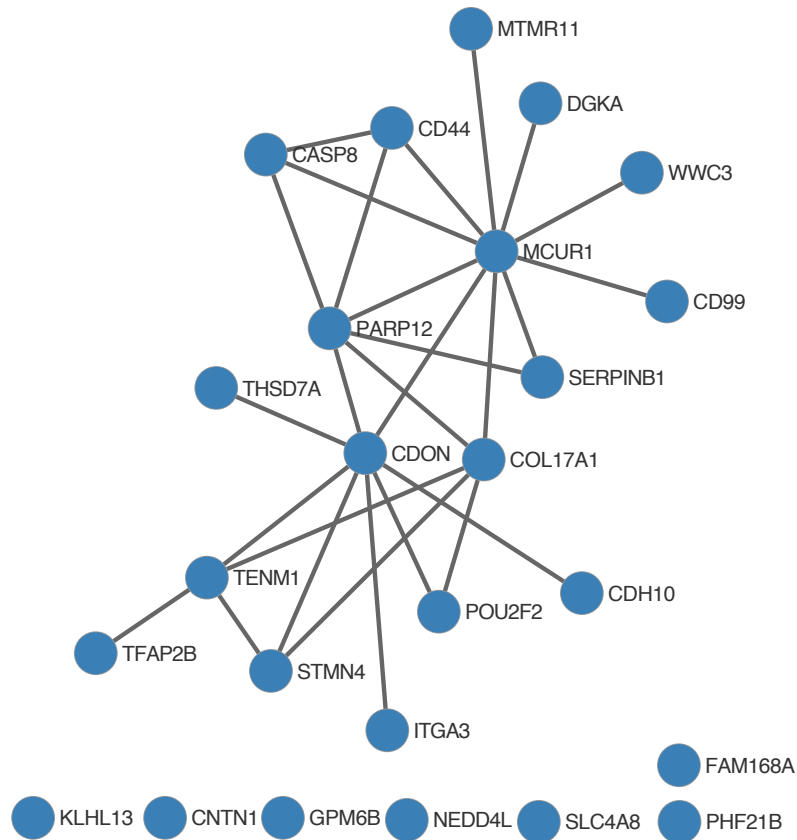


Figure 3: Undirected network inferred from the human neuronal differentiation time series data set.

Another approach to network inference from RNA-Seq data could be to further develop mutual information based methodologies with this specific problem in mind. Mutual information based methods have the benefit of being independent of any specific model of the distribution of the data, and so could help sidestep issues in parametric modelling of RNA-Seq data. However this comes at the cost of abandoning the simplifying assumptions that are made by applying a parametric model that provides a reasonable fit to the data, and presents challenges of its own in the reliable estimation of the mutual information between random variables.

Acknowledgement

T.T acknowledges the Edmond J. Safra Foundation and Lily Safra.

A Variational inference

From the results in Luts and Wand [27] the model can be written as a Poisson-Gamma mixture, so that

$$p(y_t|\lambda_t) \sim \text{Pois}(\lambda_t) \quad (9)$$

$$p(\lambda_t|x_t, \beta, \omega) \sim \text{Gamma}(\omega, \omega \exp[-X\beta]) \quad (10)$$

and the horseshoe prior on β represented using a mixture of inverse gamma distributions,

$$p(\beta_w|\zeta_w^2) \sim \mathcal{N}(0, \zeta_w^2) \quad (11)$$

$$p(\zeta_w^2|a_w) \sim \text{InvGamma}\left(\frac{1}{2}, \frac{1}{a_w}\right) \quad (12)$$

$$p(a_w|\tau^2) \sim \text{InvGamma}\left(\frac{1}{2}, \frac{1}{\tau^2}\right) \quad (13)$$

$$p(\tau^2|b) \sim \text{InvGamma}\left(\frac{1}{2}, \frac{1}{b}\right) \quad (14)$$

$$p(b|\sigma^2) \sim \text{InvGamma}\left(\frac{1}{2}, \frac{1}{\sigma^2}\right). \quad (15)$$

A.1 Mean field approximation

The mean field approximation of the posterior is then

$$\begin{aligned} \prod_i p(y_i|\lambda_i)p(\lambda_i|X_i, \beta, \omega)p(\omega) \prod_w p(\beta_w|\zeta_w^2)p(\zeta_w^2)p(\zeta_w^2|\tau)p(\tau|\sigma^2)p(\sigma^2) \\ \approx \prod_i [q(\lambda_i)]q(\beta)q(\omega) \prod_w [q(\zeta_w^2)q(a_w)]q(\tau^2)q(b). \end{aligned} \quad (16)$$

