GeoDiff Description

Lei Yang

October 11, 2021

Abstract

This report describes the models in **GeoDiff**

1 Background

We developed various models in **GeoDiff** for different scenarios based on data generating mechanism of GeoMx DSP. We are mainly concerned about modelling and testing around count data matrix $X = (x_{ij})_{I\times J}$, with samples in columns and features in rows, so x_{ij} is the observed count for the ith feature and jth sample. The features could be negative probes, positive probes or synthetic count for targets with multiple probes. The samples are usually ROIs.

2 Background modeling

2.1 Poisson Background model

The Poisson background model is the simplest model in **GeoDiff**, and it is the foundation of all other models. It takes a form of

$$X_{ij} \sim \text{Poisson}\left(\gamma_i \alpha_{0j}\right) \tag{1}$$

$$\sum_{j=1}^{J} \alpha_{0j} = 1 \tag{2}$$

Where $\gamma_i, i = 1, \cdots, I$ are feature specific factors for I features, and $\boldsymbol{\alpha}_0 = (\alpha_{0j}), j = 1, \cdots, J$ are size factors for J samples. This model assumes, aside from random noise, variation in levels of ith

feature in different ROIs is purely explained by the technical variation in the form of multiplicative size factor α_j for different samples. Therefore, this model is most suitable to features without biological variation and are subject to non-specific probe binding in the background only, thus making it most useful for modeling negative probes.

The model is intrinsic non-identifiable, i.e. different sets of parameters $c\gamma_i, \alpha_{0j}/c$ for any positive constant c yield the same model. To impose identifiability, a constraint like $\sum_{j=1}^{J} \alpha_{0j} = 1$ must be enfored. Constraints are arbitrary. However, $\sum_{j=1}^{J} \alpha_{0j} = 1$ is equivalent to $\sum_{j=1}^{J} X_{ij} \sim$ Poisson (γ_i), making γ_i as the expected value of $\sum_{j=1}^{J} X_{ij}$. This interpretation is convenient in a lot of applications.

2.2 Diagnostics and Remedy

One important metric of any Poisson model is the dispersion, i.e. the ratio of variance vs mean, which equals to 1 theoretically. In Poisson Background model, dispersion of ith feature and jth sample is

$$d_{ij} = \frac{E(x_{ij} - E(x_{ij}))^2}{E(x_{ij})} \equiv 1,$$

which can be estimated by squared Pearson residual at x_{ij}

$$\widehat{d_{ij}} = \frac{(x_{ij} - \widehat{\gamma_i}\widehat{\alpha_{0j}})^2}{\widehat{\gamma_i}\widehat{\alpha_{0j}}}.$$

To reduce variation caused by randomness of a single point, we calculate the empirical (mean) dispersion as a diagnostics metric.

$$\widehat{d} = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J} \frac{(x_{ij} - \widehat{\gamma_i}\widehat{\alpha_{0j}})^2}{\widehat{\gamma_i}\widehat{\alpha_{0j}}}}{IJ}$$

We call the data overdispersed when $\widehat{d}$ is much larger than 1, a rule of thumb cutoff can be 2. Overdispersion is a sign of extra variation not captured by model (1).

Besides empirical dispersion, a ppplot is generated as well. Let $F(x|\lambda)$ be the cumulative distribution function of the Poisson distribution, by model assumption

$$F(x_{ij}|\gamma_i\alpha_{0j}) \overset{iid}{\sim} U(0,1)$$

We calculate

$$F(x_{ij}|\widehat{\gamma_i}\widehat{\alpha_{0j}}),$$

the empirical cumulative probability function values, then sort them from the smallest to the largest, and plot the sorted vector against a even grid over $[0, 1]$ with IJ points, the theoretical cumulative probability function values. In practice, we have found the count data for negative probes could be very low, a decent amount could just be 0, and the ppplot would end up with zigzag shape on the lower end due to that. A way to mitigate that is to simulate data from the fitted model, generate empirical cumulative probability function values, sort them and use the sorted vector as the theoretical cumulative probability function values.

For a good fit, the ppplot is almost a diagonal line from 0 to 1. The most common aberration is an "S" shape, indicating overdispersion.

For biological features, fitting Poisson Background model is inadequate and will lead to overdispersion. More complicated models explained in the following sections are needed to account for their additional variation.

Even the features are only negative probes, overdispersion could still occur.

For data from multiple sources, mean expression level for each probe could be different. A quick check is to fit the alternative model below and recalculate $\hat{d}$ or generate heatmap using γ_{ik}.

$$X_{ij} \sim \text{Poisson}\left(\gamma_{is}\alpha_{0j}\right)$$
$$\sum_{j=1}^{J} \alpha_{0j} = 1$$

where γ_{is} are the mean expression level for ith feature from sth source (batch, slide,... etc.). A much smaller $\hat{d}$ and a clear clustering in the heatmap indicate batch effect.

Overdispersion could also occur due to outliers. We detect outliers by the empirical probability of each point instead of more common Cook distance.

Outliers are mostly large outliers. On the lower end, 0 is unlikely to be determined as outliers in a count matrix for negative probes, of which counts are usually low. After outliers are identified, they can be directly set as missing and Poisson background model could be refitted with missing values. Furthermore, if for ith feature or the jth sample, a lot of counts are outliers, the ith feature or the jth sample should be considered as an outlier, and be removed from the Poisson Background model. Furthermore, the jth sample outlier should be removed universally from all

downstream analysis since the underlying mechanism impact the negative probes could impact other features too.

3 Background Score Test

3.1 Without prior

After applying the Poisson Background Model to the negative probes, it is imperative to test whether a target is above the background use that as a reference. Such test is called background score test.

For kth feature, assuming it has only one probe, and it still follows the Poisson background model

$$X_{kj} \sim \text{Poisson}\left(\gamma_k \alpha_{0j}\right)$$

where the α_{0j} are background size factors estimated from Poisson background model on negative probes

$$X_{ij} \sim \text{Poisson}\left(\gamma_i \alpha_{0j}\right).$$

Let $\gamma_0 = \dfrac{\sum_{i=1}^{I} \gamma_i}{I}$, we are interested in testing

$$H_0 : \ \gamma_k \leq \gamma_0,$$
$$H_a : \ \gamma_k > \gamma_0.$$

Let $\boldsymbol{x}_k$ be the vector of x_{ik}, the observed count for kth feature in ith sample.

$$\log f\left(x_{ik}|\gamma_k\right) = x_{ik} \log\left(\alpha_{0i}\gamma_k\right) - \alpha_{0i}\gamma_k - \log\left(x_{ik}!\right)$$

and the likelihood function

$$L(\gamma_k|\boldsymbol{x}) = \sum_{i=1}^{I} \log f\left(x_{ik}|\gamma_k\right)$$

The score, i.e, the gradient of the log-likelihood function:

$$\frac{dL(\gamma_k|\boldsymbol{x})}{d\gamma_k} = \sum_{i=1}^{I}\left(\frac{x_{ik}}{\gamma_k} - \alpha_{0i}\right) = \frac{\sum_{i=1}^{I} x_i}{\gamma_k} - \sum_{i=1}^{I} \alpha_{0i}$$

