БАЙЕСОВСКИЙ ВЫВОД ДЛЯ ГАУССИАНА

Сергей Николенко

СПбГУ — Санкт-Петербург

6 ноября 2026 г.





Random facts:

- 6 ноября 1528 г. потерпевший кораблекрушение Кабеса де Вака высадился у берегов Техаса; возвращение к испанцам заняло восемь лет
- 6 ноября 1860 г. Авраам Линкольн выиграл президентские выборы в США
- 6 ноября 1888 г. Бенджамин Гаррисон победил на выборах президента США, хотя Гровер Кливленд получил больше голосов избирателей
- 6 ноября 1935 г. впервые взлетел прототип Hawker Hurricane — одного из основных британских истребителей Битвы за Британию
- 6 ноября 1947 г. на NBC впервые вышла телевизионная программа Meet the Press
- 6 ноября 1986 г. Алекс Фергюсон возглавил «Манчестер Юнайтед»; он оставался главным тренером до 2013 года

Байесовский вывод для гауссиана

Нормальное распределение: фиксируем σ

- На самом деле всё это — байесовский вывод для нормального распределения:

$$p(x_1, \dots, x_n \mid \mu, \sigma^2) \propto \frac{1}{\sigma^n} \exp\left(-\frac{1}{2\sigma^2} \sum_i (x_i - \mu)^2\right).$$

- Хотим: найти сопряжённое априорное распределение, подсчитать правдоподобие, решить задачу предсказания.
- Для начала зафиксируем σ^2 и будем в качестве параметра рассматривать только μ.

Нормальное распределение: фиксируем σ

- Сопряжённое априорное распределение для μ при фиксированном σ^2 тоже нормальное и выглядит как

$$p(\mu \mid \mu_0, \sigma_0^2) \propto \frac{1}{\sigma_0} \exp\left(-\frac{1}{2\sigma_0^2}(\mu - \mu_0)^2\right).$$

- Обычно выбирают $\mu_0 = 0$, $\sigma_0^2 \to \infty$ (порой буквально).
- Давайте рассмотрим сначала случай ровно одного наблюдения x и найдём $p(\mu \mid x)$.

Нормальное распределение: фиксируем σ

- При нашем априорном распределении у μ и x совместное нормальное распределение:

$$x = \mu + \sigma\epsilon, \quad \mu = \mu_0 + \sigma_0\delta, \quad \epsilon, \delta \sim \mathcal{N}(0, 1).$$

Упражнение. Пусть (z_1, z_2) – случайные величины с совместным нормальным распределением. Докажите, что случайная величина $z_1 \mid z_2$ распределена нормально с параметрами

$$E(z_1 \mid z_2) = E(z_1) + \frac{Cov(z_1, z_2)}{Var(z_2)}\left(z_2 - E(z_2)\right),$$

$$Var(z_1 \mid z_2) = Var(z_1) - \frac{Cov^2(z_1, z_2)}{Var(z_2)}$$

$(Var(x) = E\left[(x - Ex)^2\right], Cov(x, y) = E\left[(x - Ex)(y - Ey)\right])$.

Нормальное распределение: фиксируем σ

- В нашем случае:

$$
\begin{aligned}
x = \mu + \sigma\epsilon, \quad \mu = \mu_0 + \sigma_0\delta, \quad \epsilon, \delta \sim \mathcal{N}(0,1),\\
E(x) = \mu_0,\\
Var(x) = E(Var(x \mid \mu)) + Var(E(x \mid \mu)) = \sigma^2 + \sigma_0^2,\\
Cov(x, \mu) = E\left[(x - \mu_0)(\mu - \mu_0)\right] = \sigma_0^2.
\end{aligned}
$$

- Применив упражнение, получаем:

$$
\begin{aligned}
E(\mu \mid x) = \mu_0 + \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2}(x - \mu_0) = \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2}x + \frac{\sigma^2}{\sigma_0^2 + \sigma^2}\mu_0,\\
Var(\mu \mid x) = \frac{\sigma^2\sigma_0^2}{\sigma_0^2 + \sigma^2} = \frac{1}{\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2}}.
\end{aligned}
$$

Нормальное распределение: фиксируем σ

- Итого:

$$p(\mu \mid x) \sim \mathcal{N}\left(\frac{\sigma_0^2}{\sigma_0^2 + \sigma^2} x + \frac{\sigma^2}{\sigma_0^2 + \sigma^2}\mu_0, \left(\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2}\right)^{-1}\right).$$

- Опять же, сложные вычисления можно забыть и пользоваться этими формулами.
- Замечание: часто используют $\tau = \frac{1}{\sigma^2}$ как параметр нормального распределения (precision). Тогда

$$\tau_{\mu|x} = \tau_\mu + \tau.$$

Нормальное распределение: фиксируем σ

- А что, если данных больше, $x_1, \dots, x_n$?
- Тогда можно повторить всё то же самое, а можно заметить, что набор данных описывается своим средним.

Упражнение. Докажите, что если $p(x_i \mid \mu) \sim \mathcal{N}(\mu, \sigma^2)$ и x_i независимы, то $p(\bar{x} \mid \mu) \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$.

Нормальное распределение: фиксируем σ

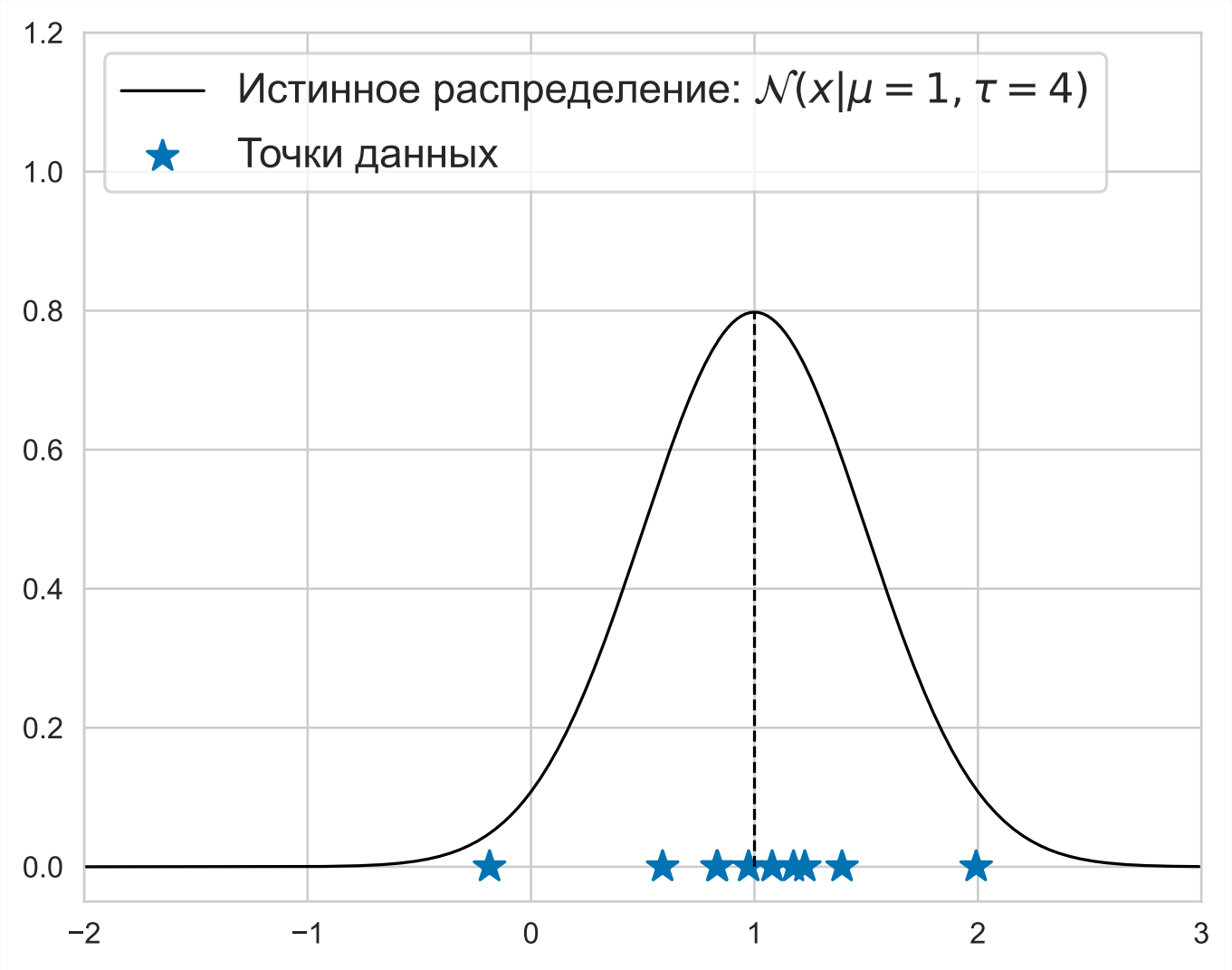
- Для апостериорной вероятности будет

$$p(\mu \mid x_1, \dots, x_n) \propto p(x_1, \dots, x_n \mid \mu) p(\mu) \propto p(\bar{x} \mid \mu) p(\mu) \propto p(\mu \mid \bar{x}).$$

