Прикладные байесовские методы

Максим Кочуров

МГУ им. М.В. Ломоносова

Лекция 2

Содержание

Сэмплирование из распределения

Сопряженные модели

- Ограниченная применимость
- Недостаточная гибкость
- Хорошая масштабируемость

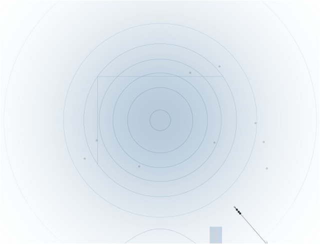


Рис.: Простое распределение

Большинство моделей

- Нет аналитического решения
- Апостериорные распределения сложные
- Плохая масштабируемость
- Гибкость

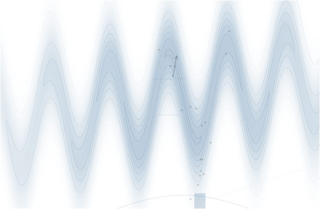


Рис.: Сложное распределение

Метод Монте-Карло с механикой Гамильтона (HMC)

В HMC сэмплирование происходит из сложных распределений

1. Берем идеи из физики
2. Считаем градиенты
3. Численно интегрируем

Откалиброванная модель HMC сходится к целевому распределению

Предупреждение

Мы обещали курс без сложной математики. Но HMC нужен для отладки ваших моделей.

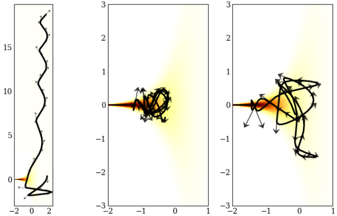


Рис.: интегрирование Leapfrog

Построение распределений через HMC

- $p(\Theta)$ - Целевое распределение, $\Theta \in \mathbb{R}^d$ (Θ или **Координаты**)
- $p(\Delta \mid \Theta)$ - Распределение импульса, $\Delta \in \mathbb{R}^d$ (Δ или **Ускорение**)

Гамильтониан вычисляется как

$$H(\Delta, \Theta) = -\log p(\Delta, \Theta)$$

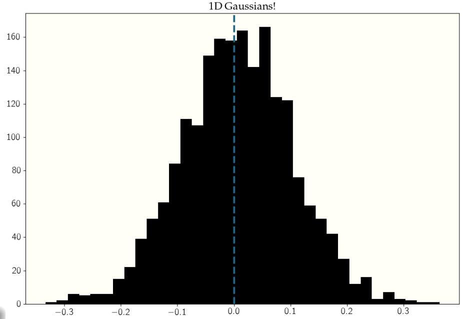


Рис.: $p(\Theta) = \text{Normal}(0, 1)$

Заметка

- $p(\Delta \mid \Theta) = \text{Normal}(0, M)$
- Δ и Θ имеют одинаковые размерности

Система дифференциальных уровнений HMC

$$
\begin{aligned}
H(\Delta, \Theta) &= -\log p(\Delta, \Theta) \\
&= -\log p(\Delta \mid \Theta) \quad -\log p(\Theta) \\
&= \underbrace{K(\Delta, \Theta)}_{\text{Kinetic E}} \quad + \underbrace{V(\Theta)}_{\text{Potential E}}
\end{aligned}
$$

Физическая система, задающая **движение**

$$
\frac{\partial \Theta}{\partial t} = \frac{\partial H}{\partial \Delta}
$$
$$
\frac{\partial \Delta}{\partial t} = -\frac{\partial H}{\partial \Theta}
$$

Движение содержит полную энергию $H(\Delta, \Theta)$

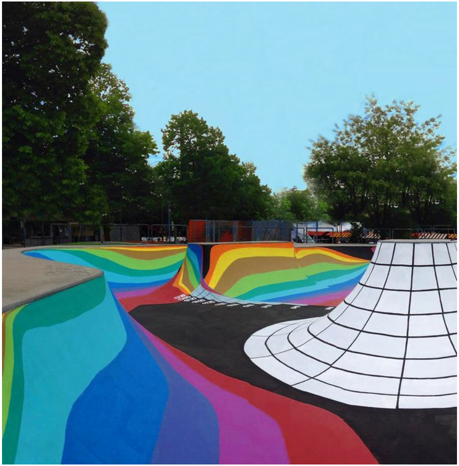


Рис.: скейт-парк и HMC

Дивергенции в HMC

Дивергенции возникают при крупной ошибке численного интегрирования в ходе решения дифференциального уравнения.

