Байесовский подход

Содержание

1 Критерия отношения правдоподобия. Лемма Неймана-Пирсона

1.1 Общая форма критерия отношения правдоподобия

Для простых гипотез $H_0 : \theta = \theta_0$ и $H_1 : \theta = \theta_1$ статистика отношения правдоподобия имеет вид:

$$\Lambda(X) = \frac{L(\theta_1|X)}{L(\theta_0|X)} = \frac{P(X|\theta_1)}{P(X|\theta_0)} \tag{1}$$

где $L(\theta|X)$ — функция правдоподобия.

1.2 Теорема Неймана-Пирсона

Для простых гипотез критерий отношения правдоподобия являетсяся асимптотически *наиболее мощным* (UMP) в частотном подходе. Для сложных гипотез (с подстановкой достаточных статистик) - равномерно наиболее мощным.

1.3 Пример: нормальное распределение

Для $X \sim N(\mu, \sigma^2)$ при известной σ^2:

$$\Lambda = \exp\left(\frac{n(\mu_1 - \mu_0)}{\sigma^2}\left(\bar{X} - \frac{\mu_0 + \mu_1}{2}\right)\right) \tag{2}$$

Пусть $\mu_1 > \mu_0$. Пороговое значение t_α — такое, что $P(\Lambda) > t_\alpha|H_0) = \alpha$.

В данном случае получаем, что критерий со статистикой Λ эквивалентен обычному критерию со статистикой $t = \bar{x}$ или, эквивалентно, $t = \frac{\sqrt{n}(\bar{x}-\mu_0)}{\sigma}$ с распределением $N(0, 1)$.

Реализация в R:

```r
# Параметры
n <- 100; mu0 <- 0; mu1 <- 0.5; sigma <- 1

# Генерация данных
set.seed(42)
x <- rnorm(n, mean=mu1, sd=sigma)

# Вычисление отношения правдоподобия
x_bar <- mean(x)
LR <- exp(n*(mu1-mu0)/sigma^2 * (x_bar - (mu0+mu1)/2))

cat("Отношение правдоподобия =", round(LR, 3), "\n")
cat("Байесовский фактор =", round(LR, 3), "\n") # Совпадает для простых гипотез
```

2 Байесовский подход к проверке простых гипотез

2.1 Теоретические основы

Байесовский подход к проверке простых гипотез основан на теореме Байеса:

$$P(H_i|\text{данные}) = \frac{P(\text{данные}|H_i)P(H_i)}{P(\text{данные})} \tag{3}$$

где:

- $P(H_i)$ — априорная вероятность гипотезы H_i
- $P(\text{данные}|H_i)$ — правдоподобие данных при гипотезе H_i
- $P(\text{данные}) = \sum_j P(\text{данные}|H_j)P(H_j)$ — нормирующая константа

Для двух гипотез H_0 и H_1 вводится **Байесовский фактор**:

$$BF_{10} = \frac{P(\text{данные}|H_1)}{P(\text{данные}|H_0)} \tag{4}$$

2.2 Определение e-value

E-value (evidence value) определяется как:

$$e := \frac{P(X|H_1)}{P(X|H_0) + P(X|H_1)} = \frac{BF_{10}}{1 + BF_{10}} \tag{5}$$

где BF_{10} – байесовский фактор в пользу H_1.

2.3 Вероятностная интерпретация

E-value допускает две вероятностные интерпретации:

2.3.1 1. Условная вероятность

При равных априорных вероятностях ($P(H_0) = P(H_1) = 0.5$):

$$e = P(H_1|X) \tag{6}$$

Таким образом, в этом частном случае e-value *совпадает* с апостериорной вероятностью.

2.3.2 2. Минимальная апостериорная вероятность

В общем случае e-value представляет *нижнюю границу* для апостериорной вероятности:

$$e \leq P(H_1|X) \quad \text{для любых} \quad P(H_1) \in (0,1) \tag{7}$$

Это означает, что e-value всегда даёт консервативную оценку поддержки H_1.

Свойство	e-value	p-value
Основа	Байесовская	Частотная
Интерпретация	$P(H_1\|X)$ (при равных априорах)	$P(X$ или более крайние$\|H_0)$
Диапазон	[0,1]	[0,1]
Зависимость от альтернативы	Да	Нет

2.4 Сравнение с p-value

2.5 Пример 1: Нормальное распределение

Рассмотрим проверку гипотез о среднем значении нормального распределения с известной дисперсией:

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_1 : \mu = \mu_1 \tag{8}$$

Правдоподобия для выборки $X = (x_1, ..., x_n)$:

$$P(X|H_i) = \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_j - \mu_i)^2}{2\sigma^2}\right) \tag{9}$$

Байесовский фактор:

$$BF_{10} = \exp\left(-\frac{n}{2\sigma^2}[(\bar{x} - \mu_0)^2 - (\bar{x} - \mu_1)^2]\right) \tag{10}$$

Пример кода на R:

```r
# Параметры нормального распределения
n <- 100        # Размер выборки
mu0 <- 0        # Среднее при H0
mu1 <- 0.5      # Среднее при H1
sigma <- 1      # Стандартное отклонение (известно)

# Генерация данных (в реальном анализе используем фактические данные)
set.seed(42)
x <- rnorm(n, mean=mu1, sd=sigma)

# Вычисление выборочного среднего
x_bar <- mean(x)

# Расчет байесовского фактора
BF10 <- exp(-n/(2*sigma^2) * ((x_bar-mu0)^2 - (x_bar-mu1)^2))

# Вывод результатов
cat("Байесовский фактор BF10 =", round(BF10, 3), "\n")
cat("Поддержка H1:", ifelse(BF10 > 3, "сильная", "недостаточная"), "\n")
```

2.6 Пример 2: Модель Бернулли

Для проверки гипотез о вероятности успеха в схеме Бернулли:

$$H_0 : p = p_0 \quad \text{vs} \quad H_1 : p = p_1 \tag{11}$$

При наблюдении k успехов в n испытаниях:

$$BF_{10} = \frac{p_1^k (1-p_1)^{n-k}}{p_0^k (1-p_0)^{n-k}} \tag{12}$$

Пример кода:

```
# Параметры биномиального распределения
n <- 100        # Количество испытаний
p0 <- 0.5       # Вероятность успеха при H0
p1 <- 0.6       # Вероятность успеха при H1

# Наблюдаемые данные (в реальном анализе используем фактические данные)
k <- 55         # Количество успехов

# Расчет байесовского фактора
BF10 <- (p1^k * (1-p1)^(n-k)) / (p0^k * (1-p0)^(n-k))

# Вывод результатов
cat("Байесовский фактор BF10 =", round(BF10, 3), "\n")
cat("Апостериорная вероятность H1:",
    round(BF10/(1+BF10), 3), "\n") # При P(H0)=P(H1)=0.5
```

2.7 Пример 3: Проверка формы распределения

Проверка гипотез о форме распределения:

$$H_0 : X \sim N(0,1) \quad \text{vs} \quad H_1 : X \sim \text{Laplace}(0,1) \tag{13}$$

Плотности распределений:

$$f_N(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \tag{14}$$

$$f_L(x) = \frac{1}{2} e^{-|x|} \tag{15}$$

Байесовский фактор:

$$BF_{10} = \prod_{i=1}^{n} \frac{f_L(x_i)}{f_N(x_i)} = 2^{-n} (2\pi)^{n/2} \exp\left(\sum_{i=1}^{n} \frac{x_i^2}{2} - |x_i| \right) \tag{16}$$

Реализация в R:

```
# Установка пакета для генерации распределения Лапласа
if (!require("VGAM")) install.packages("VGAM")

# Генерация данных (здесь из нормального распределения)
n <- 100
x <- rnorm(n)

# Функции плотности
log_dnorm <- function(x) -0.5*log(2*pi) - x^2/2
log_dlaplace <- function(x) -log(2) - abs(x)

# Вычисление логарифма байесовского фактора
log_BF10 <- sum(log_dlaplace(x) - log_dnorm(x))
BF10 <- exp(log_BF10)

```

```
# Вывод результатов
cat("Байесовский фактор BF10 =", round(BF10, 3), "\n")
cat("Интерпретация:",
    if(BF10 > 30) "очень сильная поддержка H1 (Лаплас)" else
    if(BF10 > 10) "сильная поддержка H1" else
    if(BF10 > 3) "умеренная поддержка H1" else
    "поддержка H0 (нормальное)", "\n")
```

2.8 Интерпретация результатов

Шкала интерпретации Байесовского фактора:

Диапазон BF	Поддержка H_1	Аналог p-value
$BF_{10} < 1$	Поддержка H_0	$p > 0.1$
$1 - 3$	Слабая	$0.05 - 0.1$
$3 - 10$	Умеренная	$0.01 - 0.05$
$10 - 30$	Сильная	$0.001 - 0.01$
> 30	Очень сильная	< 0.001

Апостериорные вероятности при равных априорах:

$$P(H_1|\text{данные}) = \frac{BF_{10}}{1 + BF_{10}} \tag{17}$$

3 Связь между частотным и байесовским подходами через критерий отношения правдоподобия

3.1 Общая форма критерия отношения правдоподобия

Для простых гипотез $H_0 : \theta = \theta_0$ и $H_1 : \theta = \theta_1$ статистика отношения правдоподобия имеет вид:

$$\Lambda(X) = \frac{L(\theta_1|X)}{L(\theta_0|X)} = \frac{P(X|\theta_1)}{P(X|\theta_0)} \tag{18}$$

где $L(\theta|X)$ - функция правдоподобия.