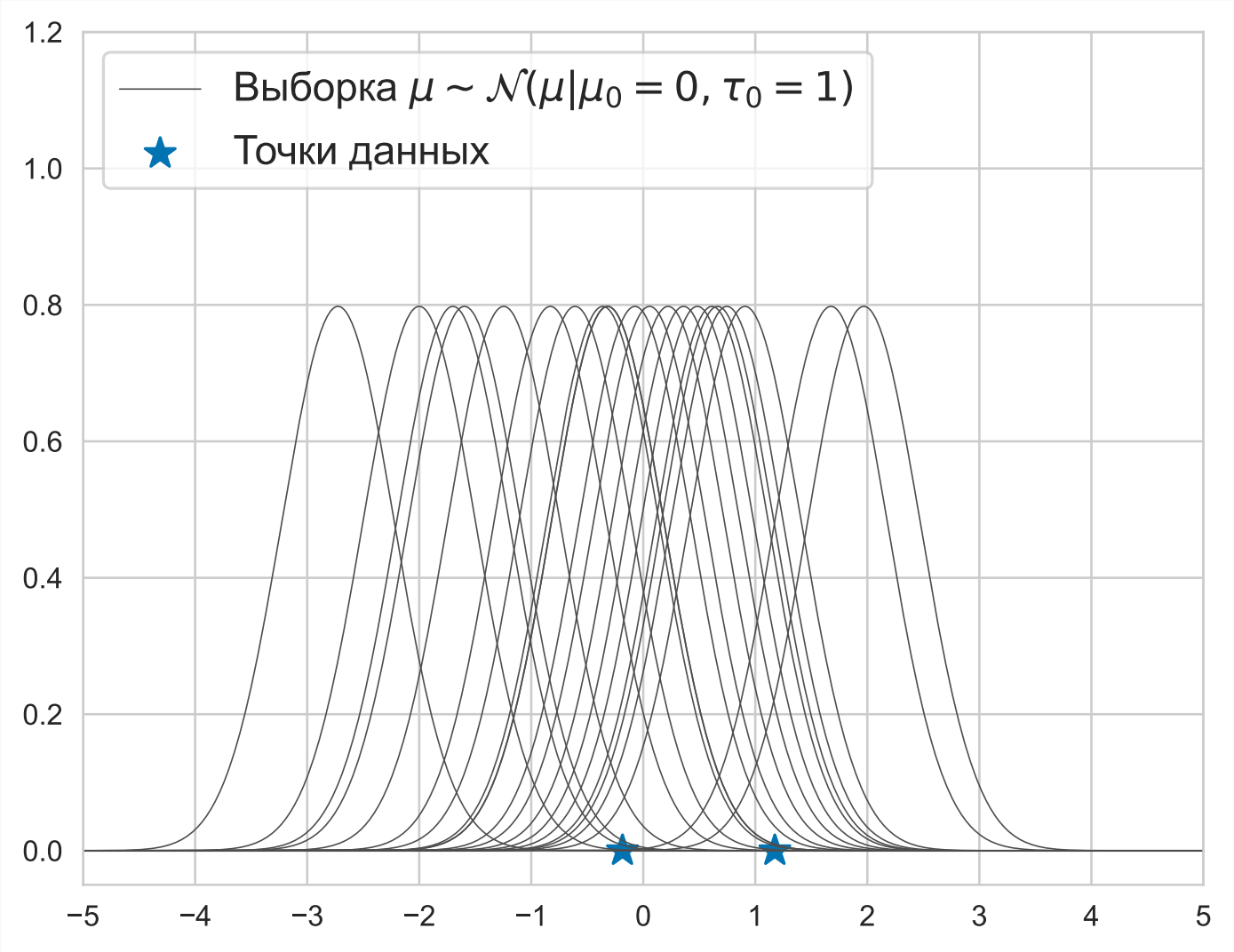
- Подставляя в наш предыдущий результат, получим:

$$p(\mu \mid x_1, \dots, x_n) \sim \mathcal{N}\left(\frac{\sigma_0^2}{\sigma_0^2 + \frac{\sigma^2}{n}} \bar{x} + \frac{\sigma^2}{n\sigma_0^2 + \sigma^2} \mu_0, \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)^{-1}\right).$$

