A probabilistic model for accurate quantification of DNA methylation modifications.

Supplemental Methods

Tarmo Äijö, Yun Huang, Henrik Mannerström, Lukas Chavez, Ageliki Tsagaratou, Anjana Rao, and Harri Lähdesmäki

March 11, 2016

On the statistical model of Lux

In this section we go through the statistical model of Lux in detail. The graphical model of Lux is depicted in Suppl. Fig. 2 and the values of the hyperprior parameters are also listed in Suppl. Table 1.

Parameters

In this subsection we describe the model parameters.

\[ \theta_{\text{control},i,j} = \left( p_{\text{control},i,j}("C") , p_{\text{control},i,j}("5mC") , p_{\text{control},i,j}("5hmC") \right) , \]

\[ i = 1, \ldots, N_{\text{control cytosines}} , \quad j = 1, \ldots, N_{\text{replicates}} \]

Description

Represents the proportions of the considered cytosine modifications for control cytosines (\( \forall i, j : \sum \theta_{\text{control},i,j} = 1 \)). Note that we model control cytosines independently across replicates, i.e. there is no hierarchical level for modeling biological variation on them.

Distribution

\[ \theta_{\text{control},i,j} \sim \text{Dir}(\alpha_{\text{control},i,j}) \]
\[ \theta^{i,j} = (p^{i,j}("C"), p^{i,j}("5mC"), p^{i,j}("5hmC")), \quad i = 1, \ldots, N_{\text{cytosines}}, \quad j = 1, \ldots, N_{\text{replicates}} \]

**Description**

Represents the proportions of the considered cytosine modifications for noncontrol cytosines (\( \forall i, j : \sum \theta^{i,j} = 1 \)).

**Distribution**

\[ \theta^{i,j} \sim \text{Dir}(\nu^{i,j} + 1) \]

\[ BS_{\text{eff}}^{j}, \quad j = 1, \ldots, N_{\text{replicates}} \]

**Description**

Represents the probability of successful bisulphite conversion, i.e. conversion of unmethylated cytosine to uracil.

**Distribution**

Is defined by parameters \( \mu_{BS_{\text{eff}}} \) and \( \sigma_{BS_{\text{eff}}} \) such that \( \logit(BS_{\text{eff}}^{j}) \sim N(\mu_{BS_{\text{eff}}}, \sigma_{BS_{\text{eff}}}) \).

\[ BS_{\ast,\text{eff}}^{j}, \quad j = 1, \ldots, N_{\text{replicates}} \]

**Description**

Represents the probability of inaccurate bisulphite conversion, i.e. conversion of 5mC or 5hmC to uracil.

**Distribution**

Is defined by parameters \( \mu_{BS_{\ast,\text{eff}}} \) and \( \sigma_{BS_{\ast,\text{eff}}} \) such that \( \logit(BS_{\ast,\text{eff}}^{j}) \sim N(\mu_{BS_{\ast,\text{eff}}}, \sigma_{BS_{\ast,\text{eff}}}) \).

\[ ox_{\text{eff}}^{j}, \quad j = 1, \ldots, N_{\text{replicates}} \]

**Description**

Represents the probability of successful oxidation, i.e. oxidation of 5hmC to 5fC.

**Distribution**

Is defined by parameters \( \mu_{OX_{\text{eff}}} \) and \( \sigma_{OX_{\text{eff}}} \) such that \( \logit(ox_{\text{eff}}^{j}) \sim N(\mu_{OX_{\text{eff}}}, \sigma_{OX_{\text{eff}}}) \).
\( \text{seq}_{\text{err}}^j, j = 1, \ldots, N_{\text{replicates}} \)

**Description**

Represents the probability of sequencing error, i.e. sequencing C instead of T or vice versa.

**Distribution**

Is defined by parameters \( \mu_{\text{seq}_{\text{err}}} \) and \( \sigma_{\text{seq}_{\text{err}}} \) such that \( \text{logit}(\text{seq}_{\text{err}}^j) \sim \mathcal{N}(\mu_{\text{seq}_{\text{err}}}, \sigma_{\text{seq}_{\text{err}}}) \).

**Hyperparameters**

In this subsection we describe the hyperparameters of the model.

\[ \alpha_{\text{control}, i, j} = (\alpha_{\text{C}, \text{control}, i, j}, \alpha_{5\text{mC}, \text{control}, i, j}, \alpha_{5\text{hmC}, \text{control}, i, j}), i = 1, \ldots, N_{\text{control cytosines}}, j = 1, \ldots, N_{\text{replicates}} \]

**Description**

Defines the priors for the proportions of the considered cytosine modifications for control cytosines, i.e. \( \theta_{\text{control}, i, j} \sim \text{Dir}(\alpha_{\text{control}, i, j}) \).

**Value** In the case of the C, 5mC and 5hmC control cytosines the values are \( \alpha_{\text{C}, \text{control}, i, j} = (1000, 1, 1) \), \( \alpha_{5\text{mC}, \text{control}, i, j} = (1, 1000, 1) \), and \( \alpha_{5\text{hmC}, \text{control}, i, j} = (6, 2, 72) \), respectively, for all \( i \) and \( j \).

\[ g^i, i = 1, \ldots, N_{\text{cytosines}} \]

**Description**

Controls the variance of the replicate-specific proportions of cytosine modifications for noncontrol cytosines \( \theta_{i, j} \).

**Distribution**

\( g^j \sim \text{Gamma}(\psi_g^1, \psi_g^2) \), where \( \psi_g^1 = 2, \psi_g^2 = 2/6 \)
\( \mu_i^i, \ i = 1, \ldots, N_{\text{cytosines}} \)

**Description**  
Represents the proportions of the considered cytosine modifications for noncontrol cytosines across replicates.

**Distribution**  
\( \mu_i^i \sim \text{Dir}(\alpha), \text{ where } \alpha = (0.8, 0.8, 0.8) \)

\( \mu^{BS_{\text{eff}}} \)

**Description**  
Controls the expected value of \( \logit(BS_{\text{eff}}) \) across replicates.

**Distribution**  
\( \mu^{BS_{\text{eff}}} \sim \mathcal{N}(\psi^{\mu^{\mu^i}}, \psi^{\mu^{\sigma^i}}), \text{ where } \psi^{\mu^{\mu^i}} = 2 \text{ and } \psi^{\mu^{\sigma^i}} = 1.29 \)

\( \sigma^{BS_{\text{eff}}} \)

**Description**  
Controls the standard deviation of \( \logit(BS_{\text{eff}}) \) across replicates.

**Distribution**  
\( \sigma^{BS_{\text{eff}}} \sim \ln\mathcal{N}(\psi^{\sigma^{\mu^i}}, \psi^{\sigma^{\sigma^i}}), \text{ where } \psi^{\sigma^{\mu^i}} = 0.4 \text{ and } \psi^{\sigma^{\sigma^i}} = 0.5 \)

\( \mu^{BS^*_{\text{eff}}} \)

**Description**  
Controls the expected value of \( \logit(BS^*_{\text{eff}}) \) across replicates.

**Distribution**  
\( \mu^{BS^*_{\text{eff}}} \sim \mathcal{N}(\psi^{\mu^{\mu^i}}, \psi^{\mu^{\sigma^i}}), \text{ where } \psi^{\mu^{\mu^i}} = -3 \text{ and } \psi^{\mu^{\sigma^i}} = 1.29 \)
\( \sigma_{\text{BS}^*_{\text{eff}}} \)

**Description**

Controls the standard deviation of logit(\( \text{BS}^*_{\text{eff}} \)) across replicates.

**Distribution**

\[ \sigma_{\text{BS}^*_{\text{eff}}} \sim \ln \mathcal{N}(\psi^\sigma_{\text{BS}^*_{\text{eff}}}, \psi^\sigma_{\text{BS}^*_{\text{eff}}}^\sigma), \text{ where } \psi^\sigma_{\text{BS}^*_{\text{eff}}} = 0.4 \text{ and } \psi^\sigma_{\text{BS}^*_{\text{eff}}}^\sigma = 0.5 \]

\( \mu_{\text{ox}^*_{\text{eff}}} \)

**Description**

Controls the expected value of logit(\( \text{ox}^*_{\text{eff}} \)) across replicates.