Когда HMC не работает

Неудачный выбор геометрического распределения вероятностей для гамильтнониана

Почему геометрическое распределение неудачно

1. Мультиколлинеарность
2. Распределения вида воронки
3. Сильное влияние функции правдоподобия
4. Неоднородное апостериорные распределение

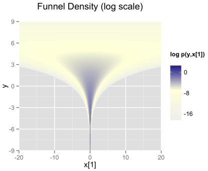


Рис.: Воронка Нила

Что почитать по HMC

Для дальнейшего погружения

1. Интерактивное демо
2. Пособие от Colin Carrol
3. Статья Michael Betancourt
4. Статья про NUTS от Matthew D. Hoffman, Andrew Gelman

Пример

Простой пример - функция Кобба-Дугласа

Скорее всего, вы знакомы с производственной функцией вида

$$Y \approx A \cdot L^{\beta}$$

Имеем:

1. 6 различных групп (иерархия)
2. Группы различимы
3. Мы знаем, что факторная производительность A различается между ними
4. Производительность труда β различается не сильно

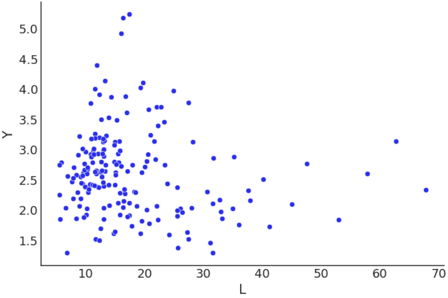


Рис.: Пример агрегированных данных

Простой пример - производство ковров

Уточним пример

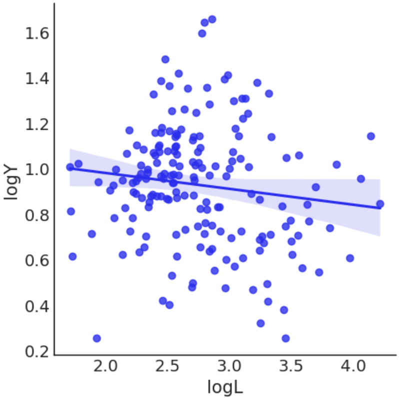
$$Y_g \approx A_g \cdot L^\beta$$

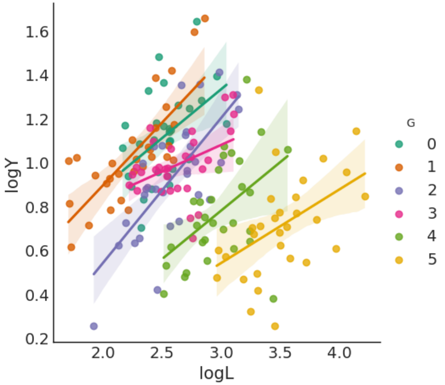
Пусть есть ателье, шьющее ковры, с 6 рабочими:

1. Работники делают разные ковры с различной факторной производительностью A
2. Производительность труда β можно воспринимать как сосредоточенность, со временем работники устают
3. Работник шьет ковер один, без чужой помощи

Рис.: Пример продукта (Y)

Парадокс Симпсона





Модель с одной группой

Несколько советов для моделирования.

- Начинать стоит с простой модели
- Убедитесь, что модель сходится
- Усложните модель

Модель с одной группой

Несколько советов для моделирования.

- Начинать стоит с простой модели
    - Рассмотрите модель с 1 группой
    - Проверьте, что априорные распределения выбраны корректно
- Убедитесь, что модель сходится
- Усложните модель

Модель с одной группой

Несколько советов для моделирования.

- Начинать стоит с простой модели
    - Рассмотрите модель с 1 группой
    - Проверьте, что априорные распределения выбраны корректно
- Убедитесь, что модель сходится
    - Если не сойдется модель с одной группой – модели с множеством групп тоже не сойдутся
    - Проведите проверку, что модель семплируется корректно
- Усложните модель

Модель с одной группой

Несколько советов для моделирования.

- Начинать стоит с простой модели
  - Рассмотрите модель с 1 группой
  - Проверьте, что априорные распределения выбраны корректно
- Убедитесь, что модель сходится
  - Если не сойдется модель с одной группой – модели с множеством групп тоже не сойдутся
  - Проведите проверку, что модель семплируется корректно
- Усложните модель
  - Попробуйте сделать параметризацию
  - Проверьте семплирование модели
  - Сравните модели (про это – поговорим далее по курсу)

Простая модель: с чего начать?

Для начала зададим себе несколько вопросов

$$Y_{g=0} \approx A_{g=0} \cdot L^{\beta}$$

1. Какое апостериорное распределение взять для A?
2. Какое апостериорное распределение взять для β?
3. Какое апостериорное предиктивное распределение для $Y_{g=0}$?