3.2 Связь с байесовским фактором

Для простых гипотез байесовский фактор *совпадает* со статистикой отношения правдоподобия:

$$BF_{10} = \frac{P(X|H_1)}{P(X|H_0)} = \Lambda(X) \tag{19}$$

Однако интерпретация различается:

- В частотном подходе Λ сравнивается с критическим значением
- В байесовском - BF интерпретируется как степень поддержки H_1

3.3 Пример: нормальное распределение

Для $X \sim N(\mu, \sigma^2)$ при известной σ^2:

$$\Lambda = \exp\left(\frac{n(\mu_1 - \mu_0)}{\sigma^2}\left(\bar{X} - \frac{\mu_0 + \mu_1}{2}\right)\right) \tag{20}$$

Реализация в R:

```r
# Параметры
n <- 100; mu0 <- 0; mu1 <- 0.5; sigma <- 1

# Генерация данных
set.seed(42)
x <- rnorm(n, mean=mu1, sd=sigma)

# Вычисление отношения правдоподобия
x_bar <- mean(x)
LR <- exp(n*(mu1-mu0)/sigma^2 * (x_bar - (mu0+mu1)/2))

cat("Отношение правдоподобия =", round(LR, 3), "\n")
cat("Байесовский фактор =", round(LR, 3), "\n") # Совпадает для простых гипотез
```

3.4 Различия в интерпретации

Критерий	Частотный подход	Байесовский подход
Основа	p-value	BF
Порог	$p < \alpha$ (обычно 0.05)	$BF > 3$ (умеренная поддержка)
Интерпретация	Вероятность ошибки I рода	Относительная поддержка гипотез

Таблица 1: Сравнение интерпретаций

3.5 Теорема Неймана-Пирсона

Для простых гипотез критерий отношения правдоподобия является *наиболее мощным* (UMP) в частотном подходе. В байесовском подходе BF минимизирует общую вероятность ошибки при равных априорных вероятностях.

3.6 Асимптотическая связь

При больших выборках и регулярных условиях:

$$2\log\Lambda \approx 2\log BF_{10} \sim \chi^2_d \quad \text{при } H_0 \tag{21}$$

где d - разница в размерности параметров H_1 и H_0.

3.7 Пример: биномиальное распределение

Для $X \sim Bin(n, p)$:

$$\Lambda = \left(\frac{p_1}{p_0}\right)^X \left(\frac{1 - p_1}{1 - p_0}\right)^{n-X} \tag{22}$$

```r
n <- 100; p0 <- 0.5; p1 <- 0.6
X <- 55 # Число успехов

LR <- (p1/p0)^X * ((1-p1)/(1-p0))^(n-X)
BF <- LR # Для простых гипотез

cat("LR =", round(LR, 3), " BF =", round(BF, 3), "\n")
```

3.8 Выводы

- Для простых гипотез LR и BF численно совпадают
- Различия проявляются в интерпретации и подходе к проверке
- В частотном подходе LR используется для построения критериев
- В байесовском - BF напрямую измеряет поддержку гипотез

4 Разница в интерпретации через байесовский фактор и апостериорные вероятности

4.1 Формальное определение

Байесовский фактор (BF) и апостериорные вероятности - два взаимосвязанных, но различных способа интерпретации результатов в байесовском анализе:

$$BF_{10} = \frac{P(X|H_1)}{P(X|H_0)} \quad \text{(Байесовский фактор)} \tag{23}$$

$$P(H_1|X) = \frac{BF_{10} \cdot P(H_1)}{BF_{10} \cdot P(H_1) + P(H_0)} \quad \text{(Апостериорная вероятность)} \tag{24}$$

4.2 Ключевые различия

Характеристика	Байесовский фактор	Апостериорная вероятность
Зависимость от априоров	Нет	Да
Диапазон значений	$(0, +\infty)$	$[0, 1]$
Интерпретация	Относительная сила доказательств	Абсолютная вероятность гипоте
Учет баланса гипотез	Нет	Да

Таблица 2: Сравнение байесовского фактора и апостериорной вероятности

4.3 Пример с нормальным распределением

Рассмотрим проверку гипотез о среднем значении:

```r
# Параметры
n <- 100
mu0 <- 0
mu1 <- 0.5
```

```r
sigma <- 1
prior_H0 <- 0.5
prior_H1 <- 1 - prior_H0

# Генерация данных
set.seed(42)
x <- rnorm(n, mean=mu1, sd=sigma)

# Вычисление байесовского фактора
x_bar <- mean(x)
BF10 <- exp(n*(mu1-mu0)/sigma^2 * (x_bar - (mu0+mu1)/2))

# Вычисление апостериорной вероятности
post_H1 <- BF10 * prior_H1 / (BF10 * prior_H1 + prior_H0)

cat("Байесовский фактор BF10 =", round(BF10, 3), "\n")
cat("Апостериорная вероятность P(H1|X) =", round(post_H1, 3), "\n")
```

Результат выполнения:

```
Байесовский фактор BF10 = 56.598
Апостериорная вероятность P(H1|X) = 0.983
```

4.4 Чувствительность к априорным вероятностям

Апостериорная вероятность зависит от априорных вероятностей гипотез, в отличие от BF:

$$\text{Если } P(H_0) \to 0, \text{ то } P(H_1|X) \to 1 \text{ независимо от данных} \tag{25}$$

Пример с разными априорами:

```r
BF10 <- 10 # Сильная поддержка H1

# Случай 1: Равные априоры
post_H1_eq <- BF10 * 0.5 / (BF10 * 0.5 + 0.5) # 0.909

# Случай 2: Скептический априор (H1 маловероятна)
post_H1_skep <- BF10 * 0.1 / (BF10 * 0.1 + 0.9) # 0.526
```

4.5 Шкалы интерпретации

4.5.1 Для байесовского фактора

Значение BF	Интерпретация
$BF_{10} < 1$	Поддержка H_0
$1 \leq BF_{10} < 3$	Слабая поддержка H_1
$3 \leq BF_{10} < 10$	Умеренная поддержка
$10 \leq BF_{10} < 30$	Сильная поддержка
$BF_{10} \geq 30$	Очень сильная поддержка

4.5.2 Для апостериорных вероятностей

Вероятность	Интерпретация
$P(H_1\|X) < 0.5$	Поддержка H_0
$0.5 \leq P(H_1\|X) < 0.75$	Слабая поддержка H_1
$0.75 \leq P(H_1\|X) < 0.95$	Умеренная поддержка
$0.95 \leq P(H_1\|X) < 0.99$	Сильная поддержка
$P(H_1\|X) \geq 0.99$	Решающая поддержка

4.6 Практические рекомендации

- Используйте **байесовский фактор**, когда:
  - Нужно сравнить две конкретные модели
  - Априорные вероятности гипотез неизвестны или неопределены
  - Важна относительная сила доказательств
- Используйте **апостериорные вероятности**, когда:
  - Известны или могут быть обоснованы априорные вероятности
  - Нужно принять решение на основе полной вероятности
  - Требуется учесть баланс между гипотезами

4.7 Выводы

- BF измеряет *относительную* поддержку данных для гипотез, независимо от их априорных вероятностей
- Апостериорная вероятность дает *абсолютную* оценку вероятности гипотезы с учетом априоров
- Для полного байесовского вывода рекомендуется использовать оба показателя
- Интерпретация должна учитывать контекст задачи и обоснованность априорных предположений