The variational updates for λ_t can be derived as

$$\begin{aligned} \log \hat{q}(\lambda_t) &= \mathbb{E}_q [\log p(y_t|\lambda_t)p(\lambda_t|X_t, \beta, \omega)] + \text{const.} \\ &= \mathbb{E}_q \left[\log \frac{\lambda_t^{y_t} e^{-\lambda_t}}{y_t!} \frac{(\omega \exp(-X_t\beta))^\omega \lambda_t^{\omega-1} e^{-\lambda_t \omega \exp(-X_t\beta)}}{\Gamma(\omega)} \right] + \text{const.} \\ &= \mathbb{E}_q [y_t \log \lambda_t - \lambda_t + (\omega - 1) \log \lambda_t - \lambda_t \omega \exp(-X_t\beta)] + \text{const.} \\ \hat{q}(\lambda_t) &\sim \text{Gamma}(y_t + \mathbb{E}_q[\omega], 1 + \mathbb{E}_q[\omega]\mathbb{E}_q[\exp(-X_t\beta)]) \end{aligned} \quad (17)$$

and due to the properties of the log-normal distribution

$$\mathbb{E}_q[\exp(-X_t\beta)] = \exp(-X_t\mathbb{E}[\beta] + \frac{1}{2}X_t\Sigma X_t^T), \quad (19)$$

where Σ is the covariance matrix of β under \hat{q} .

As derived in Luts and Wand [27], applying non-conjugate variational message passing, $\hat{q}(\beta) \sim \mathcal{N}(\mu, \Sigma)$ and the variational update for β is

$$w = \exp(-X\mu + \frac{1}{2}\text{diagonal}(X\Sigma X^T)) \quad (20)$$

$$\Sigma = [\omega X^T \text{diag}(\mathbb{E}[\lambda] \cdot w)X + M]^{-1} \quad (21)$$

$$M = \text{diag}(\mathbb{E}[\frac{1}{\sigma_w^2}]) \quad (22)$$

$$\mu = \mu + \Sigma [\omega X^T (\mathbb{E}[\lambda] \cdot w - 1) - M\mu], \quad (23)$$

and for the dispersion ω we apply numerical integration as described in Luts and Wand [27].

Then for the horseshoe prior on β , the variational updates are

$$\begin{aligned} \log \hat{q}(\zeta_w^2) &= \mathbb{E}_q [\log p(\beta_w | \zeta_w^2) p(\zeta_w^2)] + \text{const.} \\ &= \mathbb{E}_q \left[-\frac{1}{2} \log \zeta_w^2 - \frac{\beta_w^2}{2\zeta_w^2} + (-\alpha - 1) \log \zeta_w^2 - \frac{\gamma}{\zeta_w^2} \right] + \text{const.} \end{aligned} \quad (24)$$

$$\hat{q}(\zeta_w^2) \sim \text{InvGamma} \left(1, \frac{1}{2} \mathbb{E}[\beta_w^2] + \mathbb{E}_q[a_w] \right) \quad (25)$$

$$\begin{aligned} \log \hat{q}(a_w) &= \mathbb{E}_q [\log p(\zeta_w^2 | a_w) p(a_w | \tau^2)] + \text{const.} \\ &= \mathbb{E}_q \left[-\frac{1}{a_w \zeta_w^2} - \frac{1}{2} \log a_w - \frac{3}{2} \log a_w - \frac{1}{\tau^2 a_w} \right] + \text{const.} \end{aligned} \quad (26)$$

$$\hat{q}(a_w) \sim \text{InvGamma} \left(1, \mathbb{E}_q \left[\frac{1}{\zeta_w^2} \right] + \mathbb{E}_q \left[\frac{1}{\tau^2} \right] \right) \quad (27)$$

$$\begin{aligned} \log \hat{q}(\tau^2) &= \mathbb{E}_q \left[\sum_w \log p(a_w | \tau^2) + \log p(\tau^2 | b) \right] + \text{const.} \\ &= \mathbb{E}_q \left[-\sum_w \left(\frac{1}{2} \log \tau^2 + \frac{1}{a_w \tau^2} \right) - \frac{3}{2} \log \tau^2 - \frac{1}{b \tau^2} \right] + \text{const.} \end{aligned} \quad (28)$$

$$\hat{q}(\tau^2) \sim \text{InvGamma} \left(\frac{1}{2} + \frac{W}{2}, \mathbb{E}_q \left[\frac{1}{b} \right] + \sum_w \mathbb{E}_q \left[\frac{1}{a_w} \right] \right) \quad (29)$$

