Байесовские Методы. Лекция 1

БИВТ-23-9/10-ИСАД

8 сентября 2025 г.

Обо мне

- Я - Макс.
- Работаю в институте AIRI.
- Занимаюсь исследованиями ИИ в биохимии.
- Раньше - исследователь в T-Bank (неопределенность моделей) / инженер с стартапе (generative CV).

В этом курсе поговорим про байесовские методы. Затронем классические аспекты, а также важные темы для современного deep learning.

Почему байесовские методы важны

- **База Deep Learning**: байесовские методы помогают глубже понять современные подходы и модели, такие как диффузии, вариационные автоэнодеры (VAE), активное обучение и много других.

Почему байесовские методы важны

- **База Deep Learning**: байесовские методы помогают глубже понять современные подходы и модели, такие как диффузии, вариационные автоэнодеры (VAE), активное обучение и много других.
- **Оценка неопределённости**: важно знать, *когда* модель не уверена в предсказании.

Почему байесовские методы важны

- **База Deep Learning**: байесовские методы помогают глубже понять современные подходы и модели, такие как диффузии, вариационные автоэнодеры (VAE), активное обучение и много других.
- **Оценка неопределённости**: важно знать, *когда* модель не уверена в предсказании.
- **Эффективность при малых данных**: использование априорных знаний помогает обучать модели даже при небольших выборках.

Почему байесовские методы важны

- **База Deep Learning**: байесовские методы помогают глубже понять современные подходы и модели, такие как диффузии, вариационные автоэнодеры (VAE), активное обучение и много других.
- **Оценка неопределённости**: важно знать, *когда* модель не уверена в предсказании.
- **Эффективность при малых данных**: использование априорных знаний помогает обучать модели даже при небольших выборках.
- **Устойчивость к выбросам**: модель "понимает", когда данные выходят за рамки обучающего распределения.

Информация о курсе

- Курс состоит из $\approx$ **12 занятий**.

Информация о курсе

- Курс состоит из $\approx$ **12 занятий**.
- Планируется $\approx$ **4–5 домашних заданий**.

Информация о курсе

- Курс состоит из $\approx$ **12 занятий**.
- Планируется $\approx$ **4–5 домашних заданий**.
- Формат занятий:

Информация о курсе

- Курс состоит из $\approx$ **12 занятий**.
- Планируется $\approx$ **4–5 домашних заданий**.
- Формат занятий:
  - Некоторые семинары будут **лекционными**.

Информация о курсе

- Курс состоит из $\approx$ **12 занятий**.
- Планируется $\approx$ **4–5 домашних заданий**.
- Формат занятий:
  - Некоторые семинары будут **лекционными**.
  - В других — **первая половина лекция**, а **вторая половина практика**.

Информация о курсе

- Курс состоит из ≈ **12 занятий**.
- Планируется ≈ **4–5 домашних заданий**.
- Формат занятий:
  - Некоторые семинары будут **лекционными**.
  - В других — **первая половина лекция**, а **вторая половина практика**.
- **Оценка за курс** = средняя оценка за домашние задания.

Информация о курсе

- Курс состоит из $\approx$ **12 занятий**.
- Планируется $\approx$ **4–5 домашних заданий**.
- Формат занятий:
  - Некоторые семинары будут **лекционными**.
  - В других — **первая половина лекция**, а **вторая половина практика**.
- **Оценка за курс** = средняя оценка за домашние задания.
- **Автоматы возможны**: если у вас есть проекты, связанные с вероятностными методами. Детали обсуждаются индивидуально.

Информация о курсе

- Курс состоит из $\approx$ **12 занятий**.
- Планируется $\approx$ **4–5 домашних заданий**.
- Формат занятий:
  - Некоторые семинары будут **лекционными**.
  - В других — **первая половина лекция**, а **вторая половина практика**.
- **Оценка за курс** = средняя оценка за домашние задания.
- **Автоматы возможны**: если у вас есть проекты, связанные с вероятностными методами. Детали обсуждаются индивидуально.
- **Полный список тем** будет опубликован позже.

