Seminar 7: Байесовские методы

Carlos Buitrago

Table of contents

1 Байесовский подход: постановка задачи **1**
1.1 Формула Байеса . . . 2
1.2 Байесовская оценка и функция потерь . . . 2

2 Сопряжённые семейства **3**
2.1 Основные пары «модель — априорное распределение» . . . 3
2.2 Beta–Binomial . . . 3
2.3 Gamma–Poisson . . . 4
2.4 Gamma–Exponential: параметр интенсивности . . . 4
2.5 Нормальная модель с известной дисперсией . . . 4
2.6 Равномерная модель и распределение Парето . . . 4

3 Априорное распределение Джеффриса **5**
3.1 Пример: Poisson . . . 5
3.2 Пример: $N(\mu, 1)$ . . . 6

4 Состоятельность: как проверять **6**

5 Дискретные априорные распределения и сдвиг равномерного закона **6**
5.1 Двухточечный приор для нормального среднего . . . 6
5.2 Пуассоновский приор для $U(\theta, \theta + 1)$ . . . 7

6 Численные иллюстрации **7**

7 Памятка: алгоритм решения задачи **8**

1 Байесовский подход: постановка задачи

В частотном подходе неизвестный параметр θ фиксирован. В байесовском подходе параметр рассматривается как случайная величина Θ с **априорным распределением** π, а данные $X = (X_1, \dots, X_n)$ имеют условное распределение P_θ при $\Theta = \theta$.

> **Обозначения**
>
> - $\pi(\theta)$ (или $q(\theta)$) — априорная плотность / масса;
> - $L(\theta; x) = f(x \mid \theta)$ — функция правдоподобия;
> - $\pi(\theta \mid x)$ — апостериорная плотность / масса;
> - $\delta(X)$ — статистическое решающее правило (оценка).
>
> Для дискретных параметров интегралы заменяются суммами. Не следует путать априорное распределение параметра и распределение наблюдений.

1.1 Формула Байеса

Определение 1. Априорное распределение

Распределение π параметра Θ до наблюдения данных называется **априорным**.

Определение 2. Апостериорное распределение

Если $m(x) = \int_\Theta L(t;x)\pi(t)\,dt$ конечно и положительно, то

$$\boxed{\pi(\theta \mid x) = \frac{L(\theta;x)\pi(\theta)}{\int_\Theta L(t;x)\pi(t)\,dt}.}$$

Знаменатель $m(x)$ называется **маргинальной плотностью данных** (evidence). В вычислениях удобно писать $\pi(\theta \mid x) \propto L(\theta;x)\pi(\theta)$ и затем нормировать.

Для дискретного параметра:

$$P(\Theta = t \mid X = x) = \frac{P(X = x \mid \Theta = t)P(\Theta = t)}{\sum_s P(X = x \mid \Theta = s)P(\Theta = s)}.$$

1.2 Байесовская оценка и функция потерь

Определение 3. Байесовская оценка при квадратичной потере

При потере $\ell(a,\theta) = (a-\theta)^2$ байесовская оценка равна

$$\boxed{\hat\theta_B(X) = \mathbb{E}[\Theta \mid X].}$$

Для существования требуется конечный соответствующий апостериорный второй момент (либо достаточно условий, обеспечивающих конечный риск).

Определение 4. Байесовский риск и оптимальность

Для произвольной потери $\ell(a,\theta)$ частотный риск правила δ:

$$R(\theta,\delta) = \mathbb{E}_\theta[\ell(\delta(X),\theta)].$$

Байесовский риск:

$$r(\pi,\delta) = \int_\Theta R(\theta,\delta)\,\pi(d\theta).$$

Правило δ^* **байесовски оптимально**, если $r(\pi,\delta^*) \le r(\pi,\delta)$ для любого допустимого δ. Строгое неравенство для каждого другого правила, как в исходном конспекте, вообще говоря, не требуется: оптимальных правил может быть несколько.