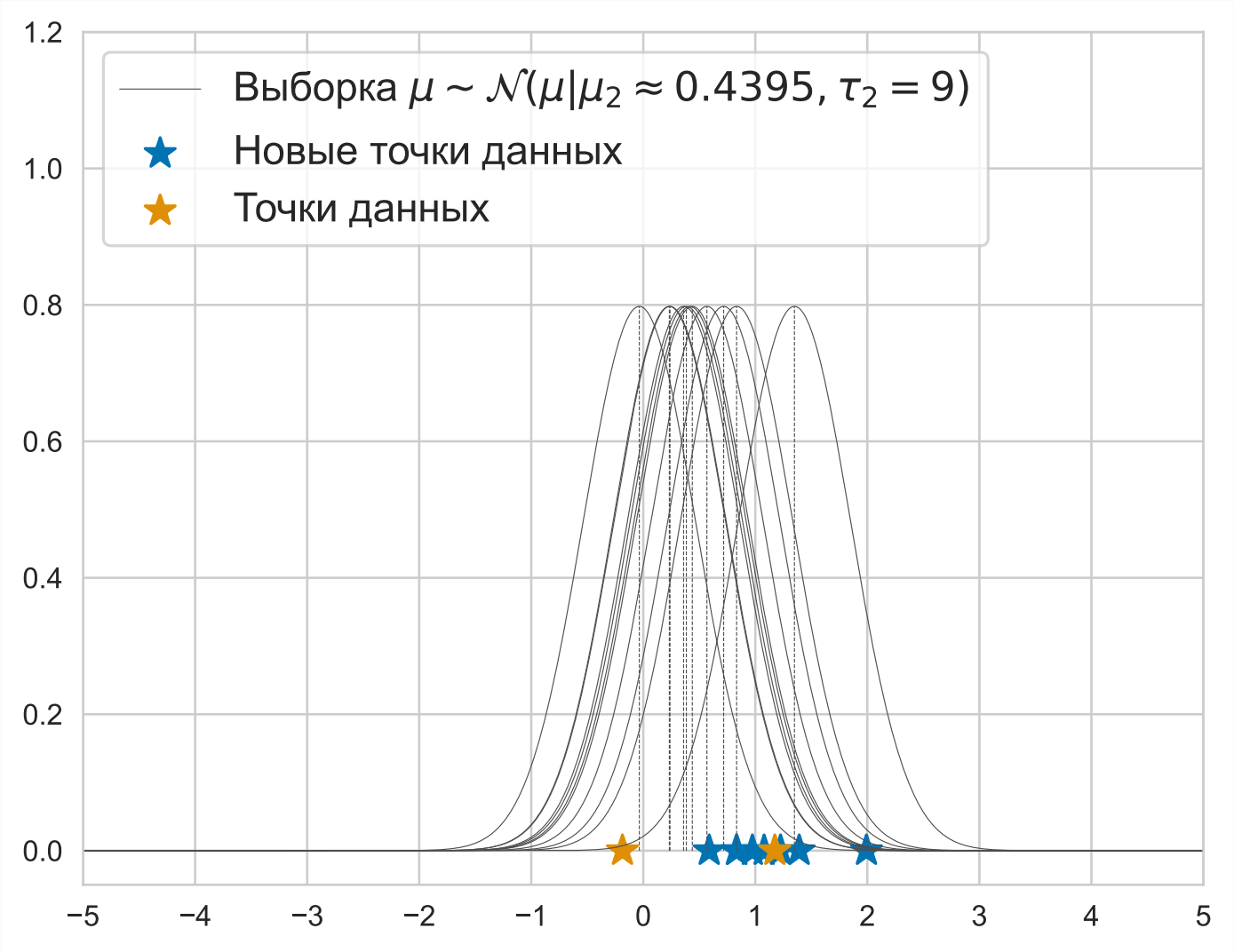
Нормальное распределение: данные



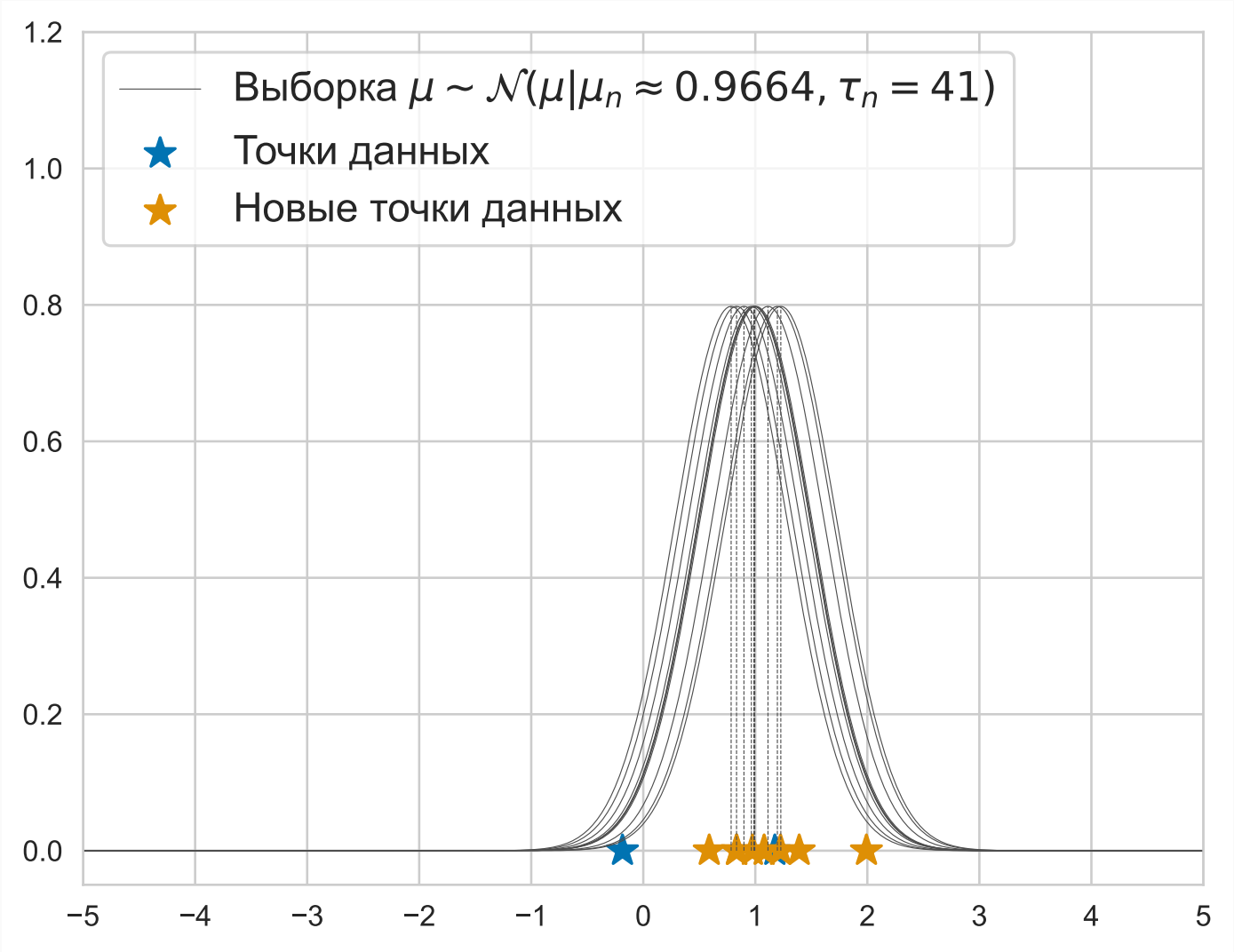
Неизвестное среднее: априорное распределение



Неизвестное среднее: два наблюдения



Неизвестное среднее: вся выборка



Нормальное распределение: фиксируем μ

- Если зафиксировать μ и менять σ^2, то сопряжённым априорным распределением будет обратное гамма-распределение:

$$p(\sigma^2 \mid \alpha, \beta) \propto IG(\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} (\sigma^2)^{-\alpha-1} \exp\left(\frac{-\beta}{\sigma^2}\right).$$

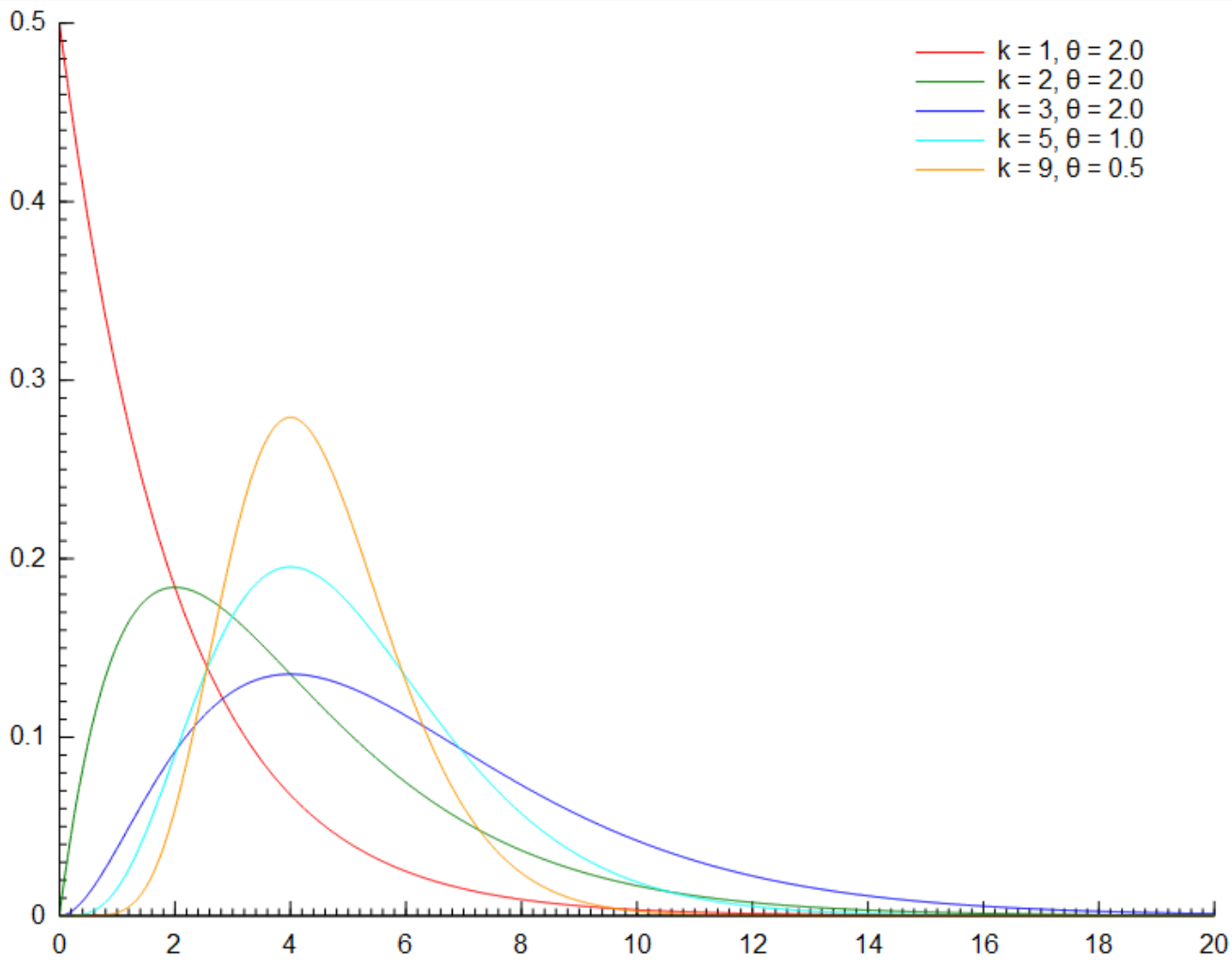
- Тогда в апостериорном распределении будет

$$p(\sigma^2 \mid x_1, \dots, x_n, \alpha, \beta) \propto IG\left(\alpha + \frac{n}{2}, \beta + \frac{1}{2}\sum (x_i - \mu)^2\right).$$