Напоминание: основы теории вероятностей

- **Случайная величина** X — переменная, принимающая значения из множества $\mathcal{X}$ с некоторым распределением вероятностей.

Напоминание: основы теории вероятностей

- **Случайная величина** X — переменная, принимающая значения из множества $\mathcal{X}$ с некоторым распределением вероятностей.
- **Вероятность события** A: $P(A) \in [0, 1]$.

Напоминание: основы теории вероятностей

- **Случайная величина** X — переменная, принимающая значения из множества $\mathcal{X}$ с некоторым распределением вероятностей.
- **Вероятность события** A: $P(A) \in [0, 1]$.
- **Совместная вероятность**: $P(A, B) = P(A \cap B)$.

Напоминание: основы теории вероятностей

- **Случайная величина** X — переменная, принимающая значения из множества $\mathcal{X}$ с некоторым распределением вероятностей.
- **Вероятность события** A: $P(A) \in [0, 1]$.
- **Совместная вероятность**: $P(A, B) = P(A \cap B)$.
- **Условная вероятность**:

$$P(A \mid B) = \frac{P(A, B)}{P(B)}, \quad P(B) > 0$$

Напоминание: основы теории вероятностей

- **Случайная величина** X — переменная, принимающая значения из множества $\mathcal{X}$ с некоторым распределением вероятностей.
- **Вероятность события** A: $P(A) \in [0, 1]$.
- **Совместная вероятность**: $P(A, B) = P(A \cap B)$.
- **Условная вероятность**:

$$P(A \mid B) = \frac{P(A, B)}{P(B)}, \quad P(B) > 0$$

- **Формула полной вероятности**:

$$P(A) = \sum_i P(A \mid B_i) \cdot P(B_i)$$

Напоминание: основы теории вероятностей

- **Случайная величина** X — переменная, принимающая значения из множества $\mathcal{X}$ с некоторым распределением вероятностей.
- **Вероятность события** A: $P(A) \in [0, 1]$.
- **Совместная вероятность**: $P(A, B) = P(A \cap B)$.
- **Условная вероятность**:

$$P(A \mid B) = \frac{P(A, B)}{P(B)}, \quad P(B) > 0$$

- **Формула полной вероятности**:

$$P(A) = \sum_i P(A \mid B_i) \cdot P(B_i)$$

- **Теорема Байеса**:

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

Частотный vs Байесовский подход

Аспект	Частотный подход	Байесовский подход
Смысл вероятности	Долгосрочная частота событий	Степень уверенности (субъективная)
Параметры θ	Фиксированные, но неизвестные константы	Случайные величины с распределениями
Данные $\mathcal{D}$	Случайные выборки из фиксированного процесса	Фиксированы после наблюдения
Цель вывода	Найти лучшее значение θ	Вычислить постериорное распределение $p(\theta \mid \mathcal{D})$
Источник неопределённости	Только из-за случайности данных	Из-за неопределённости данных и параметров
Обновление знаний	Невозможно обновить априорные представления	Обновление через теорему Байеса
Априорное распределение	Не используется	Явно задаётся как $p(\theta)$

Теорема Байеса

Формула

$$P(\theta \mid D) = \frac{P(D \mid \theta) \cdot P(\theta)}{P(D)}$$

- **Гипотеза / Параметры** θ —
  наши предположения о модели.

Теорема Байеса

Формула

$$P(\theta \mid D) = \frac{P(D \mid \theta) \cdot P(\theta)}{P(D)}$$

- **Гипотеза / Параметры** θ — наши предположения о модели.
- **Данные** D — наблюдаемые измерения.

Теорема Байеса

Формула

$$P(\theta \mid D) = \frac{P(D \mid \theta) \cdot P(\theta)}{P(D)}$$

- **Гипотеза / Параметры** θ — наши предположения о модели.
- **Данные** D — наблюдаемые измерения.
- **Априорное распределение (prior)** $P(\theta)$ — знания о параметрах *до* данных.