The Fisher information

$$\begin{aligned} I\left(\gamma_k\right) &= \operatorname{var}\left(\frac{dL(\gamma_k|\boldsymbol{x})}{d\gamma_k}\right) \\ &= \operatorname{var}\left(\frac{\sum_{i=1}^{I} x_i}{\gamma_k} - \sum_{i=1}^{I} \alpha_{0i}\right) \\ &= \operatorname{var}\left(\frac{\sum_{i=1}^{I} x_i}{\gamma_k}\right) \\ &= \frac{\sum_{i=1}^{I} \operatorname{var}\left(x_i\right)}{\gamma_k^2} = \frac{\sum_{i=1}^{I} \alpha_{0i}\gamma_k}{\gamma_k^2} = \frac{\sum_{i=1}^{I} \alpha_{0i}}{\gamma_k} \end{aligned}$$

the score statistic

$$\begin{aligned} \frac{\left(\frac{dL(\gamma_k|\boldsymbol{x})}{d\gamma_k}\right)^T \frac{dL(\gamma_k|\boldsymbol{x})}{d\gamma_k}}{I\left(\gamma_k\right)} &= \frac{\left(\frac{\sum_{i=1}^{I} x_i}{\gamma_k} - \sum_{i=1}^{I} \alpha_{0i}\right)^2}{\frac{\sum_{i=1}^{I} \alpha_{0i}}{\gamma_k}} \\ &= \left(\frac{\sum_{i=1}^{I} x_i - \sum_{i=1}^{I} \alpha_{0i}\gamma_k}{\sqrt{\sum_{i=1}^{I} \alpha_{0i}\gamma_k}}\right)^2 \end{aligned}$$

Under null hypothesis

$$\left(\frac{\sum_{i=1}^{I} x_i - \sum_{i=1}^{I} \alpha_{0i}\gamma_k}{\sqrt{\sum_{i=1}^{I} \alpha_{0i}\gamma_k}}\right)^2 \Bigg|_{\gamma_k=\gamma_0} \sim \chi_1^2$$

Or

$$\frac{\sum_{i=1}^{I} x_i - \sum_{i=1}^{I} \alpha_{0i}\gamma_0}{\sqrt{\sum_{i=1}^{I} \alpha_{0i}\gamma_0}} \sim \mathrm{N}(0,1)$$

Since this is a one-sided test, we reject the null when

$$\frac{\sum_{i=1}^{I} x_i - \sum_{i=1}^{I} \alpha_{0i}\gamma_k}{\sqrt{\sum_{i=1}^{I} \alpha_{0i}\gamma_0}} > Z_\alpha$$

where α is the significance level, default is $\alpha = 0.001$.

3.2 With priors

For kth feature, assuming it has only one probe, and it still follows the Poisson background model with Gamma prior

$$X_{kj}|\gamma_k \sim \text{Poisson}\left(\gamma_k \alpha_{0j}\right),$$
$$\gamma_k \sim \text{Gamma}\left(\frac{1}{\sigma}, \sigma\mu_k\right).$$

With α_{0j}, σ and μ_0 estimated from Poisson background model using negative probes

$$X_{ij}|\gamma_i \sim \text{Poisson}\left(\gamma_i \alpha_{0j}\right),$$
$$\gamma_i \sim \text{Gamma}\left(\frac{1}{\sigma}, \sigma\mu_0\right),$$

we are interested in testing

$$H_0 : \ \mu_k \leq \mu$$
$$H_a : \ \mu_k > \mu$$

$$\begin{aligned}
f\left(x_{ik}|\mu_k, \sigma\right) &= \int_0^\infty f\left(x_{ik}|\gamma_k\right) f\left(\gamma_k|\sigma, \mu_k\right) d\gamma_k \\
&= \int_0^\infty \frac{\left(\alpha_{0i}\gamma_k\right)^{x_{ik}} \ e^{-\alpha_{0i}\gamma_k}}{x_{ik}!} \cdot \frac{1}{\Gamma\left(\frac{1}{\sigma}\right)\left(\sigma\mu\right)^{\frac{1}{\sigma}}} {\gamma_k}^{\frac{1}{\sigma}-1} e^{-\frac{\gamma_k}{\sigma\mu_k}} d\gamma_k \\
&= \int_0^\infty \frac{\left(\sigma\mu_k\right)^{-\frac{1}{\sigma}}}{x_{ik}!\ \Gamma\left(\frac{1}{\sigma}\right)} \alpha_{0i}^{x_{ik}} {\gamma_k}^{x_{ik}+\frac{1}{\sigma}-1} e^{-(\alpha_{0i}+\frac{1}{\sigma\mu_k})\gamma_k} d\gamma_k \\
&= \frac{\left(\sigma\mu_k\right)^{-\frac{1}{\sigma}}}{x_{ik}!\ \Gamma\left(\frac{1}{\sigma}\right)} a_{0i}^{x_{ik}} \int_0^\infty \ {\gamma_k}^{x_{ik}+\frac{1}{\sigma}-1} e^{-(\alpha_{0i}+\frac{1}{\sigma\mu_k})\gamma_k} d\gamma_k \\
&= \frac{\left(\sigma\mu_k\right)^{-\frac{1}{\sigma}}}{x_{ik}!\ \Gamma\left(\frac{1}{\sigma}\right)} a_{0i}^{x_{ik}} \frac{\Gamma\left(x_{ik}+\frac{1}{\sigma}\right)}{\left(\alpha_{0i}+\frac{1}{\sigma\mu_k}\right)^{x_{ik}+\frac{1}{\sigma}}} \\
&= \frac{\Gamma\left(x_{ik}+\frac{1}{\sigma}\right)}{x_{ik}!\ \Gamma\left(\frac{1}{\sigma}\right)} \left(\frac{\frac{1}{\sigma\mu}}{\alpha_{0i}+\frac{1}{\sigma\mu_k}}\right)^{\frac{1}{\sigma}} \left(\frac{\alpha_{0i}}{\alpha_{0i}+\frac{1}{\sigma\mu_k}}\right)^{x_{ik}}
\end{aligned}$$

Thus

$$x_{ik} \sim \text{NB}(\frac{\alpha_{0i}}{\alpha_{0i}+\frac{1}{\sigma\mu_k}}, \frac{1}{\sigma})$$

$$
\begin{aligned}
\log f\left(x_{ik}|\mu_k,\sigma\right) &= -\log x_{ik}! + \log\left(\frac{\Gamma\left(x_{ik}+\frac{1}{\sigma}\right)}{\Gamma\left(\frac{1}{\sigma}\right)}\right) \\
&\quad - \left(x_{ik}+\frac{1}{\sigma}\right)\log\left(\alpha_{0i}+\frac{1}{\sigma\mu_k}\right) + x_{ik}\log\left(\alpha_{0i}\right) - \frac{1}{\sigma}\log\left(\sigma\mu_k\right) \\
&= -\log x_{ik}! + \log\left(\frac{\Gamma\left(x_{ik}+\frac{1}{\sigma}\right)}{\Gamma\left(\frac{1}{\sigma}\right)}\right) \\
&\quad - \left(x_{ik}+\frac{1}{\sigma}\right)\log\left(\alpha_{0i}\sigma\mu_k+1\right) + \left(x_{ik}+\frac{1}{\sigma}\right)\log\left(\sigma\mu_k\right) + x_{ik}\log\left(\alpha_{0i}\right) - \frac{1}{\sigma}\log\left(\sigma\mu_k\right) \\
&= -\log x_{ik}! + \log\left(\frac{\Gamma\left(x_{ik}+\frac{1}{\sigma}\right)}{\Gamma\left(\frac{1}{\sigma}\right)}\right) + x_{ik}\log(\alpha_{0i}\sigma\mu_k) \\
&\quad - \left(x_{ik}+\frac{1}{\sigma}\right)\log\left(\alpha_{0i}\sigma\mu_k+1\right)
\end{aligned}
$$

