Байесовские методы. Лекция 3.
Байесовский вывод для мультиномиальной и нормальных моделей.

Целищев М.А.

МГУ им. М. В. Ломоносова
Факультет вычислительной математики и кибернетики
Кафедра математической статистики

весна 2021

Список литературы

- J. Kruschke. Doing Bayesian Data Analysis, Second Edition. A Tutorial with R, JAGS, and Stan. Academic Press, 2014.
- R. McElreath. Statistical Rethinking. A Bayesian Course with Examples in R and Stan. Chapman and Hall CRC, 2015.
- O. Martin. Bayesian Analysis with Python. Introduction to Statistical Modeling and Probabilistic Programming using PyMC3 and ArviZ, Second Edition. Packt, 2018.
- K. P. Murphy. Machine Learning: A Probabilistic Perspective. The MIT Press, 2012.
- A. Gelman, J. Carlin, H. Stern, D. Dunson, A. Vehtari, D. Rubin. Bayesian Data Analysis, Third Edition. CRC Press, 2013.
- S. Brooks, A. Gelman, G. L. Jones and X.-L. Meng. Handbook of Markov Chain Monte Carlo. Chapman & Hall/CRC, 2011.

Напоминание

Итак, на прошлой лекции была решена задача байесовской оценки вероятности выпадения решки после n подбрасываний, среди которых было n_1 решек и $n_2 = n - n_1$ орлов.

Правдоподобие:

$$p(\mathbf{x}|\theta) = \theta^{n_1}(1-\theta)^{n_2}, \quad \mathbf{x} = (x_1, \ldots, x_n) \in \{0,1\}^n, \quad k = \sum_{i=1}^{n} x_i.$$

Априорное: $p(\theta) \sim \mathrm{Beta}(\alpha, \beta)$, $\alpha, \beta > 0$, т.е. $p(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$, $\theta \in [0,1]$.

Апостериорное: $p(\theta|\mathbf{x}) = \dfrac{p(\mathbf{x}|\theta)p(\theta)}{p(\mathbf{x})} \propto \theta^{n_1+\alpha-1}(1-\theta)^{n_2+\beta-1}$, т.е.

$$\theta|\mathbf{x} \sim \mathrm{Beta}(n_1+\alpha, n_2+\beta)$$

Напоминание

Кроме того, было сделано байесовское предсказание для следующего броска монеты (X_{new} — индикатор решки):

$$\mathbf{P}(X_{\mathrm{new}} = x_{\mathrm{new}} | X = \mathbf{x}) \equiv p(x_{\mathrm{new}} | \mathbf{x}) = \int_0^1 p(x_{\mathrm{new}} | \theta)\, p(\theta | \mathbf{x})\, d\theta.$$

В частности, для $x_{\mathrm{new}} = 1$,

$$p(1|\mathbf{x}) = \int_0^1 \theta \, \frac{\theta^{n_1+\alpha-1}(1-\theta)^{n_2+\beta-1}}{B(n_1+\alpha, n_2+\beta)}\, d\theta = \frac{n_1+\alpha}{n+\alpha+\beta},$$

а для $x_{\mathrm{new}} = 0$,

$$p(0|\mathbf{x}) = \frac{n_2+\beta}{n+\alpha+\beta}.$$

Всё хорошо считалось, потому что апостериорное распределение имело тот же функциональный вид, что и априорное.

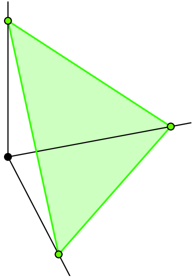
Кубик

Пусть теперь вместо монетки бросается k-гранный кубик и ξ — номер выпавшей грани. Тогда $\xi \sim \mathrm{Categorical}(\theta_1, \ldots, \theta_{k-1})$, т.е.

$$\mathbf{P}_\theta(\xi = j) \equiv p(j|\theta) = \theta_j, \quad j = 1 \ldots k.$$

Заметим, что область значений вектора параметров $\theta = (\theta_1, \ldots, \theta_k)$ есть *вероятностный симплекс*:

$$\Delta_{k-1} = \left\{ \theta \in \mathbb{R}^k \ : \theta_j \geqslant 0 \text{ и } \sum_{j=1}^{k} \theta_j = 1 \right\}$$



случай $k = 3$

и $\theta_k = 1 - \theta_1 - \ldots - \theta_{k-1}$.