Теорема Байеса

Формула

$$P(\theta \mid D) = \frac{P(D \mid \theta) \cdot P(\theta)}{P(D)}$$

- **Гипотеза / Параметры** θ — наши предположения о модели.
- **Данные** D — наблюдаемые измерения.
- **Априорное распределение (prior)** $P(\theta)$ — знания о параметрах *до* данных.
- **Правдоподобие (likelihood)** $P(D \mid \theta)$ — насколько θ объясняет D.

Теорема Байеса

Формула

$$P(\theta \mid D) = \frac{P(D \mid \theta) \cdot P(\theta)}{P(D)}$$

- **Гипотеза / Параметры** θ — наши предположения о модели.
- **Данные** D — наблюдаемые измерения.
- **Априорное распределение (prior)** $P(\theta)$ — знания о параметрах *до* данных.
- **Правдоподобие (likelihood)** $P(D \mid \theta)$ — насколько θ объясняет D.
- **Обоснование (evidence)** $P(D)$ — нормировочная константа (как бы вероятность получения данных, имея все гипотезы в целом):

$$P(D) = \int P(D \mid \theta) P(\theta)\, d\theta$$

Теорема Байеса

Формула

$$P(\theta \mid D) = \frac{P(D \mid \theta) \cdot P(\theta)}{P(D)}$$

- **Гипотеза / Параметры** θ — наши предположения о модели.
- **Данные** D — наблюдаемые измерения.
- **Априорное распределение (prior)** $P(\theta)$ — знания о параметрах *до* данных.
- **Правдоподобие (likelihood)** $P(D \mid \theta)$ — насколько θ объясняет D.
- **Обоснование (evidence)** $P(D)$ — нормировочная константа (как бы вероятность получения данных, имея все гипотезы в целом):

$$P(D) = \int P(D \mid \theta) P(\theta)\, d\theta$$

- **Постериорное распределение (posterior)** $P(\theta \mid D)$ — обновленное знание *относительно* данных.

Правдоподобие

Определение

Правдоподобие — это *интерпретация* вероятности данных как функции от параметров:

$$L(\theta; D) := P(D \mid \theta)$$

- Если фиксируем параметры $\theta \to P(D \mid \theta)$ — **вероятность данных**.

Правдоподобие

Определение

Правдоподобие — это *интерпретация* вероятности данных как функции от параметров:

$$L(\theta; D) := P(D \mid \theta)$$

- Если фиксируем параметры $\theta \to P(D \mid \theta)$ — **вероятность данных**.
- Если фиксируем данные $D \to L(\theta; D)$ — **правдоподобие параметров**.

Правдоподобие

Определение

Правдоподобие — это *интерпретация* вероятности данных как функции от параметров:

$$L(\theta; D) := P(D \mid \theta)$$

- Если фиксируем параметры $\theta \to P(D \mid \theta)$ — **вероятность данных**.
- Если фиксируем данные $D \to L(\theta; D)$ — **правдоподобие параметров**.
- **Пример:**

Правдоподобие

Определение

Правдоподобие — это *интерпретация* вероятности данных как функции от параметров:

$$L(\theta; D) := P(D \mid \theta)$$

- Если фиксируем параметры $\theta \to P(D \mid \theta)$ — **вероятность данных**.
- Если фиксируем данные $D \to L(\theta; D)$ — **правдоподобие параметров**.
- **Пример:**
  - Вероятность: «Какова вероятность получить 12 очков в 100 бросках костей, если кости честные?»

Правдоподобие

Определение

Правдоподобие — это *интерпретация* вероятности данных как функции от параметров:

$$L(\theta; D) := P(D \mid \theta)$$

- Если фиксируем параметры $\theta \to P(D \mid \theta)$ — **вероятность данных**.
- Если фиксируем данные $D \to L(\theta; D)$ — **правдоподобие параметров**.
- **Пример:**
  - Вероятность: «Какова вероятность получить 12 очков в 100 бросках костей, если кости честные?»
  - Правдоподобие: «Насколько правдоподобно, что кости честные, если в 100 бросках всегда выпало 12 очков?»