Теорема. Оптимальность апостериорного среднего

При квадратичной потере и конечных вторых моментах оптимальное правило:

$$\delta^*(X) = \mathbb{E}[\Theta \mid X].$$

Доказательство. Для фиксированного x и $m(x) = \mathbb{E}[\Theta \mid X = x]$:

$$\mathbb{E}[(a-\Theta)^2 \mid X = x] = (a - m(x))^2 + \mathrm{Var}(\Theta \mid X = x).$$

Минимум достигается при $a = m(x)$. Интегрирование по X даёт минимум байесовского риска. □

⚠ Зависимость от функции потерь

- Квадратичная потеря: апостериорное **среднее**.
- Абсолютная потеря $|a-\theta|$: апостериорная **медиана**.
- Потеря $\mathbf{1}\{a \neq \theta\}$ для дискретного параметра: апостериорная **мода** (MAP).

Поэтому формула «байесовская оценка = апостериорное среднее» предполагает квадратичную потерю.

2 Сопряжённые семейства

❗ Определение 5. Сопряжённое семейство

Семейство априорных распределений $\{\pi_\eta : \eta \in H\}$ называется **сопряжённым** к модели $\{P_\theta\}$, если апостериорное распределение принадлежит тому же семейству для всех допустимых выборок (с обновлённым гиперпараметром $\eta' = T(\eta, x)$).

2.1 Основные пары «модель — априорное распределение»

Условная модель	Априорное распределение	Апостериорное распределение
$X_i \sim \mathrm{Bern}(p)$	$p \sim \mathrm{Beta}(\alpha, \beta)$	$\mathrm{Beta}(\alpha + S, \beta + n - S)$
$X_i \sim \mathrm{Bin}(m, p)$	$p \sim \mathrm{Beta}(\alpha, \beta)$	$\mathrm{Beta}(\alpha + S, \beta + nm - S)$
$X_i \sim \mathrm{Pois}(\lambda)$	$\lambda \sim \mathrm{Gamma}(\alpha, \beta)$	$\mathrm{Gamma}(\alpha + S, \beta + n)$
$X_i \sim \mathrm{Exp}(\lambda)$, интенсивность λ	$\lambda \sim \mathrm{Gamma}(\alpha, \beta)$	$\mathrm{Gamma}(\alpha + n, \beta + S)$
$X_i \sim N(\mu, \sigma^2)$, σ^2 известно	$\mu \sim N(a, \tau^2)$	$N(m_n, v_n)$
$X_i \sim N(0, v)$, v неизвестна	$v \sim \mathrm{InvGamma}(\alpha, \beta)$	$\mathrm{InvGamma}(\alpha + n/2, \beta + \sum X_i^2/2)$
$X_i \sim U(0, \theta)$	$\theta \sim \mathrm{Pareto}(b, \alpha)$	$\mathrm{Pareto}(\max(b, M), \alpha + n)$

Здесь $S = \sum_i X_i$, $M = \max_i X_i$; Gamma использует параметризацию **shape–rate**:

$$f(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}, \quad \mathbb{E}\lambda = \frac{\alpha}{\beta}.$$

Для $\mathrm{InvGamma}(\alpha, \beta)$ плотность $\propto v^{-\alpha-1} e^{-\beta/v}$ и среднее $\beta/(\alpha-1)$ при $\alpha > 1$.

2.2 Beta–Binomial

Пусть $X_i \sim \mathrm{Bin}(m, p)$, $p \sim \mathrm{Beta}(\alpha, \beta)$, $S = \sum_i X_i$. Тогда

$$L(p; x) \propto p^S (1-p)^{nm-S}, \qquad \pi(p) \propto p^{\alpha-1}(1-p)^{\beta-1},$$

$$p \mid X \sim \mathrm{Beta}(\alpha + S, \beta + nm - S), \qquad \hat{p}_B = \frac{\alpha + S}{\alpha + \beta + nm}.$$

Если α, β фиксированы, $\hat{p}_B \xrightarrow{P} p_0$ по ЗБЧ при истинном p_0.