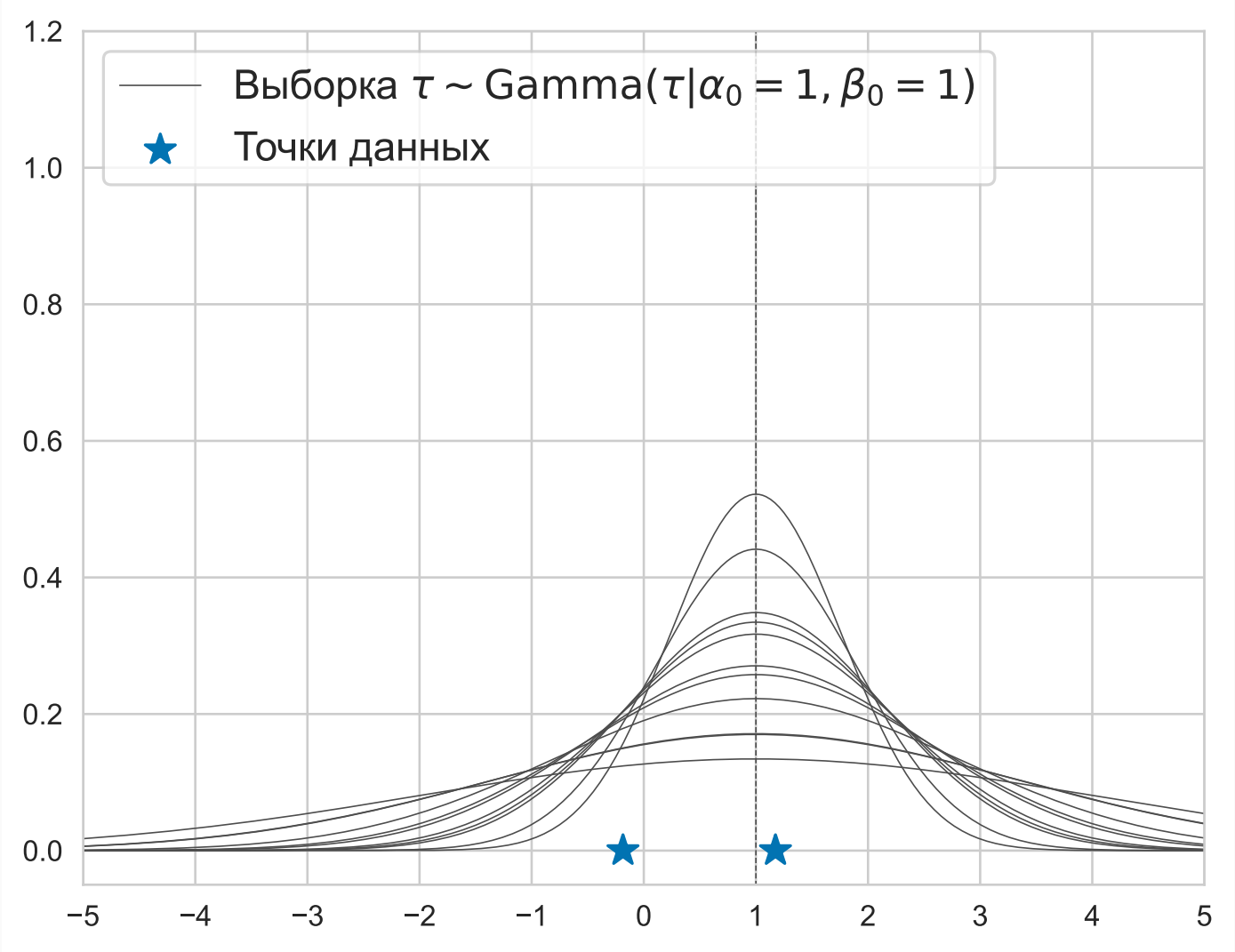
- А в терминах $\tau = \frac{1}{\sigma^2}$ будет обычное гамма-распределение:

$$p(\tau \mid x_1, \dots, x_n, \alpha, \beta) \propto \text{Gamma}\left(\alpha + \frac{n}{2}, \beta + \frac{1}{2}\sum (x_i - \mu)^2\right).$$

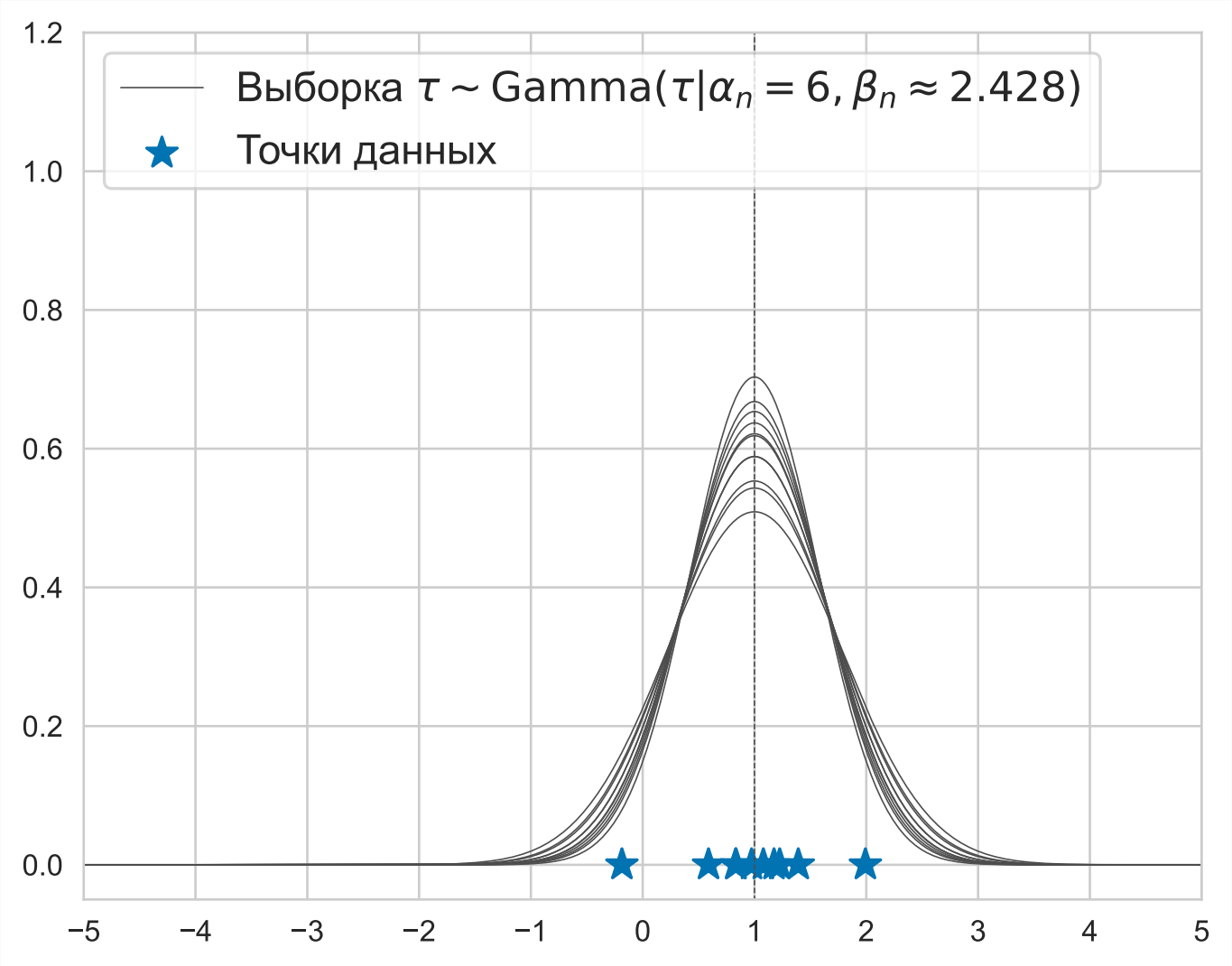
Гамма--распределение



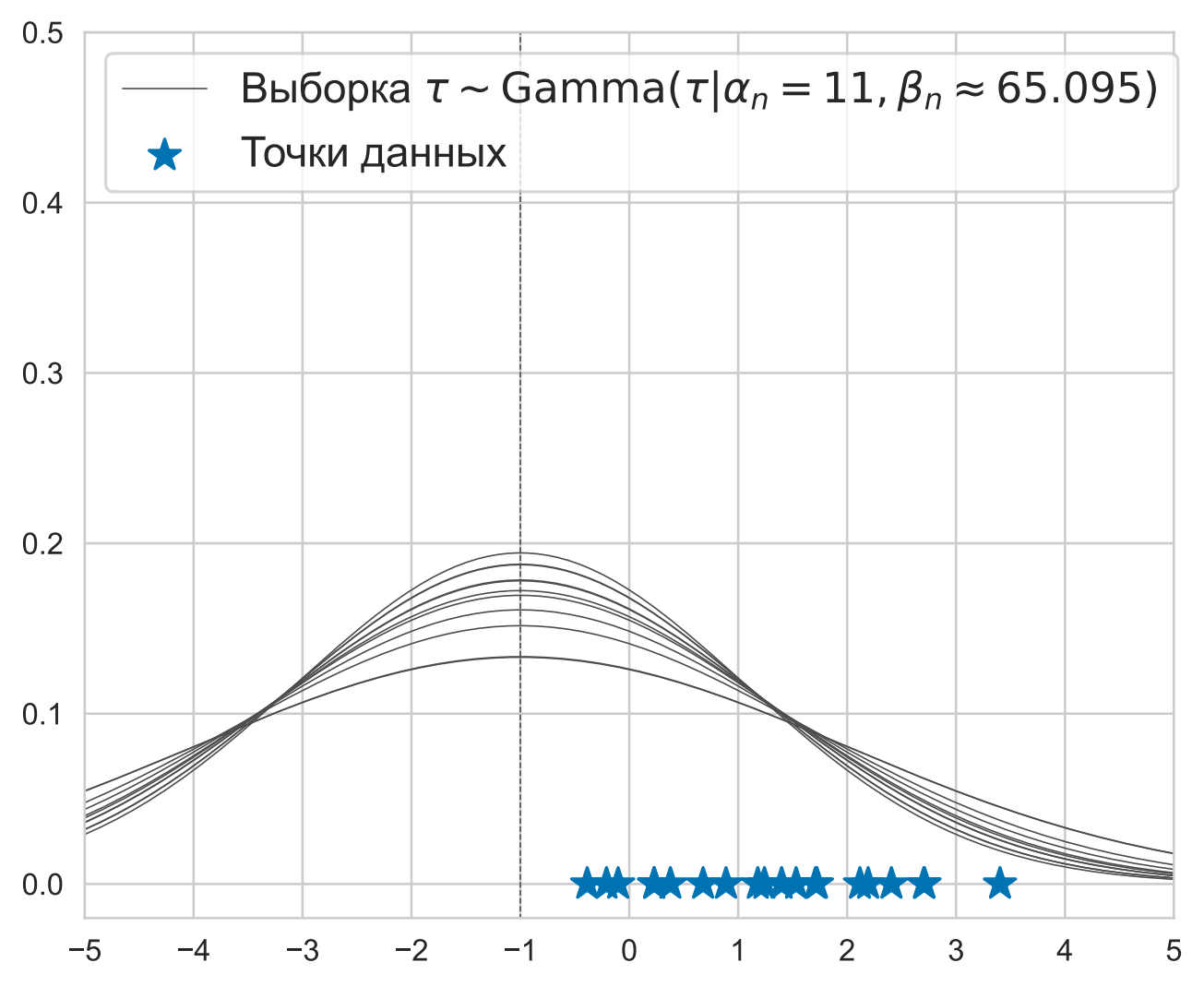
Неизвестная точность: априорное распределение