Правдоподобие

Определение

Правдоподобие — это *интерпретация* вероятности данных как функции от параметров:

$$L(\theta; D) := P(D \mid \theta)$$

- Если фиксируем параметры $\theta \to P(D \mid \theta)$ — **вероятность данных**.
- Если фиксируем данные $D \to L(\theta; D)$ — **правдоподобие параметров**.
- **Пример:**
  - Вероятность: «Какова вероятность получить 12 очков в 100 бросках костей, если кости честные?»
  - Правдоподобие: «Насколько правдоподобно, что кости честные, если в 100 бросках всегда выпало 12 очков?»
- Часто используют **лог-правдоподобие**:

$$\log P(D \mid \theta)$$

Оценка максимального правдоподобия (MLE)

Идея

В классической постановке параметры θ выбираются так, чтобы **максимизировать вероятность данных**:

$$\hat{\theta}_{\mathrm{MLE}} = \arg\max_{\theta} P(D \mid \theta)$$

- Мы выбираем те параметры, которые делают наблюдённые данные **наиболее вероятными**.

Оценка максимального правдоподобия (MLE)

Идея

В классической постановке параметры θ выбираются так, чтобы **максимизировать вероятность данных**:

$$\hat{\theta}_{\text{MLE}} = \arg\max_{\theta} P(D \mid \theta)$$

- Мы выбираем те параметры, которые делают наблюдённые данные **наиболее вероятными**.
- Эквивалентно: $\hat{\theta}_{\text{MLE}} = \arg\max_{\theta} L(\theta; D)$, где $L(\theta; D) = P(D \mid \theta)$ — функция правдоподобия.

Оценка максимального правдоподобия (MLE)

Идея

В классической постановке параметры θ выбираются так, чтобы **максимизировать вероятность данных**:

$$\hat{\theta}_{\mathrm{MLE}} = \arg\max_{\theta} P(D \mid \theta)$$

- Мы выбираем те параметры, которые делают наблюдённые данные **наиболее вероятными**.
- Эквивалентно: $\hat{\theta}_{\mathrm{MLE}} = \arg\max_{\theta} L(\theta; D)$, где $L(\theta; D) = P(D \mid \theta)$ — функция правдоподобия.
- Обычно используем **лог-правдоподобие**:

$$\hat{\theta}_{\mathrm{MLE}} = \arg\max_{\theta} \log P(D \mid \theta)$$

Оценка максимального правдоподобия (MLE)

Идея

В классической постановке параметры θ выбираются так, чтобы **максимизировать вероятность данных**:

$$\hat{\theta}_{\text{MLE}} = \arg\max_{\theta} P(D \mid \theta)$$

- Мы выбираем те параметры, которые делают наблюдённые данные **наиболее вероятными**.
- Эквивалентно: $\hat{\theta}_{\text{MLE}} = \arg\max_{\theta} L(\theta; D)$, где $L(\theta; D) = P(D \mid \theta)$ — функция правдоподобия.
- Обычно используем **лог-правдоподобие**:

$$\hat{\theta}_{\text{MLE}} = \arg\max_{\theta} \log P(D \mid \theta)$$

- С предположением, что данные независимы:

$$P(D \mid \theta) = \prod_{i=1}^{n} P(x_i \mid \theta) \quad \Rightarrow \quad \log P(D \mid \theta) = \sum_{i=1}^{n} \log P(x_i \mid \theta)$$

Maximum a Posteriori (MAP) и связь с MLE

Определение MAP

MAP-оценка выбирает параметры, которые **максимизируют постериор**:

$$\hat{\theta}_{\mathrm{MAP}} = \arg\max_{\theta} P(\theta \mid D) = \arg\max_{\theta} \frac{P(D \mid \theta)P(\theta)}{P(D)}$$