5 Проверка гипотез со сложной альтернативой в байесовском подходе

5.1 Определение сложной гипотезы

Сложная гипотеза в байесовском анализе означает, что альтернативная гипотеза H_1 содержит:

- Неопределённый параметр θ с некоторым априорным распределением
- Множество возможных значений параметров

Формально:

$$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta \sim p(\theta|H_1) \tag{26}$$

5.2 Изменение расчёта байесовского фактора

Для сложной альтернативы байесовский фактор вычисляется как:

$$BF_{10} = \frac{P(X|H_1)}{P(X|H_0)} = \frac{\int P(X|\theta)p(\theta|H_1)d\theta}{P(X|\theta_0)} \tag{27}$$

где:

- $P(X|\theta)$ - функция правдоподобия
- $p(\theta|H_1)$ - априорное распределение параметра при H_1

5.3 Пример 1: Нормальное распределение

Проверка $H_0 : \mu = 0$ vs $H_1 : \mu \neq 0$:

```
library(BayesFactor)
set.seed(42)
x <- rnorm(100, mean=0.3, sd=1)

# Простая гипотеза
BF_simple <- dnorm(mean(x), mean=0.5, sd=1/sqrt(100)) /
             dnorm(mean(x), mean=0, sd=1/sqrt(100))

# Сложная гипотеза (используем априор Коши)
ttestBF(x, mu=0, rscale=0.707)
```

5.4 Пример 2: Модель Бернулли

Проверка $H_0 : p = 0.5$ vs $H_1 : p \neq 0.5$:

$$BF_{10} = \frac{\int_0^1 p^k(1-p)^{n-k}Beta(p|a,b)dp}{0.5^k(1-0.5)^{n-k}} \tag{28}$$

```
k <- 55; n <- 100
# Априор Beta(1,1) - равномерное распределение
marg_lik_H1 <- integrate(function(p) dbinom(k,n,p)*dbeta(p,1,1), 0,1)$value
lik_H0 <- dbinom(k,n,0.5)
BF10 <- marg_lik_H1 / lik_H0
```

5.5 Влияние выбора априора

При сложной альтернативе важную роль играет выбор априорного распределения:

Тип априора	Формула	Влияние на BF
Сопряжённый	Beta(a,b)	Аналитическое решение
Референсный	Коши, нормальный	Объективность
Субъективный	Экспертные оценки	Учёт внешних знаний

5.6 Вычислительные методы

Для сложных моделей используются:

- MCMC (Markov Chain Monte Carlo)
- Интегрирование по параметрам
- Методы аппроксимации (Laplace, BIC)

Пример с MCMC:

```
library(rstan)
model_code <- "
data {
  int n;
  int k;
}
parameters {
  real<lower=0,upper=1> p;
}
model {
  p ~ beta(1,1); // Априор
  k ~ binomial(n,p);
}"

# Скомпилируем модель
stan_model <- stan_model(model_code = model_code) #долго считается

# Запустим сэмплирование
fit <- sampling(
  object = stan_model,
  data = list(n = 100, k = 55),
  chains = 4,        # Количество цепей
  iter = 200,       # Общее количество итераций на цепь
  warmup = 100,     # Количество итераций для разогрева
  seed = 42          # Для воспроизводимости
)
print(fit)
```

5.7 Интерпретация результатов

Для сложных альтернатив:

- BF зависит от выбора априора
- Необходимо проводить анализ чувствительности
- Рекомендуется использовать ROPE (Region of Practical Equivalence)

Пример анализа чувствительности:

```
# Разные априоры для нормального распределения
bf_default <- ttestBF(x, mu=0, rscale=0.707)
bf_wide <- ttestBF(x, mu=0, rscale=1)
bf_narrow <- ttestBF(x, mu=0, rscale=0.5)
```

Критерий	Простая альтернатива	Сложная альтернатива
Вычисление BF	Прямое	Интегрирование
Зависимость от априора	Нет	Сильная
Вычислительная сложность	Низкая	Высокая
Интерпретация	Прямая	Требует осторожности

5.8 Сравнение с простыми гипотезами

5.9 Практические рекомендации

- Всегда указывайте используемое априорное распределение
- Проводите анализ чувствительности к выбору априора
- Для сложных моделей используйте MCMC-методы
- Сообщайте не только BF, но и апостериорные распределения параметров

6 Источники априорных распределений для параметров

6.1 Классификация априорных распределений

В байесовском анализе априорные распределения можно классифицировать следующим образом:

Тип априора	Описание	Пример
Информативные	Основаны на экспертном знании	Beta(10,10)
Слабые	С малым числом степеней свободы	Beta(1,1)
Референсные	Объективные по умолчанию	Коши(0,1)
Сопряжённые	Сохраняют форму при обновлении	Beta для биномиального

6.2 Экспертные знания

Априоры могут основываться на предыдущих исследованиях или экспертных оценках:

$$p(\theta) = \text{Beta}(\alpha, \beta), \text{ где } \alpha, \beta \text{ - параметры, основанные на данных} \tag{29}$$

Пример для вероятности успеха в клинических испытаниях:

- Эксперт оценивает вероятность успеха как 70±10%
- Соответствующий априор: Beta(α=8.25, β=3.75)

6.3 Сопряжённые априоры

Сопряжённые априоры удобны тем, что сохраняют форму при обновлении:

Формальное определение сопряжённости:

$$p(\theta|x) \propto p(x|\theta)p(\theta) \text{ принадлежит тому же семейству, что и } p(\theta) \tag{30}$$

Модель	Сопряжённый априор
Бернулли	Beta
Нормальная (изв. дисп.)	Нормальный
Нормальная (неизв. дисп.)	Норма-Гамма
Пуассон	Гамма

6.4 Reference Prior (Референсное априорное распределение)

Референсное априорное распределение (reference prior) — это тип неинформативного априора, разработанный Бергером и Бернардо (1992) как усовершенствование априора Джеффри для многомерных случаев. Оно получило название от английского *"reference"* (эталонный), поскольку служит стандартом для сравнения при отсутствии экспертных знаний.

6.4.1 Определение и построение

Для параметра θ референсный априор определяется через максимизацию ожидаемой информации Кульбака-Лейблера между апостериорным и априорным распределениями:

$$\pi(\theta) = \underset{\pi}{\operatorname{argmax}}\, \mathbb{E}\left[D_{\mathrm{KL}}(p(\theta|X)\|\pi(\theta))\right]$$

В одномерном случае часто совпадает с априором Джеффри:

$$\pi(\theta) \propto \sqrt{I(\theta)}$$

6.4.2 Ключевые свойства

- **Инвариантность**: Сохраняется при перепараметризации $\phi = g(\theta)$
- **Согласованность**: В многомерных случаях строится итеративно с учётом порядка параметров
- **Асимптотика**: Апостериорное распределение сходится к нормальному вокруг MLE-оценки

6.4.3 Примеры

- Для нормального распределения $\mathcal{N}(\mu, \sigma^2)$:

$$\pi(\mu, \sigma) \propto \frac{1}{\sigma} \quad (\text{в отличие от } 1/\sigma^2 \text{ у Джеффри})$$

- Для биномиальной модели $\mathrm{Bin}(n, p)$:

$$\pi(p) \propto p^{-1/2}(1-p)^{-1/2} \quad (\text{как у Джеффри})$$

6.4.4 Сравнение с другими априорами

6.4.5 Преимущества и критика

+ Лучше работает в многомерных моделях, чем априор Джеффри
+ Минимизирует влияние априора на выводы
- Сложность вычисления для сложных моделей
- Требует выбора порядка параметров

Тип априора	Формула	Многомерный случай
Равномерный	$\pi(\theta) \propto 1$	Проблемы
Джеффри	$\sqrt{I(\theta)}$	Парадоксы
Референсный	Через KLD-максимизацию	Оптимален

Таблица 3: Сравнение неинформативных априоров

6.5 Эмпирические байесовские методы

Используют сами данные для определения априора:

$$\hat{p}(\theta) = \arg\max_p \int p(x|\theta)p(\theta)d\theta \tag{31}$$

Пример с нормальным распределением:

- Оцениваем гиперпараметры по данным
- Используем их для построения априора

6.6 Иерархические априоры

Для сложных моделей с несколькими уровнями:

$$\begin{aligned} \theta &\sim p(\theta|\eta) \\ \eta &\sim p(\eta) \end{aligned} \tag{32}$$

Пример:

- Отдельные параметры для каждой группы
- Общий априор для гиперпараметров

6.7 Проблемы выбора априоров

Основные сложности:

- Субъективность выбора
- Влияние на результаты (особенно при малых выборках)
- Вычислительная сложность для нестандартных априоров

6.8 Практические рекомендации

- Всегда проводите анализ чувствительности к выбору априора
- Для научных публикаций используйте референсные априоры
- В прикладных задачах учитывайте экспертные знания
- Документируйте выбор априорных распределений

7 Проверка гипотез с ограничениями на альтернативу

7.1 Постановка задачи

Рассмотрим случай, когда альтернативная гипотеза H_1 имеет ограничения вида:

$$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : |\theta - \theta_0| \geq \delta \tag{33}$$

где $\delta > 0$ - минимальное значимое отклонение.