Кубик

Задача состоит в том, чтобы по наблюдениям $\mathbf{x} = (x_1, \ldots, x_n)$ над $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \xi$ сделать прогноз для вектора параметров θ, а также прогноз для новых бросков.

Обозначим $n_j = \sum_{i=1}^{n} \mathbb{1}_{\{x_i = j\}}$ — число испытаний, на которых выпала j-ая грань, $j = 1 \ldots k$.

Тогда правдоподобием в этой задаче будет:

$$L(\mathbf{x}; \theta) \equiv p(\mathbf{x}|\theta) = p(x_1|\theta) \ldots p(x_n|\theta) = \theta_1^{n_1} \ldots \theta_{k-1}^{n_{k-1}} \theta_k^{n_k}, \quad x \in \{1..k\}^n, \ \theta \in \Delta_{k-1}.$$

Несложно убедиться в том, что классические оценки максимального правдоподобия для параметров θ будут иметь вид:

$$\hat{\theta}_j(x) = \frac{n_j}{n}, \quad j = 1..k.$$

А что с байесовским подходом?

Кубик + Байес

Как обычно, сначала следует выбрать априорное распределение на вектор параметров θ.

Для простоты возьмём равномерное распределение на подмножестве $\mathbb{R}^{k-1}$, соответствующему симплексу Δ_{k-1}:

$$p(\theta_1, \ldots, \theta_{k-1}) = \frac{1}{c_k}, \quad \theta \in \Delta_{k-1},$$

где нормирующая константа равна (ДЗ!)

$$c_k = \int_0^1 d\theta_1 \int_0^{1-\theta_1} d\theta_2 \ldots \int_0^{1-\theta_1-\ldots-\theta_{k-2}} d\theta_{k-1} = \frac{1}{(k-1)!}.$$

Иными словами, берём $\theta_1 \sim \mathrm{U}[0,1]$, $\quad (\theta_2|\theta_1) \sim \mathrm{U}[0, 1-\theta_1]$, $\ldots$, $(\theta_{k-1}|\theta_1, \ldots, \theta_{k-2}) \sim \mathrm{U}[0, 1-\theta_1-\ldots-\theta_{k-2}]$, $\quad$ и $\theta_k = 1 - \theta_1 - \ldots - \theta_{k-1}$.

Кубик + Байес

Итак, априорное распределение $p(\theta) \propto 1$, $\theta \in \Delta_{k-1}$.

А правдоподобие уже посчитано: $p(\mathbf{x}|\theta) = \theta_1^{n_1} \dots \theta_k^{n_k}$.

По теореме Байеса можно посчитать апостериорное распределение θ при известных наблюдаемых данных $\mathbf{x} \in \{1..k\}^n$:

$$p(\theta|\mathbf{x}) = \frac{p(\mathbf{x}|\theta)\, p(\theta)}{p(\mathbf{x})} \ \propto\ \theta_1^{n_1} \dots \theta_k^{n_k}, \quad \theta \in \Delta_{k-1}.$$

При этом нормировочная константа для апостериорного распределения равна:

$$\int_{\Delta_{k-1}} \theta_1^{n_1} \dots \theta_{k-1}^{n_{k-1}} (1 - \theta_1 - \dots - \theta_{k-1})^{n_k}\, d\theta_1 \dots d\theta_{k-1}.$$

Как её посчитать-то?

Интеграл Дирихле

Бета-функцию $B(\alpha_1, \alpha_2)$ можно естественным образом обобщить на случай k аргументов — это и будет нужный интеграл:

$$B(\alpha) = \int_{\Delta_{k-1}} \theta_1^{\alpha_1 - 1} \dots \theta_{k-1}^{\alpha_{k-1}-1} (1 - \theta_1 - \dots - \theta_{k-1})^{\alpha_k - 1} \, d\theta_1 \dots d\theta_{k-1},$$

где $\alpha = (\alpha_1, \dots, \alpha_k)$ и каждая $\alpha_j > 0$. При этом справедливо:

$$B(\alpha) = \frac{\Gamma(\alpha_1)\,\Gamma(\alpha_2) \dots \Gamma(\alpha_k)}{\Gamma(\alpha_1 + \alpha_2 + \dots + \alpha_k)}, \quad \alpha \in \mathbb{R}^n_+.$$