2.3 Gamma–Poisson

Для $X_i \sim \text{Pois}(\lambda)$, $\lambda \sim \text{Gamma}(\alpha, \beta)$:

$$\pi(\lambda \mid x) \propto \lambda^{\alpha+S-1} e^{-(\beta+n)\lambda},$$

$$\boxed{\hat{\lambda}_B = \frac{\alpha+S}{\beta+n}.}$$

2.4 Gamma–Exponential: параметр интенсивности

Если $f(x \mid \lambda) = \lambda e^{-\lambda x} \mathbf{1}_{x \geq 0}$, то

$$\lambda \mid X \sim \text{Gamma}(\alpha+n, \beta+S), \quad \hat{\lambda}_B = \frac{\alpha+n}{\beta+S} \xrightarrow{P} \lambda_0.$$

> ⚠ Если θ — масштаб, а не интенсивность
>
> При $f(x \mid \theta) = \theta^{-1} e^{-x/\theta}$, $\theta > 0$, сопряжённое распределение — **обратное гамма-распределение**:
>
> $$\theta \sim \text{InvGamma}(\alpha, \beta) \implies \theta \mid X \sim \text{InvGamma}(\alpha+n, \beta+S),$$
>
> $$\hat{\theta}_B = \frac{\beta+S}{\alpha+n-1}, \quad \alpha+n > 1.$$
>
> Параметризацию экспоненциального распределения всегда нужно оговаривать.

2.5 Нормальная модель с известной дисперсией

Если $X_i \mid \mu \sim N(\mu, \sigma^2)$ и $\mu \sim N(a, \tau^2)$, то

$$v_n = \left(\frac{1}{\tau^2} + \frac{n}{\sigma^2}\right)^{-1}, \qquad m_n = v_n \left(\frac{a}{\tau^2} + \frac{n\bar{X}}{\sigma^2}\right),$$

$$\mu \mid X \sim N(m_n, v_n), \qquad \hat{\mu}_B = m_n.$$

> 💡 Вывод дополнением до квадрата
>
> Логарифм апостериорной плотности с точностью до константы равен
>
> $$-\frac{(\mu-a)^2}{2\tau^2} - \frac{\sum_i (X_i-\mu)^2}{2\sigma^2} = -\frac{(\mu-m_n)^2}{2v_n} + C(x).$$
>
> При фиксированных a, τ^2, σ^2 коэффициент при $\bar{X}$ стремится к 1, поэтому $m_n \xrightarrow{P} \mu_0$.

2.6 Равномерная модель и распределение Парето

Если $X_i \mid \theta \sim U(0, \theta)$, то $L(\theta; x) = \theta^{-n} \mathbf{1}_{\theta \geq M}$ (при неотрицательных наблюдениях).

Для априорного Pareto(b, α) с плотностью

$$\pi(\theta) = \alpha b^\alpha \theta^{-\alpha-1} \mathbf{1}_{\theta \geq b},$$

получаем $\theta \mid X \sim \text{Pareto}(c, \alpha + n)$, $c = \max(b, M)$, и

$$\hat{\theta}_B = \frac{\alpha + n}{\alpha + n - 1} c \quad (\alpha + n > 1).$$

Если истинное $\theta_0 \geq b$, то $M \to \theta_0$ почти наверное, значит оценка состоятельна.

i Другой равномерный случай: априорное $U(0, 1)$

При $0 \leq M < 1$ апостериорная плотность пропорциональна $\theta^{-n} \mathbf{1}_{[M,1]}(\theta)$. Для $0 < M < 1$:

$$\hat{\theta}_B = \frac{\int_M^1 t^{1-n} dt}{\int_M^1 t^{-n} dt}.$$

Для $n \geq 3$:

$$\hat{\theta}_B = \frac{n-1}{n-2} \frac{M^{2-n} - 1}{M^{1-n} - 1}.$$

При $n = 2$: $\hat{\theta}_B = \frac{\log(1/M)}{M^{-1} - 1}$; при $n = 1$: $\hat{\theta}_B = \frac{1 - M}{\log(1/M)}$. Если $M \to \theta_0 \in (0, 1)$, то апостериорная оценка всё равно стремится к θ_0 при $n \to \infty$.

3 Априорное распределение Джеффриса

! Определение 6. Правило Джеффриса

Для одномерной регулярной модели с информацией Фишера от **одного наблюдения**

$$I(\theta) = \mathbb{E}_\theta \left[\left(\frac{\partial}{\partial \theta} \log f(X \mid \theta) \right)^2 \right],$$

априорная плотность Джеффриса задаётся (с точностью до множителя)

$$\boxed{\pi_J(\theta) \propto \sqrt{I(\theta)}.}$$

Для векторного параметра $\pi_J(\theta) \propto \sqrt{\det I(\theta)}$.