7.2 Байесовский подход

7.2.1 Модификация априорного распределения

Используем усеченное априорное распределение:

$$p(\theta|H_1) \propto p(\theta) I(|\theta - \theta_0| \geq \delta) \tag{34}$$

где $I(\cdot)$ - индикаторная функция.

7.2.2 Байесовский фактор

Вычисляется как отношение маргинальных правдоподобий:

$$BF_{10} = \frac{\int_{|\theta-\theta_0|\geq\delta} p(x|\theta)p(\theta)d\theta}{p(x|\theta_0)} \tag{35}$$

7.3 Частотный подход

7.3.1 Модифицированный критерий отношения правдоподобия

Статистика теста:

$$\Lambda = \frac{\sup_{|\theta-\theta_0|\geq\delta} p(x|\theta)}{p(x|\theta_0)} \tag{36}$$

7.3.2 Доверительные интервалы

Отвергаем H_0, если доверительный интервал уровня $(1-\alpha)$:

- Не содержит θ_0
- И его границы отстоят от θ_0 не менее чем на δ

7.4 Пример 1: Нормальное распределение

Для $X \sim N(\theta, \sigma^2)$ с известной дисперсией:

$$BF_{10} = \frac{\int_{|\theta|\geq\delta} \exp\left(-\frac{n}{2\sigma^2}(\bar{x}-\theta)^2\right) p(\theta)d\theta}{\exp\left(-\frac{n}{2\sigma^2}\bar{x}^2\right)} \tag{37}$$

Реализация в R:

```
delta <- 0.5
n <- 100; sigma <- 1; x_bar <- 0.6
prior <- function(theta) dnorm(theta, 0, 1)

# Численное интегрирование
integrand <- function(theta) {
  exp(-n/(2*sigma^2)*(x_bar-theta)^2) * prior(theta) *
  (abs(theta) >= delta)
}
BF10 <- integrate(integrand, -Inf, Inf)$value /
        exp(-n/(2*sigma^2)*x_bar^2)
```

7.5 Пример 2: Биномиальное распределение

Для $X \sim Bin(n, p)$, проверка $H_0 : p = p_0$ vs $H_1 : |p - p_0| \geq \delta$:

$$BF_{10} = \frac{\int_{|p-p_0|\geq\delta} p^k(1-p)^{n-k}Beta(p; a, b)dp}{p_0^k(1-p_0)^{n-k}} \tag{38}$$

7.6 Методы численного интегрирования

Для сложных моделей применяют:

- MCMC с ограниченной областью
- Метод Лапласа для аппроксимации интегралов
- Сэмплирование по важности

7.7 Интерпретация результатов

- При $BF_{10} > 1$: данные поддерживают значимое отклонение
- Если $BF_{10} < 1$ но $\hat{\theta} \geq \theta_0 + \delta$:
  - Либо эффект есть, но данные слабые
  - Либо априор слишком консервативен

7.8 Практические рекомендации

- Выбирайте δ на основе предметной области
- Проводите анализ чувствительности к выбору δ
- Для клинических исследований используйте δ, соответствующий минимальному клинически значимому эффекту
- Всегда сообщайте значение δ в публикациях

8 Использование ROPE и его связь с другими методами

8.1 Определение ROPE

Region of Practical Equivalence (ROPE) — это интервал вокруг нулевого значения параметра, внутри которого эффекты считаются практически незначимыми:

$$\text{ROPE} = [\theta_0 - \epsilon, \theta_0 + \epsilon] \tag{39}$$

где ϵ — порог практической значимости.

8.2 Алгоритм использования ROPE

1. Определите ROPE на основе предметных знаний
2. Постройте апостериорное распределение параметра
3. Сравните 95% HDI (Highest Density Interval) с ROPE:
   - Если HDI полностью вне ROPE — значимый эффект
   - Если HDI полностью внутри ROPE — отсутствие эффекта
   - Если HDI пересекает ROPE — неопределённость

8.3 Пример применения

Для нормального распределения с $\theta_0 = 0$, $\epsilon = 0.2$:

```
library(bayestestR)
posterior <- rnorm(1000, mean = 0.15, sd = 0.1)
rope(posterior, range = c(-0.2, 0.2))
```

Результат:

- Вероятность внутри ROPE: 12%
- Вероятность слева от ROPE: 3%
- Вероятность справа от ROPE: 85%

Вывод: значимый положительный эффект.

8.4 Связь с байесовскими факторами

Метод	Преимущества	Недостатки
ROPE	Учитывает практическую значимость	Требует выбора ϵ
Байесовский фактор	Объективное сравнение моделей	Чувствителен к априорам

Соотношение:

$$\text{Если } P(\theta \in \text{ROPE}|X) < 0.05 \text{ и } BF_{10} > 3 \Rightarrow \text{сильная поддержка } H_1 \tag{40}$$

Критерий	ROPE	p-value
Основа	Апостериорное распределение	Статистика теста
Интерпретация	Вероятность нахождения в зоне	Вероятность ошибки
Учёт ϵ	Явный	Неявный через мощность

8.5 Сравнение с частотными методами

Аналог ROPE в частотном подходе — эквивалентность тестов.

8.6 ROPE и анализ мощности

Для планирования исследования:

$$n_{\text{необх}} = \arg\min_n P(\text{HDI} \cap \text{ROPE} = \emptyset) \geq 0.8 \tag{41}$$

Пример расчёта в R:

```
# Для t-теста с delta = 0.3, sd = 1
power.t.test(delta = 0.3, sd = 1, power = 0.8)
```

8.7 ROPE в регрессионных моделях

Для коэффициентов регрессии β:

$$\text{ROPE}_\beta = [-0.1, 0.1] \text{ (стандартизированные коэффициенты)} \tag{42}$$

Пример использования:

```
model <- lm(y ~ x)
posterior <- simulate_bayes_model(model)
rope(posterior$beta_x, c(-0.1, 0.1))
```

8.8 Критика и ограничения

Проблемы метода ROPE:

- Субъективность выбора ϵ
- Зависимость от масштаба данных
- Не учитывает стоимость ошибок

8.9 Практические рекомендации

- Для стандартизированных коэффициентов используйте $\epsilon = 0.1$
- Всегда проводите анализ чувствительности к выбору ROPE
- Сочетайте с визуализацией апостериорных распределений
- Для клинических исследований определяйте ROPE через MCID (Minimal Clinically Important Difference)

9 Проблема расхождения BF и апостериорной вероятности и введение в e-value

9.1 Феномен расхождения

Наблюдается ситуация, когда:

- Байесовский фактор BF_{10} большой (например, > 100)
- Апостериорная вероятность $P(H_1|X)$ остаётся малой (например, < 0.5)

Формально:

$$BF_{10} = \frac{P(X|H_1)}{P(X|H_0)} \gg 1, \quad P(H_1|X) = \frac{BF_{10}P(H_1)}{BF_{10}P(H_1) + P(H_0)} \approx 0 \tag{43}$$

9.2 Причины расхождения

9.2.1 Экстремально малые априорные вероятности

Когда $P(H_1)$ очень мала:

$$P(H_1|X) = \frac{BF_{10}P(H_1)}{BF_{10}P(H_1) + (1 - P(H_1))} \approx BF_{10}P(H_1) \tag{44}$$

Пример:

- $BF_{10} = 1000$
- $P(H_1) = 10^{-6}$
- $P(H_1|X) \approx 0.001$

9.2.2 Парадокс Линдли

Возникает при:

- Очень больших объёмах данных
- Жёстких априорных вероятностях против H_1
- Слабой альтернативной гипотезе

9.3 Методы решения проблемы

9.3.1 Коррекция априоров

$$P_{\text{нов}}(H_1) = \frac{P_{\text{ста}}(H_1)}{P_{\text{ста}}(H_1) + k(1 - P_{\text{ста}}(H_1))} \tag{45}$$

где k - коэффициент ослабления скептицизма.

