

# БАЙЕСОВСКИЙ ВЫВОД В ГАУССИАНЕ

---

Сергей Николенко

СПбГУ — Санкт-Петербург

15 сентября 2017 г.

---

*Random facts:*

- 15 сентября 1917 г. Временное правительство выпустило из тюрьмы Троцкого и провозгласило Россию республикой; в тот же день вышел в свет первый номер журнала Forbes
- 15 сентября 1930 г. Гитлеру было отказано в праве стать депутатом рейхстага (потому что он был гражданином Австрии); пять лет спустя, 15 сентября 1935 г. немецкие евреи были лишены гражданства, а государственным символом Германии стала свастика

- 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  $\sigma^2$  и будем в качестве параметра рассматривать только  $\mu$ .

- Сопряжённое априорное распределение для  $\mu$  при фиксированном  $\sigma^2$  тоже нормальное и выглядит как

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

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

- При нашем априорном распределении у  $\mu$  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  $z_1 | z_2$  распределена нормально с параметрами

$$E(z_1 | z_2) = E(z_1) + \frac{\text{Cov}(z_1, z_2)}{\text{Var}(z_2)} (z_2 - E(z_2)),$$

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

$$(\text{Var}(x) = E[(x - Ex)^2], \text{Cov}(x, y) = E[(x - Ex)(y - Ey)]).$$

- В нашем случае:

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

$$E(x) = \mu_0,$$

$$\text{Var}(x) = E(\text{Var}(x | \mu)) + \text{Var}(E(x | \mu)) = \sigma^2 + \sigma_0^2,$$

$$\text{Cov}(x, \mu) = E[(x - \mu_0)(\mu - \mu_0)] = \sigma_0^2.$$

- Применив упражнение, получаем:

$$E(\mu | x) = \mu_0 + \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2}(x - \mu_0) = \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2}x + \frac{\sigma^2}{\sigma_0^2 + \sigma^2}\mu_0,$$

$$\text{Var}(\mu | x) = \frac{\sigma^2\sigma_0^2}{\sigma_0^2 + \sigma^2} = \frac{1}{\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2}}.$$

- Итого:

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

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

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

- 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  $p(x_i | \mu) \sim \mathcal{N}(\mu, \sigma^2)$  и  $x_i$  независимы, то  $p(\bar{x} | \mu) \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$ .

- Для апостериорной вероятности будет

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

- Подставляя в наш предыдущий результат, получим:

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



- Если зафиксировать  $\mu$  и менять  $\sigma^2$ , то сопряжённым априорным распределением будет обратное гамма-распределение:

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

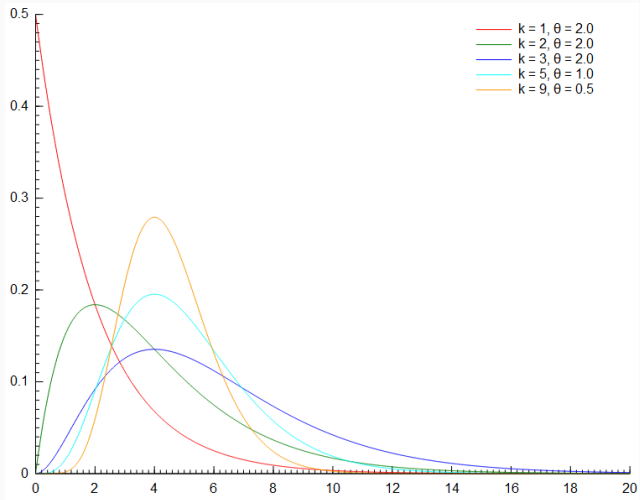
- Тогда в апостериорном распределении будет

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

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

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

# ГАММА--РАСПРЕДЕЛЕНИЕ



## КОГДА И $\mu$ , И $\sigma^2$ МЕНЯЮТСЯ

- Что делать, когда и  $\mu$ , и  $\sigma^2$  меняются?
- Можно было бы предположить, что  $\mu$  и  $\sigma^2$  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 $\mu$ , И $\sigma^2$ МЕНЯЮТСЯ

- Что делать, когда и  $\mu$ , и  $\sigma^2$  меняются?
- Можно было бы предположить, что  $\mu$  и  $\sigma^2$  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  $\mu$  и  $\sigma^2$  зависимы. :) Новая точка  $x$  вводит зависимость между ними.

- В настоящем сопряжённом априорном распределении будут:

$$\begin{aligned}x \mid \mu, \tau &\sim \mathcal{N}(\mu, \tau), \\ \mu \mid \tau &\sim \mathcal{N}(\mu_0, n_0 \tau), \\ \tau &\sim G(\alpha, \beta).\end{aligned}$$

- Давайте выясним, как изменятся параметры, и заодно докажем.

- Самое простое – это, по уже известным результатам,

$$\mu \mid x, \tau \sim \mathcal{N} \left( \frac{n\tau}{n\tau + n_0\tau} \bar{x} + \frac{n_0\tau}{n\tau + n_0\tau} \mu_0, n\tau + n_0\tau \right).$$

- Затем давайте разберёмся с  $\tau \mid x$ :

$$p(\tau, \mu \mid x) \propto p(\tau) \cdot p(\mu \mid \tau) \cdot p(x \mid \tau, \mu),$$

и мы хотим это распределение маргинализовать по  $\mu$ ...