⚠ Важные оговорки

Априорное распределение Джеффриса может быть **несобственным** (не интегрироваться в единицу). Использовать его можно, если полученное апостериорное распределение собственное. Формулы регулярной информации требуют условий дифференцируемости и интегрируемости; для модели $U(0, \theta)$ с зависящим от параметра носителем стандартные регулярные тождества могут не выполняться.

3.1 Пример: Poisson

$$\log f(x \mid \lambda) = -\lambda + x \log \lambda - \log(x!), \qquad I(\lambda) = \frac{1}{\lambda}.$$

Следовательно, $\pi_J(\lambda) \propto \lambda^{-1/2}$, то есть формально $\text{Gamma}(1/2, 0)$ (несобственное предельное обозначение).

После $n \geq 1$ наблюдений:

$$\lambda \mid X \sim \text{Gamma}(S + 1/2, n), \qquad \boxed{\hat{\lambda}_J = \frac{S + 1/2}{n} = \bar{X} + \frac{1}{2n}.}$$

3.2 Пример: $N(\mu, 1)$

$I(\mu) = 1$, значит $\pi_J(\mu) \propto 1$ на $\mathbb{R}$ (несобственное равномерное распределение). При $n \geq 1$:

$$\mu \mid X \sim N(\bar{X}, 1/n), \qquad \boxed{\hat{\mu}_J = \bar{X}.}$$

> Сравнение с сопряжёнными априорными распределениями
>
> Для Poisson с Gamma(α, β) получаем $(S+\alpha)/(n+\beta)$; при Джеффрисе — $(S+1/2)/n$. Для нормальной модели с априорным $N(a, \tau^2)$ получаем взвешенное среднее, а при Джеффрисе — $\bar{X}$.

4 Состоятельность: как проверять

> Определение
>
> Оценка $\delta_n(X_1, \dots, X_n)$ **состоятельна** в точке θ_0, если при выборке из P_{θ_0}
>
> $$\delta_n \xrightarrow[n\to\infty]{P_{\theta_0}} \theta_0.$$
>
> Состоятельность — **частотное** свойство, даже если правило построено байесовским способом.

Полезные факты: $\bar{X} \xrightarrow{P} \mathbb{E}_{\theta_0} X_1$ по ЗБЧ; для $U(0, \theta_0)$

$$P_{\theta_0}(M_n \leq t) = (t/\theta_0)^n, \quad 0 \leq t \leq \theta_0, \qquad M_n \xrightarrow{a.s.} \theta_0.$$

Для $U(\theta_0, \theta_0 + 1)$ минимум $m_n \to \theta_0$, максимум $M_n \to \theta_0 + 1$ почти наверное.

> Ограничение носителя априорного распределения
>
> Если истинное значение параметра исключено априорным распределением, байесовская оценка может быть несостоятельной. В частности, при априорном Bern(p) параметр может принимать **только 0 или 1**, поэтому состоятельность следует проверять только для $\theta_0 \in \{0, 1\}$ в рамках байесовской модели.

5 Дискретные априорные распределения и сдвиг равномерного закона

5.1 Двухточечный приор для нормального среднего

Пусть $X_i \mid \theta \sim N(\theta, 1)$ и $P(\Theta = 1) = p$, $P(\Theta = 0) = 1 - p$, $0 < p < 1$. Тогда

$$\frac{P(\Theta = 1 \mid x)}{P(\Theta = 0 \mid x)} = \frac{p}{1-p} \exp\left(S - \frac{n}{2}\right),$$

$$\boxed{\hat{\theta}_B = P(\Theta = 1 \mid X) = \frac{p e^{S - n/2}}{1 - p + p e^{S - n/2}}.}$$

При $\theta_0 = 1$ логарифм апостериорных шансов $= \log(p/(1-p)) + n/2 + O_P(\sqrt{n}) \to +\infty$, а при $\theta_0 = 0$ он стремится к $-\infty$. Значит оценка состоятельна на $\{0, 1\}$. При $p = 0$ или 1 приор вырожден и вывод меняется.