Формулируем модель

$$Y_{g=0} \approx A_{g=0} \cdot L^{\beta}$$
$$\log Y_{g=0} \approx \log A_{g=0} + \log L \cdot \beta$$

Предполагаем следующие распределения

$$\begin{aligned}
\log Y_{g=0} &\sim \mathrm{Normal}(\log A_{g=0} + \log L \cdot \beta, \epsilon) \\
\epsilon &\sim ??? \\
\beta &\sim ??? \\
A_{g=0} &\sim ???
\end{aligned}$$

Априорное распределение для β

Какие свойства мы предполагаем для производительности (или эластичности) труда β? Подумайте

1. Может ли она быть < 0?
2. Может ли она принимать большие значения?
3. Может ли она быть > 1?

Априорное распределение для β

Какие свойства мы предполагаем для производительности (или эластичности) труда β? Подумайте

1. Может ли она быть < 0? Нет
2. Может ли она принимать большие значения?
3. Может ли она быть > 1?

Априорное распределение для β

Какие свойства мы предполагаем для производительности (или эластичности) труда β? Подумайте

1. Может ли она быть < 0? Нет
2. Может ли она принимать большие значения? Нет
3. Может ли она быть > 1?

Априорное распределение для β

Какие свойства мы предполагаем для производительности (или эластичности) труда β? Подумайте

1. Может ли она быть < 0? Нет
2. Может ли она принимать большие значения? Нет
3. Может ли она быть > 1? Нет

Априорное распределение для β

Какие свойства мы предполагаем для производительности (или эластичности) труда β? Подумайте

1. Может ли она быть < 0? Нет
2. Может ли она принимать большие значения? Нет
3. Может ли она быть > 1? Нет

Заключение: она лежит в промежутке $(0, 1)$

Данные априорные предположения – субъективны!

Есть у кого-то замечания по нашим ограничениям?

Априорное распределение для β

Какие свойства мы предполагаем для производительности (или эластичности) труда β? Подумайте

1. Может ли она быть < 0? Нет
2. Может ли она принимать большие значения? Нет
3. Может ли она быть > 1? Нет

Заключение: она лежит в промежутке $(0, 1)$

Данные априорные предположения – субъективны!

Есть у кого-то замечания по нашим ограничениям?

Мы пока еще не вывели априорное распределение

Нужно подобрать распределение, которое удовлетворяет нашим условиям

Априорное распределение для β

Что имеем:

- $\beta \in (0, 1)$
- Маловероятно, что истинное значение близко к границам
- Ничего не знаем о том, где конкретно в интервале оно лежит.

Заметка

Посмотрим распределения, которые удовлетворяют требованиям

Априорное распределение для β

Что имеем:

- $\beta \in (0, 1)$
- Маловероятно, что истинное значение близко к границам
- Ничего не знаем о том, где конкретно в интервале оно лежит.

Заметка

Посмотрим распределения, которые удовлетворяют требованиям

1. $\text{Beta}(a, b)$, $a > 0, b > 0$ где a, b удовлетворяют ограничениям

Априорное распределение для β

Что имеем:

- $\beta \in (0, 1)$
- Маловероятно, что истинное значение близко к границам
- Ничего не знаем о том, где конкретно в интервале оно лежит.

Заметка

Посмотрим распределения, которые удовлетворяют требованиям

1. $\mathrm{Beta}(a, b),\ a > 0, b > 0$ где a, b удовлетворяют ограничениям
2. $\mathrm{LogitNormal}(\mu, \sigma)$ - всегда удовлетворяет ограничениям

Априорное распределение для β

Что имеем:

- $\beta \in (0, 1)$
- Маловероятно, что истинное значение близко к границам
- Ничего не знаем о том, где конкретно в интервале оно лежит.

Заметка

Посмотрим распределения, которые удовлетворяют требованиям

1. $\mathrm{Beta}(a, b),\ a > 0, b > 0$ где a, b удовлетворяют ограничениям
2. $\mathrm{LogitNormal}(\mu, \sigma)$ - всегда удовлетворяет ограничениям
3. $\mathrm{Uniform}(0, 1)$ - частный случай $\mathrm{Beta}(1, 1)$

Априорное распределение для β

Что имеем:

- $\beta \in (0, 1)$
- Маловероятно, что истинное значение близко к границам
- Ничего не знаем о том, где конкретно в интервале оно лежит.

Заметка

Посмотрим распределения, которые удовлетворяют требованиям

1. Beta(a, b), $a > 0, b > 0$ где a, b удовлетворяют ограничениям
2. LogitNormal(μ, σ) - всегда удовлетворяет ограничениям
3. Uniform$(0, 1)$ - частный случай Beta$(1, 1)$
4. Kumaraswamy(a, b), $a > 0, b > 0$ но лучше обойтись без него

Визуализируем априорное распределение

Перед написанием изобразите распределения. Какое лучше?