Распределение Дирихле

Итак, апостериорное распределение вектора параметров θ имеет вид

$$p(\theta|\mathbf{x}) = \frac{1}{B(n_1+1,\ldots,n_k+1)}\, \theta_1^{n_1}\ldots\theta_{k-1}^{n_{k-1}}\theta_k^{n_k}, \quad \theta \in \Delta_{k-1}.$$

Такое распределение известно и называется *распределением Дирихле* (CLICK ME) с параметрами $(n_1+1,\ldots,n_k+1)$:

$$\theta|\mathbf{x} \sim \mathrm{Dirichlet}(n_1+1,\ldots,n_k+1)$$

Здесь важен его функциональный вид: $p(\theta|\mathbf{x}) \ \propto\ \theta_1^{n_1}\ldots\theta_{k-1}^{n_{k-1}}\theta_k^{n_k}$,
а $B(n_1+1,\ldots,n_k+1)$ — нормировочная константа.

ДЗ: Показать, что если $\xi = (\xi_1,\ldots,\xi_k) \sim \mathrm{Dirichlet}(\alpha_1,\ldots,\alpha_k)$, $\alpha_j > 0$, то
1) $\xi_j \sim \mathrm{Beta}(\alpha_j, \sum_{s\neq j}\alpha_s)$
2) $(\xi_{\sigma(1)},\ldots,\xi_{\sigma(k)}) \sim \mathrm{Dirichlet}(\alpha_{\sigma(1)},\ldots,\alpha_{\sigma(k)})$ для произвольной перестановки σ чисел $1..k$.
3) $\xi = (\xi_1,\ldots,\xi_{k-2},\xi_{k-1}+\xi_k) \sim \mathrm{Dirichlet}(\alpha_1,\ldots,\alpha_{k-2},\alpha_{k-1}+\alpha_k)$.

Кубик + Байес

Почему мы брали в качестве априорного равномерное? Чтобы нормировочную константу посчитать аналитически. Можно ли взять другое априорное, чтобы соответствующий интеграл также легко брался?

Надо выбирать априорное так, чтобы апостериорное было из того же семейства распределений, что и априорное.

Из представления апостериорного распределения

$$p(\theta|\mathbf{x}) \;\propto\; p(\mathbf{x}|\theta)p(\theta) = \theta_1^{n_1} \dots \theta_k^{n_k} p(\theta)$$

сразу видно, что *сопряжённым* к мультиномиальному будет семейство распределений Дирихле:

$$\theta \sim \mathrm{Dirichlet}(\alpha), \quad \alpha \in \mathbb{R}_+^n,$$

т.е. $p(\theta) \propto \theta_1^{\alpha_1 - 1} \dots \theta_k^{\alpha_k - 1}$. Тогда апостериорное распределение

$$\theta|\mathbf{x} \sim \mathrm{Dirichlet}(n_1 + \alpha_1, \dots, n_k + \alpha_k)$$

Байесовский прогноз для кубика

Итак, апостериорное распределение параметра θ при наблюдаемых данных $\mathbf{x}$ есть

$$\theta|\mathbf{x} \sim \mathrm{Dirichlet}(n_1+\alpha_1, \ldots, n_k+\alpha_k),$$

где $n_j = \sum_{i=1}^{n} \mathbb{1}_{\{x_i=j\}}$ — число раз, когда выпала j-ая грань.

Хотим сделать прогноз на следующее наблюдение X_{new}.

$$\mathbf{P}(X_{\mathrm{new}}=j|X=\mathbf{x}) \equiv p(j|\mathbf{x}) = \int_{\Delta_{k-1}} p(j|\theta)\, p(\theta|\mathbf{x})\, d\theta =$$

$$= \int_{\Delta_{k-1}} \theta_j \frac{\theta_1^{(n_1+\alpha_1-1)} \ldots \theta_k^{(n_k+\alpha_k-1)}}{B(n_1+\alpha_1, \ldots, n_k+\alpha_k)} d\theta_1 \ldots d\theta_{k-1} =$$

$$= \frac{n_j+\alpha_j}{n_1+\ldots+n_k+\alpha_1+\ldots+\alpha_k}$$

для каждого $j = 1 \ldots k$.

Нормальная модель, 1

Рассмотрим теперь другую задачу. $\xi \sim \mathcal{N}(\theta, 1)$. Нужно сделать статистический вывод о параметре $\theta \in \mathbb{R}$ на основании наблюдаемых значений $\mathbf{x} = (x_1, \ldots, x_n)$ над с.в. ξ.