Given σ, the likelihood function is

$$
L(\mu_k|\boldsymbol{x}) = \sum_{i=1}^{I}\log f\left(x_k|\mu_k,\sigma\right)
$$

$$
\frac{dL(\gamma_k|\boldsymbol{x})}{d\mu_k} = \sum_{i=1}^{I}\frac{x_{ik}}{\mu_k} - \frac{x_{ik}\sigma\alpha_{0i}+\alpha_{0i}}{\alpha_{0i}\sigma\mu_k+1} = \sum_{i=1}^{I}\frac{x_{ik}-\alpha_{0i}\mu_k}{(\alpha_{0i}\sigma\mu_k+1)\mu_k}
$$

The Fisher information

$$
\begin{aligned}
I\left(\gamma_k\right) &= \operatorname{var}\left(\frac{dL(\gamma_k|\boldsymbol{x})}{d\gamma_k}\right) \\
&= \operatorname{var}\left(\sum_{i=1}^{I}\frac{x_{ik}-\alpha_{0i}\mu_k}{(\alpha_{0i}\sigma\mu_k+1)\mu_k}\right) \\
&= \operatorname{var}\left(\sum_{i=1}^{I}\frac{x_{ik}}{(\alpha_{0i}\sigma\mu_k+1)\mu_k}\right) \\
&= \sum_{i=1}^{I}\frac{\operatorname{var}\left(x_i\right)}{((\alpha_{0i}\sigma\mu_k+1)\mu_k)^2} \\
&= \sum_{i=1}^{I}\frac{\frac{\frac{\alpha_{0i}}{\alpha_{0i}+\frac{1}{\sigma\mu_k}}\frac{1}{\sigma}}{\frac{1}{\sigma\mu_k}}}{(\frac{\frac{1}{\sigma\mu_k}}{\alpha_{0i}+\frac{1}{\sigma\mu_k}})^2}}{((\alpha_{0i}\sigma\mu_k+1)\mu_k)^2} \\
&= \sum_{i=1}^{I}\frac{\alpha_{0i}}{(\alpha_{0i}\sigma\mu_k+1)\mu_k}
\end{aligned}
$$

Under the null hypothesis

$$\left.\frac{\left(\frac{dL(\gamma_k|\boldsymbol{x})}{d\mu_k}\right)^T \frac{dL(\gamma_k|\boldsymbol{x})}{d\mu_k}}{I(\mu_k)}\right|_{\mu_k=\mu_0} = \left.\left(\frac{\sum_{i=1}^I \frac{x_{ik}-\alpha_{0i}\mu_k}{(\alpha_{0i}\sigma\mu_k+1)\mu_k}}{\sqrt{\sum_{i=1}^I \frac{\alpha_{0i}}{(\alpha_{0i}\sigma\mu_k+1)\mu_k}}}\right)^2\right|_{\mu_k=\mu_0} \sim \chi_1^2$$

Or

$$\frac{\sum_{i=1}^I \frac{x_{ik}-\alpha_{0i}\mu_0}{(\alpha_{0i}\sigma\mu_0+1)\mu_0}}{\sqrt{\sum_{i=1}^I \frac{\alpha_{0i}}{(\alpha_{0i}\sigma\mu_0+1)\mu_0}}} \sim \mathrm{N}(0,1)$$

and we reject the null when

$$\frac{\sum_{i=1}^I \frac{x_{ik}-\alpha_{0i}\mu_0}{(\alpha_{0i}\sigma\mu_0+1)\mu_0}}{\sqrt{\sum_{i=1}^I \frac{\alpha_{0i}}{(\alpha_{0i}\sigma\mu_0+1)\mu_0}}} > Z_\alpha$$

where α is the significance level, default is $\alpha = 0.001$.

4 Biological feature modeling

For features distributed as background, the Poisson background model is the most appropriate. For features above the background, there are two things we need to deal with. First, the biological signal might have a different size factor as opposed to background; second, the feature might have biological variation besides difference in the size factor. We first focus on addressing the different size factor while leaving the task of explanation of the biological variation to methods of normalization and differential expression.

4.1 Poisson-Negative Binomial model

This is the most intuitive model for biological features. It simply assumes for kth feature.

$$Y_{kj} = X_{kj} + Z_{kj}$$

$$X_{kj} \sim \mathrm{Poisson}\left(\gamma_0\alpha_{0j}\right),\ Z_{kj} \sim \mathrm{NB}\left(\gamma_k\alpha_j, r_k\right)$$

The first part is the Poisson background model with γ_0 and α_{0j} both estimated from Poisson background model on negative probes, and the parameters of interest here are $\gamma_k, r_k, k = 1, \cdots, K$ and $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_J)$. The model cannot be fitted for too few features, but can be estimated in iterative fashion for $k = 1, \cdots, K$, and with the number of data larger than the number of parameters so $KJ > 2K + J$ or $K > \frac{J}{J-2}$.

As intuitive as the model is, its likelihood involves convolution and is hard to express and compute. So we propose to fit the following Negative Binomial threshold model instead.

4.2 Negative Binomial threshold model

$$Y_{kj} \sim \text{NB}(\max(\gamma_k - \gamma_t, 0)\alpha_j + \max(\gamma_i, \gamma_t)\alpha_{0j}, r_k), \tag{3}$$

$$\sum_{j=1}^{J} \alpha_j = 1 \tag{4}$$

Similar to Poisson-Negative Binomial model, we need background size factor $\boldsymbol{\alpha}_0$, but γ_t can be either set to be γ_0 or unspecified. With $K > \frac{J+1}{J-2}$ features, we can estimate parameters $\gamma_k, r_k, k = 1, \cdots, K$, $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_J)$ and γ_t. We recommend to choose a representative set of high abundance features dictated by the Background Score Test. The estimation of $\boldsymbol{\alpha}$ is stable with respect to the choice of high feature set.

This model is superior to the Poisson-Negative Binomial model in that it does not require convolution and it captures two different sources of count generation by a threshold. In additional, the threshold parameter γ_t can be estimated with relative ease, and is usually close to γ_0.

This model suffers from identifiability problems just like Poisson Background model. Specifically, $\gamma_k, k = 1, \cdots, K, \gamma_t, \boldsymbol{\alpha}_0, \boldsymbol{\alpha}$ and $\gamma_k/c, k = 1, \cdots, K, \gamma_t/c, c\boldsymbol{\alpha}_0, c\boldsymbol{\alpha}$ give the same model for any positive constant c. Poisson Background model is made identifiable by the constraint $\sum_{j=1}^{J} \alpha_{0j} = 1$, and so does this model. Applying different constraints can change the parameters by a factor, while some functions of parameters, including $\gamma_0\boldsymbol{\alpha}_0$ and $\gamma_0\boldsymbol{\alpha}$ are invariant to choice of constraints. Let $\tilde{\boldsymbol{\alpha}}_0 = \gamma_0\boldsymbol{\alpha}_0$ and $\tilde{\boldsymbol{\alpha}} = \gamma_0\boldsymbol{\alpha}$, which are parameterization(constraint) invariant size factors and are used in all following analysis.