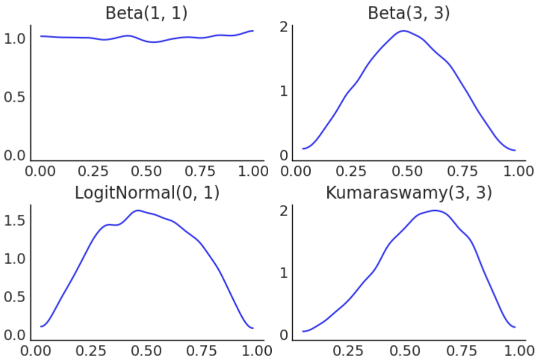


Рис.: Распределения с прошлого слайда

Выбирайте то, которое нравится и согласуется с теорией

Настраиваем априорное распределение

Для этой задачи я выберу $\mathrm{LogitNormal}(0, 1)$. У него удобная функциональная форма

Помните

- Априорное распределение при моделировании выбираете **вы**
- Этот выбор должен быть обоснован
- Выбор должен быть разумным, учитывать накопленное знание
- Вы должны иметь возможность защитить свой выбор
- **Априорное распределение – источник неопределенности**

Простая модель: что имеем на данный момент

$$
\begin{aligned}
\log Y_{g=0} &\sim \mathrm{Normal}(\log A_{g=0} + \log L \cdot \beta, \epsilon) \\
\epsilon &\sim ??? \\
\beta &\sim \mathrm{LogitNormal}(0, 1) \\
A_{g=0} &\sim ???
\end{aligned}
$$

Априорное распределение ϵ

Эмпирическое наблюдение

Ошибка обычно не велика. Но при этом не равна нулю.

В нашем случае;

- "маленькая ошибка" означает отклонение в 10-50%

Пусть

$$\epsilon \sim \mathrm{LogNormal}(-2, 1)$$

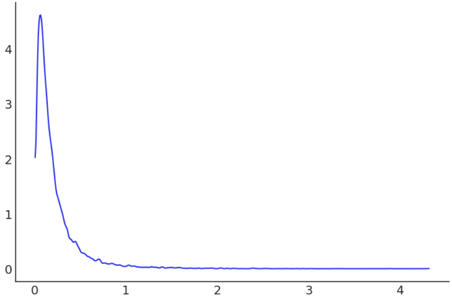


Рис.: Априорное распределение для ϵ

Априорное распределение ϵ

Эмпирическое наблюдение

Ошибка обычно не велика. Но при этом не равна нулю.

В нашем случае;

- ''маленькая ошибка'' означает отклонение в 10-50%

Пусть

$$\epsilon \sim \text{LogNormal}(-2, 1)$$

Примечание

В логарифмических моделях ошибка – величина относительная

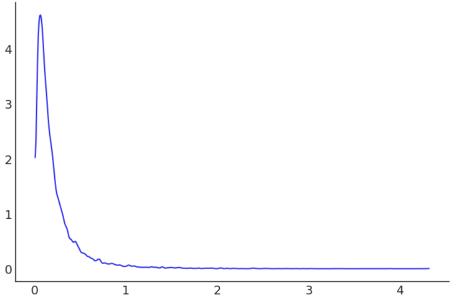
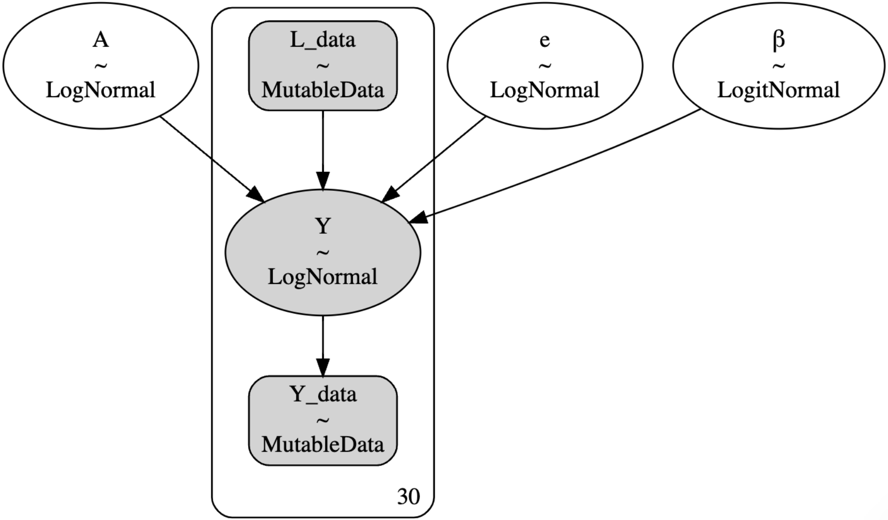