9.3.2 Использование e-value

E-value - альтернативная байесовская мера:

$$e = \frac{P(X|H_1)}{P(X|H_0) + P(X|H_1)} = \frac{BF_{10}}{1 + BF_{10}} \tag{46}$$

Свойства e-value:

- Всегда в $[0, 1]$
- Менее чувствителен к экстремальным априорам
- Интерпретация аналогична p-value (но с байесовской основой)

9.4 Сравнение мер

Мера	Формула	Диапазон	Интерпретация
BF_{10}	$\frac{P(X\|H_1)}{P(X\|H_0)}$	$(0, \infty)$	Отношение правдоподобий
$P(H_1\|X)$	$\frac{BF_{10}P(H_1)}{1+BF_{10}P(H_1)}$	$[0, 1]$	Полная вероятность
e-value	$\frac{BF_{10}}{1+BF_{10}}$	$[0, 1]$	"Байесовский p-value"

9.5 Практические рекомендации

- При больших BF_{10} и малых $P(H_1|X)$:
  1. Проверьте априорные вероятности гипотез
  2. Рассмотрите e-value как дополнение к анализу
  3. Проведите анализ чувствительности к априорам
- Для публикаций:
  - Всегда сообщайте и BF_{10}, и $P(H_1|X)$
  - Указывайте использованные априорные вероятности
  - Для консервативных выводов используйте e-value

9.6 Пример расчёта

Для $BF_{10} = 100$ при разных априорах:

```
# Традиционный байесовский подход
P_H1 <- c(0.5, 0.1, 0.01, 1e-6)
posterior <- (100*P_H1)/(100*P_H1 + (1-P_H1))

# E-value подход
e_value <- 100/(1+100) \approx 0.9901
```

Результаты:

$P(H_1)$	$P(H_1\|X)$	e-value
0.5	0.9901	0.9901
0.1	0.9174	0.9901
0.01	0.5025	0.9901
10^{-6}	0.0001	0.9901

9.7 Выводы

- Большой BF_{10} не гарантирует высокую $P(H_1|X)$
- E-value предоставляет устойчивую альтернативу
- Все три меры (BF_{10}, $P(H_1|X)$, e-value) дополняют друг друга
- Критически важно анализировать чувствительность к априорам

10 Вероятностная интерпретация e-value

10.1 Определение e-value

E-value (evidence value) определяется как:

$$e := \frac{P(X|H_1)}{P(X|H_0) + P(X|H_1)} = \frac{BF_{10}}{1 + BF_{10}} \tag{47}$$

где BF_{10} – байесовский фактор в пользу H_1.

10.2 Вероятностная интерпретация

E-value допускает две вероятностные интерпретации:

10.2.1 1. Условная вероятность

При равных априорных вероятностях ($P(H_0) = P(H_1) = 0.5$):

$$e = P(H_1|X) \tag{48}$$

Таким образом, в этом частном случае e-value *совпадает* с апостериорной вероятностью.

10.2.2 2. Минимальная апостериорная вероятность

В общем случае e-value представляет *нижнюю границу* для апостериорной вероятности:

$$e \leq P(H_1|X) \quad \text{для любых} \quad P(H_1) \in (0, 1) \tag{49}$$

Это означает, что e-value всегда даёт консервативную оценку поддержки H_1.

Свойство	e-value	p-value
Основа	Байесовская	Частотная
Интерпретация	$P(H_1\|X)$ (при равных априорах)	$P(X$ или более крайние$\|H_0)$
Диапазон	[0,1]	[0,1]
Зависимость от альтернативы	Да	Нет

10.3 Сравнение с p-value

10.4 Пример интерпретации

Для $BF_{10} = 19$:

$$e = \frac{19}{1+19} = 0.95 \tag{50}$$

Интерпретация:

- При $P(H_1) = 0.5$: $P(H_1|X) = 0.95$
- При $P(H_1) = 0.01$: $P(H_1|X) \approx 0.16$
- Но всегда $P(H_1|X) \geq 0.95$ *невозможно* (ограничение снизу 0.95)

10.5 Преимущества e-value

- Более устойчив к экстремальным априорам, чем $P(H_1|X)$
- Сохраняет байесовскую интерпретацию
- Обеспечивает нижнюю границу для апостериорной вероятности
- Легко рассчитывается из байесовского фактора

10.6 Ограничения

- Не является полноценной заменой апостериорной вероятности
- Может быть слишком консервативным
- Требует аккуратной интерпретации в контексте конкретной задачи

11 Применение байесовского подхода к A/B тестам и оценка денежных потерь

Байесовский подход предлагает принципиально иную интерпретацию результатов A/B тестов по сравнению с классической частотной статистикой. Вместо p-значений и доверительных интервалов мы работаем с апостериорными распределениями и прямыми вероятностными утверждениями.

11.1 Базовый алгоритм байесовского A/B тестирования

1. Выбираем априорное распределение для конверсии (обычно Beta распределение)
2. Собираем данные по вариантам A и B
3. Обновляем апостериорные распределения по формуле Байеса
4. Анализируем разницу между распределениями

Формально, если $X_A \sim Beta(\alpha_A, \beta_A)$ и $X_B \sim Beta(\alpha_B, \beta_B)$, то апостериорные распределения после наблюдения данных:

$$X_A|data \sim Beta(\alpha_A + conversions_A, \beta_A + failures_A) \tag{51}$$

$$X_B|data \sim Beta(\alpha_B + conversions_B, \beta_B + failures_B) \tag{52}$$

11.2 Пример расчета на R

```
# Установка априорных параметров (например, Beta(1,1) - равномерное распределение)
alpha_prior <- 1
beta_prior <- 1

# Данные по варианту A: 150 конверсий из 1000 показов
conversions_A <- 150
trials_A <- 1000

# Данные по варианту B: 170 конверсий из 1000 показов
conversions_B <- 170
trials_B <- 1000

# Расчет апостериорных распределений
posterior_A <- rbeta(100000, alpha_prior + conversions_A,
                    beta_prior + trials_A - conversions_A)
posterior_B <- rbeta(100000, alpha_prior + conversions_B,
                    beta_prior + trials_B - conversions_B)

# Вероятность того, что вариант B лучше A
prob_B_better <- mean(posterior_B > posterior_A)
```

11.3 Перевод результатов в денежные потери

Для оценки денежных потерь необходимо:

1. Определить средний доход с одной конверсии (ARPU)
2. Рассчитать ожидаемую потерю дохода при выборе неоптимального варианта

Формула ожидаемой потери:

$$ExpectedLoss = ARPU \times \max(0, E[X_B] - E[X_A]) \times N \tag{53}$$

где N - количество пользователей в будущем периоде.

```
# Продолжение R кода
ARPU <- 500 # Средний доход с конверсии в рублях
future_users <- 10000 # Ожидаемое количество пользователей в следующем месяце

# Расчет ожидаемых конверсий
expected_A <- mean(posterior_A)
expected_B <- mean(posterior_B)

# Ожидаемая потеря при выборе варианта A вместо B
expected_loss <- ifelse(expected_B > expected_A,
                       ARPU * (expected_B - expected_A) * future_users,
                       0)
```

11.4 Интерпретация результатов

- Вероятность превосходства B над A: `prob_B_better`
- Ожидаемая потеря при выборе A: `expected_loss` рублей
- Рекомендация: если ожидаемая потеря превышает стоимость переключения на вариант B, следует выбрать B

11.5 Преимущества байесовского подхода

Преимущество	**Описание**
Прямая интерпретация	Вероятность того, что вариант B лучше A
Учет априорной информации	Можно использовать данные предыдущих тестов
Оценка потерь	Прямой расчет ожидаемых финансовых последствий
Гибкость	Возможность остановить тест при достижении определенной уверенности

12 Jeffreys Prior (Общий случай и примеры)

12.1 Определение

Распределение Джеффри (Jeffreys prior) — это **неинформативное априорное распределение**, предложенное Гарольдом Джеффрисом (Harold Jeffreys) в рамках байесовского подхода. Оно строится на основе **информации Фишера** и обладает свойством **инвариантности** относительно параметризации модели.