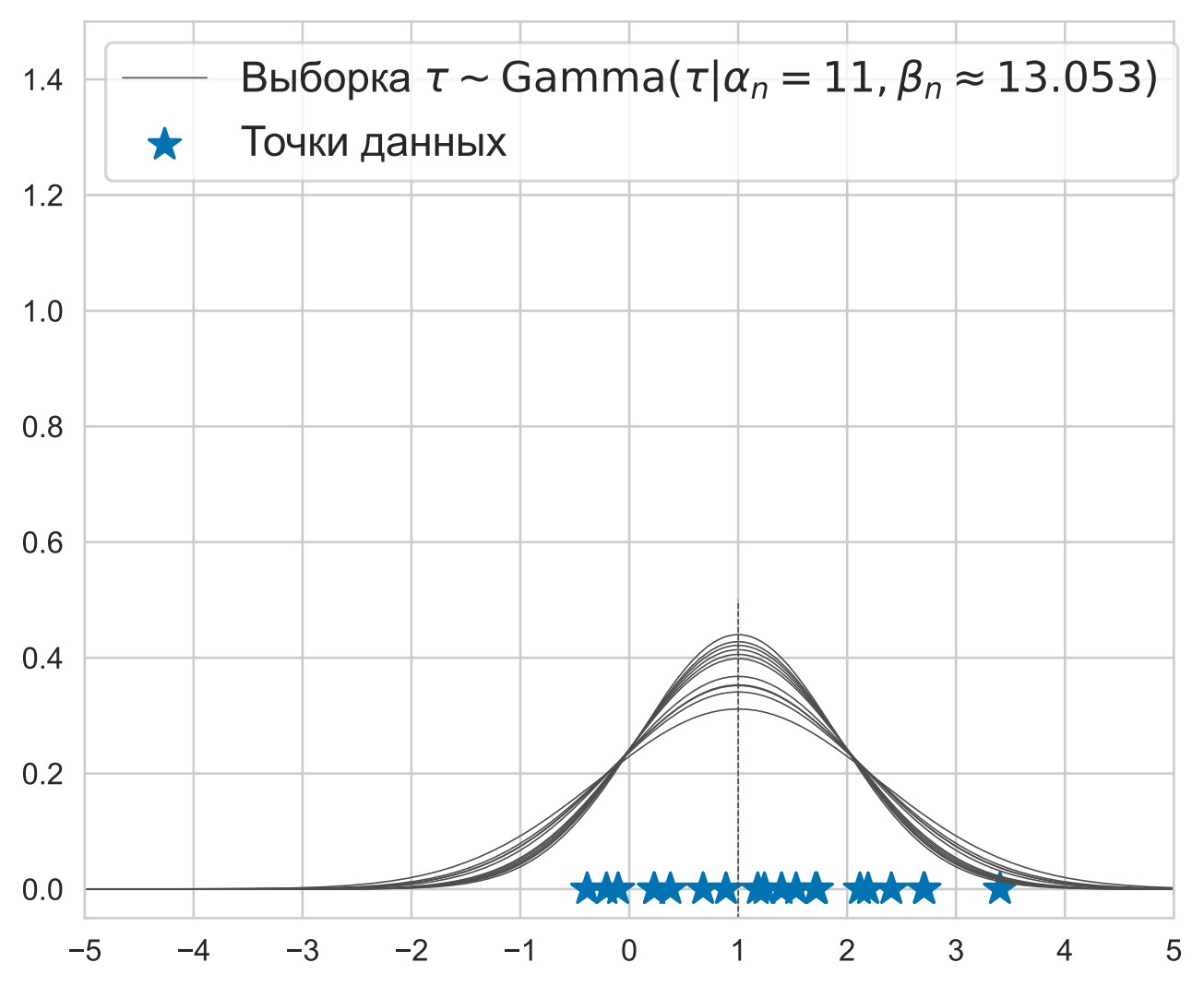
Неизвестная точность: апостериорное распределение



Неизвестная точность: среднее зафиксировано неверно



Неизвестная точность: среднее зафиксировано верно



Когда и μ, и σ^2 меняются

- Что делать, когда и μ, и σ^2 меняются?
- Можно было бы предположить, что μ и σ^2 независимы; тогда просто априорное распределение будет

$$p(\mu, \sigma \mid \mu_0, \sigma_0, \alpha, \beta) \propto \mathcal{N}(\mu_0, \sigma_0^2) \cdot IG(\alpha, \beta).$$

- К сожалению, это распределение не будет сопряжённым к нормальному. Почему?

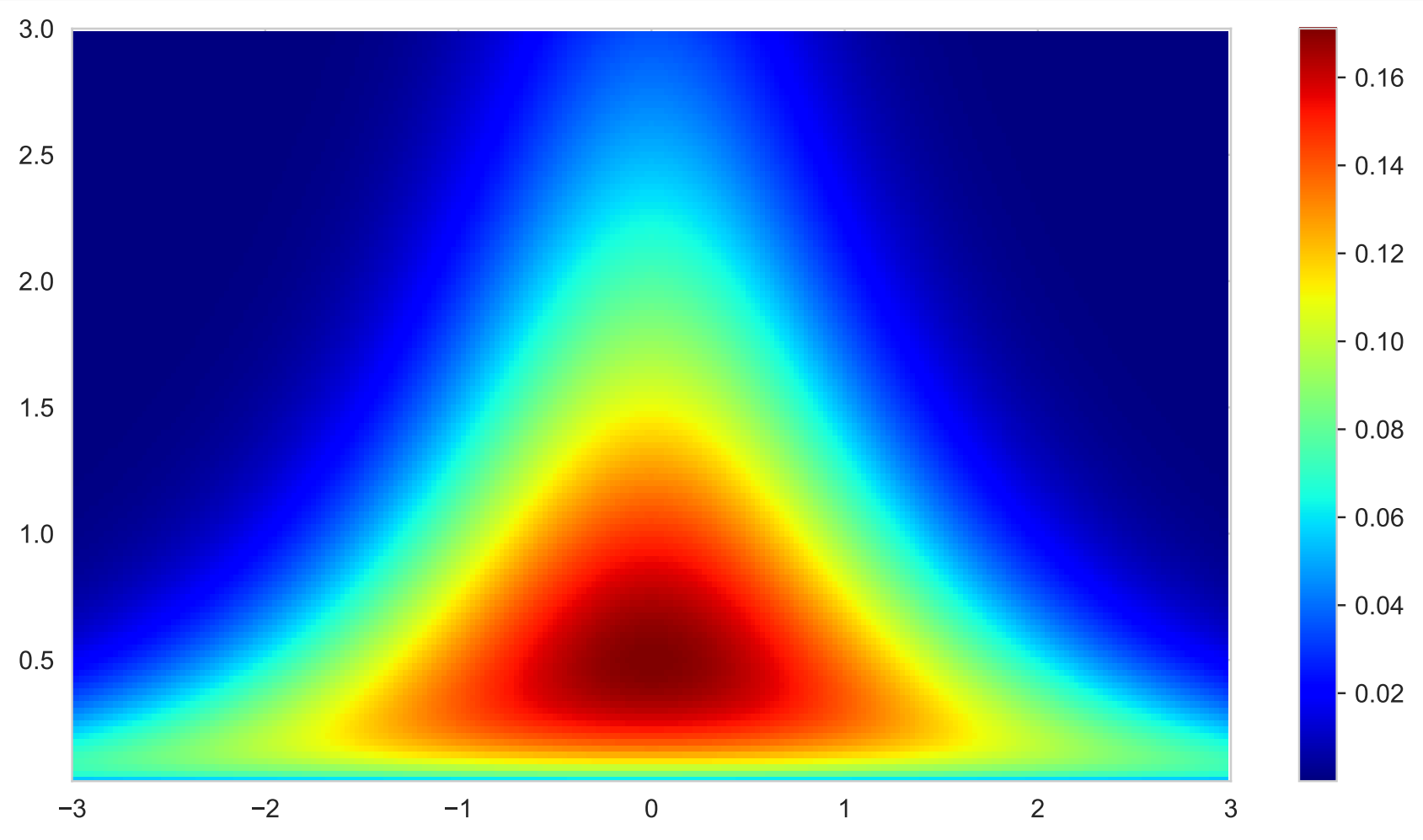
Когда и μ, и σ^2 меняются

- Что делать, когда и μ, и σ^2 меняются?
- Можно было бы предположить, что μ и σ^2 независимы; тогда просто априорное распределение будет