Рис.: Априорное распределение для ϵ

Простая модель: что имеем на данный момент

$$
\begin{aligned}
\log Y_{g=0} &\sim \text{Normal}(\log A_{g=0} + \log L \cdot \beta, \epsilon) \\
\epsilon &\sim \text{LogNormal}(-2, 1) \\
\beta &\sim \text{LogitNormal}(0, 1) \\
A_{g=0} &\sim ???
\end{aligned}
$$

Визуализиуруем модель



Априорное предиктивное распределение

Подобрать априорное распределение β было несложно. А для $A_{g=0}$?

Априорное предиктивное распределение

Подобрать априорное распределение β было несложно. А для $A_{g=0}$?

- Нет никакой информацией об априорном распределении

Априорное предиктивное распределение

Подобрать априорное распределение β было несложно. А для $A_{g=0}$?

- Нет никакой информацией об априорном распределении
- Но кое-что знаем про Y

Априорное предиктивное распределение

Подобрать априорное распределение β было несложно. А для $A_{g=0}$?

- Нет никакой информацией об априорном распределении
- Но кое-что знаем про Y
- Y положительно, соответственно A тоже положительно

Априорное предиктивное распределение

Подобрать априорное распределение β было несложно. А для $A_{g=0}$?

- Нет никакой информацией об априорном распределении
- Но кое-что знаем про Y
- Y положительно, соответственно A тоже положительно
- Есть наблюдаемые значения для Y, значит можно вывести значения для A?

Априорное предиктивное распределение

Подобрать априорное распределение β было несложно. А для $A_{g=0}$?

- Нет никакой информацией об априорном распределении
- Но кое-что знаем про Y
- Y положительно, соответственно A тоже положительно
- Есть наблюдаемые значения для Y, значит можно вывести значения для A?
- Степени 10 подойдут как значения для Y

Априорное предиктивное распределение

Подобрать априорное распределение β было несложно. А для $A_{g=0}$?

- Нет никакой информацией об априорном распределении
- Но кое-что знаем про Y
- Y положительно, соответственно A тоже положительно
- Есть наблюдаемые значения для Y, значит можно вывести значения для A?
- Степени 10 подойдут как значения для Y
- Для проверки – построим априорное предиктивное распределение

Априорное предиктивное распределение

Подобрать априорное распределение β было несложно. А для $A_{g=0}$?

- Нет никакой информацией об априорном распределении
- Но кое-что знаем про Y
- Y положительно, соответственно A тоже положительно
- Есть наблюдаемые значения для Y, значит можно вывести значения для A?
- Степени 10 подойдут как значения для Y
- Для проверки – построим априорное предиктивное распределение

Определение

Априорное предиктивное распределение строится как симуляция распределения без учета данных.

Горькая правда

Подбор априорных распределений – основа нашей непростой работы

Подбор распределения ''наобум''

Почему бы не предположить, например

$$A \sim \mathrm{LogNormal}(0, 1)$$

Подбор распределения "наобум"

Почему бы не предположить, например

$$A \sim \mathrm{LogNormal}(0, 1)$$

Контринтуитивный результат

Никакой работник не пошьет вручную 800 ковров за неделю

Потому что (см. график ниже)

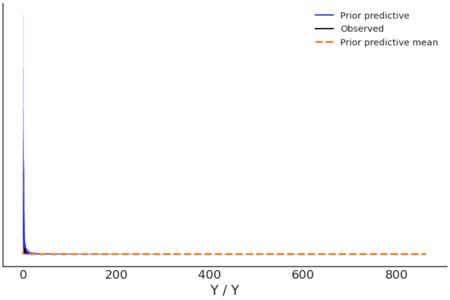


Рис.: Априорное предиктивное распределение Y и распределение реальных значений

Анализируем априорные предиктивные распределения

Еще раз взглянем на нашу модель

$$
\begin{aligned}
\log Y_{g=0} &\sim \text{Normal}(\log A_{g=0} + \log L \cdot \beta, \epsilon) \\
\epsilon &\sim \text{LogNormal}(-2, 1) \\
\beta &\sim \text{LogitNormal}(0, 1) \\
A_{g=0} &\sim \text{LogNormal}(0, 1)
\end{aligned}
$$

- Наблюдаем избыточную дисперсию
- Она может возникать из-за A или ϵ

Что делать?