**Distribution**

\[ \mu_{\text{ox}^*_{\text{eff}}} \sim \mathcal{N}(\psi^\mu_{\text{ox}^*_{\text{eff}}}, \psi^\mu_{\text{ox}^*_{\text{eff}}}^\sigma), \text{ where } \psi^\mu_{\text{ox}^*_{\text{eff}}} = 2 \text{ and } \psi^\mu_{\text{ox}^*_{\text{eff}}}^\sigma = 1.29 \]

\( \sigma_{\text{ox}^*_{\text{eff}}} \)

**Description**

Controls the standard deviation of logit(\( \text{ox}^*_{\text{eff}} \)) across replicates.

**Distribution**

\[ \sigma_{\text{ox}^*_{\text{eff}}} \sim \ln \mathcal{N}(\psi^\sigma_{\text{ox}^*_{\text{eff}}}, \psi^\sigma_{\text{ox}^*_{\text{eff}}}^\sigma), \text{ where } \psi^\sigma_{\text{ox}^*_{\text{eff}}} = 0.4 \text{ and } \psi^\sigma_{\text{ox}^*_{\text{eff}}}^\sigma = 0.5 \]

\( \mu_{\text{seq}^*_{\text{err}}} \)

**Description**

Controls the expected value of logit(\( \text{seq}^*_{\text{err}} \)) across replicates.

**Distribution**

\[ \mu_{\text{seq}^*_{\text{err}}} \sim \mathcal{N}(\psi^\mu_{\text{seq}^*_{\text{err}}}, \psi^\mu_{\text{seq}^*_{\text{err}}}^\sigma), \text{ where } \psi^\mu_{\text{seq}^*_{\text{err}}} = -3 \text{ and } \psi^\mu_{\text{seq}^*_{\text{err}}}^\sigma = 1.29 \]

\( \sigma_{\text{seq}^*_{\text{err}}} \)

**Description**
Controls the standard deviation of logit(seqerr) across replicates.

**Distribution**

\[ \sigma_{\text{seqerr}} \sim \ln \mathcal{N}(\psi^{\sigma \mu}_{\text{seqerr}}, \psi^{\sigma \sigma}_{\text{seqerr}}) \], where \( \psi^{\sigma \mu}_{\text{seqerr}} = 0.4 \) and \( \psi^{\sigma \sigma}_{\text{seqerr}} = 0.5 \)

**Data**

In this subsection we describe the data.

\[ N^{\text{control,}i,j}_{\text{BS,C}}, \ i = 1, \ldots, N^{\text{control cytosines}}, \ j = 1, \ldots, N^{\text{replicates}} \]

**Description**

The count of “C” read-outs for the \( i^{\text{th}} \) control cytosine in the \( j^{\text{th}} \) BS-seq experiment.

\[ N^{\text{control,}i,j}_{\text{BS}}, \ i = 1, \ldots, N^{\text{control cytosines}}, \ j = 1, \ldots, N^{\text{replicates}} \]

**Description**

The count of total read-outs for the \( i^{\text{th}} \) control cytosine in the \( j^{\text{th}} \) BS-seq experiment.

\[ N^{\text{control,}i,j}_{\text{oxBS,C}}, \ i = 1, \ldots, N^{\text{control cytosines}}, \ j = 1, \ldots, N^{\text{replicates}} \]

**Description**

The count of “C” read-outs for the \( i^{\text{th}} \) control cytosine in the \( j^{\text{th}} \) oxBS-seq experiment.

\[ N^{\text{control,}i,j}_{\text{oxBS}}, \ i = 1, \ldots, N^{\text{control cytosines}}, \ j = 1, \ldots, N^{\text{replicates}} \]

**Description**

The count of total read-outs for the \( i^{\text{th}} \) control cytosine in the \( j^{\text{th}} \) oxBS-seq experiment.

\[ N^{i,j}_{\text{BS,C}}, \ i = 1, \ldots, N^{\text{cytosines}}, \ j = 1, \ldots, N^{\text{replicates}} \]

**Description**

The count of “C” read-outs for the \( i^{\text{th}} \) noncontrol cytosine in the \( j^{\text{th}} \) BS-seq experiment.
\( N_{i;j}^{BS}, \ i = 1, \ldots, N_{\text{cytosines}}, \ j = 1, \ldots, N_{\text{replicates}} \)

**Description**

The count of total read-outs for the \(i^{\text{th}}\) noncontrol cytosine in the \(j^{\text{th}}\) BS-seq experiment.

\( N_{i;j}^{oxBS,C}, \ i = 1, \ldots, N_{\text{cytosines}}, \ j = 1, \ldots, N_{\text{replicates}} \)

**Description**

The count of “C” read-outs for the \(i^{\text{th}}\) noncontrol cytosine in the \(j^{\text{th}}\) oxBS-seq experiment.

\( N_{i;j}^{oxBS}, \ i = 1, \ldots, N_{\text{cytosines}}, \ j = 1, \ldots, N_{\text{replicates}} \)

**Description**

The count of total read-outs for the \(i^{\text{th}}\) noncontrol cytosine in the \(j^{\text{th}}\) oxBS-seq experiment.

**Likelihood**

In this subsection we start with the combination of BS-seq and oxBS-seq data sets and state the likelihood in the general form, i.e. for replicates, and control and noncontrol cytosines. First, let us denote

\[
D^{\text{control}} = \left\{ N_{i;j}^{\text{BS},C}^{\text{control}}, N_{i;j}^{\text{BS}}^{\text{control}}, N_{i;j}^{\text{oxBS},C}^{\text{control}}, N_{i;j}^{\text{oxBS}}^{\text{control}} \right\}_{i=1,\ldots,N_{\text{cytosines}}}^{j=1,\ldots,N_{\text{replicates}}} \]

\[
D = \left\{ N_{i;j}^{\text{BS},C}, N_{i;j}^{\text{BS}}, N_{i;j}^{\text{oxBS},C}, N_{i;j}^{\text{oxBS}} \right\}_{i=1,\ldots,N_{\text{cytosines}}}^{j=1,\ldots,N_{\text{replicates}}} \]

\[
G^{\text{control}} = \left\{ \theta_{i;j}^{\text{control}} \right\}_{i=1,\ldots,N_{\text{cytosines}}}^{j=1,\ldots,N_{\text{replicates}}} \]

\[
\Theta = \left\{ \theta_{i;j} \right\}_{i=1,\ldots,N_{\text{cytosines}}}^{j=1,\ldots,N_{\text{replicates}}} \]

and

\[
\Omega = \left\{ BS_{i;j}^{\text{eff}}, BS_{i;j}^{\text{eff}}, ox_{i;j}^{\text{eff}}, seq_{i;j}^{\text{err}} \right\}_{j=1,\ldots,N_{\text{replicates}}} \]
Then the likelihood of the data \( D_{control} \) and \( D \) under the binomial model, assuming conditional independency, and given the parameters \( \Theta_{control} \), \( \Theta \), and \( \Omega \) is

\[
p(D_{control}, D|\Theta_{control}, \Theta, \Omega) = p(D_{control}|\Theta_{control}, \Omega)p(D|\Theta, \Omega)
\]

\[
= \prod_{j=1}^{N_{replicates}} \left[ \frac{N_{control \text{ cytosines}}}{N_{control \text{ cytosines}}} \prod_{i=1}^{N_{control \text{ oxBS}}} \left( \frac{N_{control \text{ i,j}}}{} \frac{N_{control \text{ i,j}} \frac{p_{BS}}{p_{BS}} \frac{C_{i,j}}{C_{i,j}} \frac{D_{i,j}}{D_{i,j}}}{N_{control \text{ i,j}}} \right) \prod_{i=1}^{N_{control \text{ oxBS}}} \left( \frac{N_{control \text{ oxBS},C}}{} \frac{p_{oxBS}}{p_{oxBS}} \frac{C_{i,j}}{C_{i,j}} \frac{D_{i,j}}{D_{i,j}}}{N_{control \text{ oxBS},C}} \right) \right]
\]

\[(1)\]