$$\begin{aligned} \log \hat{q}(b) &= \mathbb{E}_q [\log p(\tau^2 | b) p(b | \sigma^2)] + \text{const.} \\ &= \mathbb{E}_q \left[-\frac{1}{2} \log b - \frac{1}{\tau^2 b} - \frac{3}{2} \log b - \frac{1}{\sigma^2 b} \right] + \text{const.} \end{aligned} \quad (30)$$

$$\hat{q}(b) \sim \text{InvGamma} \left(1, \mathbb{E}_q \left[\frac{1}{\tau^2} \right] + \mathbb{E}_q \left[\frac{1}{\sigma^2} \right] \right) \quad (31)$$

$$\begin{aligned} \log \hat{q}(\sigma^2) &= \mathbb{E}_q [\log p(b | \sigma^2) p(\sigma^2)] + \text{const.} \\ &= \mathbb{E}_q \left[-\frac{1}{2} \log \sigma^2 - \frac{1}{b \sigma^2} - \log \sigma^2 \right] + \text{const.} \end{aligned} \quad (32)$$

$$\hat{q}(\sigma^2) \sim \text{InvGamma} \left(\frac{1}{2}, \mathbb{E}_q \left[\frac{1}{b} \right] \right). \quad (33)$$

References

- [1] T. Äijö, V. Butty, Z. Chen, V. Salo, S. Tripathi, C. B. Burge, R. Lahesmaa, and H. Lähdesmäki. Methods for time series analysis of RNA-seq data with application to human Th17 cell differentiation. *Bioinformatics*, 30(12):i113–i120, June 2014.
- [2] G. I. Allen and Z. Liu. A Local Poisson Graphical Model for inferring networks from sequencing data. *NanoBioscience, IEEE Transactions on*, 12(3):189–198, Aug. 2013.
- [3] S. Anders and W. Huber. Differential expression analysis for sequence count data. *Genome biology*, 11(10):R106, Oct. 2010.
- [4] D. Barber. *Bayesian Reasoning and Machine Learning*. Cambridge University Press, Feb. 2012.
- [5] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer Verlag, Aug. 2006.
- [6] C. M. Carvalho, N. G. Polson, and J. G. Scott. Handling Sparsity via the Horseshoe. *AISTATS*, 2009.
- [7] C. M. Carvalho, N. G. Polson, and J. G. Scott. The horseshoe estimator for sparse signals. *Biometrika*, 97(2):465–480, June 2010.
- [8] L. Collado-Torres, A. Nellore, K. Kammers, S. E. Ellis, M. A. Taub, K. D. Hansen, A. E. Jaffe, B. Langmead, and J. T. Leek. Reproducible RNA-seq analysis using recount2. *Nature biotechnology*, 35(4):319–321, Apr. 2017.
- [9] J. Davis and M. Goadrich. *The relationship between Precision-Recall and ROC curves*. ACM, New York, New York, USA, June 2006.
- [10] L. Fagerberg, B. M. Hallström, P. Oksvold, C. Kampf, D. Djureinovic, J. Odeberg, M. Habuka, S. Tahmasebpoor, A. Danielsson, K. Edlund, A. Asplund, E. Sjöstedt, E. Lundberg, C. A.-K. Szgyarto, M. Skogs, J. O. Takanen, H. Berling, H. Tegel, J. Mulder, P. Nilsson, J. M. Schwenk, C. Lindskog, F. Danielsson, A. Mardinoglu, Å. Sivertsson, K. von Feilitzen, M. Forsberg, M. Zwahlen, I. Olsson, S. Navani, M. Huss, J. Nielsen, F. Pontén, and M. Uhlén. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Molecular & Cellular Proteomics*, 13(2):397–406, Feb. 2014.
- [11] J. J. Faith, B. Hayete, J. T. Thaden, I. Mogno, J. Wierzbowski, G. Cottarel, S. Kasif, J. J. Collins, and T. S. Gardner. Large-Scale Mapping and Validation of Escherichia coli Transcriptional Regulation from a Compendium of Expression Profiles. *PLoS Biology*, 5(1):e8, Jan. 2007.
- [12] J. Friedman, T. Hastie, and R. Tibshirani. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of statistical software*, 33(1):1–22, 2010.
- [13] M. Gallopin, A. Rau, and F. Jaffrézic. A hierarchical poisson log-normal model for network inference from RNA sequencing data. *PLoS ONE*, 8(10):e77503–e77503, Jan. 2013.