1. Попробуем уменьшить дисперсию A
2. Попробуем уменьшить дисперсию ϵ

Что получаем

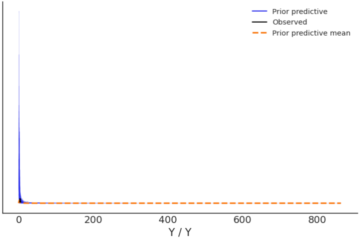


Рис.: Априорное предиктивное распределение Y и распределение реальных значений

Пример хорошего априорного предиктивного распределения

На семинаре

Разберем схожий пример на семинаре.

Хорошее априорное предиктивное можно задать следующим образом:

$$\log Y_{g=0} \sim \text{Normal}(\log A_{g=0} + \log L \cdot \beta, \epsilon)$$
$$\epsilon \sim \text{LogNormal}(-2, 0.1)$$
$$\beta \sim \text{LogitNormal}(0, 1)$$
$$A_{g=0} \sim \text{LogNormal}(-0.5, 0.1)$$

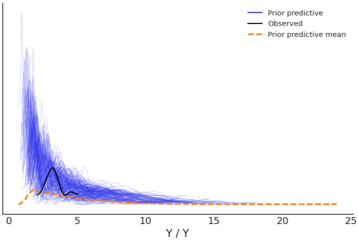


Рис.: Априорное предиктивное распределение Y и распределение реальных значений

Как понять, что априорное предиктивное распределение удачное?

- Распределение **достаточно** соответствует данным.

- **Данные – ориентир**, не самоцель

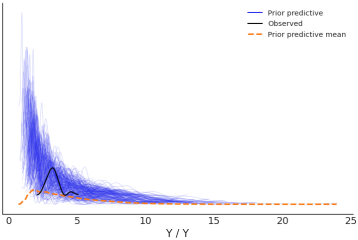


Рис.: Априорное предиктивное распределение Y и распределение реальных значений

Как понять, что априорное предиктивное распределение удачное?

- Распределение **достаточно** соответствует данным.
  - нет астрономических скоростей
  - нет маленьких расстояний
  - нет очень больших расстояний
  - нет работников-стахановцев
- **Данные – ориентир**, не самоцель

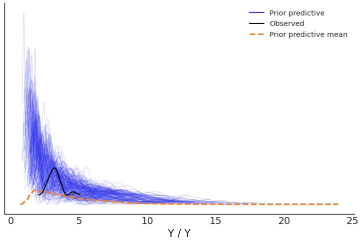


Рис.: Априорное предиктивное распределение Y и распределение реальных значений

Как понять, что априорное предиктивное распределение удачное?

- Распределение **достаточно** соответствует данным.
  - нет астрономических скоростей
  - нет маленьких расстояний
  - нет очень больших расстояний
  - нет работников-стахановцев
- **Данные – ориентир**, не самоцель
  - не стоит подгонять распределение точь в точь под данные
  - в 90% данные не нужны
  - в 90% достаточно здравого смысла
  - в 10% можно спросить у экспертов
  - смотреть на данные – крайняя мера

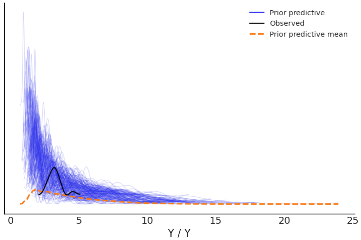


Рис.: Априорное предиктивное распределение Y и распределение реальных значений

Пора применять HMC

Сэмплирование

После всех проведенных проверок – можно начинать сэмплирование.

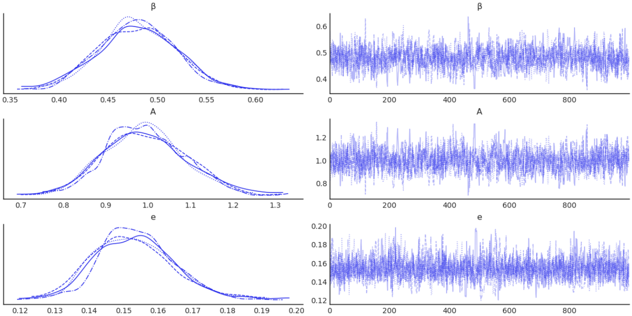


Рис.: Апострериорное распределение, полученное MCMC

Иерархии

Иерархии

В исходных данных есть группы. Как это учесть?

$$\begin{aligned}
\log Y_{\mathbf{g}} &\sim \text{Normal}(\log A_{\mathbf{g}} + \log L \cdot \beta, \epsilon) \\
\epsilon &\sim \text{LogNormal}(-2, 0.1) \\
\beta &\sim \text{LogitNormal}(0, 1) \\
A_{\mathbf{g}} &\sim ???
\end{aligned}$$

Что такое иерархии?