- MAP сочетает:

$$\underbrace{P(D \mid \theta)}_{\text{правдоподобие}} \times \underbrace{P(\theta)}_{\text{априор}}$$

Maximum a Posteriori (MAP) и связь с MLE

Определение MAP

MAP-оценка выбирает параметры, которые **максимизируют постериор**:

$$\hat{\theta}_{\text{MAP}} = \arg\max_{\theta} P(\theta \mid D) = \arg\max_{\theta} \frac{P(D \mid \theta)P(\theta)}{P(D)}$$

- MAP сочетает:

$$\underbrace{P(D \mid \theta)}_{\text{правдоподобие}} \times \underbrace{P(\theta)}_{\text{априор}}$$

- В логарифмах:

$$\hat{\theta}_{\text{MAP}} = \arg\max_{\theta} \left[\log P(D \mid \theta) + \log P(\theta)\right]$$

Maximum a Posteriori (MAP) и связь с MLE

Определение MAP

MAP-оценка выбирает параметры, которые **максимизируют постериор**:

$$\hat{\theta}_{\text{MAP}} = \arg\max_{\theta} P(\theta \mid D) = \arg\max_{\theta} \frac{P(D \mid \theta)P(\theta)}{P(D)}$$

- MAP сочетает:

$$\underbrace{P(D \mid \theta)}_{\text{правдоподобие}} \times \underbrace{P(\theta)}_{\text{априор}}$$

- В логарифмах:

$$\hat{\theta}_{\text{MAP}} = \arg\max_{\theta} [\log P(D \mid \theta) + \log P(\theta)]$$

- **MLE как частный случай**: если априор $P(\theta)$ равномерный, тогда $\hat{\theta}_{\text{MAP}} = \hat{\theta}_{\text{MLE}}$.

Rich Man's Bayes vs Poor Man's Bayes

Постериор

$$P(\theta \mid D) = \frac{P(D \mid \theta)P(\theta)}{P(D)}, \quad \text{где } P(D) = \int P(D \mid \theta)P(\theta)\, d\theta$$

Rich Man's Bayes

- Считаем $P(D)$ точно.

Poor Man's Bayes

Rich Man's Bayes vs Poor Man's Bayes

Постериор

$$P(\theta \mid D) = \frac{P(D \mid \theta)P(\theta)}{P(D)}, \quad \text{где } P(D) = \int P(D \mid \theta)P(\theta)\, d\theta$$

Rich Man’s Bayes

- Считаем $P(D)$ точно.
- Получаем **нормированное** распределение $P(\theta \mid D)$.

Poor Man’s Bayes

Rich Man's Bayes vs Poor Man's Bayes

Постериор

$$P(\theta \mid D) = \frac{P(D \mid \theta)P(\theta)}{P(D)}, \quad \text{где } P(D) = \int P(D \mid \theta)P(\theta)\, d\theta$$

Rich Man's Bayes

- Считаем $P(D)$ точно.
- Получаем **нормированное** распределение $P(\theta \mid D)$.
- Используем при малом числе параметров или простых моделях.

Poor Man's Bayes

Rich Man's Bayes vs Poor Man's Bayes

Постериор

$$P(\theta \mid D) = \frac{P(D \mid \theta)P(\theta)}{P(D)}, \quad \text{где } P(D) = \int P(D \mid \theta)P(\theta)\, d\theta$$

Rich Man's Bayes

- Считаем $P(D)$ точно.
- Получаем **нормированное** распределение $P(\theta \mid D)$.
- Используем при малом числе параметров или простых моделях.
- Пример: аналитические задачи, conjugate priors.

Poor Man's Bayes

Rich Man's Bayes vs Poor Man's Bayes

Постериор

$$P(\theta \mid D) = \frac{P(D \mid \theta)P(\theta)}{P(D)}, \quad \text{где } P(D) = \int P(D \mid \theta)P(\theta)\, d\theta$$

Rich Man's Bayes

- Считаем $P(D)$ точно.
- Получаем **нормированное** распределение $P(\theta \mid D)$.
- Используем при малом числе параметров или простых моделях.
- Пример: аналитические задачи, conjugate priors.