5.2 Пуассоновский приор для $U(\theta, \theta + 1)$

Пусть $\Theta \sim \text{Pois}(1)$, так что $\Theta = k \in \{0, 1, 2, ...\}$ с весами $e^{-1}/k!$. Для наблюдений x правдоподобие при k равно

$$L(k; x) = \mathbf{1}\{k \le m,\ M \le k + 1\}, \qquad m = \min_i x_i, \quad M = \max_i x_i.$$

Следовательно,

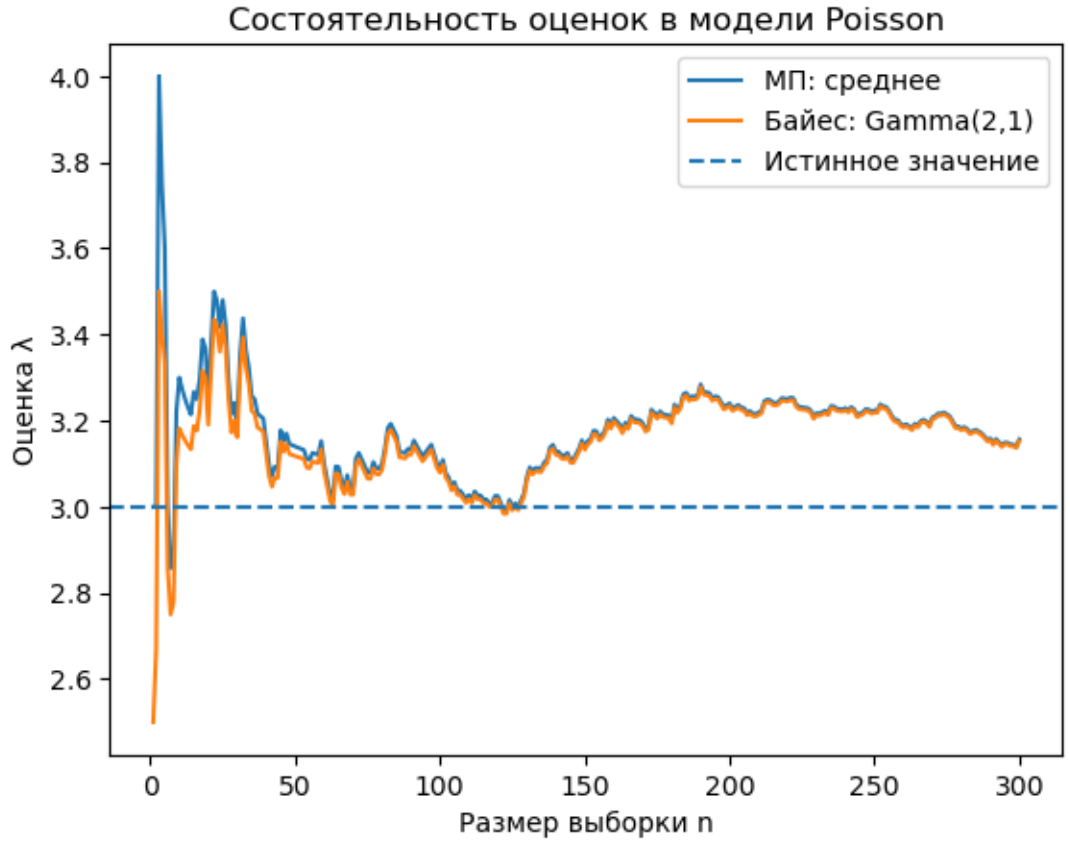
$$P(\Theta = k \mid x) = \frac{\frac{1}{k!}\mathbf{1}\{M - 1 \le k \le m\}}{\sum_{j \ge 0} \frac{1}{j!}\mathbf{1}\{M - 1 \le j \le m\}},$$