4.3 Q90 range

We want to filter out ROI with low signal. $\tilde{\boldsymbol{\alpha}}$ represent the signal size factor after accounting for the background, making it useful for this task. However, $\tilde{\boldsymbol{\alpha}}$ is estimated by iterative model estimation, and could be impacted by different choice of high feature set. Thus it is not intuitive and straightforward for sample QC purpose. We propose to use the following metric named "Q90 range" as a proxy of $\tilde{\boldsymbol{\alpha}}$:

$$\text{Q90r} = \text{Q90} - \tilde{\boldsymbol{\alpha}}_0.$$

Q90r is just the difference between 90% quantile and the average background level. Similarly one can compute quantile range for other quantiles like Q75r. But using higher quantile is more

conservative in determining signal and tend to filter out less samples. We usually use the Q90r as default. Quantile ranges are shown to be better correlated with $\tilde{\boldsymbol{\alpha}}$ than quantiles.

4.4 Normalization with Poisson threshold model

Assume $\boldsymbol{\beta}_k = (\beta_{k1}, \cdots, \beta_{kJ})$ to be the vector of $\log 2$ expression of kth feature for J samples without technical variability. The goal of normalization is to optimally estimate $\boldsymbol{\beta}_k$ with available information. Treating $\boldsymbol{\beta}_k$ as the parameter, we will have a saturated model where the number of parameters(J) is identical or more than the number of obervations(J), for which we have complete information and therefore no overdispersion in theory. So we use Poisson distribution for normalization

$$Y_{kj} \sim \text{Poisson}(\mu_{kj}).$$

Let $\boldsymbol{\mu}_k = (\mu_{k1}, \cdots, \mu_{kJ})$, knowing background size factor $\boldsymbol{\alpha}_0$, signal size factor $\boldsymbol{\alpha}$, and mean background feature factor γ_0, its relation with respect to $\boldsymbol{\beta}_k$ can be expressed as

$$\boldsymbol{\mu}_k = \boldsymbol{\alpha} 2^{\boldsymbol{\beta}_k} + \boldsymbol{\alpha}_0 \gamma_k,$$
$$\gamma_k \sim \text{N}(\gamma_0, \sigma_\gamma^2),\ \boldsymbol{\beta}_k \sim \text{N}(\mathbf{0}, \Sigma_{\boldsymbol{\beta}}).$$

However, one issue with this model is that the $\boldsymbol{\alpha}_0$ and $\boldsymbol{\alpha}$ are subject to constraint, so as the normalized expression $\boldsymbol{\beta}_k$. To eliminate the impact of constraints, we replace $\boldsymbol{\alpha}$ by $\tilde{\boldsymbol{\alpha}}$ and $\boldsymbol{\alpha}_0$ by $\tilde{\boldsymbol{\alpha}}_0$, and fix γ_j at 1 by setting a heavy penalty like $\sigma_\gamma^2 = \frac{1}{10000}$.

$$Y_{kj} \sim \text{Poisson}(\mu_{kj}),$$
$$\boldsymbol{\mu}_k = \tilde{\boldsymbol{\alpha}}_j 2^{\boldsymbol{\beta}_k} + \tilde{\boldsymbol{\alpha}}_0 \gamma_k,$$
$$\gamma_k \sim \text{N}(1, \sigma_\gamma^2),\ \boldsymbol{\beta}_k \sim \text{N}(\mathbf{0}, \Sigma_{\boldsymbol{\beta}}),$$

with $\sigma_\gamma^2 = \frac{1}{10000}$.

The precision matrix Σ^{-1} is determined by Empirical Bayes approach in 2 steps.

1. Solving this model using subset of high abundance genes using prior, the default is $\Sigma_{\boldsymbol{\beta}}^{-1} = \frac{1}{\sigma^2} B^T B$ Where $B = \left(\frac{1}{n}, \frac{1}{n}, \cdots, \frac{1}{n}\right)$. This means we are only adding a penalty of $\frac{1}{\sigma^2}$ to the mean of each of these high abundance feature, the default is $\sigma = 5$. This is a weak penalty helps with the numerical stability.

2. Using the $\hat{\beta}$ estimated from high expressors. Calculate $\Sigma^{-1} = kA^T\Sigma_c^{-1}A + \frac{1}{\sigma^2}B^TB$ Where $B = \left(\frac{1}{n}, \frac{1}{n}, \cdots, \frac{1}{n}\right)$, and $A_{(n-1)\times n}$ is any full rank matrix satisfying $AB^T = \mathbf{0}$. By definition, $A_{(n-1)\times n}$ consists of vectors of contrasts. When Σ_c is estimated as the empirical covariance matrix of $A\boldsymbol{\beta}_k$, $A^T\Sigma_c^{-1}A$ is invariant with respect to choices of A. The strength parameter $k \in (0, 1)$ can adjust the strength of the contrast precision. This form of EB prior is based on the idea of decomposing the precision matrix into the orthogonal space of precision matrix of mean and precision matrix of contrasts, which comes from the belief contrast information of the high expressors should be passed to the parameter estimation of low expressors, but the covariance between the mean gene expression and the contrasts shouldn't.

It is more natural to assume only the ROIs in the same slide are correlated, so in the case of multiple slides data, it is advisable to apply this normalization function on each slide separately, which is implemented in **GeoDiff** as default for normalization of multiple slides data.

4.5 Differential Expression analysis with Negative Binomial threshold DE model

To perform differential expression analysis, we fit the following Negative Binomial threshold DE model:

$$Y_{kj} \sim \mathrm{NB}(\mu_{kj}, r_k)$$
$$\boldsymbol{\mu} = \boldsymbol{\alpha} 2^{X\boldsymbol{\beta}_k} + \boldsymbol{\alpha}_0\gamma_j$$
$$\gamma_j \sim \mathrm{N}(\gamma_0, \sigma_\gamma^2),\ \beta_j \sim \mathrm{N}(0, \Sigma_{\boldsymbol{\beta}})$$

where

$$X = \begin{pmatrix} 1 & 0 & \dots & 1 \\ 1 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix}_{n\times p}$$

is the design matrix.

However, one issue with this model is that the constraint $\sum_{j=1}^{J}\alpha_j = 1$ in the Poisson Background model will have an impact on the parameter α_j, so the parameter estimation of this model is also affected. To make the model invariant for parameterization choice of Poisson Background model, we usually fit the following model instead

$$Y_{kj} \sim \text{NB}(\mu_{kj}, r_k)$$
$$\boldsymbol{\mu} = \tilde{\boldsymbol{\alpha}} 2^{X\boldsymbol{\beta}_k} + \tilde{\boldsymbol{\alpha}}_0 \gamma_j$$
$$\gamma_j \sim \text{N}(\gamma_0, \sigma_\gamma^2),\ \beta_j \sim \text{N}(0, \Sigma_{\boldsymbol{\beta}})$$

with $\sigma_\gamma^2 = \frac{1}{10000}$.

The covariance matrix Σ^{-1} is determined by Empirical Bayes approach in 2 steps.

1. Solving this model using subset of high abundance genes using prior, the default is $\Sigma_\beta^{-1} = \frac{1}{\sigma^2} X^T B^T B X$ Where $B = \left(\frac{1}{n}, \frac{1}{n}, \cdots, \frac{1}{n}\right)$. This means we are only adding a penalty of $\frac{1}{\sigma^2}$ to the mean of each of these high abundance feature, the default is $\sigma = 5$. This is a weak penalty helps with the numerical stability. However, we do see in cases we need a strong prior to mitigate the bias of DE when the variable of interest is confounded by the disparity between $\boldsymbol{\alpha}_0$ and $\boldsymbol{\alpha}$.

