Лекция 11. Методы Монте-Карло. Фильтр частиц.

Д. П. Ветров[1]

[1]МГУ, ВМиК, каф. ММП

Спецкурс «Структурные методы анализа изображений и сигналов»

План

План

Идея метода Монте-Карло

- Метод Монте-Карло применяется для решения задач численного моделирования, в частности взятия интегралов

$$\int_a^b f(x)dx \approx \frac{b-a}{n}\sum_{i=1}^n f(x_i) = \hat{f}, \quad x_i \sim U[a,b]$$

- Можно показать, что при весьма общих предположениях $\hat{f} \to \int_a^b f(x)dx$ при $n \to \infty$
- Точность оценки интегралов не зависит от размерности пространства d, а определяется исключительно дисперсией самой функции

$$\mathbb{D}\hat{f} = \frac{1}{n}\left[(b-a)\int f^2(x)dx - \left(\int f(x)dx\right)^2\right]$$

- Для численной оценки вероятностных интегралов необходимы специальные методы

Вероятностные интегралы

- В дальнейшем будем рассматривать интегралы вида

$$\mathbb{E}f = \int f(\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x}$$

- К ним сводятся многие интегралы, возникающие при байесовском обучении, в частности обоснованность

$$Evidence = \mathbb{E}_{\boldsymbol{w}}p(\boldsymbol{t}|\boldsymbol{w}) = \int p(\boldsymbol{t}|\boldsymbol{w})p(\boldsymbol{w})d\boldsymbol{w}$$

и голосование по апостериорному распределению

$$p(t_{new}|\boldsymbol{t}) = \mathbb{E}_{\boldsymbol{w}}p(t_{new}|\boldsymbol{w}) = \int p(t_{new}|\boldsymbol{w})p(\boldsymbol{w}|\boldsymbol{t})d\boldsymbol{w}$$

Особенности вероятностных интегралов

- Классическая выборка из равномерного распределения для взятия таких интегралов, т.е. формула

$$\int_D f(\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x} \approx \frac{|D|}{n}\sum f(\boldsymbol{x}_i)p(\boldsymbol{x}_i), \quad \boldsymbol{x} \sim U(D),$$

  крайне неэффективна, так как в большей части области интегрирования плотность, а, следовательно, и подынтегральная функция близка к нулю
- Для взятия интегралов вида $\int f(\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x}$ нужно уметь проводить выборку из распределения $p(\boldsymbol{x})$
- В этом случае интеграл может быть оценен конечной суммой

$$\int f(\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x} \approx \frac{1}{n}\sum f(\boldsymbol{x}_i), \quad \boldsymbol{x} \sim p(\boldsymbol{x})$$

Метод обратной функции

- В некоторых случаях можно свести задачу генерации выборки из некоторого распределения к генерации выборки из равномерного распределения
- Пусть $F(x) = P(X < x) = \int_{-\infty}^{x} p(\xi)d\xi$ — функция распределения случайной величины X
- Легко показать (Упр.), что $Y = F(X) \sim U(0,1)$, тогда $X \sim F^{-1}(U(0,1))$
- Так удается сгенерировать выборку из показательного распределения и распределения Коши (Упр.)

План

Схема с весами

- В дальнейшем полагаем, что нам в каждой точке известна плотность распределения величины с точностью до множителя, т.е.

$$p(\boldsymbol{x}) = \frac{1}{Z_p}\tilde{p}(\boldsymbol{x}),$$

причем Z_p неизвестна, а $\tilde{p}(\boldsymbol{x})$ может быть легко подсчитана в любой точке

- Введем распределение $q(\boldsymbol{x})$, из которого легко сгенерировать выборку, тогда

$$\mathbb{E}_p f = \int f(\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x} = \frac{1}{Z_p}\int f(\boldsymbol{x})\frac{\tilde{p}(\boldsymbol{x})}{q(\boldsymbol{x})}q(\boldsymbol{x})d\boldsymbol{x} \approx$$

$$\frac{1}{nZ_p}\sum_{i=1}^{n} f(\boldsymbol{x}_i)\frac{\tilde{p}(\boldsymbol{x}_i)}{q(\boldsymbol{x}_i)} = \frac{1}{n\sum_{i=1}^{n} r_i}\sum_{i=1}^{n} f(\boldsymbol{x}_i)r_i, \qquad \boldsymbol{x} \sim q(\boldsymbol{x})$$