Плотность ξ: $\quad p_\theta(x_1) \equiv p(x_1|\theta) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x_1-\theta)^2}{2}\right)$, $x_1 \in \mathbb{R}$.

Правдоподобие:

$$L(\mathbf{x}; \theta) \equiv p(\mathbf{x}|\theta) = \prod_{i=1}^{n} p(x_i|\theta) = (2\pi)^{-n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i - \theta)^2\right).$$

В курсе мат. статистики было доказано, что оценка

$$\overline{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

является оценкой максимального правдоподобия параметра θ (а также она является оптимальной в среднем квадратическом среди несмещённых оценок).

Нормальная модель, 1

А как решить эту задачу по-байесовски?

Как обычно, нужно выбрать априорное распределение $p(\theta)$ и, воспользовавшись теоремой Байеса, вычислить апостериорное распределение $p(\theta|\mathbf{x})$.

Как выбрать априорное распределение? Надо выбрать некоторое семейство распределений, так чтобы при умножении любого распределения из этого семейства на правдоподобие получили бы снова функциональный вид этого семейства.

Правдоподобие:

$$p(\mathbf{x}|\theta) = (2\pi)^{-n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i - \theta)^2\right).$$

Относительно θ — это парабола под экспонентой с отрицательным коэффициентом при θ^2.

Нормальная модель, 1

Правдоподобие:

$$p(\mathbf{x}|\theta) = (2\pi)^{-n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)^2\right).$$

Заметим, что при умножении двух экспонент с отрицательным коэффициентом при θ^2 снова получим параболу под экспонентой с отрицательным коэффициентом при θ^2.

А какое распределение имеет такой вид? **Нормальное!**

Итак, если взять $\theta \sim \mathcal{N}(a_0, \sigma_0^2)$, то есть

$$p(\theta) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{(\theta-a_0)^2}{2\sigma_0^2}\right) \propto \exp\left(-\frac{(\theta-a_0)^2}{2\sigma_0^2}\right), \quad \theta \in \mathbb{R},$$

то $p(\theta|\mathbf{x}) \propto p(\theta)p(\mathbf{x}|\theta) \propto \exp\left(-\frac{1}{2}\left[\frac{(\theta-a_0)^2}{\sigma_0^2} + \sum_{i=1}^{n}(x_i-\theta)^2\right]\right).$

Нормальная модель, 1

Далее,

$$p(\theta|\mathbf{x}) \propto \exp\left(-\frac{1}{2}\left[\frac{(\theta-a_0)^2}{\sigma_0^2}+\sum_{i=1}^{n}(x_i-\theta)^2\right]\right) \propto$$

$$\propto \exp\left(-\frac{1}{2}\left[\left(\frac{1}{\sigma_0^2}+n\right)\theta^2-2\left(\frac{a_0}{\sigma_0^2}+n\overline{\mathbf{x}}\right)\theta\right]\right) =$$

$$= \exp\left(-\frac{1}{2}\left(\frac{1}{\sigma_0^2}+n\right)\cdot\left[\theta^2-2\left(\frac{\frac{a_0}{\sigma_0^2}+n\overline{\mathbf{x}}}{\frac{1}{\sigma_0^2}+n}\right)\theta\right]\right) \propto$$

$$\propto \exp\left(-\frac{1}{2}\left(\frac{1}{\sigma_0^2}+n\right)\cdot\left[\theta-\left(\frac{\frac{a_0}{\sigma_0^2}+n\overline{\mathbf{x}}}{\frac{1}{\sigma_0^2}+n}\right)\right]^2\right),$$

то есть $\theta|\mathbf{x} \sim \mathcal{N}\left(\frac{\frac{a_0}{\sigma_0^2}+n\overline{\mathbf{x}}}{\frac{1}{\sigma_0^2}+n}, \left(\frac{1}{\sigma_0^2}+n\right)^{-1}\right).$

Нормальная модель, 1

Итого, если брать априорное

$$\theta \sim \mathcal{N}(a_0, \sigma_0^2),$$

то апостериорным будет

$$\theta | \mathbf{x} \sim \mathcal{N}\left(\frac{\frac{a_0}{\sigma_0^2} + n\overline{\mathbf{x}}}{\frac{1}{\sigma_0^2} + n}, \left(\frac{1}{\sigma_0^2} + n\right)^{-1}\right).$$

Иными словами, *сопряжённым* к нормальному распределению с известной дисперсией будет семейство нормальных распределений.