- [14] J. Grau, I. Grosse, and J. Keilwagen. PRROC: computing and visualizing precision-recall and receiver operating characteristic curves in R. *Bioinformatics*, 31(15):2595–2597, Aug. 2015.
- [15] M. Grzegorzczak and D. Husmeier. Improvements in the reconstruction of time-varying gene regulatory networks: dynamic programming and regularization by information sharing among genes. *Bioinformatics*, 27(5):693–699, Mar. 2011.
- [16] T. J. Hardcastle and K. A. Kelly. baySeq: Empirical Bayesian methods for identifying differential expression in sequence count data. *BMC Bioinformatics*, 11(1):422, Aug. 2010.
- [17] T. Hastie and B. Efron. *lars: Least Angle Regression, Lasso and Forward Stagewise*, 2013. URL <https://CRAN.R-project.org/package=lars>. R package version 1.2.
- [18] D. Husmeier and A. V. Werhli. Bayesian integration of biological prior knowledge into the reconstruction of gene regulatory networks with Bayesian networks. *Computational systems bioinformatics / Life Sciences Society. Computational Systems Bioinformatics Conference*, 6:85–95, Feb. 2007.
- [19] D. I. Inouye, E. Yang, G. I. Allen, and P. Ravikumar. A review of multivariate distributions for count data derived from the Poisson distribution. *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(3):e1398, Mar. 2017.
- [20] M.-H. Jeong, S.-M. Ho, T. A. Vuong, S.-B. Jo, G. Liu, S. A. Aaronson, Y.-E. Leem, and J.-S. Kang. Cdo suppresses canonical Wnt signalling via interaction with Lrp6 thereby promoting neuronal differentiation. *Nature Communications*, 5:ncomms6455, Nov. 2014.
- [21] K. Jo, H.-B. Kwon, and S. Kim. Time-series RNA-seq analysis package (TRAP) and its application to the analysis of rice, *Oryza sativa* L. ssp. Japonica, upon drought stress. *Methods*, 67(3):364–372, June 2014.
- [22] D. A. Knowles and T. Minka. Non-conjugate Variational Message Passing for Multinomial and Binary Regression. In *Proceedings of the th International Conference on Neural Information Processing Systems*, pages 1701–1709, 2011.
- [23] D. Koller and N. Friedman. *Probabilistic Graphical Models*. MIT Press, Cambridge, MA.
- [24] S. Lèbre. *Inferring dynamic Bayesian network with low order independencies*. *Statistical Applications in Genetics and Molecular Biology*, 8(1), 2009.
- [25] S. Lèbre, J. Becq, F. Devaux, M. P. Stumpf, and G. Lelandais. Statistical inference of the time-varying structure of gene-regulation networks. *BMC Systems Biology*, 4, 2010.

- [26] M. I. Love, W. Huber, and S. Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, 15(12):550, Dec. 2014.
- [27] J. Luts and M. P. Wand. Variational Inference for Count Response Semi-parametric Regression. *Bayesian Analysis*, 10(4):991–1023, Dec. 2015.
- [28] D. J. C. MacKay. Developments in Probabilistic Modelling with Neural Networks — Ensemble Learning. In *Machine Learning*, pages 191–198. Springer London, London, 1995.
- [29] D. J. C. MacKay. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, Sept. 2003.
- [30] K. Mallilankaraman, C. Cárdenas, P. J. Doonan, H. C. Chandramoorthy, K. M. Irrinki, T. Golenár, G. Csordás, P. Madireddi, J. Yang, M. Müller, R. Miller, J. E. Kolesar, J. Molgó, B. Kaufman, G. Hajnóczky, J. K. Foskett, and M. Madesh. MCUR1 is an essential component of mitochondrial Ca²⁺ uptake that regulates cellular metabolism. *Nature Cell Biology*, 14(12):1336–1343, Nov. 2012.
- [31] D. Marbach, T. Schaffter, C. Mattiussi, and D. Floreano. Generating realistic in silico gene networks for performance assessment of reverse engineering methods. *Journal of Computational Biology*, 16(2):229–239, Feb. 2009.
- [32] D. Marbach, R. J. Prill, T. Schaffter, C. Mattiussi, D. Floreano, and G. Stolovitzky. Revealing strengths and weaknesses of methods for gene network inference. *Proceedings of the National Academy of Sciences*, 107(14):6286–6291, Apr. 2010.
- [33] A. A. Margolin, I. Nemenman, K. Basso, C. Wiggins, G. Stolovitzky, R. D. Favera, and A. Califano. ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context. *BMC Bioinformatics*, 7(1):S7, Mar. 2006.
- [34] P. E. Meyer, K. Kontos, F. Lafitte, and G. Bontempi. Information-Theoretic Inference of Large Transcriptional Regulatory Networks. *EURASIP Journal on Bioinformatics and Systems Biology*, 2007(1):79879–9, 2007.
- [35] P. E. Meyer, F. Lafitte, and G. Bontempi. minet: A R/Bioconductor Package for Inferring Large Transcriptional Networks Using Mutual Information. *BMC Bioinformatics*, 9(1):461, 2008.
- [36] K. P. Murphy. *Machine Learning A Probabilistic Perspective*. MIT Press, Sept. 2012.
- [37] J.-E. Oh, G.-U. Bae, Y.-J. Yang, M.-J. Yi, H.-J. Lee, B.-G. Kim, R. S. Krauss, and J.-S. Kang. Cdo promotes neuronal differentiation via activation of the p38 mitogen-activated protein kinase pathway. *The FASEB Journal*, 23(7):2088–2099, July 2009.