where

\[
p_{BS}^{control, i,j \ "C\"} = p_{BS}^{control, i,j \ "C\"} \left[ \left( 1 - BS_{eff}^{i,j} \right) \left( 1 - seq_{err}^{i,j} \right) + BS_{eff}^{i,j} \cdot seq_{err}^{i,j} \right] +
\]

\[
p_{control, i,j \ "5mC\"} \left[ \left( 1 - BS_{eff}^{i,j} \right) \left( 1 - seq_{err}^{i,j} \right) + BS_{eff}^{i,j} \cdot seq_{err}^{i,j} \right] +
\]

\[
p_{control, i,j \ "5hmC\"} \left[ \left( 1 - BS_{eff}^{i,j} \right) \left( 1 - seq_{err}^{i,j} \right) + BS_{eff}^{i,j} \cdot seq_{err}^{i,j} \right],
\]

\[(2)\]

\[
p_{oxBS}^{control, i,j \ "C\"} = p_{oxBS}^{control, i,j \ "C\"} \left[ \left( 1 - BS_{eff}^{i,j} \right) \left( 1 - seq_{err}^{i,j} \right) + BS_{eff}^{i,j} \cdot seq_{err}^{i,j} \right] +
\]

\[
p_{control, i,j \ "5mC\"} \left[ \left( 1 - BS_{eff}^{i,j} \right) \left( 1 - seq_{err}^{i,j} \right) + BS_{eff}^{i,j} \cdot seq_{err}^{i,j} \right] +
\]

\[
p_{control, i,j \ "5hmC\"} \left[ ox_{eff}^{i,j} \left( 1 - BS_{eff}^{i,j} \right) \left( 1 - seq_{err}^{i,j} \right) + BS_{eff}^{i,j} \cdot seq_{err}^{i,j} \right] +
\]

\[
(1 - ox_{eff}^{i,j}) \left[ \left( 1 - BS_{eff}^{i,j} \right) \left( 1 - seq_{err}^{i,j} \right) + BS_{eff}^{i,j} \cdot seq_{err}^{i,j} \right],
\]

\[(3)\]

\[
p_{BS}^{i,j \ "C\"} = p_{BS}^{i,j \ "C\"} \left[ \left( 1 - BS_{eff}^{i,j} \right) \left( 1 - seq_{err}^{i,j} \right) + BS_{eff}^{i,j} \cdot seq_{err}^{i,j} \right] +
\]

\[
p_{i,j \ "5mC\"} \left[ \left( 1 - BS_{eff}^{i,j} \right) \left( 1 - seq_{err}^{i,j} \right) + BS_{eff}^{i,j} \cdot seq_{err}^{i,j} \right] +
\]

\[
p_{i,j \ "5hmC\"} \left[ \left( 1 - BS_{eff}^{i,j} \right) \left( 1 - seq_{err}^{i,j} \right) + BS_{eff}^{i,j} \cdot seq_{err}^{i,j} \right],
\]

\[(4)\]
and

\[ p^{i,j}_{\text{oxBS}}(\text{“C”}) = p^{i,j}(\text{“C”}) \left[ (1 - \text{BS}_{\text{eff}}^j)(1 - \text{seq}_r^j) + \text{BS}_{\text{eff}}^j \text{seq}_r^j \right] + \\
\quad p^{i,j}(\text{“5mC”}) \left[ (1 - \text{BS}_{\text{eff}}^j)(1 - \text{seq}_r^j) + \text{BS}_{\text{eff}}^j \text{seq}_r^j \right] + \\
\quad (1 - \text{ox}_r^j) \left[ (1 - \text{BS}_{\text{eff}}^j)(1 - \text{seq}_r^j) + \text{BS}_{\text{eff}}^j \text{seq}_r^j \right]. \]  

(5)

These propagated probabilities Equations (2), (3), (4), and (5) are derived in Fig. 1b-c and Suppl. Fig. 1.

**Integrative analysis of BS-seq, TAB-seq, fCAB-seq, CAB-seq, redBS-seq, and MAB-seq data**

Let BS$_{\text{eff}}$, BS$_{\text{eff}}^*$, lab$_{\text{eff}}$, ox$_{\text{eff}}$, pro$_{\text{eff}}$, EDC$_{\text{eff}}$, red$_{\text{eff}}$, M.Sssl$_{\text{eff}}$, and seq$_{\text{err}}$ denote bisulphite conversion efficiency, inaccurate bisulphite conversion efficiency, labeling efficiency, oxidation efficiency, Et-ONH$_2$ protection efficiency, EDC protection efficiency, NaBH$_4$ reduction efficiency, M.Sssl efficiency and sequencing error, respectively. All the parameters have the same prior distributions and hyper parameters as described at the beginning of this document. Furthermore, let $i$ and $j$ denote cytosine and replicate, respectively. In the following subsections we state the emission probabilities, $p_{\text{BS}}(\text{“C”})$, $p_{\text{TAB}}(\text{“C”})$, $p_{\text{fCAB}}(\text{“C”})$, $p_{\text{CAB}}(\text{“C”})$, $p_{\text{redBS}}(\text{“C”})$ and $p_{\text{MAB}}(\text{“C”})$, in the terms of the aforementioned experimental parameters and methylation proportions,

\[ \theta = [p(\text{“C”}), p(\text{“5mC”}), p(\text{“5hmC”}), p(\text{“5fC”}), p(\text{“5caC”})]. \]
BS-seq

\[ p_{BS}^{i,j}("C") = p_{BS}^{i,j}("C") \left[ (1 - BS_{eff}^{i,j})(1 - seq_{err}^{i,j}) + BS_{eff}^{i,j}seq_{err}^{i,j} \right] + \]

\[ = p_{BS}^{i,j}("C"|C) \]

\[ p_{BS}^{i,j}("5mC") \left[ (1 - BS_{eff}^{i,j})(1 - seq_{err}^{i,j}) + BS_{eff}^{i,j}seq_{err}^{i,j} \right] + \]

\[ = p_{BS}^{i,j}("C"|5mC) \]

\[ p_{BS}^{i,j}("5hmC") \left[ (1 - BS_{eff}^{i,j})(1 - seq_{err}^{i,j}) + BS_{eff}^{i,j}seq_{err}^{i,j} \right] + \]

\[ = p_{BS}^{i,j}("C"|5hmC) \]

\[ p_{BS}^{i,j}("5fC") \left[ (1 - BS_{eff}^{i,j})(1 - seq_{err}^{i,j}) + BS_{eff}^{i,j}seq_{err}^{i,j} \right] + \]

\[ = p_{BS}^{i,j}("C"|5fC) \]

\[ p_{BS}^{i,j}("5caC") \left[ (1 - BS_{eff}^{i,j})(1 - seq_{err}^{i,j}) + BS_{eff}^{i,j}seq_{err}^{i,j} \right] + \]

\[ = p_{BS}^{i,j}("C"|5caC) \]

\[ (6) \]

TAB-seq

\[ p_{TAB}^{i,j}("C") = p_{TAB}^{i,j}("C") \left[ (1 - BS_{eff}^{i,j})(1 - seq_{err}^{i,j}) + BS_{eff}^{i,j}seq_{err}^{i,j} \right] + \]

\[ = p_{TAB}^{i,j}("C"|C) \]

\[ p_{TAB}^{i,j}("5mC") \left[ (1 - BS_{eff}^{i,j})(1 - seq_{err}^{i,j}) + BS_{eff}^{i,j}seq_{err}^{i,j} \right] + \]

\[ = p_{TAB}^{i,j}("C"|5mC) \]

\[ p_{TAB}^{i,j}("5hmC") \left[ (1 - BS_{eff}^{i,j})(1 - seq_{err}^{i,j}) + BS_{eff}^{i,j}seq_{err}^{i,j} \right] + \]

\[ = p_{TAB}^{i,j}("C"|5hmC) \]

\[ p_{TAB}^{i,j}("5fC") \left[ (1 - BS_{eff}^{i,j})(1 - seq_{err}^{i,j}) + BS_{eff}^{i,j}seq_{err}^{i,j} \right] + \]

\[ = p_{TAB}^{i,j}("C"|5fC) \]