- Подсчитаем:

$$\begin{aligned} p(\tau, \mu | x) &\propto p(\tau) \cdot p(\mu | \tau) \cdot p(x | \tau, \mu) \\ &\propto \tau^{\alpha-1} e^{-\tau\beta} \cdot \tau^{\frac{1}{2}} e^{-\frac{n_0\tau}{2}(\mu-\mu_0)^2} \cdot \tau^{\frac{n}{2}} e^{-\frac{\tau}{2} \sum (x_i-\mu)^2} \\ &\propto \tau^{\alpha+\frac{n}{2}-\frac{1}{2}} e^{-\tau(\beta+\frac{1}{2} \sum (x_i-\bar{x})^2)} e^{-\frac{\tau}{2} (n_0(\mu-\mu_0)^2+n(\bar{x}-\mu)^2)} \end{aligned}$$

(простой трюк:  $x_i - \mu = x_i - \bar{x} + \bar{x} - \mu$ ).

- Теперь надо проинтегрировать

$$\int_{\mu} e^{-\frac{\tau}{2}(n_0(\mu-\mu_0)^2+n(\bar{x}-\mu)^2)} d\mu.$$

Упражнение. Проинтегрируйте. :) Должна получиться нормировочная константа

$$\tau^{-\frac{1}{2}} e^{\frac{-nn_0\tau}{2(n+n_0)}(\bar{x}-\mu_0)^2}.$$



- Таким образом, получается апостериорное распределение

$$p(\tau | x) \propto \tau^{\alpha + \frac{n}{2} - 1} e^{-\tau \left( \beta + \frac{1}{2} \sum (x_i - \bar{x})^2 + \frac{nn_0}{2(n+n_0)} (\bar{x} - \mu_0)^2 \right)}.$$

- Итого результаты такие:

$$\begin{aligned} \mu | \tau, x &\sim \mathcal{N} \left( \frac{n\tau}{n\tau + n_0\tau} \bar{x} + \frac{n_0\tau}{n\tau + n_0\tau} \mu_0, n\tau + n_0\tau \right), \\ \tau | x &\sim G \left( \alpha + \frac{n}{2}, \beta + \frac{1}{2} \sum (x_i - \bar{x})^2 + \frac{nn_0}{2(n+n_0)} (\bar{x} - \mu_0)^2 \right). \end{aligned}$$

- Теперь предсказание нового  $x_{\text{new}}$ :

$$\begin{aligned}
 p(x_{\text{new}} | x) &= \int \int \underbrace{\text{Gamma}}_{\tau|x} \cdot \underbrace{\text{Gaussian}}_{\mu|\tau,x} \cdot \underbrace{\text{Gaussian}}_{x_{\text{new}}|\tau,\mu} d\tau d\mu = \\
 &= \int \underbrace{\text{Gamma}}_{\tau|x} \int \underbrace{\text{Gaussian}}_{\mu|\tau,x} \cdot \underbrace{\text{Gaussian}}_{x_{\text{new}}|\tau,\mu} d\tau d\mu = \\
 &= \int \underbrace{\text{Gamma}}_{\tau|x} \cdot \underbrace{\text{Gaussian}}_{x_{\text{new}}|\tau,x} d\tau = \dots
 \end{aligned}$$

- В результате получится распределение Стьюдента.
- Упражнение.** Проведите эти вычисления.

- Вообще говоря, всё, о чём мы говорили – частные случаи *экспоненциального семейства* распределений:

$$p(\mathbf{x} | \eta) = h(\mathbf{x})g(\eta)e^{\eta^T \mathbf{u}(\mathbf{x})}.$$

- $\eta$  называются *естественными параметрами* (natural parameters).

- Например, распределение Бернулли:

$$\begin{aligned} p(x | \mu) &= \mu^x (1 - \mu)^{1-x} = e^{x \ln \mu + (1-x) \ln(1-\mu)} = \\ &= (1 - \mu) e^{\ln\left(\frac{\mu}{1-\mu}\right)x}, \end{aligned}$$

и естественный параметр получился  $\eta = \ln\left(\frac{\mu}{1-\mu}\right)$ :

$$p(x | \eta) = \sigma(-\eta) e^{-\eta x},$$

где  $\sigma(y) = \frac{1}{1+e^{-y}}$  – сигмоид-функция.

- Для мультиномиального распределения с параметрами  $\mu_1, \dots, \mu_{M-1}$  получаются

$$\eta_k = \ln \left( \frac{\mu_k}{1 - \sum_j \mu_j} \right) \text{ и}$$

$$p(\mathbf{x} | \eta) = \left( 1 + \sum_{k=1}^{M-1} e^{\eta_k} \right)^{-1} e^{\eta^\top \mathbf{x}}.$$

Упражнение. Проверьте!

- Так вот, для распределений из экспоненциального семейства

$$p(\mathbf{x} | \eta) = h(\mathbf{x})g(\eta)e^{\eta^\top \mathbf{u}(\mathbf{x})}$$

можно сразу оптом найти сопряжённые априорные распределения:

$$p(\eta | \chi, \nu) = f(\chi, \nu)g(\eta)^\nu e^{\nu\eta^\top \chi},$$

где  $\chi$  – гиперпараметры, а  $g$  то же самое, что в исходном распределении.

**Упражнение.** Проверьте это и получите вышеописанные примеры как частные случаи.

Спасибо за внимание!