- Если распределение $q(\boldsymbol{x})$ сильно отличается от $p(\boldsymbol{x})$, большинство весов r_i близки к нулю, и метод становится неустойчивым

Марковская цепь

- Методы Монте-Карло, использующие Марковские цепи (Monte Carlo Markov chain, MCMC) являются более эффективными средствами получения выборки из заданного распределения
- При использовании MCMC каждая очередная точка выборки $\boldsymbol{x}_i$ зависит некоторым образом от предыдущей точки $\boldsymbol{x}_{i-1}$
- Методы этой группы позволяют «нащупать» области с высоким значением плотности и проводить выборку из них
- Полученная выборка $(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n)$ не является выборкой независимых одинаково распределенных случайных величин, но вполне подходит для взятия интеграла

Схема Гиббса

Вход: Многомерное распределение $p(\boldsymbol{x})$;

Выход: Выборка из распределения $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$

Инициализация $\boldsymbol{x}_0 = (x_1^0, \ldots, x_d^0)$;

для $i = 1, \ldots, n$

    Сгенерировать x_1^i из распределения $p(x_1 | x_2^{i-1}, x_3^{i-1}, \ldots, x_d^{i-1})$;

    Сгенерировать x_2^i из распределения $p(x_2 | x_1^{i}, x_3^{i-1}, \ldots, x_d^{i-1})$;

    $\ldots$

    Сгенерировать x_d^i из распределения $p(x_d | x_2^{i}, x_3^{i}, \ldots, x_{d-1}^{i})$;

    $\boldsymbol{x}_i := (x_1^i, \ldots, x_d^i)$;

План

- Гибридные методы используют информацию не только о значении плотности $p(\boldsymbol{x})$, но и о градиенте ее логарифма $\frac{\partial \log p(\boldsymbol{x})}{\partial \boldsymbol{x}}$
- Для этого используются аналогии с аналитической механикой

  Аналитическая механика была разработана в первой половине 19 в. ирландским математиком Гамильтоном. В ее основе лежит идея замены одного дифференциального уравнения второго порядка во втором законе Ньютона на систему двух дифференциальных уравнений первого порядка
- Считая $\boldsymbol{x}$ переменными состояния, введем потенциальную энергию системы

$$E(\boldsymbol{x}) = -\log p(\boldsymbol{x}) + C$$

- Здесь используется принцип минимальной потенциальной энергии, гласящий, что состояние системы тем более вероятно, чем меньше ее потенциальная энергия

Аналитическая механика

- Введем дополнительные переменные, называемые моментами
$$\boldsymbol{r} = \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t}$$
- Кинетическая энергия системы является функцией моментов $K(\boldsymbol{r}) = 0.5\|\boldsymbol{r}\|^2$, а полная энергия системы (гамильтониан) равна
$$H(\boldsymbol{x}, \boldsymbol{r}) = E(\boldsymbol{x}) + K(\boldsymbol{r})$$
- Уравнения Гамильтона являются записью второго закона Ньютона через переменные состояния и моменты
$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = \frac{\partial H}{\partial \boldsymbol{r}}$$
$$\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t} = -\frac{\partial H}{\partial \boldsymbol{x}}$$

Интегрирование уравнений Гамильтона

- При динамическом изменении замкнутой системы гамильтониан H является постоянным по времени (закон сохранения энергии)
- Изменение системы описывается функциями $\boldsymbol{x}(t)$ и $\boldsymbol{r}(t)$, связанными уравнениями Гамильтона
- При численном решении уравнений получаем

$$\boldsymbol{r}(t+\varepsilon/2) = \boldsymbol{r}(t) - \frac{\varepsilon}{2}\frac{\partial E}{\partial \boldsymbol{x}}(\boldsymbol{x}(t))$$

$$\boldsymbol{x}(t+\varepsilon) = \boldsymbol{x}(t) + \varepsilon \boldsymbol{r}(t+\varepsilon/2)$$

$$\boldsymbol{r}(t+\varepsilon) = \boldsymbol{r}(t+\varepsilon/2) - \frac{\varepsilon}{2}\frac{\partial E}{\partial \boldsymbol{x}}(\boldsymbol{x}(t+\varepsilon))$$