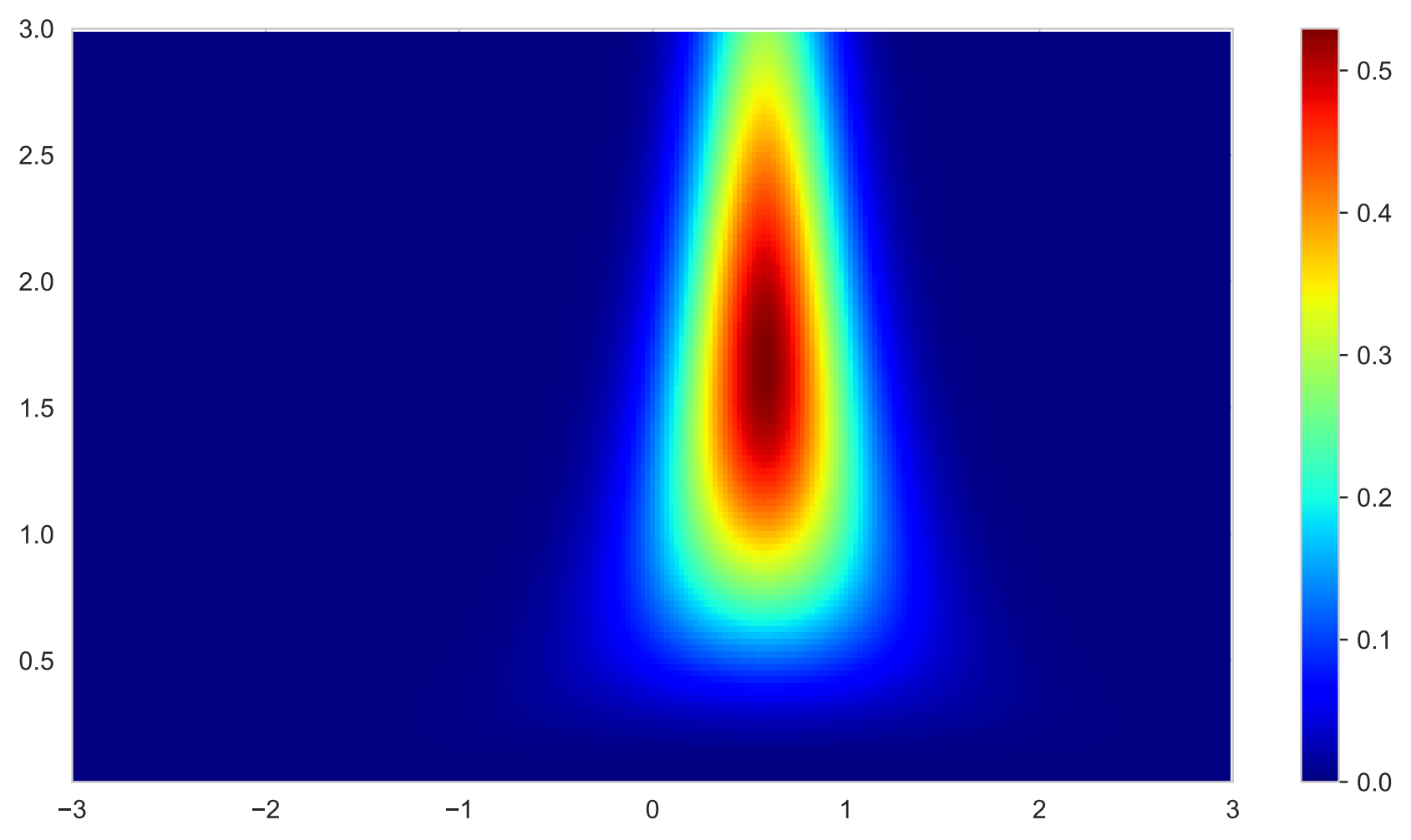
$$p(\mu, \sigma \mid \mu_0, \sigma_0, \alpha, \beta) \propto \mathcal{N}(\mu_0, \sigma_0^2) \cdot IG(\alpha, \beta).$$

- К сожалению, это распределение не будет сопряжённым к нормальному. Почему?
- Потому что μ и σ^2 зависимы. :) Новая точка x вводит зависимость между ними.
- В результате получается распределение Стьюдента.