Poor Man's Bayes

- Игнорируем $P(D)$, работаем с $\tilde{P}(\theta \mid D) \propto P(D \mid \theta)P(\theta)$.

Rich Man's Bayes vs Poor Man's Bayes

Постериор

$$P(\theta \mid D) = \frac{P(D \mid \theta)P(\theta)}{P(D)}, \quad \text{где } P(D) = \int P(D \mid \theta)P(\theta)\, d\theta$$

Rich Man's Bayes

- Считаем $P(D)$ точно.
- Получаем **нормированное** распределение $P(\theta \mid D)$.
- Используем при малом числе параметров или простых моделях.
- Пример: аналитические задачи, conjugate priors.

Poor Man's Bayes

- Игнорируем $P(D)$, работаем с $\tilde{P}(\theta \mid D) \propto P(D \mid \theta)P(\theta)$.
- Не нормируем постериор.

Rich Man's Bayes vs Poor Man's Bayes

Постериор

$$P(\theta \mid D) = \frac{P(D \mid \theta)P(\theta)}{P(D)}, \quad \text{где } P(D) = \int P(D \mid \theta)P(\theta)\, d\theta$$

Rich Man's Bayes

- Считаем $P(D)$ точно.
- Получаем **нормированное** распределение $P(\theta \mid D)$.
- Используем при малом числе параметров или простых моделях.
- Пример: аналитические задачи, conjugate priors.

Poor Man's Bayes

- Игнорируем $P(D)$, работаем с $\tilde{P}(\theta \mid D) \propto P(D \mid \theta)P(\theta)$.
- Не нормируем постериор.
- Достаточно для MAP, MCMC, вариационного вывода.

Rich Man's Bayes vs Poor Man's Bayes

Постериор

$$P(\theta \mid D) = \frac{P(D \mid \theta)P(\theta)}{P(D)}, \quad \text{где } P(D) = \int P(D \mid \theta)P(\theta)\, d\theta$$

Rich Man's Bayes

- Считаем $P(D)$ точно.
- Получаем **нормированное** распределение $P(\theta \mid D)$.
- Используем при малом числе параметров или простых моделях.
- Пример: аналитические задачи, conjugate priors.

Poor Man's Bayes

- Игнорируем $P(D)$, работаем с $\tilde{P}(\theta \mid D) \propto P(D \mid \theta)P(\theta)$.
- Не нормируем постериор.
- Достаточно для MAP, MCMC, вариационного вывода.
- Популярен при высокоразмерных моделях.

Сопряжённые априоры (Conjugate Priors)

Определение

Априор $P(\theta)$ называется **сопряжённым** для правдоподобия $P(D \mid \theta)$, если **постериор** $P(\theta \mid D)$ принадлежит *тому же семейству распределений*, что и априор:

$$P(\theta \mid D) \propto P(D \mid \theta)\, P(\theta), \quad \text{где } P(\theta),\, P(\theta \mid D) \in \mathcal{F}.$$

- Главная идея: выбираем такой априор, чтобы после обновления мы получили распределение **того же типа**.

Сопряжённые априоры (Conjugate Priors)

Определение

Априор $P(\theta)$ называется **сопряжённым** для правдоподобия $P(D \mid \theta)$, если **постериор** $P(\theta \mid D)$ принадлежит *тому же семейству распределений*, что и априор:

$$P(\theta \mid D) \propto P(D \mid \theta)\, P(\theta), \quad \text{где } P(\theta),\, P(\theta \mid D) \in \mathcal{F}.$$

- Главная идея: выбираем такой априор, чтобы после обновления мы получили распределение **того же типа**.
- Это даёт:

Сопряжённые априоры (Conjugate Priors)

Определение

Априор $P(\theta)$ называется **сопряжённым** для правдоподобия $P(D \mid \theta)$, если **постериор** $P(\theta \mid D)$ принадлежит *тому же семейству распределений*, что и априор:

$$P(\theta \mid D) \propto P(D \mid \theta)\, P(\theta), \quad \text{где } P(\theta),\, P(\theta \mid D) \in \mathcal{F}.$$

- Главная идея: выбираем такой априор, чтобы после обновления мы получили распределение **того же типа**.
- Это даёт:
  - Закрытую аналитическую форму постериора.