\[ p_{TAB}^{i,j}("5caC") \left[ (1 - BS_{eff}^{i,j})(1 - seq_{err}^{i,j}) + BS_{eff}^{i,j}seq_{err}^{i,j} \right] + \]

\[ = p_{TAB}^{i,j}("C"|5caC) \]

\[ (7) \]
\[ fCAB-seq \]

\[
p^{i,j}_{\text{fCAB}}(\text{"C"}) = p^{i,j}_{\text{fCAB}}(\text{"C"}) \left[ (1 - BS_{\text{eff}}^{i,j})(1 - \text{seq}_{\text{err}}^{j}) + BS_{\text{eff}}^{i,j}\text{seq}_{\text{err}}^{j} \right] + \]

\[
= p^{i,j}_{\text{fCAB}}(\text{"C"}|\text{C}) \]

\[
p^{i,j}_{\text{fCAB}}(\text{"5mC"}) \left[ (1 - BS_{\text{eff}}^{i,j})(1 - \text{seq}_{\text{err}}^{j}) + BS_{\text{eff}}^{i,j}\text{seq}_{\text{err}}^{j} \right] + \]

\[
= p^{i,j}_{\text{fCAB}}(\text{"C"}|\text{5mC}) \]

\[
p^{i,j}_{\text{fCAB}}(\text{"5hmC"}) \left[ (1 - BS_{\text{eff}}^{i,j})(1 - \text{seq}_{\text{err}}^{j}) + BS_{\text{eff}}^{i,j}\text{seq}_{\text{err}}^{j} \right] + \]

\[
= p^{i,j}_{\text{fCAB}}(\text{"C"}|\text{5hmC}) \]

\[
p^{i,j}_{\text{fCAB}}(\text{"5fC"}) \left[ \left( 1 - \text{pro}_{\text{eff}}^{i,j} \right) \left[ (1 - BS_{\text{eff}}^{i,j})(1 - \text{seq}_{\text{err}}^{j}) + BS_{\text{eff}}^{i,j}\text{seq}_{\text{err}}^{j} \right] \right] + \]

\[
\left( 1 - \text{pro}_{\text{eff}}^{i,j} \right) \left[ (1 - BS_{\text{eff}}^{i,j})(1 - \text{seq}_{\text{err}}^{j}) + BS_{\text{eff}}^{i,j}\text{seq}_{\text{err}}^{j} \right] \]

\[
= p^{i,j}_{\text{fCAB}}(\text{"C"}|\text{5fC}) \]

\[
p^{i,j}_{\text{fCAB}}(\text{"5caC"}) \left[ (1 - BS_{\text{eff}}^{i,j})(1 - \text{seq}_{\text{err}}^{j}) + BS_{\text{eff}}^{i,j}\text{seq}_{\text{err}}^{j} \right] \]

\[
= p^{i,j}_{\text{fCAB}}(\text{"C"}|\text{5caC}) \]

\[
(8) \]

\[
CAB-seq \]

\[
p^{i,j}_{\text{CAB}}(\text{"C"}) = p^{i,j}_{\text{CAB}}(\text{"C"}) \left[ (1 - BS_{\text{eff}}^{i,j})(1 - \text{seq}_{\text{err}}^{j}) + BS_{\text{eff}}^{i,j}\text{seq}_{\text{err}}^{j} \right] + \]

\[
= p^{i,j}_{\text{CAB}}(\text{"C"}|\text{C}) \]

\[
p^{i,j}_{\text{CAB}}(\text{"5mC"}) \left[ (1 - BS_{\text{eff}}^{i,j})(1 - \text{seq}_{\text{err}}^{j}) + BS_{\text{eff}}^{i,j}\text{seq}_{\text{err}}^{j} \right] + \]

\[
= p^{i,j}_{\text{CAB}}(\text{"C"}|\text{5mC}) \]

\[
p^{i,j}_{\text{CAB}}(\text{"5hmC"}) \left[ (1 - BS_{\text{eff}}^{i,j})(1 - \text{seq}_{\text{err}}^{j}) + BS_{\text{eff}}^{i,j}\text{seq}_{\text{err}}^{j} \right] + \]

\[
= p^{i,j}_{\text{CAB}}(\text{"C"}|\text{5hmC}) \]

\[
p^{i,j}_{\text{CAB}}(\text{"5fC"}) \left[ (1 - BS_{\text{eff}}^{i,j})(1 - \text{seq}_{\text{err}}^{j}) + BS_{\text{eff}}^{i,j}\text{seq}_{\text{err}}^{j} \right] + \]

\[
= p^{i,j}_{\text{CAB}}(\text{"C"}|\text{5fC}) \]

\[
p^{i,j}_{\text{CAB}}(\text{"5caC"}) \left[ (1 - BS_{\text{eff}}^{i,j})(1 - \text{seq}_{\text{err}}^{j}) + BS_{\text{eff}}^{i,j}\text{seq}_{\text{err}}^{j} \right] \]

\[
= p^{i,j}_{\text{CAB}}(\text{"C"}|\text{5caC}) \]

\[
(9) \]

\[11\]
redBS-seq

\[
p_{\text{redBS}}^{i,j} (\text{"C"}) = p_{\text{redBS}}^{i,j} (\text{"C"}) \left[ (1 - \text{BS}_\text{eff}^{i,j})(1 - \text{seq}_\text{err}^{i,j}) + \text{BS}_\text{eff}^{i,j}\text{seq}_\text{err}^{i,j} \right] + \\
= p_{\text{redBS}}^{i,j} (\text{"C"}|\text{C}) \\
p_{\text{redBS}}^{i,j} (\text{"5mC"}) \left[ (1 - \text{BS}_\text{eff}^{i,j})(1 - \text{seq}_\text{err}^{i,j}) + \text{BS}_\text{eff}^{i,j}\text{seq}_\text{err}^{i,j} \right] + \\
= p_{\text{redBS}}^{i,j} (\text{"5mC"}|\text{5mC}) \\
p_{\text{redBS}}^{i,j} (\text{"5hmC"}) \left[ (1 - \text{BS}_\text{eff}^{i,j})(1 - \text{seq}_\text{err}^{i,j}) + \text{BS}_\text{eff}^{i,j}\text{seq}_\text{err}^{i,j} \right] + \\
= p_{\text{redBS}}^{i,j} (\text{"5hmC"}|\text{5hmC}) \\
p_{\text{redBS}}^{i,j} (\text{"5fC"}) \left[ (1 - \text{BS}_\text{eff}^{i,j})(1 - \text{seq}_\text{err}^{i,j}) + \text{BS}_\text{eff}^{i,j}\text{seq}_\text{err}^{i,j} \right] + \\
= p_{\text{redBS}}^{i,j} (\text{"5fC"}|\text{5fC}) \\
p_{\text{redBS}}^{i,j} (\text{"5caC"}) \left[ (1 - \text{BS}_\text{eff}^{i,j})(1 - \text{seq}_\text{err}^{i,j}) + \text{BS}_\text{eff}^{i,j}\text{seq}_\text{err}^{i,j} \right] + \\
= p_{\text{redBS}}^{i,j} (\text{"5caC"}|\text{5caC}) \\
\]

