Gene Interaction Networks using the Gibbs Sampling Technique

A gene interaction network (GIN) in this method is the same as the one defined in Toyoshiba et.al. [1]. Briefly, a GIN consists of a collection of P genes, denoted G_1, G_2,… G_P, with observed values g_1,g_2,….g_p. Define β_{ij} (i,j=1,2,…, P) as parameters in the log-linear function form describing the linkage from gene *i* to gene *j*. Mathematically, this is written as

$$E[\log(G_j)] = \sum_{i=1,\neq j}^{P} I_{ij}\beta_{ij}\log(g_i) \qquad (1)$$

where $E[\log(G_j)]$ represents the expectation for the natural logarithm of G_j and I_{ij} (i,j = 1,2, … ,P) is an indicator function that equals 1 if gene G_i has a link to gene G_j, otherwise it equals 0. If a gene has a regulatory effect on gene G_i, then that gene is referred to as a "Parent of gene G_i" and we refer to it as belonging to the set $Pa(G_i)$.

Define ***T*** to be the transition matrix whose (i, j) element consists of the indicator function I_{ij} and let ***B*** represents the matrix whose (i, j) elements consists of β_{ij}

Prior distributions

In the log-linear model, the parameters are the variance for the distribution for each gene, denoted σ_j^2, parameters in the log-linear form, β_{ij} (i,j=1,2,..P, $i\neq j$), and the

indicator functions defining the network, I_{ij}. Prior distributions can be assumed for each parameter using either informative or uninformative priors. Let $h(\sigma_j^2)$ denote the prior distribution for σ_j^2. Define $h(\beta_{ij}|I_{ij}=1)$ to be the prior distribution for β_{ij} when, in a given network structure, $I_{ij}=1$ (*i.e.* there is a linkage from G_i to G_j). Similarly, define $h(\beta_{ij}|I_{ij}=0)$. Finally, define $h(I_{ij})$ to be the prior distribution for I_{ij}, $h(I_{ij})$ is a discrete distribution taking on only two values, 0 or 1.

In what follows, the prior distribution for the variance, $h(\sigma_j^2)$, is assumed to be inverse Gamma with parameters a_1 and a_2 (i.e. $h(\sigma_j^2)$ is $inv\Gamma(a_1/2,a_2/2)$). For β_{ij}, we assume $h(\beta_{ij}|I_{ij})$ is normal with mean 0 and variance $\sigma(\beta)^2(I_{ij}+(1-I_{ij})*C)^2$ (*i.e.* $h(\beta_{ij}|I_{ij}=1)$ is $N(0, \sigma(\beta)^2(I_{ij}+(1-I_{ij})*C)^2)$) where C is some constant to keep β_{ij} close to zero when $I_{ij}=0$. This is the key to initially guide the Gibbs sampler in choosing between networks with $I_{ij}=1$ and those with $I_{ij}=0$. In general, denoting $\boldsymbol{\beta}_{.j}$ as the j^{th} column vector of $\boldsymbol{B}$ with j^{th} element is absent, the prior distribution of $\boldsymbol{\beta}_{.j}$ can be written by P-1 dimension multivariate normal distribution $N_{P-1}(0, \boldsymbol{CC}|\boldsymbol{I}_j)$, where $\boldsymbol{C}$ is a diagonal matrix with $\sigma(\beta)(I_{ij}+(1-I_{ij})*C)$ is the (i, i) elements and $\boldsymbol{I}_j$ is the j^{th} column vector of $\boldsymbol{T}$ with the j^{th} element is absent. The prior distribution for I_{ij} was assumed to be a Bernoulli distribution with success ($I_{ij}=1$) probability p_{ij}. In the uninformative case, p_{ij} could be set to 0.5 and if there is some expectation that I_{ij} is not equal to zero, the prior

probability could be set higher or the correlation information between some linkages could be possible as well in more advance.

With these prior distributions, and assuming that the natural log of G_j follows a normal distribution with mean $\sum_{i=1,\neq j}^{P} \beta_{ij} \cdot \log(g_i)$ and standard deviation σ_j, posterior distributions for each parameter can be estimated. It should be noted that the mean function $\sum_{i=1,\neq j}^{P} \beta_{ij} \cdot \log(g_i)$ is slightly different from the form shown in Equation (1). In the algorithm used here, β_{ij} is sampled even if I_{ij}=0. Hence, in the Gibbs sampling application, the summation range will be all genes except for gene G_j itself. The development presented here parallels the methods developed by George and McCulloch [2] using Gibbs sampling to select models for linear regression. Other prior setting had been studied and detail comparison had been made bythe other researchers [3,4,5]. In our case, the model selection for linear regression will be extended to the model selection for Bayesian networks. Hence, some improvement to overcome the acyclic condition on Bayesian network will be needed to extend the method and that will be given in the below.