Иерархии

Если в данных есть похожие группы, значит есть иерархии.

Примеры:

1. Страны, Регионы
2. Разделение пользователей по полу, возрасту и тд.
3. Группы воздействия
4. Временные эффекты
5. Панельные данные

Наш пример

Рабочие делают разные ковры и обладают совокупной производительностью A

Как учитывать иерархии

Классический эконометрический подход:

1. Если все группы независимы. **Регрессия пула**

$$y_{k,i} = \alpha + \beta x_{k,i} + \varepsilon_{i,k}$$

2. Если в группах есть различия.
Регрессия с фиксированными эффектами

$$y_{k,i} = \alpha_k + \beta x_{k,i} + \varepsilon_{i,k}$$

3. Разница между группами случайная, незначительная.
Регрессия со случайными эффектами

$$y_{k,i} = \alpha + \beta x_{k,i} + u_k + \varepsilon_{i,k}$$

Где

$$\mathbb{E}u_{k,i} = 0, \quad \mathbb{E}\varepsilon_{k,i} = 0$$

Байесовский подход к иерархиям

Перепишем уравнение

$$y_{k,i} = \alpha + \beta x_{k,i} + u_k + \varepsilon_{i,k}$$

Следующим образом

$$y_{k,i} = (\alpha + u_k) + \beta x_{k,i} + \varepsilon_{i,k}$$

- α - среднее по выборке
- $\alpha_k = \alpha + u_k$ - среднее по группе

Байесовский подход требует априорных распределений. Например

$$\begin{aligned} \alpha &\sim \text{Normal}(\bar{\mu}, \bar{\sigma}) \\ u_k &\sim \text{Normal}(0, 1) \\ \alpha_k &= \alpha + u_k \cdot \sigma \end{aligned}$$

$$\begin{aligned} \alpha &\sim \text{Normal}(\bar{\mu}, \bar{\sigma}) \\ \alpha_k &\sim \text{Normal}(\alpha, \sigma) \end{aligned}$$

Больше про априорные распределения

Нецентрированная параметризация

$$\begin{aligned} \alpha &\sim \text{Normal}(\bar{\mu}, \bar{\sigma}) \\ u_k &\sim \text{Normal}(0, 1) \\ \alpha_k &= \alpha + u_k \cdot \sigma \end{aligned}$$

Group specific parameter u_k is disentangled

Центрированная параметризация

$$\begin{aligned} \alpha &\sim \text{Normal}(\bar{\mu}, \bar{\sigma}) \\ \alpha_k &\sim \text{Normal}(\alpha, \sigma) \end{aligned}$$

σ отражает разницу между группами

1. $\sigma \to 0$: Модель пула
2. Небольшая σ: Случайные эффекты / Частичный пул
3. Большая σ: Фиксированные эффекты / Необъединенная модель

σ присутствует в каждом типе моделей

Деградация моделей

Для центрированной параметризации имеем

$$\alpha \sim \text{Normal}(\bar{\mu}, \bar{\sigma})$$
$$\alpha_k \sim \text{Normal}(\alpha, \sigma)$$

Предупреждение

Для малого объема данных центрирование приводит к появлению воронок

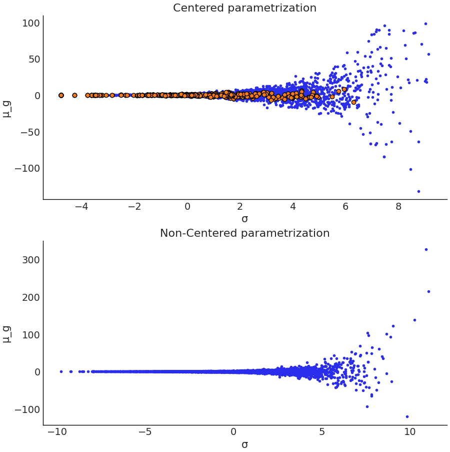


Рис.: Центрирование и расхождения

Как появляются воронки?

Геометрия распределения имеет значение

1. При сэмплировании шаг учитывает предыдущие итерации
2. При неудачно выбранном распределении сэмплеру сложно предсказать следующий шаг