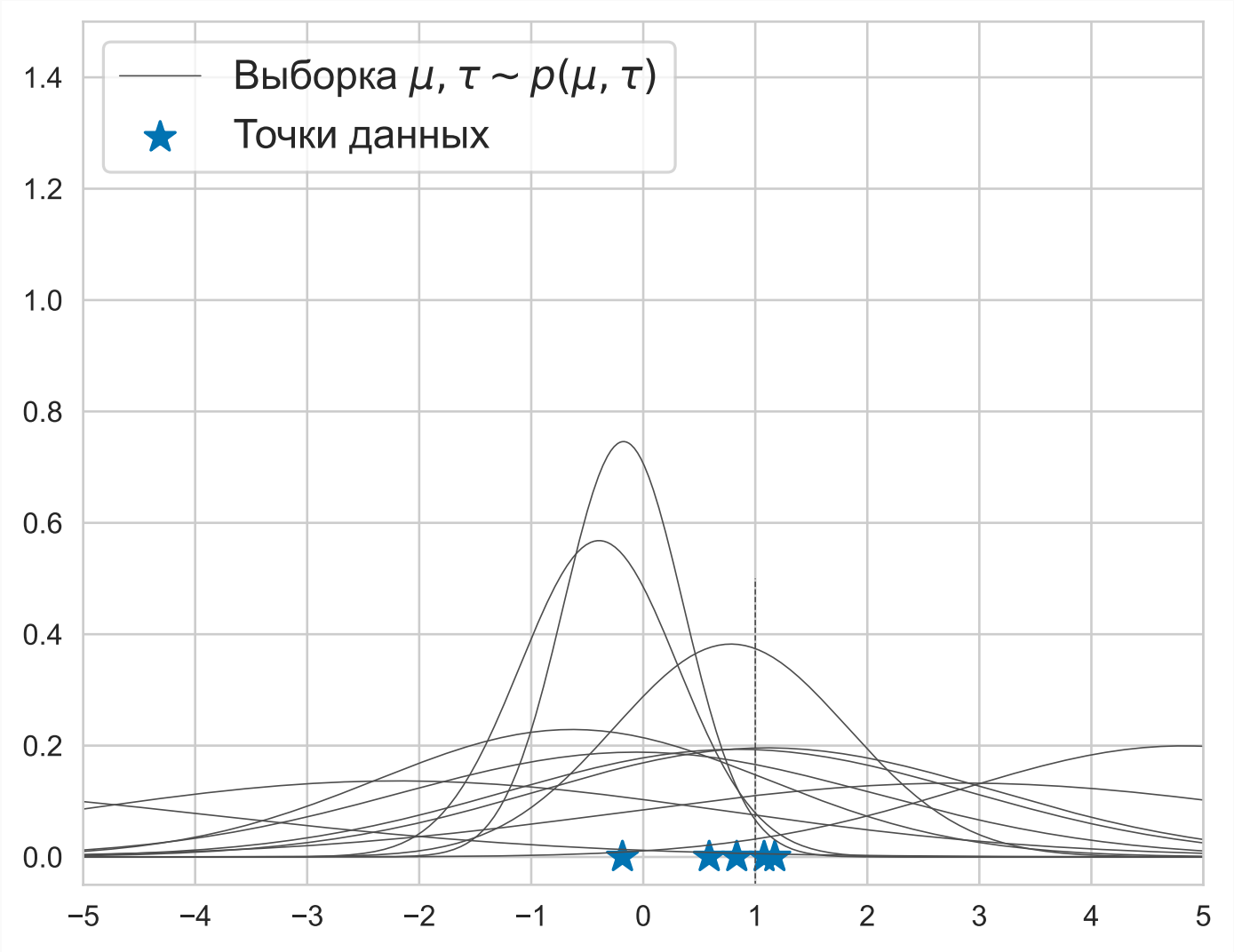
Совместное априорное распределение



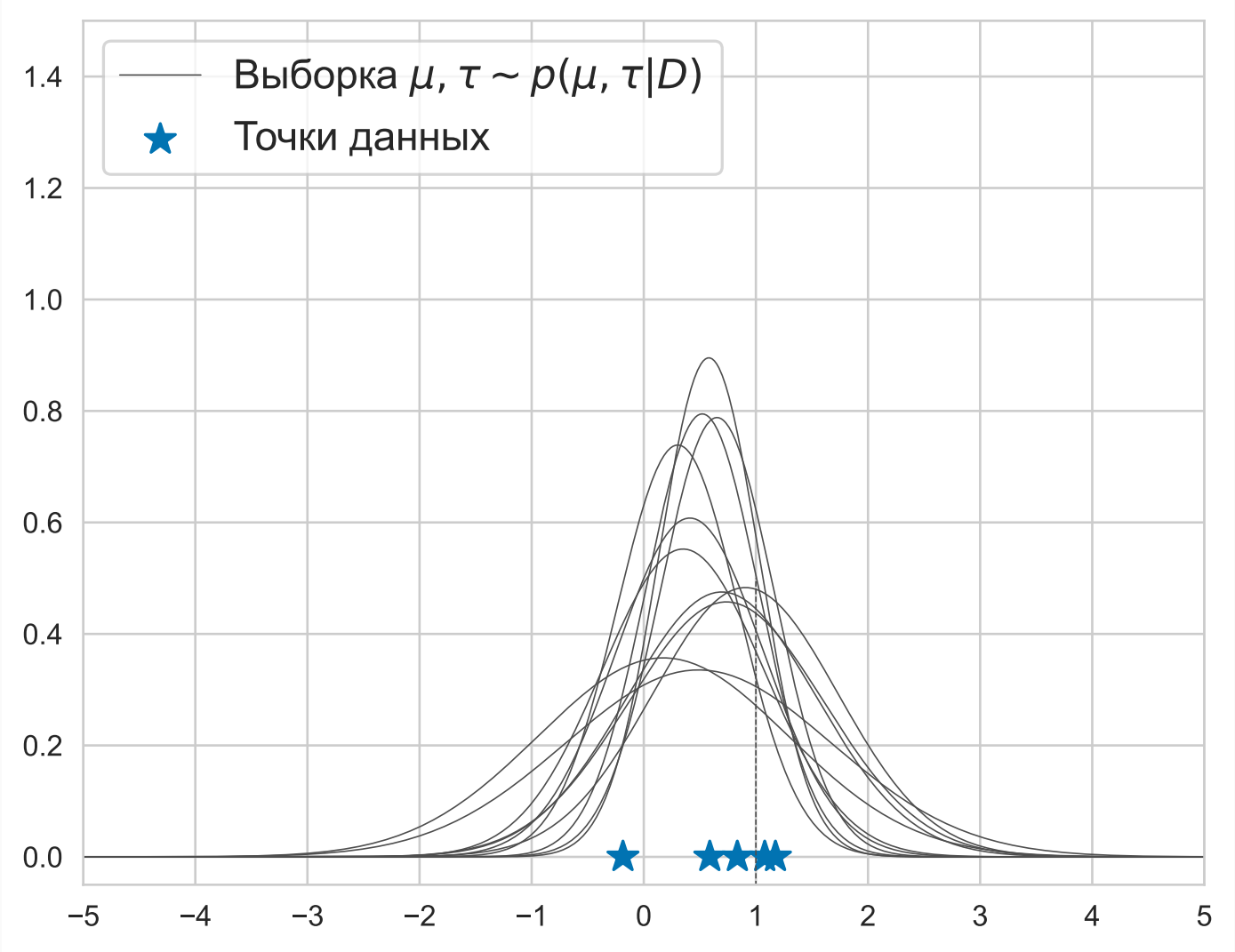
Совместное апостериорное распределение



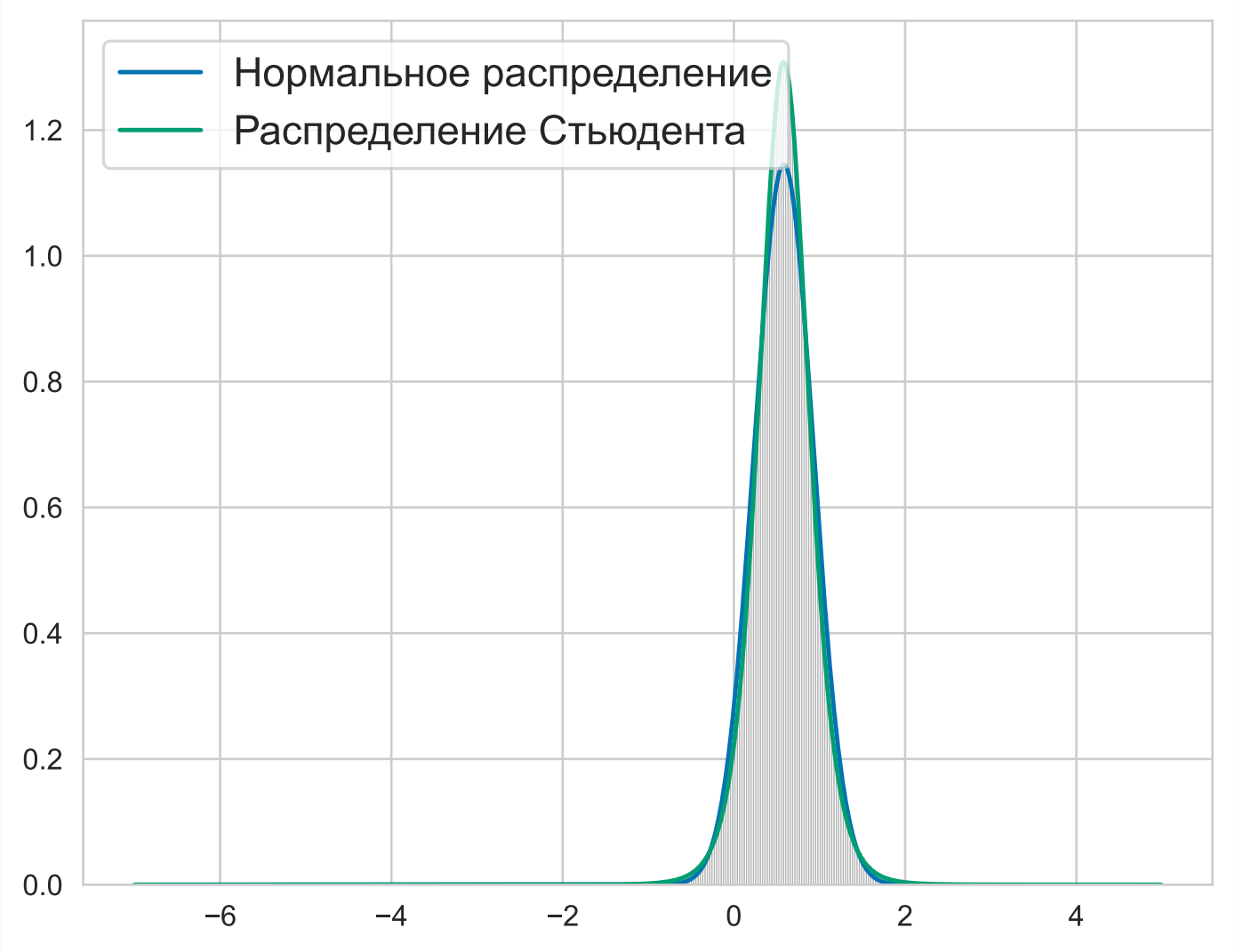
Выборка из совместного априорного распределения



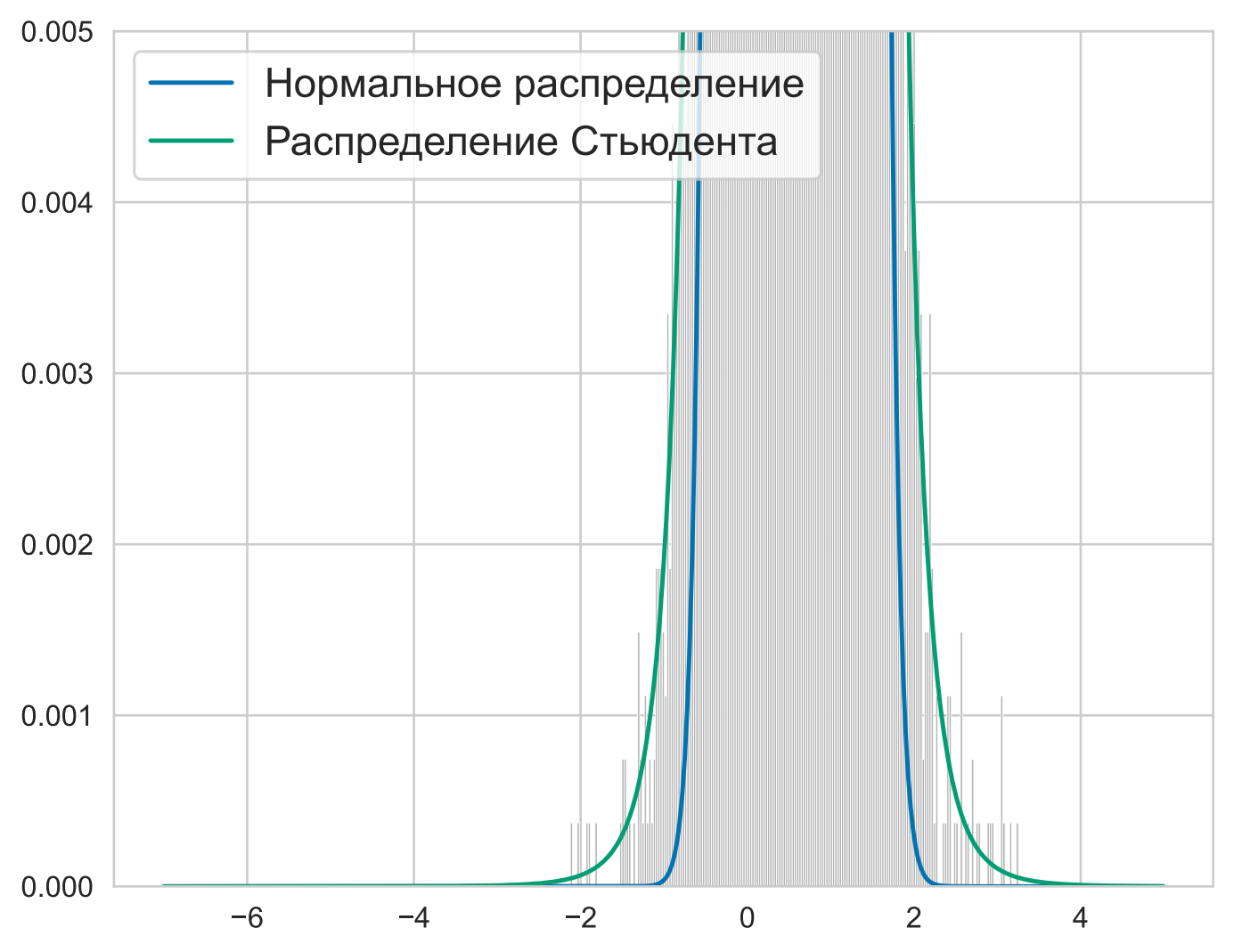
Выборка из совместного апостериорного распределения