Сопряжённые априоры (Conjugate Priors)

Определение

Априор $P(\theta)$ называется **сопряжённым** для правдоподобия $P(D \mid \theta)$, если **постериор** $P(\theta \mid D)$ принадлежит *тому же семейству распределений*, что и априор:

$$P(\theta \mid D) \propto P(D \mid \theta)\, P(\theta), \quad \text{где } P(\theta),\, P(\theta \mid D) \in \mathcal{F}.$$

- Главная идея: выбираем такой априор, чтобы после обновления мы получили распределение **того же типа**.
- Это даёт:
  - Закрытую аналитическую форму постериора.
  - Простое обновление параметров.

Сопряжённые априоры (Conjugate Priors)

Определение

Априор $P(\theta)$ называется **сопряжённым** для правдоподобия $P(D \mid \theta)$, если **постериор** $P(\theta \mid D)$ принадлежит *тому же семейству распределений*, что и априор:

$$P(\theta \mid D) \propto P(D \mid \theta)\, P(\theta), \quad \text{где } P(\theta),\, P(\theta \mid D) \in \mathcal{F}.$$

- Главная идея: выбираем такой априор, чтобы после обновления мы получили распределение **того же типа**.
- Это даёт:
  - Закрытую аналитическую форму постериора.
  - Простое обновление параметров.
  - Упрощение байесовского вывода.

Сопряжённые априоры (Conjugate Priors)

Определение

Априор $P(\theta)$ называется **сопряжённым** для правдоподобия $P(D \mid \theta)$, если **постериор** $P(\theta \mid D)$ принадлежит *тому же семейству распределений*, что и априор:

$$P(\theta \mid D) \propto P(D \mid \theta)\, P(\theta), \quad \text{где } P(\theta),\, P(\theta \mid D) \in \mathcal{F}.$$

- Главная идея: выбираем такой априор, чтобы после обновления мы получили распределение **того же типа**.
- Это даёт:
  - Закрытую аналитическую форму постериора.
  - Простое обновление параметров.
  - Упрощение байесовского вывода.
- Если априор **не сопряжён**, постериор не имеет аналитической формы:

$$P(\theta \mid D) = \frac{P(D \mid \theta)\, P(\theta)}{\int P(D \mid \theta)\, P(\theta)\, d\theta}$$

→ сложный интеграл, как на картинке.

Сопряжённые априоры (Conjugate Priors)

Определение

Априор $P(\theta)$ называется **сопряжённым** для правдоподобия $P(D \mid \theta)$, если **постериор** $P(\theta \mid D)$ принадлежит *тому же семейству распределений*, что и априор:

$$P(\theta \mid D) \propto P(D \mid \theta)\, P(\theta), \quad \text{где } P(\theta),\, P(\theta \mid D) \in \mathcal{F}.$$

- Главная идея: выбираем такой априор, чтобы после обновления мы получили распределение **того же типа**.
- Это даёт:
  - Закрытую аналитическую форму постериора.
  - Простое обновление параметров.
  - Упрощение байесовского вывода.
- Если априор **не сопряжён**, постериор не имеет аналитической формы:

$$P(\theta \mid D) = \frac{P(D \mid \theta)\, P(\theta)}{\int P(D \mid \theta)\, P(\theta)\, d\theta}$$

  → сложный интеграл, как на картинке.
- Примеры сопряжённых пар:

Сопряжённые априоры (Conjugate Priors)