2. Using the $\hat{\beta}$ estimated from high expressors. Calculate $\Sigma_\beta^{-1} = kA^T\Sigma_c^{-1}A + \frac{1}{\sigma^2} X^T B^T B X$ Where $B = \left(\frac{1}{n}, \frac{1}{n}, \cdots, \frac{1}{n}\right)$, and $A_{(p-1)\times p}$ is any full rank matrix satisfying $AX^T B^T = \mathbf{0}$. By definition, $A_{(p-1)\times p}$ consists of vectors of contrasts. When Σ_c is estimated as the empirical covariance matrix of $A\boldsymbol{\beta}_k$, $A^T\Sigma_c^{-1}A$ is invariant with respect to choices of A. The strength parameter $k \in (0, 1)$ can adjust the strength of the contrast precision. This form of EB prior is based on the idea of decomposing the precision matrix into the orthogonal space of precision matrix of mean and precision matrix of contrasts, which comes from the belief contrast information of the high expressors should be passed to the parameter estimation of low expressors, but the covariance between the mean gene expression and the contrasts shouldn't.

This way of constructing prior seems complicated, but it is mathematically intuitive and parameterization invariant: the model will stay the same no matter how X is specified, while this is not the case for DE model in **DEseq2**. One may argue different groups of features may have different systematic behavior, thus requiring different EB priors, and this type of analysis could be carried out by fitting the models separately on each group of features.

4.6 Differential Expression analysis with Negative Binomial threshold mixed DE model

In case of non-independent data, random effect need to be added to the model

$$\begin{aligned} Y_{kj} &\sim \mathrm{NB}(\mu_{kj}, r_k), \\ \boldsymbol{\mu}_k &= \boldsymbol{\alpha} 2^{X\boldsymbol{\beta}_k + Z\boldsymbol{U}_k} + \boldsymbol{\alpha}_0 \gamma_k, \\ \gamma_k \sim \mathrm{N}(\gamma_0, \sigma_\gamma^2),\ \boldsymbol{\beta}_k &\sim \mathrm{N}(0, \Sigma_{\boldsymbol{\beta}}),\ \boldsymbol{U}_k \sim \mathrm{N}(0, \Sigma_{\boldsymbol{U}}(\boldsymbol{\theta}_k)). \end{aligned}$$

One can also assume $\Sigma_{\boldsymbol{U}}(\boldsymbol{\theta})$ follows an Inverse Wishart prior

$$\Sigma_{\boldsymbol{U}}(\boldsymbol{\theta}) \sim IW(\Lambda, \nu).$$

Again in practice we often use

$$\begin{aligned} Y_{kj} &\sim \mathrm{NB}(\mu_{kj}, r_k), \\ \boldsymbol{\mu}_k &= \tilde{\boldsymbol{\alpha}} 2^{X\boldsymbol{\beta}_k + Z\boldsymbol{U}_k} + \tilde{\boldsymbol{\alpha}}_0 \gamma_k, \\ \gamma_k \sim \mathrm{N}(1, \frac{1}{10000}),\ \beta_j &\sim \mathrm{N}(0, \Sigma_{\boldsymbol{\beta}}),\ U_j \sim \mathrm{N}(0, \Sigma_{\boldsymbol{U}}(\boldsymbol{\theta}_k)). \end{aligned}$$

with optional

$$\Sigma_{\boldsymbol{U}}(\boldsymbol{\theta}) \sim IW(\Lambda, \nu).$$

Parameter estimation of this model involves computational intensive procedure called MCEM(Monte Carlo Expectation–maximization). So instead of using the 2 step approach to come up with a EB prior for regression coefficients, we usually just set up a default prior as $\Sigma_\beta^{-1} = \frac{1}{\sigma^2} B^T B$ Where $B = \left(\frac{1}{n}, \frac{1}{n}, \cdots, \frac{1}{n}\right)$, or use the estimated Σ_β^{-1} from the equivalent form of Negative Binomial threshold DE model.

5 Multiple probe target

Some DSP GeoMx panels use multiple probes for one target. For such case, we assume for kth feature(target), there are $v = 1, \cdots, V$ probes, and each of them should follow the same distribution since they measure the same thing. With the most general form of Negative Binomial distribution for each probe, we have

$$Y_{kjv} \sim NB(\mu_{kj}, r_k)$$

Then

$$\sum_{v=1}^{V} Y_{kjv} \sim NB(V\mu_{kj}, r_k).$$

So any inference with respect to that target should be based on the sum of all probes, and this impact every aspect of the workflow including:

1. In Background Score Test, testing against $V\gamma_0$ rather than γ_0,
2. In Negative Binomial threshold model, Normalization with Poisson threshold model and DE with Negative Binomial threshold DE model, replace γ_0 by $V\gamma_0$.

With this change all methods still apply to the multiple probe case using $\sum_{v=1}^{V} Y_{kjv}$ as the count data.

6 Numerics

All models are solved by maximum likelihood(MLE). Only Poisson Background model has closed form MLE solution (in presence of missing values, closed form MLE solution are conditioned on other parameters fixed) and are solved by directly using these forms. Other models are solved numerically by L-BFGS-B implemented by optim function. The threshold parameter γ_k, the size factors $\boldsymbol{\alpha}$ and parameter r in Negative Binomial distribution are all bounded below by 0. It is very convenient to incorporate such box contraint in L-BFGS-B.

6.1 Poisson Background model

The log-likelihood function

$$L(\gamma_i, \alpha_j|(x_{ij})_{I\times J}) = \sum_{i=1}^{I}\sum_{j=1}^{J}(x_{ij}\log(\gamma_i\alpha_j) - \gamma_i\alpha_j - \log(x_{ij}!))$$

$$\frac{\partial L(\gamma_i, \alpha_j)}{\partial \gamma_i} = \sum_{i=1}^{I}\sum_{j=1}^{J}(\frac{x_{ij}}{\gamma_i} - \alpha_j) = 0$$

$$\frac{\partial L(\gamma_i, \alpha_j)}{\partial \alpha_j} = \sum_{i=1}^{I}\sum_{j=1}^{J}(\frac{x_{ij}}{\alpha_j} - \gamma_i) = 0$$

$$\gamma_i = \frac{\sum_{j=1}^{J} x_{ij}}{\sum_{j=1}^{J} \alpha_j}$$

$$\alpha_j = \frac{\sum_{i=1}^{I} x_{ij}}{\sum_{i=1}^{I} \gamma_i}$$

With the constraint $\sum_{j=1}^{J} \alpha_j = 1$,

$$\gamma_i = \sum_{j=1}^{J} x_{ij}$$

$$\alpha_j = \frac{\sum_{i=1}^{I} x_{ij}}{\sum_{i=1}^{I}\sum_{j=1}^{J} x_{ij}}$$

Thus the Poisson Background modeling has closed form solution with complete data. However, we often refit this model after identify a set of outliers. With a subset Ω of (x_{ij}) of missing values, estimation of γ_i and α_j are iterative.

With the same constraint $\sum_{i=1}^{I} \alpha_i = 1$, iterate the following steps until converge

Starting value $\alpha_j = \sum_{i=1}^{I} x_{ij} I(x_{ij} \notin \Omega)$ and rescaled so that $\sum_{i=1}^{I} \alpha_i = 1$.

1. For $i = 1, \cdots, I$ calculate
$$\gamma_i = \frac{\sum_{j=1}^{J} x_{ij} I(x_{ij} \notin \Omega)}{\sum_{j=1}^{J} \alpha_j I(x_{ij} \notin \Omega)},$$
2. For $j = 1, \cdots, J$ calculate
$$\alpha_j = \frac{\sum_{i=1}^{I} x_{ij} I(x_{ij} \notin \Omega)}{\sum_{I=1}^{I} \gamma_i I(x_{ij} \notin \Omega)},$$
3. with $C = \sum_{j=1}^{J} \alpha_j$, update
$$\alpha_j = \frac{\alpha_j}{C}$$
for $j = 1, \cdots, J$.