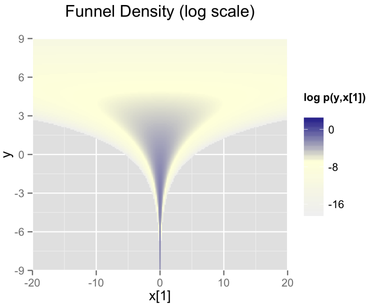


Рис.: Пример воронки

Рекомендация

Про репараметризацию можно почитать здесь

Инвертированная воронкообразная деградация

''Удачная'' параметризация тоже имеет свои особенности

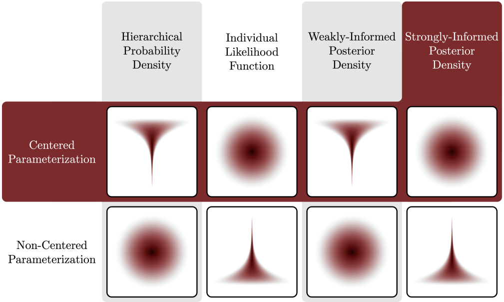


Рис.: Пример инвертированной воронкообразной деградации

Для дальнейшего погружения

Почитать про это можно здесь

Настраиваем иерархии в априорных распределениях

1. Начать лучше с модели с одной группой или модели пула
2. Добавляем иерархии

Настраиваем иерархии в априорных распределениях

1. Начать лучше с модели с одной группой или модели пула
   - Понимаем размерность и особенности параметров
   - Структурируем модель
   - Пока не думаем о предсказательной способности модели
2. Добавляем иерархии

Настраиваем иерархии в априорных распределениях

1. Начать лучше с модели с одной группой или модели пула
   - Понимаем размерность и особенности параметров
   - Структурируем модель
   - Пока не думаем о предсказательной способности модели
2. Добавляем иерархии
   - Выбираем параметр, для которого реализуем иерархии
   - Выбираем, допустима ли вариативность остальных параметров
   - Поправляем отклонения, если нужно – повторно подбираем параметры

Настраиваем иерархии в априорных распределениях

1. Начать лучше с модели с одной группой или модели пула
   - Понимаем размерность и особенности параметров
   - Структурируем модель
   - Пока не думаем о предсказательной способности модели
2. Добавляем иерархии
   - Выбираем параметр, для которого реализуем иерархии
   - Выбираем, допустима ли вариативность остальных параметров
   - Поправляем отклонения, если нужно – повторно подбираем параметры

Практический совет

Не надо хардкодить параметризацию, делайте свой код гибким

Пример функции Кобба-Дугласа

Модель с единственной группой

$$\begin{aligned}
\log Y_0 &\sim \mathrm{Normal}(\log A_0 + \log L \cdot \beta, \epsilon) \\
\epsilon &\sim \mathrm{LogNormal}(-2, 0.1) \\
\beta &\sim \mathrm{LogitNormal}(0, 1) \\
A_0 &\sim \mathrm{LogNormal}(-0.5, 0.1)
\end{aligned}$$

Иерархическая модель

$$\begin{aligned}
\log Y_k &\sim \mathrm{Normal}(\log A_k + \log L \cdot \beta, \epsilon) \\
\epsilon &\sim \mathrm{LogNormal}(-2, 0.1) \\
\beta &\sim \mathrm{LogitNormal}(0, 1) \\
A_k &\sim \mathrm{LogNormal}(\log A_{\mathrm{pop}}, \sigma_A) \\
A_{\mathrm{pop}} &\sim \mathrm{LogNormal}(-0.5, 0.1) \\
\sigma_A &\sim \mathrm{LogNormal}(-2, 0.1)
\end{aligned}$$

Пример функции Кобба-Дугласа

Модель с единственной группой

$$
\begin{aligned}
\log Y_0 &\sim \mathrm{Normal}(\log A_0 + \log L \cdot \beta, \epsilon) \\
\epsilon &\sim \mathrm{LogNormal}(-2, 0.1) \\
\beta &\sim \mathrm{LogitNormal}(0, 1) \\
A_0 &\sim \mathrm{LogNormal}(-0.5, 0.1)
\end{aligned}
$$

Иерархическая модель

$$
\begin{aligned}
\log Y_k &\sim \mathrm{Normal}(\log A_k + \log L \cdot \beta, \epsilon) \\
\epsilon &\sim \mathrm{LogNormal}(-2, 0.1) \\
\beta &\sim \mathrm{LogitNormal}(0, 1) \\
A_k &\sim \mathrm{LogNormal}(\log A_{\mathrm{pop}}, \sigma_A) \\
A_{\mathrm{pop}} &\sim \mathrm{LogNormal}(-0.5, 0.1) \\
\sigma_A &\sim \mathrm{LogNormal}(-2, 0.1)
\end{aligned}
$$

Подсказка

Некоторые параметры можно переиспользовать, нужно лишь добавить немного вариативности с помощью σ_A

Обсуждаем итоги

При подборе априорных распределений стоит помнить следующее

- Не всегда доступны экспертные знания
- Параметризация не всегда помогает подобрать хорошие распределения
- Иногда априорные предиктивные распределения зависят от многих параметров
- Вы часто скованы временем
- Можно применять гипераприорные распределения