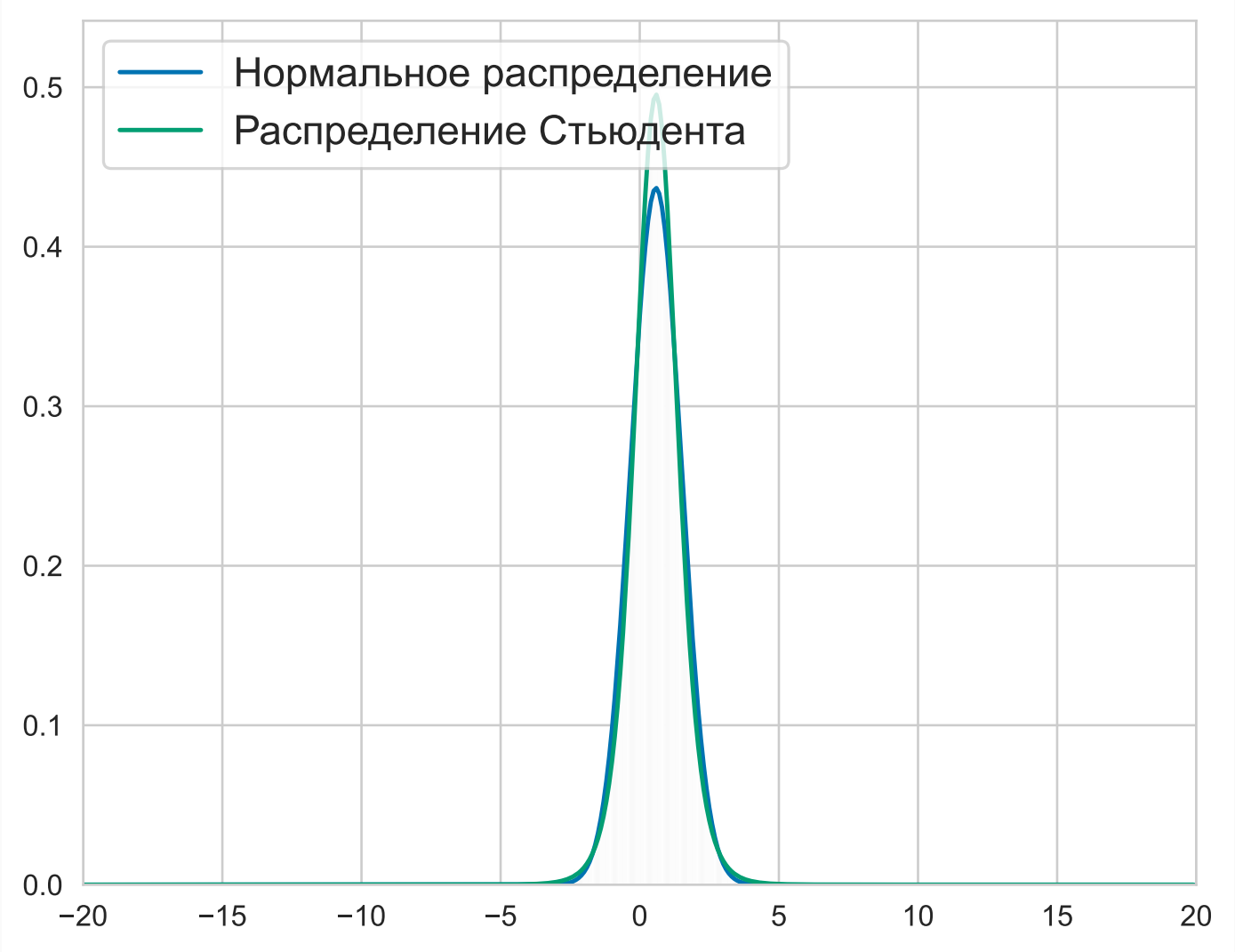
Распределение среднего: нормальное и Стьюдента



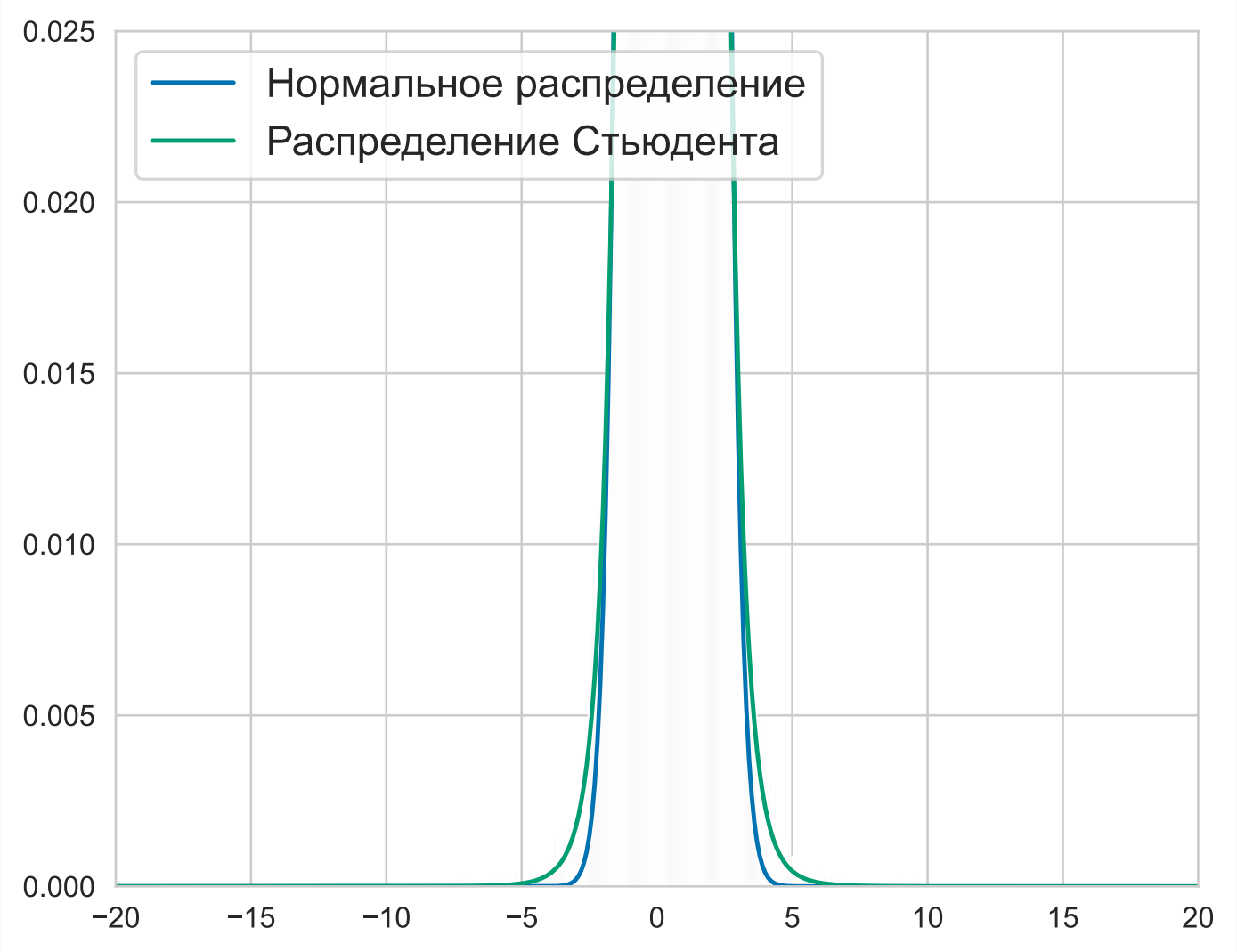
Распределение среднего: хвосты



Предсказательное распределение: нормальное и Стьюдента



Предсказательное распределение: хвосты



Спасибо за внимание!