MAB-seq

\[
p_{\text{MAB}}^{i,j} (\text{"C"}) = p_{\text{MAB}}^{i,j} (\text{"C"}) \left[ (1 - \text{BS}_\text{eff}^{i,j})(1 - \text{seq}_\text{err}^{i,j}) + \text{BS}_\text{eff}^{i,j}\text{seq}_\text{err}^{i,j} \right] + \\
= p_{\text{MAB}}^{i,j} (\text{"C"}|\text{C}) \\
p_{\text{MAB}}^{i,j} (\text{"5mC"}) \left[ (1 - \text{BS}_\text{eff}^{i,j})(1 - \text{seq}_\text{err}^{i,j}) + \text{BS}_\text{eff}^{i,j}\text{seq}_\text{err}^{i,j} \right] + \\
= p_{\text{MAB}}^{i,j} (\text{"5mC"}|\text{5mC}) \\
p_{\text{MAB}}^{i,j} (\text{"5hmC"}) \left[ (1 - \text{BS}_\text{eff}^{i,j})(1 - \text{seq}_\text{err}^{i,j}) + \text{BS}_\text{eff}^{i,j}\text{seq}_\text{err}^{i,j} \right] + \\
= p_{\text{MAB}}^{i,j} (\text{"5hmC"}|\text{5hmC}) \\
p_{\text{MAB}}^{i,j} (\text{"5fC"}) \left[ (1 - \text{BS}_\text{eff}^{i,j})(1 - \text{seq}_\text{err}^{i,j}) + \text{BS}_\text{eff}^{i,j}\text{seq}_\text{err}^{i,j} \right] + \\
= p_{\text{MAB}}^{i,j} (\text{"5fC"}|\text{5fC}) \\
p_{\text{MAB}}^{i,j} (\text{"5caC"}) \left[ (1 - \text{BS}_\text{eff}^{i,j})(1 - \text{seq}_\text{err}^{i,j}) + \text{BS}_\text{eff}^{i,j}\text{seq}_\text{err}^{i,j} \right] + \\
= p_{\text{MAB}}^{i,j} (\text{"5caC"}|\text{5caC}) \\
\]

For a particular combination of different BS-seq and "oxi-mC"-seq data sets, the likelihood will be computed as in Equation (10) using the emission probabilities defined in Equations (6-11).

For a particular combination of different BS-seq and "oxi-mC"-seq data sets, the likelihood will be computed as in Equation (10) using the emission probabilities defined in Equations (6-11).
On the difference of two independent Dirichlet random variables

Let $X_1$ and $X_2$ be independent Dirichlet random variables of order 3 with parameters $\alpha_1$ and $\alpha_2$, respectively. Consequently, the probability density functions of $X_1$ and $X_2$ are

$$f_{X_j}(\mathbf{x}_j; \alpha_j) = \frac{1}{\mathcal{B}(\alpha_j)} \prod_{i=1}^{3} x_{j,i}^{\alpha_{j,i}-1}, \quad (12)$$

where $\mathcal{B}(\alpha_j) = \left( \prod_{i=1}^{3} \Gamma(\alpha_{j,i}) \right) / \left( \Gamma \left( \sum_{i=1}^{3} \alpha_{j,i} \right) \right)$ and $j \in \{1, 2\}$.

Alternatively, because the following statement $\sum_{i=1}^{3} x_{j,i} = 1$ holds Equation $(12)$ can be equivalently stated as

$$f_{X_j}(\mathbf{x}_j; \alpha_j) = \frac{1}{\mathcal{B}(\alpha_j)} x_{j,1}^{\alpha_{j,1}-1} x_{j,2}^{\alpha_{j,2}-1} (1 - x_{j,1} - x_{j,2})^{\alpha_{j,3}-1}, \quad (13)$$

where $j \in \{1, 2\}$.

Now let us consider the difference of the random variables $X_1$ and $X_2$, i.e., the random variable $Y = X_1 - X_2$. The probability density function of $Y$ is the convolution of the probability density functions of $X_1$ and $X_2$

$$f_{Y}(\mathbf{y}) = \int_{\mathcal{S}} f_{X_1}(\mathbf{x}; \alpha) f_{X_2}(\mathbf{x} - \mathbf{y}; \alpha) \, d\mathbf{x} \quad (14)$$

Let us follow the procedure described by Wolpert and Wolf to solve the integral. If we denote $h_i(x_i) = x_i^{\alpha_{i,1}-1}(x_i - y_i)^{\alpha_{i,2}-1}$, then Equation $(14)$ can be stated as

$$f_{Y}(\mathbf{y}) = \int_{\mathcal{S}} f_{X_1}(\mathbf{x}; \alpha) f_{X_2}(\mathbf{x} - \mathbf{y}; \alpha) \, d\mathbf{x}$$

$$= \frac{1}{\mathcal{B}(\alpha_1)} \frac{1}{\mathcal{B}(\alpha_2)} \int_{0}^{1} \int_{0}^{1-x_1} h_1(x_1) h_2(x_2) h_3(1 - x_1 - x_2) \, dx_2 \, dx_1 \quad (15)$$

Now let us define $\tau_1$, $\tau_2$ and $\tau_3$ recursively as $\tau_1 = \sum_{i=1}^{3} x_i = 1$ and $\tau_k = \tau_{k-1} - x_{k-1}$, then $\tau_k =$
\[ \tau_1 - \sum_{i=1}^{k-1} x_i. \] Thus, the integral of Equation (15) can be equivalently stated as

\[
f_Y(y) = \frac{1}{B(\alpha_1) B(\alpha_2)} \int_0^{\tau_1} h_1(x_1) \int_0^{\tau_2} h_2(x_2) h_3(\tau_2 - x_2) \, dx_2 \, dx_1
\]

\[
= \frac{1}{B(\alpha_1) B(\alpha_2)} \int_0^{\tau_1} h_1(x_1) (h_2 * h_3) (\tau_2) \, dx_1
\]

\[
= \frac{1}{B(\alpha_1) B(\alpha_2)} \int_0^{\tau_1} h_1(x_1) (h_2 * h_3) (\tau_1 - x_1) \, dx_1
\]

\[
= \frac{1}{B(\alpha_1) B(\alpha_2)} (h_1 * h_2 * h_3) (\tau_1)
\]

\[
= \frac{1}{B(\alpha_1) B(\alpha_2)} (h_1(x_1) * h_2(x_2) * h_3(x_3)) (1).
\]  

We know that the Laplace transformation has the property \( \mathcal{L} [h_1(x_1) * h_2(x_2) * h_3(x_3)] = \prod_{i=1}^{3} \mathcal{L} [h_i(x_i)]. \) Since we are only interested on the probability density at the origin, we may assume \( y_i = 0, \ i \in \{1, 2, 3\}, \) then

\[
\mathcal{L}_{x_i} [h_i(x_i)] = \mathcal{L}_{x_i} \left[ x_i^{\alpha_{1,i}-1} x_i^{\alpha_{2,i}-1} \right] = s^{\alpha_{1,i}-\alpha_{2,i}+1} \Gamma (\alpha_{1,i} + \alpha_{2,i} - 1),
\]  

where \( i \in \{1, 2, 3\}. \) Then by applying the inverse Laplace transformation to \( \prod_{i=1}^{3} \mathcal{L} [h_i(x_i)] \) we get

\[
\mathcal{L}^{-1} \left[ \prod_{i=1}^{3} \mathcal{L}[h_i(x_i)] \right] = \prod_{i=1}^{3} \Gamma (\alpha_{1,i} + \alpha_{2,i} - 1) \mathcal{L}^{-1} \left[ s^{\sum_{i=1}^{3} (\alpha_{1,i}-\alpha_{2,i}+1)} \right] (1)
\]

\[
\mathcal{L}^{-1} \left[ \prod_{i=1}^{3} \mathcal{L}[h_i(x_i)] \right] = \frac{\prod_{i=1}^{3} \Gamma (\alpha_{1,i} + \alpha_{2,i} - 1)}{\Gamma \left( \sum_{i=1}^{3} (\alpha_{1,i} + \alpha_{2,i} - 1) \right)} (18)
\]

Finally, we obtain the probability density at the origin of the difference of two Dirichlet random variables with the concentration parameters \( \alpha_1 \) and \( \alpha_2 \)

\[
f_Y(0) = \frac{1}{B(\alpha_1) B(\alpha_2)} \frac{\prod_{i=1}^{3} \Gamma (\alpha_{1,i} + \alpha_{2,i} - 1)}{\Gamma \left( \sum_{i=1}^{3} (\alpha_{1,i} + \alpha_{2,i} - 1) \right)}
\]

\[
= \frac{\Gamma \left( \sum_{i=1}^{3} (\alpha_{1,i}) \right) \prod_{i=1}^{3} \Gamma (\alpha_{2,i}) \Gamma \left( \sum_{i=1}^{3} (\alpha_{1,i} + \alpha_{2,i} - 1) \right)}{\prod_{i=1}^{3} \Gamma (\alpha_{1,i}) \prod_{i=1}^{3} \Gamma (\alpha_{2,i})}
\]  

(19)

On the approximation of the beta distribution with the reparameterized versions

The Beta distribution has very intuitive properties when specifying the distribution of the parameter \( p \) of a Bernoulli trial \( \text{Be}(p) \) or a binomial experiment \( \text{B}(n, p) \). For example, to define the prior knowledge of
n trials with k successes, we use a prior of \( p \sim \text{Beta}(k, n - k) \). However, when implementing a sampler for a hierarchical probability model, the Beta distribution has the disadvantage of being restricted to the \([0, 1]\) support and lacking a noncentered parametrization. As shown in Papaspiliopoulos et al. (2007), the noncentered parametrization enables significantly better sampling in hierarchical models. In the Stan sampler, variables with restricted support are automatically transformed into unrestricted equivalents in order to remove boundaries that would need to be treated separately (Stan Development Team 2014). Currently, however, there is no automatic noncentering transformation. Therefore, to model the parameter \( p \) of a binomial experiment \( B(n, p) \), we chose to use a logistic-normal distribution (Aitchison and Shen 1980) that has a noncentered variant. To enjoy the best of both worlds, we initially elicit the required priors as Beta distributions and then choose the logistic-normal distribution that best approximates them.