When data is complete, thus Ω is empty, the results of this iterative algorithm are identical to closed form solution. Therefore, this algorithm is implemented to handle all scenarios.

For Poisson Background model with multiple sources, the algorithm is very similar. Let Φ be the function mapping an sample to its source, i.e. $\Phi(j) = s$ if jth sample is from sth source. Starting value $\alpha_j = \sum_{i=1}^{I} x_{ij} I(x_{ij} \notin \Omega)$ and rescaled so that $\sum_{i=1}^{I} \alpha_i = 1$.

1. For $s = 1, \cdots, S$ calculate
$$\gamma_{is} = \frac{\sum_{\Phi(j)=s} x_{ij} I(x_{ij} \notin \Omega)}{\sum_{\Phi(j)=s}^{J} \alpha_j I(x_{ij} \notin \Omega)}$$
2. For $j = 1, \cdots, J$ calculate
$$\alpha_j = \frac{\sum_{i=1}^{I} x_{ij} I(x_{ij} \notin \Omega)}{\sum_{I=1}^{I} \gamma_i I(x_{ij} \notin \Omega)}$$

3. with $C = \sum_{j=1}^{J} \alpha_j$, update

$$\alpha_j = \frac{\alpha_j}{C}$$

for $j = 1, \cdots, J$.

6.2 Negative Binomial threshold model

In the Negative Binomial threshold model

$$Y_{kj} \sim \text{NB}(\max(\gamma_k - \gamma_t, 0)\alpha_j + \max(\gamma_i, \gamma_t)\alpha_{0j}, r_k),$$

parameters γ_t, $\gamma_k, r_k, k = 1, \cdots, K$ and $\alpha_j, j = 1, \cdots, J$ are solved in the iterative fashion:

1. Optimize

$$\gamma_k, r_k, k = 1, \cdots, K$$

sequentially by fixing γ_t and $\alpha_j, j = 1, \cdots, J$

2. Optimize

$$\alpha_j, j = 1, \cdots, J$$

sequentially by fixing γ_t and $\gamma_k, r_k, k = 1, \cdots, K$

3. (Optional) Optimize γ_t by fixing $\gamma_k, r_k, k = 1, \cdots, K$ and $\alpha_j, j = 1, \cdots, J$

6.3 Poisson threshold Normalization

The probability mass function for Poisson distribution with mean μ is

$$f(Y = y|\mu) = \frac{\mu^y e^{-\mu}}{y!}.$$

To make it similar to DE models, we let X identity matrix and write $\boldsymbol{\beta} = X\boldsymbol{\beta}$. We use vector form to keep notation succinct. For a feature with observed count vector $\boldsymbol{y}$, the likelihood is

$$L(\boldsymbol{\beta}, \gamma|\boldsymbol{y}) = f(\boldsymbol{y}|\boldsymbol{\beta}, \gamma) = \mathbf{1}^T \left(\boldsymbol{y} \log \boldsymbol{\mu} - \boldsymbol{\mu} - \log(\boldsymbol{x}!)\right),$$
$$\text{where } \boldsymbol{\mu} = \boldsymbol{\alpha} 2^{X\boldsymbol{\beta}} + \boldsymbol{\alpha}_0 \gamma.$$

and the objective function to minimize under the priors is

$$W(\boldsymbol{\beta}, \gamma|\boldsymbol{y}, \Sigma_{\boldsymbol{\beta}}, \sigma_\gamma) = -L(\boldsymbol{\beta}, \gamma|\boldsymbol{y}) + \frac{1}{2}\boldsymbol{\beta}^T \Sigma_{\boldsymbol{\beta}}^{-1} \boldsymbol{\beta} + \frac{1}{2}\frac{(\gamma - \gamma_0)^2}{\sigma_\gamma^2}$$

$$\frac{\partial L}{\partial \boldsymbol{\mu}} = (\frac{\boldsymbol{y}}{\boldsymbol{\mu}} - \mathbf{1})^T$$
$$\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}} = (\log 2)\text{diag}(\boldsymbol{\alpha} 2^{X\boldsymbol{\beta}})X$$
$$\frac{\partial \boldsymbol{\mu}}{\partial \gamma} = \boldsymbol{\alpha}_0$$

$$\begin{aligned}
\frac{\partial W}{\partial \boldsymbol{\beta}} &= -\frac{\partial L}{\partial \boldsymbol{\mu}}\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}} + \boldsymbol{\beta}^T \Sigma_{\boldsymbol{\beta}}^{-1} \\
&= -(\log 2)\left((\frac{\boldsymbol{y}}{\boldsymbol{\mu}} - \mathbf{1})\boldsymbol{\alpha} 2^{X\boldsymbol{\beta}}\right)^T X + \boldsymbol{\beta}_k^T \Sigma_{\boldsymbol{\beta}}^{-1} \\
\frac{\partial W}{\partial \gamma} &= -\frac{\partial L}{\partial \boldsymbol{\mu}}\frac{\partial \boldsymbol{\mu}}{\partial \gamma} + \frac{(\gamma - \gamma_0)}{\sigma_\gamma^2} \\
&= -(\frac{\boldsymbol{y}}{\boldsymbol{\mu}} - \mathbf{1})^T \boldsymbol{\alpha}_0 + \frac{(\gamma - \gamma_0)}{\sigma_\gamma^2}
\end{aligned}$$

6.4 Negative Binomial threshold DE model

Since probability mass function for Negative Binomial distribution with mean μ and rate r is

$$f(Y = y|\mu, r) = \frac{\Gamma(y + r)}{\Gamma(r)\Gamma(y + 1)}\left(\frac{1}{1 + \mu/r}\right)^r \left(\frac{\mu/r}{1 + \mu/r}\right)^y,$$

the likelihood for Negative Binomial threshold DE model is

$$L(\boldsymbol{\beta}, \gamma, r|\boldsymbol{y}) = f(\boldsymbol{y}|\boldsymbol{\beta}, \gamma) = \mathbf{1}^T(\log\Gamma(\boldsymbol{y} + r) - \log\Gamma(r) - \log\Gamma(\boldsymbol{y} + 1)$$
$$-r_k \log(1 + \boldsymbol{\mu}/r) + \boldsymbol{y}\log(\boldsymbol{\mu}/r) - \boldsymbol{y}\log(1 + \boldsymbol{\mu}/r)),$$
$$\text{where } \boldsymbol{\mu} = \boldsymbol{\alpha} 2^{X\boldsymbol{\beta}} + \boldsymbol{\alpha}_0\gamma.$$

and the objective function to minimize under the priors is

$$W(\boldsymbol{\beta}, \gamma, r|\boldsymbol{y}, \Sigma_{\boldsymbol{\beta}}, \sigma_\gamma) = -L(\boldsymbol{\beta}_k, \gamma, r|\boldsymbol{y}) + \frac{1}{2}\boldsymbol{\beta}^T \Sigma_\beta^{-1}\boldsymbol{\beta} + \frac{1}{2}\frac{(\gamma - \gamma_0)^2}{\sigma_\gamma^2}$$

$$\frac{\partial L}{\partial \boldsymbol{\mu}} = \left(\frac{\boldsymbol{y}/\boldsymbol{\mu} - \mathbf{1}}{\mathbf{1} + \frac{1}{r}\boldsymbol{\mu}}\right)^T$$
$$\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}} = (\log 2)\text{diag}(\boldsymbol{\alpha} 2^{X\boldsymbol{\beta}})X$$
$$\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\gamma}} = \boldsymbol{\alpha}_0$$