Определение

Априор $P(\theta)$ называется **сопряжённым** для правдоподобия $P(D \mid \theta)$, если **постериор** $P(\theta \mid D)$ принадлежит *тому же семейству распределений*, что и априор:

$$P(\theta \mid D) \propto P(D \mid \theta)\, P(\theta), \quad \text{где } P(\theta), P(\theta \mid D) \in \mathcal{F}.$$

- Главная идея: выбираем такой априор, чтобы после обновления мы получили распределение **того же типа**.
- Это даёт:
  - Закрытую аналитическую форму постериора.
  - Простое обновление параметров.
  - Упрощение байесовского вывода.
- Если априор **не сопряжён**, постериор не имеет аналитической формы:

$$P(\theta \mid D) = \frac{P(D \mid \theta)\, P(\theta)}{\int P(D \mid \theta)\, P(\theta)\, d\theta}$$

  → сложный интеграл, как на картинке.
- Примеры сопряжённых пар:
  - Бета ↔ биномиальное правдоподобие.

Сопряжённые априоры (Conjugate Priors)

Определение

Априор $P(\theta)$ называется **сопряжённым** для правдоподобия $P(D \mid \theta)$, если **постериор** $P(\theta \mid D)$ принадлежит *тому же семейству распределений*, что и априор:

$$P(\theta \mid D) \propto P(D \mid \theta)\, P(\theta), \quad \text{где } P(\theta), P(\theta \mid D) \in \mathcal{F}.$$

- Главная идея: выбираем такой априор, чтобы после обновления мы получили распределение **того же типа**.
- Это даёт:
  - Закрытую аналитическую форму постериора.
  - Простое обновление параметров.
  - Упрощение байесовского вывода.
- Если априор **не сопряжён**, постериор не имеет аналитической формы:

$$P(\theta \mid D) = \frac{P(D \mid \theta)\, P(\theta)}{\int P(D \mid \theta)\, P(\theta)\, d\theta}$$

→ сложный интеграл, как на картинке.

- Примеры сопряжённых пар:
  - Бета ↔ биномиальное правдоподобие.
  - Нормальное ↔ нормальное.

Сопряжённые априоры (Conjugate Priors)

Определение

Априор $P(\theta)$ называется **сопряжённым** для правдоподобия $P(D \mid \theta)$, если **постериор** $P(\theta \mid D)$ принадлежит *тому же семейству распределений*, что и априор:

$$P(\theta \mid D) \propto P(D \mid \theta)\, P(\theta), \quad \text{где } P(\theta),\, P(\theta \mid D) \in \mathcal{F}.$$

- Главная идея: выбираем такой априор, чтобы после обновления мы получили распределение **того же типа**.
- Это даёт:
  - Закрытую аналитическую форму постериора.
  - Простое обновление параметров.
  - Упрощение байесовского вывода.
- Если априор **не сопряжён**, постериор не имеет аналитической формы:

$$P(\theta \mid D) = \frac{P(D \mid \theta)\, P(\theta)}{\int P(D \mid \theta)\, P(\theta)\, d\theta}$$

  → сложный интеграл, как на картинке.
- Примеры сопряжённых пар:
  - Бета ↔ биномиальное правдоподобие.
  - Нормальное ↔ нормальное.
  - Гамма ↔ пуассон.

Пример: биномиальное правдоподобие + бета-априор

Задача

Подбрасываем монетку n раз, в k случаях выпал орёл.

Правдоподобие: биномиальное

$$P(D \mid \theta) = \binom{n}{k} \theta^k (1-\theta)^{n-k}$$

где θ — вероятность орла.

Априор: бета-распределение

$$P(\theta) = \mathrm{Beta}(\theta;\ \alpha, \beta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha, \beta)}$$

Постериор: снова бета!

$$P(\theta \mid D) \propto \theta^k (1-\theta)^{n-k} \cdot \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$$\boxed{P(\theta \mid D) = \mathrm{Beta}\big(\alpha + k,\ \beta + n - k\big)}$$