Правило Маферона

16 февраля 2021 г.

К. О. Вайсер[1]

Работа посвящена эффективному сэмплированию из гауссовских процессов. Стандартная схема с репараметризацией требует минимум $O(m^3)$ операций для сэмплирования m значений. Было разработано несколько методов представления гауссовского процесса для увеличения скорости сэмплирования. Ключевой идеей является разложение апостериорного распределения в сумму двух слагаемых —априорного и некоторого обновления распределения. Представление этих слагаемых в определенном виде позволяет увеличить эффективность сэмлирования. Теоретическим обоснованием возможности подобного разложения служит предлагаемая к рассмотрению теорема.

Теорема 1. *(Правило Маферона). Пусть* $\mathbf{a}$ *и* $\mathbf{b}$ *- имеют совместное нормальное многомерное распределение. Тогда условное распределение на* $\mathbf{a}$ *при* $\mathbf{b} = \beta$ *вычисляется как*

$$(\boldsymbol{a} \mid \boldsymbol{b} = \boldsymbol{\beta}) \stackrel{\mathrm{d}}{=} \boldsymbol{a} + \operatorname{Cov}(\boldsymbol{a}, \boldsymbol{b}) \operatorname{Cov}(\boldsymbol{b}, \boldsymbol{b})^{-1}(\boldsymbol{\beta} - \boldsymbol{b})$$

Доказательство. Доказательство можно провести двумя способами. Первый носит сугубо технический характер и представляет собой прямое вычисление матожидания и ковариационных матриц обоих частей по определению. В силу свойств нормального распределения, это будет означать совпадение функций плотности. Подробные вычисления читатель может увидеть здесь

Мы приведем более изящное доказательство. Для удобства проведем переобозначения : $\Sigma_{11} = \operatorname{Cov}(\mathbf{a}, \mathbf{a}),\ \Sigma_{12} = \operatorname{Cov}(\mathbf{a}, \mathbf{b}),\ \Sigma_{21} = \operatorname{Cov}(\mathbf{b}, \mathbf{a}),\ \Sigma_{22} = \operatorname{Cov}(\mathbf{b}, \mathbf{b})$. Пусть также $E\mathbf{a} = \boldsymbol{\mu}_1,\ E\mathbf{b} = \boldsymbol{\mu}_2$

Обозначим $\mathbf{z} = \mathbf{a} + \mathbf{A}\mathbf{b}$, где $\mathbf{A} = -\Sigma_{12}\Sigma_{22}^{-1}$. Тогда

$$\begin{aligned}
\operatorname{Cov}(\mathbf{z}, \mathbf{b}) &= \operatorname{Cov}(\mathbf{a}, \mathbf{b}) + \operatorname{Cov}(\mathbf{A}\mathbf{b}, \mathbf{b}) = \Sigma_{12} + \mathbf{A} \operatorname{var}(\mathbf{b}) \\
&= \Sigma_{12} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{22} = 0
\end{aligned}$$

[1]Московский физико-технический институт, vajser.ko@phystech.edu

Это значит, что $\mathbf{z}$ и $\mathbf{b}$ некореллированны и, вследствие их нормальности, независимы. Тогда $E\mathbf{z} = \boldsymbol{\mu}_1 + \mathbf{A}\boldsymbol{\mu}_2$. Из этого следует, что

$$\begin{aligned} E(\mathbf{a} \mid \mathbf{b}) &= E(\mathbf{z} - \mathbf{A}\mathbf{b} \mid \mathbf{b}) \\ &= E(\mathbf{z} \mid \mathbf{b}) - E(\mathbf{A}\mathbf{b} \mid \mathbf{b}) \\ &= E(\mathbf{z}) - \mathbf{A}\mathbf{b} \\ &= \boldsymbol{\mu}_1 + \mathbf{A}(\boldsymbol{\mu}_2 - \mathbf{b}) \\ &= \boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{b} - \boldsymbol{\mu}_2) \end{aligned}$$

Получается, что матожидания левой и правой части искомого тождества совпадают. Осталось рассмотреть ковариацию:

$$\begin{aligned} \operatorname{var}(\mathbf{a} \mid \mathbf{b}) &= \operatorname{var}(\mathbf{z} - \mathbf{A}\mathbf{b} \mid \mathbf{b}) \\ &= \operatorname{var}(\mathbf{z} \mid \mathbf{b}) + \operatorname{var}(\mathbf{A}\mathbf{b} \mid \mathbf{b}) - \mathbf{A}\operatorname{Cov}(\mathbf{b}, \mathbf{z}) - \operatorname{Cov}(\mathbf{z}, \mathbf{b})\mathbf{A}^\top \\ &= \operatorname{var}(\mathbf{z} \mid \mathbf{b}) \\ &= \operatorname{var}(\mathbf{z}) \end{aligned}$$

Раскроем получившееся выражение:

$$\operatorname{var}(\mathbf{a} \mid \mathbf{b}) = \operatorname{var}(\mathbf{z}) = \operatorname{var}(\mathbf{a} + \mathbf{A}\mathbf{b})$$

$$\begin{aligned} &= \operatorname{var}(\mathbf{a}) + \mathbf{A}\operatorname{var}(\mathbf{b})\mathbf{A}^\top + \mathbf{A}\operatorname{Cov}(\mathbf{b}, \mathbf{a}) + \operatorname{Cov}(\mathbf{a}, \mathbf{b})\mathbf{A}^\top \\ &= \Sigma_{11} + \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{22}\Sigma_{22}^{-1}\Sigma_{21} - 2\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \\ &= \Sigma_{11} + \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} - 2\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \\ &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \end{aligned}$$

Ковариации левой и правой части так же совпадают. В силу свойств нормального распределения отсюда следует искомое утверждение. □

Эта теорема находит множество применений, в том числе в представленной работе. Из правила Маферона следует важное

Утверждение 1. *Пусть $f \sim GP(0, k)$, $\mathbf{Z}$ - множество точек размера m и $f_m = f(\mathbf{Z})$. Тогда*

$$\underbrace{(f \mid f_m = \boldsymbol{u})(\cdot)}_{\textit{posterior}} \stackrel{\mathrm{d}}{=} \underbrace{f(\cdot)}_{\textit{prior}} + \underbrace{k(\cdot, \mathbf{Z})\mathbf{K}_{m,m}^{-1}(\boldsymbol{u} - \boldsymbol{f}_m)}_{\textit{update}}$$

Находя впоследствие удобные представления для априорной и обновленнной части распределения можно добиться большой эффективности сэмплирования.