The best approximation \( P \) of a distribution \( Q \) in the information theoretic sense is the one that minimizes the Kullback–Liebler distance, or discrepancy (Bernardo and Smith 2000 page 76), from \( Q \) to \( P \).

\[
\delta(P \mid Q) = \mathbb{E}_Q \left[ \log \frac{Q}{P} \right].
\] (20)

Given the logistic map \( L \) and its inverse \( L^{-1} \)

\[
L(x) = \frac{1}{1 + e^{-x}} \quad \text{and} \quad L^{-1}(y) = \log \left( \frac{y}{1 - y} \right),
\] (21)

the logistic-normal distribution with parameters \( \mu \) and \( \sigma \) is defined as

\[
\text{LN}(x \mid \mu, \sigma) = \frac{1}{x(1-x)} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{1}{2} \left( \frac{L^{-1}(x) - \mu}{\sigma} \right)^2 \right].
\] (22)

The Beta distribution with parameters \( \alpha \) and \( \beta \) is defined as

\[
\text{Beta}(x \mid \alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\text{B}(\alpha, \beta)}.
\] (23)
The discrepancy from \( \text{LN}(x; \mu, \sigma) \) to \( \text{Beta}(x; \alpha, \beta) \) is

\[
\delta(\text{LN}(x) \mid \text{Beta}(x)) = \int_0^1 \log \left( \frac{\text{Beta}(x; \alpha, \beta)}{\text{LN}(x; \mu, \sigma)} \right) \text{Beta}(x; \alpha, \beta) \, dx
= \mathbb{E} \left[ \log \left( \frac{\text{Beta}(x; \alpha, \beta)}{\text{LN}(x; \mu, \sigma)} \right) \right]
= \mathbb{E} \log \left[ \frac{x^\alpha (1 - x)^\beta}{B(\alpha, \beta)} \sqrt{2\pi \sigma^2} \exp \left( \frac{1}{2} \frac{(\text{L}^{-1}(x) - \mu)^2}{\sigma^2} \right) \right]
= \mathbb{E} \left[ \alpha \log x + \beta \log(1-x) - \log B(\alpha, \beta) + \log \sqrt{2\pi \sigma^2} + \frac{1}{2} \frac{(\text{L}^{-1}(x) - \mu)^2}{\sigma^2} \right].
\]  

(24)

where all the expectations are taken with respect to the \( \text{Beta}(x; \alpha, \beta) \) distribution. To find the values of the parameters \( \mu \) and \( \sigma \) that minimize the discrepancy, we solve the roots of the derivatives:

\[
D_\mu \delta(\text{LN}(x) \mid \text{Beta}(x)) = \mathbb{E} \left[ \frac{1}{2} \frac{(\text{L}^{-1}(x) - \mu)^2}{\sigma^2} \right]
= \frac{1}{2\sigma^2} \mathbb{E} \left[ D_\mu (\text{L}^{-1}(x) - \mu)^2 \right]
= \frac{1}{2\sigma^2} \mathbb{E} \left[ -2(\text{L}^{-1}(x) - \mu) \right]
= -\frac{1}{\sigma^2} \mathbb{E} \left[ \text{L}^{-1}(x) \right] + \frac{1}{\sigma^2} \mathbb{E}[\mu]
= -\frac{1}{\sigma^2} \left[ \psi(\alpha) - \psi(\beta) \right] + \frac{1}{\sigma^2} \mu
= 0,
\]

(25)

where \( \psi \) is the digamma function. Equation (25) uses the identities in (28). The solution of the equation is \( \mu = \psi(\alpha) - \psi(\beta) \).
\[ D_{\sigma^2} \delta (\ln(x) | \text{Beta}(x)) = D_{\sigma^2} E \left[ \log \sqrt{2\pi\sigma^2} + \frac{1}{2} \left( \frac{L^{-1}(x) - \mu}{\sigma^2} \right)^2 \right] \]

\[ = E \left[ \frac{1}{2} D_{\sigma^2} \log 2\pi\sigma^2 + \frac{1}{2} D_{\sigma^2} \left( \frac{L^{-1}(x) - \mu}{\sigma^2} \right)^2 \right] \]

\[ = E \left[ \frac{1}{2} \sigma^{-2} - \frac{1}{2} \sigma^{-4} (L^{-1}(x) - \mu)^2 \right] \]

\[ = \frac{1}{2} \sigma^{-2} - \frac{1}{2} \sigma^{-4} E \left[ (L^{-1}(x) - \mu)^2 \right] \]

\[ = \frac{1}{2} \sigma^{-2} - \frac{1}{2} \sigma^{-4} E \left[ \log^2 x + \log^2 (1 - x) + \mu^2 \right. \]

\[ - 2 \log x \log (1 - x) + 2\mu \log x + 2\mu \log (1 - x) \]

\[ \left. = \frac{1}{2} \sigma^{-2} - \frac{1}{2} \sigma^{-4} \left[ (\psi(\alpha) - \psi(\alpha + \beta))^2 + \psi_1(\alpha) - \psi_1(\alpha + \beta) \right. \right. \]

\[ + (\psi(\beta) - \psi(\alpha + \beta))^2 + \psi_1(\beta) - \psi_1(\alpha + \beta) \]

\[ + \mu^2 \]

\[ - 2(\psi(\alpha) - \psi(\alpha + \beta))(\psi(\beta) - \psi(\alpha + \beta)) + 2\psi_1(\alpha + \beta) \]

\[ - 2\mu\psi(\alpha) + 2\mu\psi(\alpha + \beta) + 2\mu\psi(\beta) - 2\mu\psi(\alpha + \beta) \]

\[ = \frac{1}{2} \sigma^{-2} - \frac{1}{2} \sigma^{-4} \left[ \psi_1(\alpha) + \psi_1(\beta) \right. \]

\[ + \psi^2(\alpha) + \psi^2(\alpha + \beta) - 2\psi(\alpha)\psi(\alpha + \beta) \]

\[ + \psi^2(\beta) + \psi^2(\alpha + \beta) - 2\psi(\beta)\psi(\alpha + \beta) \]

\[ + \psi^2(\alpha) + \psi^2(\beta) - 2\psi(\alpha)\psi(\beta) \]

\[ - 2\psi(\alpha)\psi(\beta) + 2\psi(\alpha)\psi(\alpha + \beta) + 2\psi(\beta)\psi(\alpha + \beta) - 2\psi^2(\alpha + \beta) \]

\[ - 2\psi^2(\alpha) + 2\psi(\beta)\psi(\alpha) - 2\psi^2(\beta) + 2\psi(\alpha)\psi(\beta) \]

\[ = \frac{1}{2} \sigma^{-2} - \frac{1}{2} \sigma^{-4} \left[ \psi_1(\alpha) + \psi_1(\beta) \right] \]

\[ = 0, \]

where \( \psi_1 \) is the trigamma function. Equation (26) uses the identities in (28). The solution of the equation is \( \sigma^2 = \psi_1(\alpha) + \psi_1(\beta) \).
Conclusion