Для параметра θ априор Джеффри задаётся в виде:

$$\pi(\theta) \propto \sqrt{\det I(\theta)},$$

где $I(\theta)$ — **матрица информации Фишера**:

$$I(\theta)_{ij} = -\mathbb{E}\left[\frac{\partial^2 \log p(x|\theta)}{\partial\theta_i \partial\theta_j}\right].$$

В одномерном случае:

$$\pi(\theta) \propto \sqrt{I(\theta)}, \quad I(\theta) = -\mathbb{E}\left[\frac{\partial^2 \log p(x|\theta)}{\partial\theta^2}\right].$$

12.2 Свойства

- **Инвариантность**: Если $\phi = g(\theta)$ — перепараметризация, то априорное распределение для ϕ согласовано:
$$\pi(\phi) = \pi(\theta)\left|\frac{d\theta}{d\phi}\right|.$$
- **Неинформативность**: Минимизирует влияние априорных предположений, особенно полезно при отсутствии экспертных знаний.
- **Автоматическое определение**: Зависит только от вида правдоподобия $p(x|\theta)$.

12.3 Примеры

12.3.1 Бернулли-распределение ($X \sim$ **Bern**(p))

Правдоподобие:

$$p(x|p) = p^x(1-p)^{1-x}, \quad x \in \{0, 1\}.$$

Логарифмическое правдоподобие:

$$\log p(x|p) = x \log p + (1-x)\log(1-p).$$

Вторая производная:

$$\frac{\partial^2 \log p(x|p)}{\partial p^2} = -\frac{x}{p^2} - \frac{1-x}{(1-p)^2}.$$

Информация Фишера:

$$I(p) = -\mathbb{E}\left[\frac{\partial^2 \log p(x|p)}{\partial p^2}\right] = \frac{1}{p(1-p)}.$$

Априор Джеффри:

$$\pi(p) \propto \sqrt{I(p)} = p^{-1/2}(1-p)^{-1/2} \implies p \sim \mathrm{Beta}(1/2, 1/2).$$

12.3.2 Нормальное распределение ($X \sim \mathcal{N}(\mu, \sigma^2)$)

- **Случай μ (известно σ)**:

$$I(\mu) = \frac{1}{\sigma^2} \implies \pi(\mu) \propto 1 \quad \text{(равномерный априор)}.$$

- **Случай σ (известно μ)**:

$$I(\sigma) = \frac{2}{\sigma^2} \implies \pi(\sigma) \propto \frac{1}{\sigma}.$$

- **Оба параметра неизвестны**:

$$\pi(\mu, \sigma) \propto \frac{1}{\sigma^2}.$$

12.3.3 Экспоненциальное распределение ($X \sim \mathbf{Exp}(\lambda)$)

$$p(x|\lambda) = \lambda e^{-\lambda x}, \quad x \geq 0.$$

Информация Фишера:

$$I(\lambda) = \frac{1}{\lambda^2} \implies \pi(\lambda) \propto \frac{1}{\lambda}.$$

12.4 Критика

- Может быть **несобственным** (не интегрируемым), например, $\pi(\mu) \propto 1$ для нормального распределения.
- В многомерных случаях иногда приводит к **парадоксальным** результатам.
- Альтернативы: reference priors, maximum entropy priors.

12.5 Связь с частотными методами

Априор Джеффри тесно связан с классическими частотными (frequentist) методами статистики через понятие **информации Фишера** $I(\theta)$. Эта связь проявляется в нескольких аспектах:

12.5.1 Эффективность оценок

В частотной статистике информация Фишера определяет нижнюю границу дисперсии несмещённых оценок (**неравенство Крамера–Рао**):

$$\mathrm{Var}(\hat{\theta}) \geq \frac{1}{I(\theta)}.$$

Априор Джеффри $\pi(\theta) \propto \sqrt{I(\theta)}$ автоматически "взвешивает"параметр θ в соответствии с его **предполагаемой точностью** в данных. Это согласуется с байесовским принципом учёта информации: области, где данные дают больше информации о θ (больше $I(\theta)$), получают более "плотное"априорное распределение.

12.5.2 Асимптотика байесовских доверительных интервалов

При больших объёмах выборки ($n \to \infty$) апостериорное распределение для θ с априором Джеффри становится близким к нормальному:

$$\theta \mid X \approx \mathcal{N}\left(\hat{\theta}_{\text{ML}}, \frac{1}{I(\hat{\theta}_{\text{ML}})}\right),$$

где $\hat{\theta}_{\text{ML}}$ — оценка максимального правдоподобия. Таким образом, байесовские **доверительные интервалы** с априором Джеффри асимптотически совпадают с частотными **интервалами на основе MLE**:

$$\hat{\theta}_{\text{ML}} \pm z_{\alpha/2} \cdot \frac{1}{\sqrt{I(\hat{\theta}_{\text{ML}})}}.$$

12.5.3 Инвариантность и параметризация

В частотной статистике MLE инвариантен относительно перепараметризации: если $\phi = g(\theta)$, то $\hat{\phi}_{\text{ML}} = g(\hat{\theta}_{\text{ML}})$. Априор Джеффри сохраняет это свойство:

$$\pi(\phi) = \pi(\theta)\left|\frac{d\theta}{d\phi}\right| \propto \sqrt{I(\phi)}.$$

Например, для модели $\text{Bern}(p)$:

- При $\theta = p$: $\pi(p) \propto p^{-1/2}(1-p)^{-1/2}$.
- При $\phi = \log \frac{p}{1-p}$ (логит): $\pi(\phi) \propto 1$ (равномерный априор).

12.5.4 Сравнение с другими неинформативными априорами

- **Равномерный априор** $\pi(\theta) \propto 1$ не учитывает $I(\theta)$ и может приводить к неинвариантным выводам.
- **Reference priors** (Berger–Bernardo) обобщают априор Джеффри для многомерных случаев.

13 Сопряжённость нормального распределения для среднего нормальной выборки с известной дисперсией

13.1 Постановка задачи

Пусть имеется выборка данных:

$$X_1, X_2, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2),$$

где σ^2 известно, а μ – неизвестный параметр.

Выберем априорное распределение:

$$\mu \sim \mathcal{N}(\mu_0, \tau_0^2),$$

где μ_0 и τ_0^2 – гиперпараметры.

13.2 Доказательство

Функция правдоподобия:

$$P(\mathbf{X} \mid \mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(X_i-\mu)^2}{2\sigma^2}\right) \propto \exp\left(-\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{2\sigma^2}\right)$$

Априорное распределение:

$$P(\mu) \propto \exp\left(-\frac{(\mu-\mu_0)^2}{2\tau_0^2}\right)$$

Апостериорное распределение пропорционально произведению правдоподобия и априорного:

$$P(\mu \mid \mathbf{X}) \propto P(\mathbf{X} \mid \mu)P(\mu) \propto \exp\left(-\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{2\sigma^2} - \frac{(\mu-\mu_0)^2}{2\tau_0^2}\right)$$

Выделим полный квадрат относительно μ в показателе экспоненты. Показатель экспоненты:

$$-\frac{1}{2}\left[\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} + \frac{(\mu-\mu_0)^2}{\tau_0^2}\right]$$

Раскроем квадраты:

$$\sum_{i=1}^{n}(X_i-\mu)^2 = n\mu^2 - 2n\bar{X}\mu + \sum_{i=1}^{n} X_i^2,$$

где $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Таким образом, показатель экспоненты принимает вид:

$$-\frac{1}{2}\left[\left(\frac{n}{\sigma^2} + \frac{1}{\tau_0^2}\right)\mu^2 - 2\left(\frac{n\bar{X}}{\sigma^2} + \frac{\mu_0}{\tau_0^2}\right)\mu + \mathrm{const}\right]$$

Это квадратичная форма относительно μ, что соответствует нормальному распределению. Обозначим:

$$\tau_n^{-2} = \frac{n}{\sigma^2} + \frac{1}{\tau_0^2}, \quad \mu_n = \tau_n^2\left(\frac{n\bar{X}}{\sigma^2} + \frac{\mu_0}{\tau_0^2}\right)$$

Тогда апостериорное распределение:

$$P(\mu \mid \mathbf{X}) \propto \exp\left(-\frac{(\mu-\mu_n)^2}{2\tau_n^2}\right)$$

Следовательно:

$$\mu \mid \mathbf{X} \sim \mathcal{N}(\mu_n, \tau_n^2)$$

Что и доказывает сопряжённость нормального априорного распределения.