Заметим, что $\dfrac{a_0\sigma_0^{-2} + n\overline{\mathbf{x}}}{\sigma_0^{-2} + n} = \dfrac{\sigma_0^{-2}}{\sigma_0^{-2} + n}\, a_0 + \dfrac{n}{\sigma_0^{-2} + n}\, \overline{\mathbf{x}}$ — компромисс между априорным средним a_0 и оценкой макс. правдоподобия $\overline{\mathbf{x}}$, и что $\mathbf{D}(\theta|\mathbf{x}) = \left(\frac{1}{\sigma_0^2} + n\right)^{-1} \to 0$ при $n \to \infty$.

Нормальная модель, 1

Как сделать предсказания для нового наблюдения x_{new}? Предсказание — это, как и ранее, условное распределение на X_{new} при условии, что первые n наблюдений $\mathbf{x} = (x_1, \ldots, x_n)$ известны.

$$p(x_{\text{new}}|\mathbf{x}) = \int\limits_{-\infty}^{+\infty} p(x_{\text{new}}|\theta)\, p(\theta|\mathbf{x})\, d\theta =$$

$$= \int\limits_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(x_{\text{new}} - \theta)^2}{2}\right] \cdot \frac{1}{\sqrt{2\pi\,\mathbf{D}(\theta|\mathbf{x})}} \exp\left[-\frac{(\theta - \mathbf{E}(\theta|\mathbf{x}))^2}{2\,\mathbf{D}(\theta|\mathbf{x})}\right] d\theta =$$

$$= \ldots$$

Поэтому $(X_{\text{new}}|\mathbf{x}) \sim \ldots$ (дз).

Нормальная модель, 2

Пусть теперь $\xi \sim \mathcal{N}(0, \theta^2)$, $\theta > 0$. Нужно сделать статистический вывод о параметре θ на основании наблюдаемых значений $\mathbf{x} = (x_1, \ldots, x_n)$ над с.в. ξ.

Плотность ξ: $\quad p_\theta(x_1) \equiv p(x_1|\theta^2) = \frac{1}{\sqrt{2\pi\theta^2}} \exp\left(-\frac{x_1^2}{2\theta^2}\right), \quad x_1 \in \mathbb{R}.$

Правдоподобие:

$$L(\mathbf{x};\theta) \equiv p(\mathbf{x}|\theta^2) = \prod_{i=1}^{n} p(x_i|\theta^2) = (2\pi\theta^2)^{-n/2} \exp\left(-\frac{1}{2\theta^2}\sum_{i=1}^{n} x_i^2\right).$$

В курсе мат. статистики было доказано, что оценка

$$\frac{1}{n}\sum_{i=1}^{n} x_i^2$$

является оценкой максимального правдоподобия дисперсии θ^2 (а также она является оптимальной в среднем квадратическом среди несмещённых оценок).

Нормальная модель, 2

Правдоподобие:

$$L(\mathbf{x};\theta) \equiv p(\mathbf{x}|\theta^2) = \prod_{i=1}^{n} p(x_i|\theta^2) = (2\pi\theta^2)^{-n/2} \exp\left(-\frac{1}{2\theta^2}\sum_{i=1}^{n} x_i^2\right).$$

Теперь по Байесу: какое априорное $p(\theta^2)$ надо выбрать, чтобы при умножении на правдоподобие сохранялась та же функциональная форма?

$$p(\theta^2|\mathbf{x}) \;\propto\; p(\mathbf{x}|\theta^2)p(\theta^2) \;\propto\; (\theta^2)^{-n/2} \exp\left(-\frac{1}{2\theta^2}\sum_{i=1}^{n} x_i^2\right) p(\theta^2)$$

Нормальная модель, 2

Сделаем замену $\tau = \dfrac{1}{\theta^2}$. Параметр τ называется *точностью* («precision»).

Правдоподобие: $p(\mathbf{x}|\tau) = \left(\dfrac{\tau}{2\pi}\right)^{n/2} \exp\left(-\dfrac{\tau}{2}\sum_{i=1}^{n} x_i^2\right)$.