We have shown that to approximate, in the K–L sense, a Beta distribution as defined in (23) with a logistic-normal distribution as defined in (22) one should choose the parameters as follows:

\[ \mu = \psi(\alpha) - \psi(\beta), \]
\[ \sigma^2 = \psi_1(\alpha) + \psi_1(\beta). \]  

(27)

Logarithmic moments of the Beta distribution

\[ E[\ln(X)] = \psi(\alpha) - \psi(\alpha + \beta), \]
\[ E[\ln(1 - X)] = \psi(\beta) - \psi(\alpha + \beta), \]
\[ E[\ln^2(X)] = (\psi(\alpha) - \psi(\alpha + \beta))^2 + \psi_1(\alpha) - \psi_1(\alpha + \beta), \]
\[ E[\ln^2(1 - X)] = (\psi(\beta) - \psi(\alpha + \beta))^2 + \psi_1(\beta) - \psi_1(\alpha + \beta), \]
\[ E[\ln(X) \ln(1 - X)] = (\psi(\alpha) - \psi(\alpha + \beta))(\psi(\beta) - \psi(\alpha + \beta)) - \psi_1(\alpha + \beta). \]  

(28)

Maximum likelihood estimators

In this section we will consider frequentist estimation for methylation levels from BS-seq/oxBS-seq and BS-seq/TAB-seq/fCAB-seq data. We show that the estimators considered in Booth (2012) and in Wang et al. (2014) for BS-seq/oxBS-seq and BS-seq/TAB-seq/fCAB-seq data, respectively, correspond to (unconstrained) maximum likelihood estimators.

BS-seq and oxBS-seq

Let us consider a single cytosine for the sake of notation simplicity. Assume the proportions of unmethylated C (C), methylated C (5mC), and hydroxymethylated C (5hmC) are governed by the probabilities \( p(\text{“C”}), p(\text{“mC”}), \) and \( p(\text{“5hmC”}). \) Moreover, let \( \theta = (p(\text{“C”}), p(\text{“5mC”}), p(\text{“5hmC”})). \) denote a vector composed by the these unknown probabilities.

Using the conversion chart and \( \theta \), we can state the probabilities of the BS-seq and oxBS-seq
outcomes with as follows

\[ p_{\text{BS}}(\text{"C"}|\theta) = 1 - p(\text{"C"}) = p(\text{"5mC"}) + p(\text{"5hmC"}) \]
\[ p_{\text{BS}}(\text{"T"}|\theta) = p(\text{"C"}) = 1 - p(\text{"5mC"}) + p(\text{"5hmC"}) \]
\[ p_{\text{oxBS}}(\text{"C"}|\theta) = p(\text{"5mC"}) = 1 - p(\text{"C"}) + p(\text{"5hmC"}) \]
\[ p_{\text{oxBS}}(\text{"T"}|\theta) = p(\text{"5mC"}) = p(\text{"C"}) + p(\text{"5hmC"}). \]

(29)

Let us denote BS-seq and oxBS-seq data as follows

\[ D = \{ N_{\text{BS},C}, N_{\text{BS}}, N_{\text{oxBS},C}, N_{\text{oxBS}} \}, \]

(30)

where \( N_{\text{BS},C} \) and \( N_{\text{oxBS},C} \) are the numbers of \( C \) reads out of \( N_{\text{BS}} > 0 \) (\( C \) reads and \( T \) reads from BS-seq) and \( N_{\text{oxBS}} > 0 \) (\( C \) reads and \( T \) reads from oxBS-seq) total reads, respectively.

Then the likelihood of \( \theta \), given \( D \) is

\[ L(\theta|D) \propto p(\text{"C"})^{N_{\text{BS}} - N_{\text{BS},C}} (1 - p(\text{"C"}))^{N_{\text{BS},C}} \]
\[ (1 - p(\text{"5mC"}))^{N_{\text{oxBS}} - N_{\text{oxBS},C}} p(\text{"5mC"})^{N_{\text{oxBS},C}}, \]

with the following constraints \( p(\text{"C"}) + p(\text{"5mC"}) + p(\text{"5hmC"}) = 1 \) (and \( p(\text{"C"}), p(\text{"5mC"}), p(\text{"5hmC"}) \in [0, 1] \)).

Next, let us calculate the gradient

\[
\nabla L(\theta|D) = \begin{cases}
\frac{\partial L(\theta|D)}{\partial p(\text{"C"})} = p(\text{"C"})^{N_{\text{BS}} - N_{\text{BS},C}} (1 - p(\text{"C"}))^{N_{\text{BS},C}} \\
(1 - p(\text{"5mC"}))^{N_{\text{oxBS}} - N_{\text{oxBS},C}} p(\text{"5mC"})^{N_{\text{oxBS},C}} \\
(N_{\text{BS}} - N_{\text{BS},C})(p(\text{"C"}) - 1) + N_{\text{BS},C} p(\text{"C"}) \]
\frac{\partial L(\theta|D)}{\partial p(\text{"5mC"})} = p(\text{"5mC"})^{N_{\text{BS}} - N_{\text{BS},C}} (1 - p(\text{"C"}))^{N_{\text{BS},C}} \\
(1 - p(\text{"5mC"}))^{N_{\text{oxBS}} - N_{\text{oxBS},C}} p(\text{"5mC"})^{N_{\text{oxBS},C}} \\
(N_{\text{oxBS}} - N_{\text{oxBS},C})(p(\text{"5mC"}) + N_{\text{oxBS},C} p(\text{"5mC"}) - 1))
\end{cases}.
\]

(32)

Then we set \( \nabla L(\theta|D) = 0 \). We will not consider trivial solutions. The nontrivial solution is \( p(\text{"C"}) = (N_{\text{BS}} - N_{\text{BS},C})/N_{\text{BS}} \wedge p(\text{"5mC"}) = N_{\text{oxBS},C}/N_{\text{oxBS}} \).

Note that we have assumed that \( p(\text{"C"}) + p(\text{"5mC"}) + p(\text{"5hmC"}) = 1 \). Then we can state the
(unconstrained) maximum likelihood estimator (MLE) of $\theta$

$$\hat{\theta} = \begin{pmatrix} \hat{p}("C") \\ \hat{p}("5mC") \\ \hat{p}("5hmC") \end{pmatrix} = \begin{pmatrix} \frac{N_{BS} - N_{BS,C}}{N_{BS}} \\ \frac{N_{oxBS,C}/N_{oxBS}}{1 - (N_{BS} - N_{BS,C})/N_{BS} - N_{oxBS,C}/N_{oxBS}} \end{pmatrix} = \begin{pmatrix} N_{BS,T}/N_{BS} \\ N_{oxBS,C}/N_{oxBS} \end{pmatrix}. \quad (33)$$

**BS-seq, TAB-seq and fCAB-seq**

Similarly as in the previous section, let us assume the proportions of C, 5mC, 5hmC, and 5fC are governed by the probabilities $p("C"), p("5mC"), p("5hmC")$ and $p("5fC")$. Moreover, let $\theta = (p("C"), p("5mC"), p("5hmC"), p("5fC"))^T$ denote a vector composed by the these unknown probabilities.