- Полученные решения приблизительно описывают одну из линий уровня функции Гамильтона

Графическая иллюстрация эволюции системы

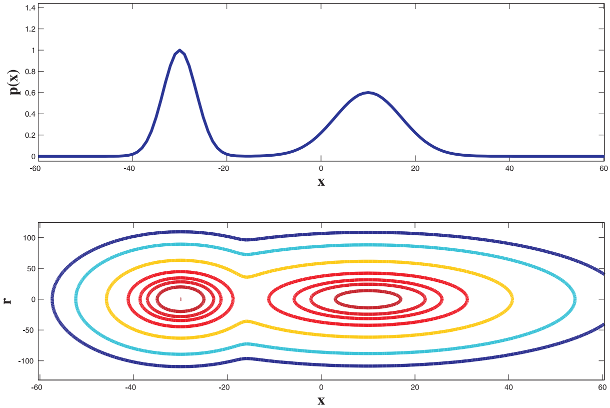


Схема генерации выборки

- Точки $(\boldsymbol{x}(t_1), \dots, \boldsymbol{x}(t_n))$ представляют собой равномерную выборку из множества $\{\boldsymbol{x}|p(\boldsymbol{x}) \geq C_0\}$
- Чтобы получить выборку из распределения $p(\boldsymbol{x})$ через каждые $m \ll n$ итераций значение моментов берется из распределения $p(\boldsymbol{r}) = \frac{1}{Z_r}\exp(-K(\boldsymbol{r})) = \mathcal{N}(\boldsymbol{r}|0, I)$
- Такая схема генерации выборки позволяет быстро найти области с большим значением $p(\boldsymbol{x})$ и получить репрезентативную выборку из этих областей

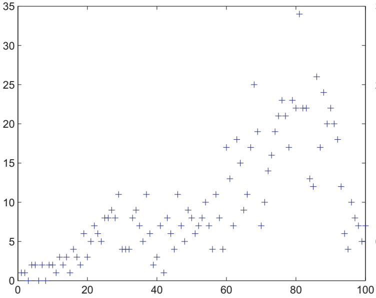
Нелинейная фильтрация

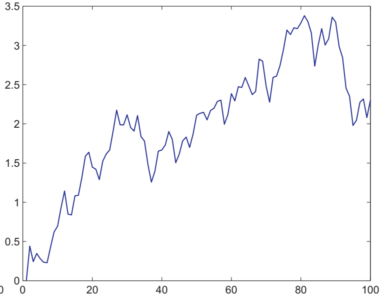
- Вернемся к задаче фильтрации сигналов
- До сих пор мы рассматривали ситуацию, когда атомарные распределения задавались гауссианами
- Только в этом случае мы можем гарантировать два важных свойства
  - Неусложнение распределения скрытой компоненты по мере увеличения длины сигнала
  - Аналитическое взятие интегралов при применении алгоритма sum-product, позволяющего нам оценить условное распределение скрытой компоненты

Пример задачи нелинейной фильтрации

Рассмотрим такую задачу: имеется некоторый гауссовский процесс с известными свойствами, а мы наблюдаем его косвенные характеристики — величину, распределенную по закону Пуассона с параметром $\lambda_n = \exp(t_n)$, т.е.