Posterior distributions

Let $\boldsymbol{g} = [\boldsymbol{g_1},\boldsymbol{g_2},\ldots,\boldsymbol{g_P}]$ represent an n x P matrix containing the gene expression data after normalization and taking log-scale, where $\boldsymbol{n}$ is the sample size and $\boldsymbol{g_i}$ is a vector of n observations from G_i, and define the complementation set $\boldsymbol{g(i)}= [\boldsymbol{g_1},\boldsymbol{g_2}..\boldsymbol{g_{i-1}},\boldsymbol{g_{i+1}}\ldots,\boldsymbol{g_P}]$. The prior distribution assumed in the previous, the conditional posterior distribution for σ_j^2 is given by

$$\sigma_j^{\,2} \sim f(\sigma_j^{\,2} | \boldsymbol{g(j)},\beta_{\cdot j},\boldsymbol{I}_{\cdot j}) = Inv\Gamma\left(\frac{n + a_1}{2}, \frac{|\,\boldsymbol{g}_j - \boldsymbol{g(j)} \cdot \beta_{\cdot j}\,| + a_2}{2}\right) \qquad (2)$$

where $\boldsymbol{\beta}_{\cdot j}$ is the same as the above. In this equation, a_1 and a_2 could be chosen as zero for the non-informative prior. For $\boldsymbol{\beta}_{\cdot j}$, the conditional posterior is given by

$$\boldsymbol{\beta}_{\cdot j} \sim f(\boldsymbol{\beta}_{\cdot j} | \boldsymbol{g}, \sigma_j, \boldsymbol{I}_{\cdot j}) = N_P(\boldsymbol{A} \cdot \sigma_j^{-2} \cdot \boldsymbol{g}(j)^t \boldsymbol{g}_j, \boldsymbol{A}) \qquad (3)$$

where the matrix $\boldsymbol{A}$ is defined by its inverse as

$$\boldsymbol{A}^{-1} = \sigma_j^{-2} \cdot \boldsymbol{g}(j)^t \cdot \boldsymbol{g}(j) + (\boldsymbol{CRC}\quad)^{-1} \qquad (4)$$

where $\boldsymbol{R}$ could be $\boldsymbol{g}(\mathrm{j})^{\mathrm{t}}*\boldsymbol{g}(\mathrm{j})$, or P-1 dimension identical matrix, $\boldsymbol{E}$(P-1) and $\boldsymbol{C}$ is the same as above used in the covariance matrix for $\boldsymbol{\beta}_{\cdot j}$. Finally, I_{ij} is obtained by sampling consecutively from the conditional distribution,

$$I_{ij} \sim f(I_{ij} | \boldsymbol{g}, \beta_{\cdot j}, \sigma_j, \boldsymbol{T}/I_{ij}) = f(I_{ij} | \beta_{\cdot j}, \sigma_j, \boldsymbol{T}/I_{ij}), \qquad (6)$$

where $\boldsymbol{T}/I_{ij}$ represent the other all elements of $\boldsymbol{T}$ except for I_{ij}. Here, it should be

noted that the posterior distribution of I_{ij} is conditioned by $\boldsymbol{T}/I_{ij}$ in Bayesian model selection. Because, if I_{ij} =1 makes cycle in a network, the success probability for I_{ij} will be zero. Hence, the following one step is put into the sampling process. Each distribution is Bernoulli with probability

$$P(I_{ij}=1|\beta_{\cdot j},\sigma_j,\boldsymbol{T}/I_{ij})=\frac{a}{a+b}, \tag{7}$$

where a is equal to 0 and b=1 if I_{ij}=1 produces a cyclic in a network, else

$$a=f(\beta_{\cdot j}|\boldsymbol{T}/I_{ij},I_{ij}=1)\times f(\sigma_j|\boldsymbol{T}/I_{ij},I_{ij}=1)\times f(I_{ij}=1) \tag{8}$$

$$b=f(\beta_{\cdot j}|\boldsymbol{T}/I_{ij},I_{ij}=0)\times f(\sigma_j|\boldsymbol{T}/I_{ij},I_{ij}=0)\times f(I_{ij}=0). \tag{9}$$