Using the conversion chart and $\theta$, we can state the probabilities of the BS-seq, TAB-seq, and fCAB-seq outcomes as follows

$$p_{BS}("C|\theta") = 1 - p("C") - p("5fC") = p("5mC") + p("5hmC")$$

$$p_{BS}("T|\theta") = p("C") + p("5fC") = 1 - p("5mC") - p("5hmC")$$

$$p_{TAB}("C|\theta") = p("5hmC") = 1 - p("C") - p("5mC") - p("5fC")$$

$$p_{TAB}("T|\theta") = 1 - p("5hmC") = p("C") + p("5mC") + p("5fC")$$

$$p_{fCAB}("C|\theta") = p("5mC") + p("5hmC") + p("5fC") = 1 - p("C")$$

$$p_{fCAB}("T|\theta") = 1 - p("5mC") - p("5hmC") - p("5fC") = p("C"). \quad (34)$$

Let us denote BS-seq, TAB-seq, and fCAB-seq data as follows

$$\mathcal{D} = \{N_{BS,C}, N_{BS}, N_{TAB,C}, N_{TAB}, N_{fCAB,C}, N_{fCAB}\}, \quad (35)$$

where $N_{BS,C}$, $N_{TAB,C}$ and $N_{fCAB,C}$ are the numbers of C reads out of $N_{BS} > 0$ (C reads and T reads from BS-seq), $N_{TAB} > 0$ (C reads and T reads from TAB-seq), and $N_{fCAB} > 0$ (C reads and T reads from fCAB-seq) total reads, respectively.
Then the likelihood of $\theta$, given $D$ is

$$
\mathcal{L}(\theta|D) = (p(\text{"C"}) + p(\text{"5fC"}))^{N_{\text{BS}\rightarrow\text{BS,C}}} (1 - p(\text{"C"}) - p(\text{"5fC"}))^{N_{\text{BS,C}}} \\
(1 - p(\text{"5hmC"}))^{N_{\text{TAB}\rightarrow\text{N}_{\text{TAB,C}}} p(\text{"5hmC"})^{N_{\text{TAB,C}}}} \\
p(\text{"C"})^{N_{\text{ICAB}\rightarrow\text{N}_{\text{ICAB,C}}} (1 - p(\text{"C"}))^{N_{\text{ICAB,C}}} ,}
$$

with the following constraints $p(\text{"C"}) + p(\text{"5mC"}) + p(\text{"5hmC"}) + p(\text{"5fC"}) = 1$ and $p(\text{"C"}), p(\text{"5mC"}), p(\text{"5hmC"}), p(\text{"5fC"}) \in [0, 1]$.  

Next, let us calculate the gradient

$$
\nabla \mathcal{L}(\theta|D) = \begin{cases} 
\frac{\partial \mathcal{L}(\theta|D)}{\partial p(\text{"C"})} = (1 - p(\text{"5hmC"}))^{N_{\text{TAB}\rightarrow\text{N}_{\text{TAB,C}}} p(\text{"5hmC"})^{N_{\text{TAB,C}}}} \\
(1 - p(\text{"5hmC"}))^{N_{\text{TAB}\rightarrow\text{N}_{\text{TAB,C}}} p(\text{"5hmC"})^{N_{\text{TAB,C}}}} \\
p(\text{"C"})^{N_{\text{ICAB}\rightarrow\text{N}_{\text{ICAB,C}}} (1 - p(\text{"C"}))^{N_{\text{ICAB,C}}} ,}
\end{cases}
$$

$$
\nabla \mathcal{L}(\theta|D) = \begin{cases} 
\frac{\partial \mathcal{L}(\theta|D)}{\partial p(\text{"5hmC"})} = (1 - p(\text{"5hmC"}))^{N_{\text{TAB}\rightarrow\text{N}_{\text{TAB,C}}} p(\text{"5hmC"})^{N_{\text{TAB,C}}}} \\
(1 - p(\text{"5hmC"}))^{N_{\text{TAB}\rightarrow\text{N}_{\text{TAB,C}}} p(\text{"5hmC"})^{N_{\text{TAB,C}}}} \\
p(\text{"C"})^{N_{\text{ICAB}\rightarrow\text{N}_{\text{ICAB,C}}} (1 - p(\text{"C"}))^{N_{\text{ICAB,C}}} ,}
\end{cases}
$$

Next, let us set $\nabla \mathcal{L}(\theta|D) = 0$. Moreover, let us skip the trivial solutions.

Firstly, we observe that $\frac{\partial \mathcal{L}(\theta|D)}{\partial p(\text{"5hmC"})} = 0$ when $p(\text{"5hmC"}) = N_{\text{TAB,C}}/N_{\text{TAB}}$.

Secondly, we observe that $\frac{\partial \mathcal{L}(\theta|D)}{\partial p(\text{"5fC"})} = 0$ when $p(\text{"5fC"}) + p(\text{"5C"}) = (N_{\text{BS}} - N_{\text{BS,C}})/N_{\text{BS}}$. Then, using the identity $1 - p(\text{"C"}) - p(\text{"5fC"}) = p(\text{"5mC"}) + p(\text{"5hmC"})$ we get $p(\text{"5mC"}) = 1 - (N_{\text{BS}} - N_{\text{BS,C}})/N_{\text{BS}} - N_{\text{TAB,C}}/N_{\text{TAB}} = N_{\text{BS,C}}/N_{\text{BS}} - N_{\text{TAB,C}}/N_{\text{TAB}}$.  

21
Then, let us check when the following term is zero

\[
(N_{BS} - N_{BS,C})(p(“C”) - 1)p(“C”)(p(“5fC”) + p(“C”) - 1) + \\
(p(“5fC”) + p(“C”))(N_{BS,C}(p(“C”) - 1)p(“C”)) + \\
(p(“5fC”) + p(“C”) - 1)((N_{ICAB} - N_{ICAB,C})(p(“C”) - 1) + N_{ICAB,C}p(“C”)) = 0
\]  

(38)

Next, let us make the substitution \(p(“C”) + p(“5fC”) = (N_{BS} - N_{BS,C})/N_{BS} = N_{BS,T}/N_{BS}\)

\[
(N_{BS} - N_{BS,C})(p(“C”) - 1)p(“C”)(N_{BS,T}/N_{BS} - 1) + \\
N_{BS,T}/N_{BS}(N_{BS,C}(p(“C”) - 1)p(“C”)) + \\
(N_{BS,T}/N_{BS} - 1) \left( (N_{ICAB} - N_{ICAB,C})(p(“C”) - 1) + N_{ICAB,C}p(“C”) \right) = 0
\]

\[
= N_{ICAB}(“C”) - N_{ICAB,T}
\]  

(39)

Let us then rewrite some of the terms and simplify

\[
- (N_{BS,T}/N_{BS})N_{BS,C}(p(“C”) - 1)p(“C”) + \\
(N_{BS,T}/N_{BS})N_{BS,C}(p(“C”) - 1)p(“C”) - \\
(N_{BS,T}/N_{BS})(N_{BS,C}/N_{BS})(N_{ICAB}p(“C”) - N_{ICAB,T}) = 0,
\]

which can be simplified to

\[
N_{BS,T}N_{BS,C}p(“C”) = N_{BS,T}N_{BS,C}(N_{ICAB,T}/N_{ICAB}).
\]  

(41)

Then by assuming \(N_{BS,T}N_{BS,C} \neq 0\) we get

\[
p(“C”) = N_{ICAB,T}/N_{ICAB}.
\]  

(42)

By combining \(p(“C”) = N_{ICAB,T}/N_{ICAB}\) and \(p(“5fC”) + p(“5C”) = (N_{BS} - N_{BS,C})/N_{BS}\)

\[
p(“5fC”) = (N_{BS} - N_{BS,C})/N_{BS} - N_{ICAB,T}/N_{ICAB}.
\]  

(43)
Finally, we can state the (unconstrained) MLE of $\theta$

$$\hat{\theta} = \begin{pmatrix} \hat{\rho}(\text{"C"}) \\ \hat{\rho}(\text{"5mC"}) \\ \hat{\rho}(\text{"5hmC"}) \\ \hat{\rho}(\text{"5fC"}) \end{pmatrix} = \begin{pmatrix} (N_{\text{ICAB}} - N_{\text{ICAB},C})/N_{\text{ICAB}} \\ N_{\text{BS},C}/N_{\text{BS}} - N_{\text{TAB},C}/N_{\text{TAB}} \\ N_{\text{TAB},C}/N_{\text{TAB}} \\ (N_{\text{BS}} - N_{\text{BS},C})/N_{\text{BS}} - (N_{\text{ICAB}} - N_{\text{ICAB},C})/N_{\text{ICAB}} \end{pmatrix} = \begin{pmatrix} N_{\text{ICAB,T}}/N_{\text{ICAB}} \\ N_{\text{BS},C}/N_{\text{BS}} - N_{\text{TAB},C}/N_{\text{TAB}} \\ N_{\text{TAB},C}/N_{\text{TAB}} \\ N_{\text{BS},T}/N_{\text{BS}} - N_{\text{ICAB,T}}/N_{\text{ICAB}} \end{pmatrix}. \quad (44)$$

References