$$
\begin{aligned}
\frac{\partial W}{\partial \boldsymbol{\beta}} &= -\frac{\partial L}{\partial \boldsymbol{\mu}}\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}} + \boldsymbol{\beta}^T\Sigma_{\boldsymbol{\beta}}^{-1} \\
&= -(\log 2)\left(\left(\frac{\boldsymbol{y}/\boldsymbol{\mu}-\mathbf{1}}{\mathbf{1}+\frac{1}{r}\boldsymbol{\mu}}\right)\boldsymbol{\alpha}2^{X\boldsymbol{\beta}}\right)^T X + \boldsymbol{\beta}^T\Sigma_{\boldsymbol{\beta}}^{-1} \\
\frac{\partial W}{\partial \gamma} &= -\frac{\partial L}{\partial \boldsymbol{\mu}}\frac{\partial \boldsymbol{\mu}}{\partial \gamma} + \frac{(\gamma-\gamma_0)}{\sigma_\gamma^2} \\
&= -\left(\frac{\boldsymbol{y}/\boldsymbol{\mu}_k-\mathbf{1}}{\mathbf{1}+\frac{1}{r}\boldsymbol{\mu}}\right)^T\boldsymbol{\alpha}_0 + \frac{(\gamma-\gamma_0)}{\sigma_\gamma^2} \\
\frac{\partial W}{\partial r} &= -\sum_{j=1}^{J}\left(-\log(1+\frac{1}{r}\mu_j) + \sum_{l=0}^{y_j}\frac{1}{l+r} - \frac{y_j-\mu_j}{r+\mu_j}\right)
\end{aligned}
$$

6.5 Negative Binomial threshold mixed DE model

For mixed model, treating the random effect $\boldsymbol{u}$ as missing value, let $\boldsymbol{w} = (\boldsymbol{y}, \boldsymbol{u})$ be the complete variable, then

$$L_{\boldsymbol{w}}(\boldsymbol{\beta}, \gamma, r, \boldsymbol{\theta}) = f_{\boldsymbol{Y}|\boldsymbol{U}}(\boldsymbol{y}|\boldsymbol{\beta}, \gamma, r, \boldsymbol{u}) f_{\boldsymbol{U}}(\boldsymbol{u}|\boldsymbol{\theta}),$$

to perform MLE, parameters $\boldsymbol{\beta}, \gamma, r, \boldsymbol{\theta}$ need to be estimated by optimizing

$$L_{\boldsymbol{y}}(\boldsymbol{\beta}, \gamma, r, \boldsymbol{\theta}) = \int_{\mathbb{R}^V} f_{\boldsymbol{Y}|\boldsymbol{U}}(\boldsymbol{y}|\boldsymbol{\beta}, \gamma, r, \boldsymbol{u}) f_{\boldsymbol{U}}(\boldsymbol{u}|\boldsymbol{\theta}) d\boldsymbol{u}.$$

This model is solved by MCEM algorithm consist of following steps.

Initialize $\boldsymbol{\beta}_0, \gamma_0, r_0, \boldsymbol{\theta}_0$,

1. At step n, sample $\boldsymbol{u}_{nm}, m = 1, \cdots, M_n$ from distribution

$$f_{\boldsymbol{U}|\boldsymbol{Y}}(\boldsymbol{u}|\boldsymbol{y}, \boldsymbol{\beta}_{n-1}, \gamma_{n-1}, r_{n-1}, \boldsymbol{\theta}_{n-1}) \propto f_{\boldsymbol{Y}|\boldsymbol{U}}(\boldsymbol{y}|\boldsymbol{u}, \boldsymbol{\beta}_{n-1}, \gamma_{n-1}, r_{n-1}) f_{\boldsymbol{U}}(\boldsymbol{u}|\boldsymbol{\theta}_{n-1}).$$

2. Obtain $\boldsymbol{\theta}_n$ by optimizing

$$\hat{Q}_n(\boldsymbol{\theta}) = \frac{1}{M}\sum_{m=1}^{M_n}\log f_{\boldsymbol{U}}(\boldsymbol{u}_{nm}|\boldsymbol{\theta}).$$

3. Obtain $\boldsymbol{\beta}_n, \gamma_n, r_n$ by optimizing

$$\hat{Q}_n(\boldsymbol{\beta}, \gamma, r) = \sum_{m=1}^{M_n}\log f_{\boldsymbol{Y}|\boldsymbol{U}}(\boldsymbol{y}, \boldsymbol{u}_{nm}|\boldsymbol{\beta}, \gamma, r).$$

Or, optimizing the following object function with thinned samples $\Psi = \{g, 2g, \cdots, [\frac{M_n}{g}]g\}$, so about the same results can be achieved with less computation

$$\hat{Q}_n(\boldsymbol{\beta}, \gamma, r) = \sum_{m\in\Psi}\log f_{\boldsymbol{Y}|\boldsymbol{U}}(\boldsymbol{y}, \boldsymbol{u}_{nm}|\boldsymbol{\beta}, \gamma, r).$$

For step 1, there are various ways to sample $\boldsymbol{u}_{nm}$ from distribution $f_{\boldsymbol{U}|\boldsymbol{Y}}(\boldsymbol{u}|\boldsymbol{y}, \boldsymbol{\beta}_{n-1}, \gamma_{n-1}, r_{n-1}, \boldsymbol{\theta}_{n-1})$, here we implement Gibbs sampler and Metropolis Hasting algorithm. We update each element of $\boldsymbol{u}$ by the proposal distribution $h_{\boldsymbol{U}}(\boldsymbol{u}^*|\boldsymbol{u}) = f_U(\boldsymbol{u}^*|\boldsymbol{\theta})$, then the acceptance function is calculated as

$$\begin{aligned} A_v(\boldsymbol{u}, \boldsymbol{u}^*) &= \frac{f_{\boldsymbol{U}|\boldsymbol{Y}}(\boldsymbol{u}^*|\boldsymbol{y}, \boldsymbol{\beta}, \gamma, r) h_{\boldsymbol{U}}(\boldsymbol{u}|\boldsymbol{u}^*)}{f_{\boldsymbol{U}|\boldsymbol{Y}}(\boldsymbol{u}|\boldsymbol{y}, \boldsymbol{\beta}, \gamma, r) h_{\boldsymbol{U}}(\boldsymbol{u}^*|\boldsymbol{u})} \\ &= \frac{f_{\boldsymbol{Y}|\boldsymbol{U}}(\boldsymbol{y}|\boldsymbol{u}^*, \boldsymbol{\beta}, \gamma, r) f_{\boldsymbol{U}}(\boldsymbol{u}^*|\boldsymbol{\theta}) f_{\boldsymbol{U}}(\boldsymbol{u}|\boldsymbol{\theta})}{f_{\boldsymbol{Y}|\boldsymbol{U}}(\boldsymbol{y}|\boldsymbol{u}, \boldsymbol{\beta}, \gamma, r) f_{\boldsymbol{U}}(\boldsymbol{u}|\boldsymbol{\theta}) f_{\boldsymbol{U}}(\boldsymbol{u}^*|\boldsymbol{\theta})} \\ &= \frac{f_{\boldsymbol{Y}|\boldsymbol{U}}(\boldsymbol{y}|\boldsymbol{u}^*, \boldsymbol{\beta}, \gamma, r)}{f_{\boldsymbol{Y}|\boldsymbol{U}}(\boldsymbol{y}|\boldsymbol{u}, \boldsymbol{\beta}, \gamma, r)}. \end{aligned}$$