- [38] R. Opgen-Rhein and K. Strimmer. From correlation to causation networks: a simple approximate learning algorithm and its application to high-dimensional plant gene expression data. *BMC Systems Biology*, 1, Aug. 2007.
- [39] R. J. Prill, D. Marbach, J. Saez-Rodriguez, P. K. Sorger, L. G. Alexopoulos, X. Xue, N. D. Clarke, G. Altan-Bonnet, and G. Stolovitzky. Towards a Rigorous Assessment of Systems Biology Models: The DREAM3 Challenges. *PLoS ONE*, 5(2):e9202, Feb. 2010.
- [40] T. Rharass, H. Lemcke, M. Lantow, S. A. Kuznetsov, D. G. Weiss, and D. Panáková. Ca²⁺-mediated mitochondrial reactive oxygen species metabolism augments Wnt/ β -catenin pathway activation to facilitate cell differentiation. *Journal of Biological Chemistry*, 289(40):27937–27951, Oct. 2014.
- [41] M. D. Robinson, D. J. McCarthy, and G. K. Smyth. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1):139–140, Jan. 2010.
- [42] M. Sauvageau, L. A. Goff, S. Lodato, B. Bonev, A. F. Groff, C. Gerhardinger, D. B. Sanchez-Gomez, E. Hacisuleyman, E. Li, M. Spence, S. C. Liapis, W. Mallard, M. Morse, M. R. Swerdel, M. F. D Ecclessis, J. C. Moore, V. Lai, G. Gong, G. D. Yancopoulos, D. Frendewey, M. Kellis, R. P. Hart, D. M. Valenzuela, P. Arlotta, and J. L. Rinn. Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *eLife*, 2:360, Dec. 2013.
- [43] T. Schaffter, D. Marbach, and D. Floreano. GeneNetWeaver: in silico benchmark generation and performance profiling of network inference methods. *Bioinformatics*, 27(16):2263–2270, Aug. 2011.
- [44] T. Thorne and M. P. H. Stumpf. Inference of temporally varying Bayesian networks. *Bioinformatics*, 28(24):3298–3305, Dec. 2012.
- [45] T. Thorne, P. Fratta, M. G. Hanna, A. Cortese, V. Plagnol, E. M. Fisher, and M. P. H. Stumpf. Graphical modelling of molecular networks underlying sporadic inclusion body myositis. *Mol. BioSyst.*, 9(7):1736–1742, June 2013.
- [46] T. Wang, Z. Ren, Y. Ding, Z. Fang, Z. Sun, M. L. MacDonald, R. A. Sweet, J. Wang, and W. Chen. FastGGM: An Efficient Algorithm for the Inference of Gaussian Graphical Model in Biological Networks. *PLoS Computational Biology*, 12(2):e1004755, Feb. 2016.
- [47] A. V. Werhli, M. Grzegorzcyk, and D. Husmeier. Comparative evaluation of reverse engineering gene regulatory networks with relevance networks, graphical Gaussian models and Bayesian networks. *Bioinformatics*, 22(20):2523–2531, Oct. 2006.
- [48] W. Zhang, M.-J. Yi, X. Chen, F. Cole, R. S. Krauss, and J.-S. Kang. Cortical thinning and hydrocephalus in mice lacking the immunoglobulin superfamily member CDO. *Molecular and Cellular Biology*, 26(10):3764–3772, May 2006.