$$P(x_n = k|t_n) = \frac{\lambda_n^k}{k!} \exp(-\lambda_n)$$



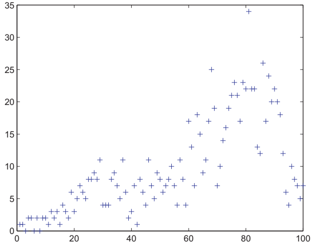


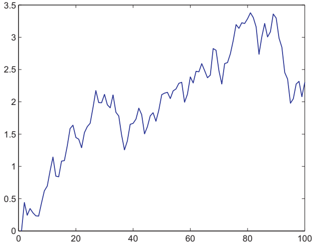
Оценка распределения скрытой компоненты

- Как обычно в задаче фильтрации сигналов необходимо оценить $p(t_k|X_k)$, где $X_k = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k)$ — значения наблюдаемой компоненты сигнала в первые k моментов времени
- Формально применяя алгоритм sum-product и делая проход вперед получаем формулу пересчета

$$p(t_k|X_k) = \frac{1}{Z}p(\boldsymbol{x}_k|t_k)\int p(t_k|t_{k-1})p(t_{k-1}|X_{k-1})dt_{k-1}$$

- Проблема в том, что интеграл мы в общем случае взять не можем





Предиктор-корректорная схема

Как и в случае калмановской фильтрации будем решать задачу в два этапа:

- Прогноз: оценка распределения

$$p(t_{k+1}|X_k) = \int p(t_{k+1}|t_k)p(t_k|X_k)dt_k$$

- Коррекция: оценка распределения

$$p(t_{k+1}|X_{k+1}) = \frac{1}{Z}p(\boldsymbol{x}_{k+1}|t_{k+1})p(t_{k+1}|X_k)$$

  Здесь мы воспользовались свойством условной независимости X_k и $\boldsymbol{x}_{k+1}$ при известном t_{k+1}

Для оценивания интеграла воспользуемся схемой Монте-Карло

Применение метода Монте-Карло

- Нам нужно оценить значение интеграла

$$p(t_{k+1}|X_k) = \int p(t_{k+1}|t_k)p(t_k|X_k)dt_k$$

  для каждого t_{k+1}
- Мы умеем генерировать выборку из $p(t_k|X_{k-1})$
- Воспользуемся этим, чтобы сгенерировать выборку из $p(t_{k+1}|X_k)$ вместо того, чтобы честно оценивать значение интеграла для всех t_k

Пересчет выборок

Допустим, у нас есть выборка $(t_k^{(1)}, \dots, t_k^{(L)}) \sim p(t_k|X_{k-1})$. Справедлива следующая цепочка равенств

$$p(t_{k+1}|X_k) = \int p(t_{k+1}|t_k)p(t_k|X_k)dt_k =$$

$$= \int p(t_{k+1}|t_k)p(t_k|x_k, X_{k-1})dt_k =$$

$$= \frac{\int p(t_{k+1}|t_k)p(x_k|t_k, X_{k-1})p(t_k|X_{k-1})dt_k}{\int p(x_k|t_k, X_{k-1})p(t_k|X_{k-1})dt_k} \approx$$

$$\approx \frac{\sum_{l=1}^L p(t_{k+1}|t_k^{(l)})p(x_k|t_k^{(l)})}{\sum_{l=1}^L p(x_k|t_k^{(l)})} = \sum_{l=1}^L w_k^{(l)} p(t_{k+1}|t_k^{(l)})$$