Therefore, given $\boldsymbol{u} = (u_1, \cdots, u_V)$ and denote $\boldsymbol{u}$ with vth element removed as $\boldsymbol{u}_{-v}$, at each iteration m we update each element of $\boldsymbol{u}$ by

1. At vth element, generate $u_v^* \sim f(u_v|\boldsymbol{u}_{-v}, \boldsymbol{\theta})$, and let $\boldsymbol{u}^* = (u_1, u_2, \cdots, u_{v-1}, u_v^*, u_{v+1} \cdots, u_V)$,
2. Accept u_v^* by the probability of

$$A_v(\boldsymbol{u}, \boldsymbol{u}^*) = \min\left(1, \frac{f_{\boldsymbol{Y}|\boldsymbol{U}}(\boldsymbol{y}|\boldsymbol{u}^*, \boldsymbol{\beta}, \gamma, r)}{f_{\boldsymbol{Y}|\boldsymbol{U}}(\boldsymbol{y}|\boldsymbol{u}, \boldsymbol{\beta}, \gamma, r)}\right).$$

For step 2, $\boldsymbol{\theta}$ is estimated by optimizing

$$\begin{aligned} W(\boldsymbol{\theta}|\boldsymbol{u}_1, \cdots, \boldsymbol{u}_M) &= -L(\boldsymbol{\theta}|\boldsymbol{u}_1, \cdots, \boldsymbol{u}_M) \\ &= \frac{M}{2} \det \Sigma_{\boldsymbol{U}}(\boldsymbol{\theta}) + \frac{1}{2} \sum_{m=1}^{M} \boldsymbol{u}_m^T \Sigma_{\boldsymbol{U}}(\boldsymbol{\theta})^{-1} \boldsymbol{u}_m \end{aligned}$$

With inverse Wishart prior,

$$\Sigma_{\boldsymbol{U}}(\boldsymbol{\theta}) \sim IW(\Lambda, \nu).$$

$\boldsymbol{\theta}$ is estimated by MAP, thus optimizing

$$\begin{aligned} W(\boldsymbol{\theta}|\boldsymbol{u}_1, \cdots, \boldsymbol{u}_M, \Lambda, \nu) &= -L(\boldsymbol{\theta}|\boldsymbol{u}_1, \cdots, \boldsymbol{u}_M) - \log f(\Sigma|\Lambda, \nu) \\ &= \frac{M + \nu + V + 1}{2} \det \Sigma_{\boldsymbol{U}}(\boldsymbol{\theta}) + \frac{1}{2} \sum_{m=1}^{M} \boldsymbol{u}_m^T \Sigma_{\boldsymbol{U}}(\boldsymbol{\theta})^{-1} \boldsymbol{u}_m + \frac{1}{2} \mathrm{tr}(\Sigma_{\boldsymbol{U}}(\boldsymbol{\theta})^{-1} \Lambda) \end{aligned}$$

For step 3, $\boldsymbol{\beta}, \gamma, r$ is estimated in the similar manner as the fixed model. The only difference is the we need to optimize the average log-likelihood function instead of single log-likelihood, with $X\boldsymbol{\beta} + Z\boldsymbol{u}_m$ in the place of $X\boldsymbol{\beta}$.

6.6 Inference

Assume $\boldsymbol{\eta}_{p\times 1}$ represent all parameters in a statistical model, $I(\boldsymbol{\eta})$ is the Fisher information matrix at $\boldsymbol{\eta}$, then for contrast $C_{q\times p}$ and $\boldsymbol{h} = C\boldsymbol{\eta}$, under certain regularity condition, we have $\widehat{I(\boldsymbol{\eta})}$ a good estimator of Information matrix, and $\hat{\boldsymbol{\eta}}$, MLE of $\boldsymbol{\eta}$, satisfies

$$(C\hat{\boldsymbol{\eta}} - \boldsymbol{h}) \left(C\widehat{I(\boldsymbol{\eta})}^{-1} C^T \right) (C\hat{\boldsymbol{\eta}} - \boldsymbol{h}) \overset{d}{\sim} \chi^2(q).$$

Let C be a vector $\boldsymbol{c}_{p\times 1}$, then

$$\frac{\boldsymbol{c}\hat{\boldsymbol{\eta}} - h}{\sqrt{\boldsymbol{c}\widehat{I(\boldsymbol{\eta})}^{-1}\boldsymbol{c}^T}} \overset{d}{\sim} N(0,1).$$

When $c = (0, \cdots, 0, 1, 0, \cdots, 0,)$ and $h = 0$, it is testing single coefficient against 0 as we often see in any standard model output.

This is called Wald Test and is implemented for testing every regression coefficients against 0 as well as any user designated contrast in **GeoDiff**.

6.6.1 Differential Expression analysis with Negative Binomial threshold DE model

It is well known that

$$J(\boldsymbol{y}, \hat{\boldsymbol{\eta}}) = -\frac{\partial^2 L(\boldsymbol{\eta}|\boldsymbol{y})}{\partial \boldsymbol{\eta}^2}\bigg|_{\hat{\boldsymbol{\eta}}}$$

the Hessian of negative log likelihood at the MLE is the observed Fisher information matrix, a good estimator of $I(\boldsymbol{\eta})$.

6.6.2 Differential Expression analysis with Negative Binomial threshold mixed DE model

In the Negative Binomial threshold mixed DE model, the parameter $\boldsymbol{\eta} = (\boldsymbol{\beta}, \gamma, r, \boldsymbol{\theta})$ and the complete variable is $\boldsymbol{w} = (\boldsymbol{y}, \boldsymbol{u})$. Let

$$S(\boldsymbol{w}, \boldsymbol{\eta}) = \frac{\partial L_{\boldsymbol{w}}(\boldsymbol{\eta}|\boldsymbol{w})}{\partial \boldsymbol{\eta}} \text{ and } J(\boldsymbol{w}, \boldsymbol{\eta}) = -\frac{\partial^2 L_{\boldsymbol{w}}(\boldsymbol{\eta}|\boldsymbol{w})}{\partial \boldsymbol{\eta}^2}.$$

Then the observed information is

$$I_{\boldsymbol{y}}(\hat{\boldsymbol{\eta}}) = E_{\boldsymbol{u}|\boldsymbol{y},\hat{\boldsymbol{\eta}}} J(\boldsymbol{w}, \boldsymbol{\eta}) - E_{\boldsymbol{u}|\boldsymbol{y},\hat{\boldsymbol{\eta}}} \left(S(\boldsymbol{w}, \boldsymbol{\eta}) S^T(\boldsymbol{w}, \boldsymbol{\eta}) \right),$$

and is estimated using the MC samples $\{\boldsymbol{u}_m, m = 1, \cdots, M\}$ as

$$\widehat{I_{\boldsymbol{y}}(\hat{\boldsymbol{\eta}})} = \frac{1}{M} \sum_{m=1}^{M} J((\boldsymbol{y}, \boldsymbol{u}_m), \hat{\boldsymbol{\eta}}) - \frac{1}{M_n} \sum_{m=1}^{M} \left(S\left((\boldsymbol{y}, \boldsymbol{u}_m), \hat{\boldsymbol{\eta}}\right) S^T\left((\boldsymbol{y}, \boldsymbol{u}_m), \hat{\boldsymbol{\eta}}\right) \right).$$

References