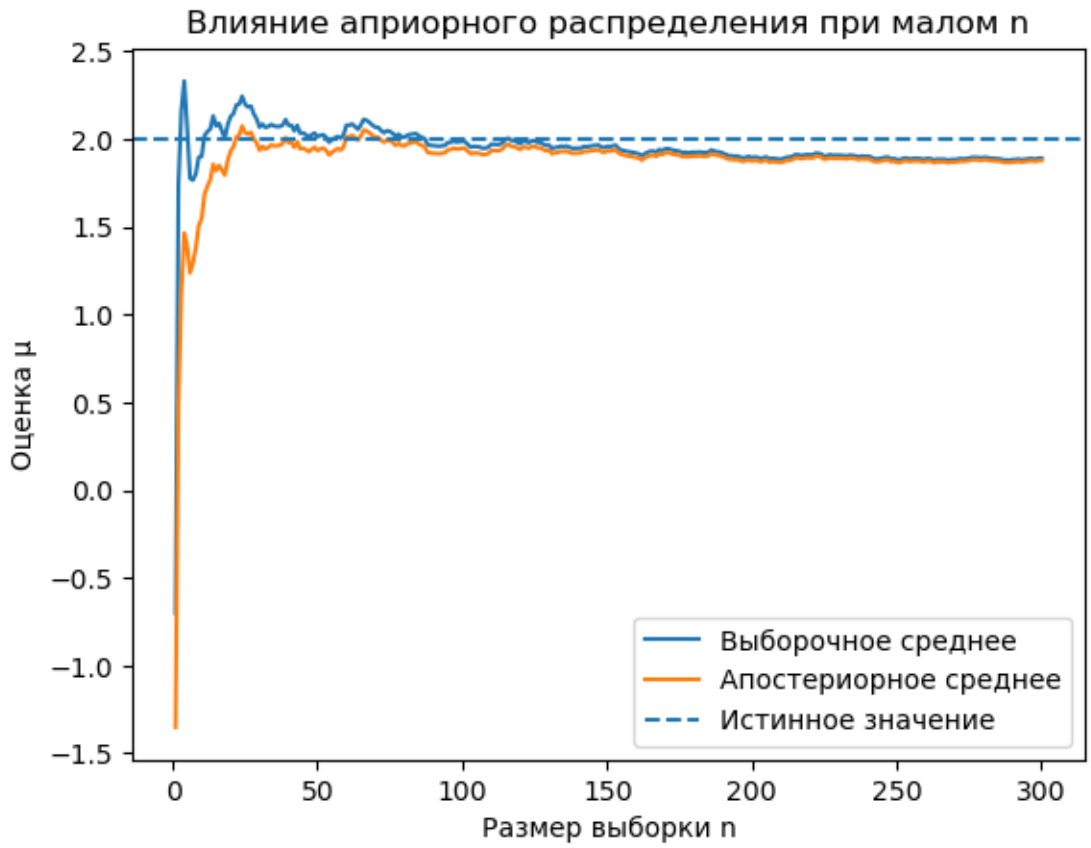
$$\boxed{\hat{\theta}_B = \frac{\sum_{k \ge 0} \frac{k}{k!}\mathbf{1}\{M - 1 \le k \le m\}}{\sum_{k \ge 0} \frac{1}{k!}\mathbf{1}\{M - 1 \le k \le m\}}.}$$

Интервал $[M - 1, m]$ имеет длину $1 - (M - m)$, которая стремится к нулю. При истинном целом $\theta_0 \ge 0$ апостериорное распределение концентрируется в θ_0 и оценка состоятельна. Для нецелого истинного параметра модель приора неверно специфицирована, и универсальная состоятельность не гарантируется.

6 Численные иллюстрации

В следующих экспериментах сравним байесовскую оценку и оценку максимального правдоподобия. Численный эксперимент иллюстрирует, но не заменяет доказательство состоятельности.





7 Памятка: алгоритм решения задачи

1. **Уточнить параметризацию** распределения и пространство параметров.
2. **Выписать правдоподобие** $L(\theta; x)$, не забывая индикаторы носителя.
3. **Умножить на априорную плотность** и определить нормирующую константу.
4. **Распознать семейство** апостериорного распределения или вычислить интегралы / суммы.
5. **Найти апостериорное среднее** для квадратичной потери и проверить его существование.
6. **Проверить состоятельность** при фиксированном истинном θ_0 с учётом носителя приора.
7. Для **приора Джеффриса** сначала найти информацию Фишера и проверить собственность апостериорного распределения.