By equation (2) and the prior distribution assumed for I_{ij}, (8) and (9) can be given by,

$$a=f(\beta_{\cdot j}|\boldsymbol{T}/I_{ij},I_{ij}=1)p_{ij}, \tag{10}$$

$$b=f(\beta_{\cdot j}|T/I_{ij},I_{ij}=0)(1-p_{ij}). \tag{11}$$

Continuing the sampling method for all values of j results in a single observation from the posterior distribution for each parameter. The summary of the posterior distributions for $\boldsymbol{I}_{ij}$s can be given as a matrix and provide an indication of the strongest or weakest possible linkages between genes.

Sampling Algorithm in Bayesian network

Gibbs sampling using the posterior distributions defined above could work in Bayesian network setting. The posterior distribution of the Bayesian network defined by $\boldsymbol{T}$, $\boldsymbol{B}$ and $\boldsymbol{S}$=[σ_1, σ_2,…σ_P] given $\boldsymbol{g}$ = [$\boldsymbol{g_1}$,$\boldsymbol{g_2}$,…,$\boldsymbol{g_P}$] can be written in the following,

$$\begin{aligned} f(T,B,S \mid g) &= f(g \mid T,B,S) * h(T) * h(B \mid T) * h(S) \\ &= \prod_{j=1}^{P} f(g_j \mid I_{\cdot j}, \beta_{\cdot j}, \sigma_j) h(\sigma_j) h(\beta_{\cdot j} \mid I_{\cdot j}) h(I_{\cdot j}) \end{aligned} \quad . \qquad (12)$$

It is well known that the whole posterior distribution over every parameter can be decomposed into sub-networks consisting of each G_j and its parents.

From the decomposition in equation (12), it is obvious that σ_j and $\boldsymbol{\beta}_{\cdot j}$ can be sampled from equation (2) and (3) because the posterior distributions for these parameters are not conditioned by other parameters except for $\boldsymbol{I}_{\cdot j}$ $\boldsymbol{\beta}_{\cdot j}$ $\boldsymbol{\sigma}_j$. In the below, these are called j^{th} column parameters in $\boldsymbol{T}$, $\boldsymbol{B}$, and $\boldsymbol{S}$. In sampling $\boldsymbol{I}_{\cdot j}$, some caution should be needed, because the posterior distribution of $\boldsymbol{I}_{\cdot j}$ is conditioned by $\boldsymbol{T}/\boldsymbol{I}_{\cdot j}$ which is other column parameters in T. This comes from the acyclic condition of Bayesian networks.

Hence, in sampling I_{ij} in $\boldsymbol{I}_{.j}$, if I_{ij}=1 makes a cycle in the network, then the probability for I_{ij}=1 is set to be zero and if I_{ij}=1 does not make cycles, then equation (7) can be obtained by removing out the same terms in equation (12) which is the full posterior distribution and used for sampling I_{ij}. After sampling all $\boldsymbol{I}_{.j}$ elements (I_{ij}, i=1,2,3,…P) then the other column parameters in $\boldsymbol{T}$, $\boldsymbol{B}$ and $\boldsymbol{S}$ will be sampled as the same manner as explained above. This is one step of the Gibbs sampling for Bayesian network.

References

1. Toyoshiba H, Yamanaka T, Sone H, Parham FM, Walker NJ, Martinez J, et al. 2004. Gene interaction network suggests dioxin induces a significant linkage between aryl hydrocarbon receptor and retinoic acid receptor beta. Environ Health Perspect 112(12):1217-1224.
2. George EI, McCulloch. RE. 1993. Variable Selection Via Gibbs Sampling. Journal of the American Statistics Association 88(423):881-889.
3. Dellaportas P, Forster JJ, Ntzoufras I. 2000. Bayesian Variable Selection Using the Gibbs Sampler Generalized Linear Models: A Bayesian Perspective (D. K. Dey, S. Ghosh, and B. Mallick, eds.). New York: Marcel Dekker.271-286.
4. Dellaportas P, Forster JJ, Ntzoufras I. 2002. On Bayesian Model and Variable Selection Using MCMC. Statistics and Computing 12:27-36.
5. Ntzoufras I. 2002. Gibbs Variable Selection Using BUGS. Journal of Statistical Software 7(7):1-19.

.