**Beta mixture model**

1. **Empirical Bayesian estimation**

For a given a region from cell, there are CpG sites. For each CpG site, we assume that the methylated count follows binomial distribution with a common methylation probability. We further assume that there are a total ofread counts for the CpG site (). Then, on this CpG site, we have the methylated count

. (1)

Denote , the joint probability function can be written as

. (2)

Since the true methylation probability is unknown, we treat as a random variable to follow beta distribution,

, (3)

Therefore, we have the posterior distribution of that is also Beta distribution

. (4)

The parameters of the prior distribution are unknown. In order to estimate them, first, the beta distribution may be reparameterized by its mean and precision, that is

.

According to the previous assumptions of distributions, the marginal distribution of methylated counts is then given by Beta-binomial distribution. Second, the parameters and of the beta-binomial distribution are estimated uing an empirical Bayesian method [[1](#_ENREF_1)]. Consequently, we obtain an estimation based on the method of moments:

, (5) here

which is the weighted mean of observed methylation level . An estimation of precisionmay be obtained via [[2](#_ENREF_2)]:

. (6)

The total weighted sampled variance is

.

Therefore, based on (5) and (6), andare estimated as follows

, (7)

. (8)

It is worth noting that becausecan be negative, we assignin this case [[1](#_ENREF_1)]. In addition, for CpG sites with missing methylation data, we set their methylated counts and total counts as zeros.

1. **Methylation variance of cell to cell**

To understand methylation heterogeneity driven by cell subset specific methylation, we evaluate the cell-to-cell methylation variance. To this end, we employ random effect model to describe the variances across single cells. According to the posterior estimations of methylation probabilities above, we have the expectations and variances of methylation probabilities of

, (9)

. (10)

Also, we assume that is the abstract methylation probability across single cells. Furthermore, is defined as the variance of population; is defined as the deviation from the average methylation probability across single cells; as well as is a random effect. The observed methylation probabilities with the corresponding variance for the region from cell are considered to be a function of the abstract methylation probability, and

, (11)

To resolve the random effect model, based on a non-iteration algorithm [[2](#_ENREF_2)], we estimateas follows

, (12)

In which represents the number of cells and is defined

, (13) here.

Therefore, is estimated as a weighted mean of the observed methylation probabilities

, (14)

whereis defined as(15). Also, the estimator of the cell-to-cell methylation variance is

(16).

Here the 95% confidence interval of is obtained based on 1000 Bootstrap samplings.

1. **Clustering of single cell subpopulations**

Suppose that there are *K* methylation states in a given region. As the composition of methylation state is unknown, a mixture model is employed to decompose the mixture methylation states. To this end, we focus on some candidate regions with methylation variation across cells. For a given region, we assume that the proportion of the subgroup over the cell population is, where. As mentioned above, we assume that the number of methylated counts for each CpG site in a given region follows Binomial distribution and in the region the consistent methylation probability follows Beta distribution. Then, we obtain the posterior distribution of methylation probability in the region from cell :

,

in which . Since cells are grouped in the region, the methylation probability of the cells from a subgroup keep consistent. Let denote the methylation probability of group *k*. Then, the probability for the observed methylation in cell *i* is:

.

Also, according to the posterior distribution of methylation probability, the conditional probability observed cell from subgroup is obtained:

,

where is Gamma function.

Therefore, the joint likelihood function:

(17),

where. The parameters may be estimated by maximizing the log likelihood function:

(18),

The optimized problem may be resolved by Expectation-Maximization (EM) algorithm by a latent random variable which denotes the membership of cell *i*: . Let the probability of denote: .

Hence, a rewriting log likelihood is

, (19)

In (19), the equality holds if and only if

, (20)

In which we have actuallyand then combined with (20), we obtain a posterior estimation of given the observed cell and parameters:

. (21)

Then the optimization problem (18) may be estimated by maximizing, that is,

(22). In order to maximize (22), we have

(23). Consequently, we obtain

, (24) here.

In summary, we iteratively estimated all parameters based on EM algorithm. Also, when the function value of meet the threshold we expected, the iterative process will be stopped,

E-step:

,

M-step:

, here.

1. Ziller MJ, Gu H, Muller F, Donaghey J, Tsai LT, Kohlbacher O, et al. Charting a dynamic DNA methylation landscape of the human genome. Nature. 2013;500(7463):477-81.

2. Martuzzi M, Elliott P. Empirical Bayes estimation of small area prevalence of non-rare conditions. Stat Med. 1996;15(17-18):1867-73.