Здесь

$$w_k^{(l)} = \frac{p(x_k|t_k^{(l)})}{\sum_{m=1}^L p(x_k|t_k^{(m)})}$$

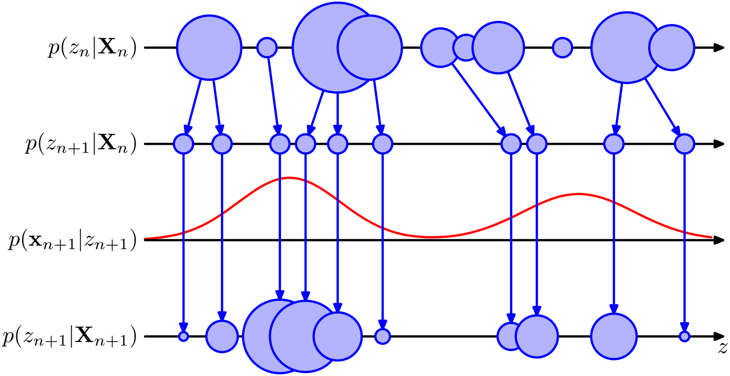
Генерация выборки $\{t_{k+1}^{(l)}\}$

- Итак, приближенное значение плотности $p(t_{k+1}|X_k)$ равно
$$p(t_{k+1}|X_k) \approx \sum_{l=1}^{L} w_k^{(l)} p(t_{k+1}|t_k^{(l)})$$
- Но нам для следующего шага (вычисления $p(t_{k+2}|X_{k+1})$) понадобится не значение плотности, а выборка $\{t_{k+1}^{(l)}\}_{l=1}^{L}$ из распределения $p(t_{k+1}|X_k)$, представляющая собой взвешенную смесь из L распределений $p(t_{k+1}|t_k^{(l)})$ с весами $w_k^{(l)}$
- Генерируем $\{t_{k+1}^{(l)}\}_{l=1}^{L}$ в два этапа: сначала с учетом весов $w_k^{(m)}$ разыгрываем номер m компоненты смеси, а затем генерируем выборку размера один из этой компоненты $t_{k+1}^{(l)} \sim p(t_{k+1}|t_k^{(m)})$ и так L раз

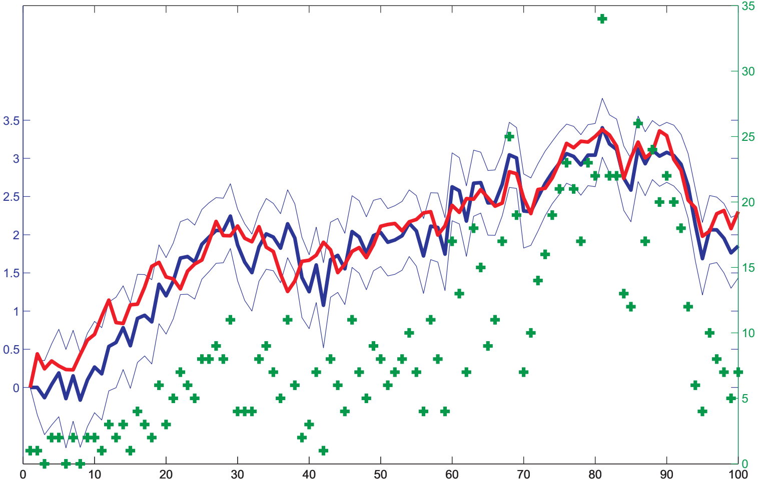
Подсчет распределения $p(t_k|X_k)$

- Таким образом осуществляется переход от выборки из $p(t_k|X_{k-1})$ к выборке из $p(t_{k+1}|X_k)$
- Также для каждого момента времени мы знаем веса
$$w_k^{(l)} = \frac{p(x_k|t_k^{(l)})}{\sum_{m=1}^L p(x_k|t_k^{(m)})},$$
определяющие изменения репрезентативности выборки из $p(t_k|X_{k-1})$ по отношению к выборке из $p(t_k|X_k)$
- В частности для оценки мат. ожидания $\mathbb{E}_{t_k|X_k} t_k$ достаточно взять взвешенное среднее
$$\mathbb{E}_{t_k|X_k} t_k \approx \sum_{l=1}^L w_k^{(l)} t_k^{(l)}$$

Схема работы



Пример использования в задаче пуассоновской фильтрации



Заключительные замечания

- Вместо того, чтобы подсчитывать и хранить распределения произвольной формы, мы храним выборки из этих распределений и эффективно трансформируем одну выборку в другую с учетом их вероятностных взаимосвязей
- При использовании фильтров частиц необходимо уметь эффективно генерировать выборку из распределения $p(t_k|t_{k-1})$, поэтому рекомендуется в качестве такового брать простые распределения, допускающие эффективную генерацию выборки
- Подход, аналогичный фильтрам частиц, может быть использован и при расчете более сложных графических моделей