14 Сопряжённость бета-распределения для распределения Бернулли

14.1 Постановка задачи

Пусть имеется выборка данных:

$$X_1, X_2, \ldots, X_n \sim \text{Bernoulli}(\theta),$$

где θ – неизвестный параметр.

Выберем априорное распределение:

$$\theta \sim \text{Beta}(\alpha, \beta),$$

где $\alpha, \beta > 0$ – гиперпараметры.

14.2 Доказательство

Функция правдоподобия:

$$P(\mathbf{X} \mid \theta) = \prod_{i=1}^{n} \theta^{X_i}(1-\theta)^{1-X_i} = \theta^{\sum X_i}(1-\theta)^{n-\sum X_i}$$

Априорное распределение:

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

Апостериорное распределение пропорционально произведению правдоподобия и априорного:

$$P(\theta \mid \mathbf{X}) \propto P(\mathbf{X} \mid \theta)P(\theta) \propto \theta^{\sum X_i}(1-\theta)^{n-\sum X_i} \cdot \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

Объединяя степени, получаем:

$$P(\theta \mid \mathbf{X}) \propto \theta^{\alpha+\sum X_i-1}(1-\theta)^{\beta+n-\sum X_i-1}$$

Это пропорционально плотности бета-распределения:

$$\theta \mid \mathbf{X} \sim \text{Beta}(\alpha + \sum_{i=1}^{n} X_i, \beta + n - \sum_{i=1}^{n} X_i)$$

Что и доказывает сопряжённость бета-априорного распределения для модели Бернулли.

15 Сравнение байесовского и частотного подходов к построению доверительных интервалов

15.1 Философские основания

- **Частотный подход** рассматривает параметр как фиксированную величину, а доверительный интервал – как случайную область
- **Байесовский подход** трактует параметр как случайную величину с апостериорным распределением

15.2 Формальные определения

15.2.1 Частотный доверительный интервал

Для параметра θ строится интервал $[L(X), U(X)]$ такой, что:

$$P(L(X) \leq \theta \leq U(X)|\theta) = 1 - \alpha \tag{54}$$

где вероятность вычисляется по распределению данных X.

15.2.2 Байесовский кредитный интервал

Для апостериорного распределения $p(\theta|X)$ выбирается область C:

$$P(\theta \in C|X) = \int_C p(\theta|X)d\theta = 1 - \alpha \tag{55}$$

15.3 Вычислительные методы

Метод	Частотный подход	Байесовский подход
Основа	Выборочное распределение статистики	Апостериорное распределе
Нормальное приближение	$\hat{\theta} \pm z_{\alpha/2}SE$	Использование квантилей а
Бутстреп	Многократная перевыборка данных	Не требуется
MCMC	Неприменимо	Основной метод для сложн

15.4 Пример: интервал для доли

Для данных $X \sim Bin(n, p)$:

15.4.1 Частотный подход (Вальд)

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \tag{56}$$

15.4.2 Байесовский подход (Beta-априор)

С априором $Beta(a, b)$ апостериор:

$$p|X \sim Beta(a + x, b + n - x) \tag{57}$$

Интервал строится по квантилям этого распределения.

15.5 Ключевые различия

15.6 Рекомендации по выбору априоров

При отсутствии экспертной информации:

- Для пропорций: $Beta(1, 1)$ (равномерный) или $Beta(0.5, 0.5)$ (Джеффриса)
- Для нормального среднего: $N(0, 10\sigma^2)$
- Для дисперсии: $InvGamma(0.001, 0.001)$

Характеристика	Сравнение подходов
Интерпретация	• Байесовский: Вероятность для параметра • Частотный: Доля покрытий в повторных выборках
Учет априорной информации	• Байесовский: Явный через априорное распределение • Частотный: Отсутствует
Вычислительная сложность	• Байесовский: Требует интегрирования/MCMC • Частотный: Часто аналитические решения
Поведение при малых выборках	• Байесовский: Более стабилен (с хорошим априором) • Частотный: Может давать бессмысленные интервалы

Таблица 5: Сравнение байесовского и частотного подходов к доверительным интервалам

15.7 Пример в R

```
# Частотный подход (95% ДИ для доли)
prop.test(x=15, n=100)$conf.int

# Байесовский подход
bayes_prop_ci <- function(x, n, alpha = 0.05, a_prior = 1, b_prior = 1) {
  a_post <- a_prior + x
  b_post <- b_prior + n - x

  ci <- qbeta(c(alpha/2, 1 - alpha/2), a_post, b_post)
  list(
    estimate = a_post / (a_post + b_post),
    ci = ci,
    a_post = a_post,
    b_post = b_post
  )
}

# Пример использования
result <- bayes_prop_ci(x=n.p, n=n, a_prior = 1, b_prior = 1)
```

```
print(result$ci)
```

15.8 Преимущества и недостатки

- **Байесовские интервалы**:
  - \+ Естественная интерпретация
  - \+ Учет априорной информации
  - \- Зависимость от априоров
  - \- Вычислительная сложность
- **Частотные интервалы**:
  - \+ Объективность
  - \+ Стандартизированные методы
  - \- Контр-интуитивная интерпретация
  - \- Проблемы при малых выборках

15.9 Выводы

- Байесовские интервалы дают прямую вероятностную интерпретацию параметра
- При отсутствии априорной информации можно использовать слабоинформативные априоры
- Для сложных моделей байесовский подход часто более гибок
- В больших выборках оба подхода дают схожие результаты

16 Совпадение доверительного и байесовского интервалов

Существует классический и важный пример, когда доверительный интервал (ДИ) и байесовский достоверный интервал (Credible Interval) полностью совпадают по форме, но коренным образом различаются по интерпретации. Это случай оценки среднего значения нормального распределения с известной дисперсией.

16.1 Постановка задачи

Пусть имеется выборка данных:

$$X_1, X_2, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2),$$

где:

- μ – неизвестное среднее значение (параметр, который нужно оценить),
- σ^2 – известная дисперсия.

16.2 Частотный подход (Доверительный Интервал)

- **Оценка:** Выборочное среднее $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ является точечной оценкой для μ.
- **Стандартная ошибка:** $SE = \frac{\sigma}{\sqrt{n}}$.
- **95% Доверительный Интервал (ДИ)** строится по формуле:

$$\left[\bar{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}, \quad \bar{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right]$$

Интерпретация: "Если бы мы многократно повторяли эксперимент и каждый раз строили такой интервал, то в 95% случаев эти интервалы накрыли бы истинное значение μ."

16.3 Байесовский подход (Достоверный Интервал)

- **Априорное распределение:** Выбирается неинформативное априорное распределение, отражающее полную неопределённость:

$$P(\mu) \propto 1.$$

- **Апостериорное распределение:** Можно показать, что апостериорное распределение для μ будет нормальным:

$$\begin{aligned} \text{Среднее апостериорного:} \quad & \bar{X}, \\ \text{Дисперсия апостериорного:} \quad & \frac{\sigma^2}{n}. \end{aligned}$$

- **95% Достоверный Интервал (Credible Interval)** для μ:

$$\left[\bar{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}, \quad \bar{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right]$$

Интерпретация: "Исходя из наших данных, мы на 95% уверены, что истинное значение μ лежит в этом интервале."

16.4 Сравнительная таблица

Аспект	**Частотный подход**	**Байесовский подход**
Формула	$\bar{X} \pm 1.96 \cdot \frac{\sigma}{\sqrt{n}}$	**Та же самая:** $\bar{X} \pm 1.96 \cdot \frac{\sigma}{\sqrt{n}}$
Численный результат	Интервал $[a, b]$	**Тот же интервал** $[a, b]$
Интерпретация	О долгосрочных свойствах **метода**	О степени уверенности в **конкретном значении параметра** μ
Объект вероятности	**Процедура** (доверительный интервал)	**Параметр** μ

16.5 Вывод

В данном примере численные значения интервалов полностью совпадают. Вся разница заключается исключительно в их интерпретации:

- Частотный подход отказывается говорить о вероятности нахождения параметра μ в рассчитанном интервале, описывая лишь свойства процедуры его построения.
- Байесовский подход даёт прямую вероятностную интерпретацию для конкретного рассчитанного интервала.

Это наглядно демонстрирует, что спор между подходами часто является не спором о вычислениях, а спором о значении и интерпретации этих вычислений.

17 Априорные распределения для математического ожидания

17.1 Случай известной дисперсии

При известной дисперсии σ^2 рассматриваем модель:

$$X_i \sim N(\mu, \sigma^2), \quad i = 1, \dots, n \tag{58}$$

17.2 Выбор априоров

- **Равномерный априор** (неинформативный):

$$p(\mu) \propto 1 \tag{59}$$

- **Нормальный априор** (сопряжённый):

$$\mu \sim N(\mu_0, \tau^2) \tag{60}$$

где μ_0 – априорное среднее, τ^2 – априорная дисперсия

- **Априор Джеффриса**:

$$p(\mu) \propto 1 \tag{61}$$

(совпадает с равномерным для этого случая)

17.3 Апостериорное распределение

Для нормального априора апостериорное распределение:

$$\mu | X \sim N\left(\frac{\frac{n\bar{X}}{\sigma^2} + \frac{\mu_0}{\tau^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}, \left(\frac{n}{\sigma^2} + \frac{1}{\tau^2}\right)^{-1}\right) \tag{62}$$

Для равномерного априора ($\tau^2 \to \infty$):

$$\mu | X \sim N\left(\bar{X}, \frac{\sigma^2}{n}\right) \tag{63}$$

18 Сравнение подходов

18.1 Частотный доверительный интервал

$$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \tag{64}$$

18.2 Байесовский кредитный интервал

$$\mu_{\text{post}} \pm z_{\alpha/2} \sqrt{\text{Var}(\mu|X)} \tag{65}$$

18.3 Код на R для сравнения

```
compare_mean_intervals <- function(x, sigma, alpha = 0.05,
                                   mu_prior = 0, tau_prior = Inf) {
  n <- length(x)
  x_bar <- mean(x)

  # Частотный подход
  z <- qnorm(1 - alpha/2)
  freq_ci <- c(
    lower = x_bar - z * sigma / sqrt(n),
    upper = x_bar + z * sigma / sqrt(n)
  )

  # Байесовский подход
  if (is.infinite(tau_prior)) {
    # Равномерный априор
    post_mean <- x_bar
    post_var <- sigma^2 / n
  } else {
    # Нормальный априор
    post_precision <- n/sigma^2 + 1/tau_prior^2
    post_mean <- (n*x_bar/sigma^2 + mu_prior/tau_prior^2) / post_precision
    post_var <- 1 / post_precision
  }

  bayes_ci <- c(
    lower = qnorm(alpha/2, post_mean, sqrt(post_var)),
    upper = qnorm(1 - alpha/2, post_mean, sqrt(post_var))
  )

  # Результаты
  data.frame(
    Method = c("Частотный", "Байесовский"),
    Estimate = c(x_bar, post_mean),
    Lower = c(freq_ci["lower"], bayes_ci["lower"]),
    Upper = c(freq_ci["upper"], bayes_ci["upper"]),
    Width = c(diff(freq_ci), diff(bayes_ci))
```

```
  )
}

# Пример использования
set.seed(42)
data <- rnorm(30, mean = 5, sd = 2)
sigma <- 2  # Известная дисперсия

# 1. Равномерный априор
compare_mean_intervals(data, sigma)

# 2. Информативный априор (mu_prior = 0, tau = 1)
compare_mean_intervals(data, sigma, mu_prior = 0, tau_prior = 1)

# 3. Визуализация влияния априора
library(ggplot2)
tau_values <- c(0.5, 1, 2, 5, 10, Inf)
results <- lapply(tau_values, function(tau) {
  res <- compare_mean_intervals(data, sigma, mu_prior = 0, tau_prior = tau)
  res$TauPrior <- ifelse(is.infinite(tau), "Inf (равномерный)", as.character(tau))
  res
})
results_df <- do.call(rbind, results)

ggplot(results_df[results_df$Method == "Байесовский",],
       aes(x = TauPrior, y = Estimate)) +
  geom_point() +
  geom_errorbar(aes(ymin = Lower, ymax = Upper), width = 0.2) +
  labs(title = "Влияние априорной дисперсии на байесовский интервал",
       x = "Априорное SD (tau)", y = "Оценка среднего") +
  geom_hline(yintercept = mean(data), linetype = "dashed",
             color = "red", size = 1) +
  annotate("text", x = 3, y = mean(data)+0.2,
           label = "Выборочное среднее", color = "red")
```

19 Выводы

- При равномерном априоре байесовский интервал совпадает с частотным
- Информативные априоры "сдвигают"оценку в сторону априорного среднего
- Чем меньше τ, тем сильнее влияние априора
- При больших выборках влияние априора уменьшается

20 Критический анализ проблемы априорных распределений в байесовском подходе

Ваша критика затрагивает самый уязвимый элемент байесовской методологии: зависимость выводов от субъективного выбора априорного распределения, особенно в условиях малых данных.

20.1 Объективная проблема: Смещение, вносимое априором

Фундаментальная проблема заключается в следующем: байесовский вывод производится по формуле:

$$P(\theta \mid \text{Данные}) \propto P(\text{Данные} \mid \theta) \cdot P(\theta)$$

где $P(\theta)$ — априорное распределение. Если априорное распределение выбрано неудачно, оно становится источником систематической ошибки, которая **не исчезает** с ростом количества данных, а лишь асимптотически подавляется. В случае малых выборок это смещение может доминировать в окончательном выводе.

Это не теоретическая абстракция, а практическая проблема: два исследователя, работающие с одними и теми же данными, но придерживающиеся разных априорных убеждений, могут прийти к статистически значимо различным выводам.

20.2 Некорректность утверждений о преимуществе при малых данных

Вы совершенно правы, критикуя фразу о «практичных результатах при малых данных». Это утверждение верно лишь с огромной оговоркой: байесовский подход даёт **чёткий и интерпретируемый** результат при малых данных, но этот результат **может быть сильно смещён**, если априор выбран poorly. Таким образом, «уверенность» в результате является в этом случае в большей степени отражением первоначальных убеждений исследователя, нежели объективной информации, извлечённой из данных.

Чем меньше данных, тем **большее влияние** на финальный вывод оказывает произвол в выборе априорного распределения. Следовательно, утверждение о практической полезности подхода в условиях малых данных без указания на эту фундаментальную проблему является некорректным.

20.3 Сравнение с частотным подходом: Два разных ответа на один вопрос

Вы правы, что данные одни и те же, и оба подхода оперируют с одной и той же информацией. Разница заключается в том, **какой вопрос они задают** и **какую форму ответа выдают**.

- **Частотный подход** отвечает на вопрос: «Каковы были бы свойства моего алгоритма оценки, если бы эксперимент повторялся многократно?». Ответ — доверительный интервал — это характеристика **метода**.

- **Байесовский подход** отвечает на вопрос: «Как мне следует обновить свои убеждения о параметре в свете новых данных?». Ответ — апостериорное распределение — это обновлённое состояние **знания** (или незнания) исследователя.

Прямое сравнение «точности» здесь некорректно, так как эти подходы производprincipally разные продукты. Один производит алгоритм с гарантированными долгосрочными свойствами, другой — сиюминутную количественную оценку субъективной уверенности.

20.4 Объективные минусы байесовского подхода

1. **Субъективизм**. Выводы зависят от выбора априора, который часто произволен. Анализ чувствительности лишь констатирует эту зависимость, но не устраняет её.

2. **Вычислительная сложность**. Для сколь-либо сложных моделей получение апостериорного распределения требует применения методов Монте-Карло (MCMC), что computatio expensive и introduces свои источники ошибок (сходимость, автокорреляция).

3. **Иллюзия точности**. Красивые графики апостериорных распределений могут создавать ложное впечатление точности и объективности, маскируя то, что эта «точность» могла быть заложена в априор изначально.

20.5 Заключение: Не преимущество, а компромисс

Таким образом, байесовский подход не является «более точным» или «более практичным» по умолчанию. Он представляет собой **методологический компромисс**:

В обмен на возможность инкорпорировать предварительные знания (когда они действительно есть) и получать прямые вероятностные утверждения о параметрах, исследователь:

1. Вносит в анализ элемент субъективизма.

2. Принимает риск получения смещённых оценок при неудачном выборе априора.

3. Берет на себя burden доказательства того, что его выводы робастны к выбору априорного распределения.

Сила байесовского подхода проявляется не тогда, когда априор «предполагается», а тогда, когда он осмыслен и обоснован (например, прошлыми исследованиями или механизмом генерации данных). Во всех остальных случаях его использование является не строгим преимуществом, а сознательным выбором в пользу одной из парадигм статистического вывода со всеми её inherent limitations.