Какое априорное $p(\tau)$ надо выбрать, чтобы при умножении на правдоподобие оставалась бы та же функциональная форма по τ?

$$p(\tau|\mathbf{x}) \;\propto\; p(\mathbf{x}|\tau)\, p(\tau) \;\propto\; \tau^{n/2} \exp\left(-\frac{\tau}{2}\sum_{i=1}^{n} x_i^2\right) p(\tau)$$

Ответ:

$$p(\tau) \;\propto\; \tau^{\alpha-1} \exp(-\beta\tau), \quad \tau > 0$$

Это **гамма-распределение** (link): $\tau \sim \Gamma(\alpha, \beta)$.

$$p(\tau) = \frac{\beta^\alpha}{\Gamma(\alpha)} \tau^{\alpha-1} \exp(-\beta\tau), \quad \tau > 0.$$

Нормальная модель, 2

Итак, при априорном $\tau \sim \Gamma(\alpha, \beta)$, $\alpha, \beta > 0$, апостериорное на τ выглядит так:

$$p(\tau|\mathbf{x}) \;\propto\; p(\mathbf{x}|\tau)\, p(\tau) \;\propto\; \tau^{n/2+\alpha-1} \exp\left(-\left[\beta + \frac{1}{2}\sum_{i=1}^{n} x_i^2\right]\tau\right), \quad \tau > 0,$$

то есть

$$(\tau|\mathbf{x}) \sim \Gamma\left(\alpha + \frac{n}{2}, \beta + \frac{1}{2}\sum_{i=1}^{n} x_i^2\right), \quad \tau > 0.$$

Нормальная модель, 2

Поскольку $\tau = 1/\theta^2$, то запись $\tau \sim \Gamma(\alpha, \beta)$ часто обозначают следующим образом:

$$\theta^2 \sim \mathrm{InvGamma}(\alpha, \beta), \quad \alpha, \beta > 0$$

и называют это распределение *обратным гамма-распределением* (link).

При этом апостериорное

$$(\theta^2|\mathbf{x}) \sim \mathrm{InvGamma}\left(\alpha + \frac{n}{2}, \beta + \frac{1}{2}\sum_{i=1}^{n} x_i^2\right).$$

Итак, семейство обратных гамма-распределений является *сопряжённым* к нормальному распределению с неизвестной дисперсией.

ДЗ: построить прогноз для следующего наблюдения, т.е. посчитать $p(x_{\mathrm{new}}|\mathbf{x})$.

Нормальная модель, 3

Пусть теперь $\xi \sim \mathcal{N}(\mu, \tau^{-1})$, $\mu \in \mathbb{R}$, $\tau > 0$. Нужно сделать статистический вывод о векторе $\theta = (\mu, \tau)$ на основании наблюдаемых значений $\mathbf{x} = (x_1, \ldots, x_n)$ над с.в. ξ.

Плотность ξ: $\quad p_\theta(x_1) \equiv p(x_1|\mu, \tau) = \sqrt{\frac{\tau}{2\pi}} \exp\left(-\frac{\tau(x_1-\mu)^2}{2}\right)$.

Правдоподобие:

$$L(\mathbf{x}; \theta) \equiv p(\mathbf{x}|\mu, \tau) = \prod_{i=1}^{n} p(x_i|\mu, \tau) = \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left(-\frac{\tau}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right).$$

В курсе мат. статистики было доказано, что оценки

$$\overline{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{\mathbf{x}})^2$$

являются оценками максимального правдоподобия для среднего μ и дисперсии τ^{-1} соответственно.

Нормальная модель, 3

Байес. Правдоподобие:

$$p(\mathbf{x}|\mu,\tau) = \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left(-\frac{\tau}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right), \quad \mu \in \mathbb{R}, \tau > 0.$$

Как выбрать априорное $p(\mu,\tau)$? По той же схеме...

Апостериорное:

$$p(\mu,\tau|\mathbf{x}) \;\propto\; p(\mathbf{x}|\mu,\tau)\, p(\mu,\tau) \;\propto\; \tau^{n/2} \exp\left(-\frac{\tau}{2}\left[n\mu^2 - 2\mu n\overline{\mathbf{x}} + \sum_{i=1}^{n} x_i^2\right]\right) p(\mu,\tau)$$

Сразу заметим, что априорное $p(\mu,\tau) = p(\mu)p(\tau)$ (с независимыми компонентами) не подходит, так как правдоподобие не факторизуется...

Апостериорное:

$$p(\mu,\tau|\mathbf{x}) \ \propto \ \tau^{n/2}\exp\left(-\frac{\tau}{2}\left[n\mu^2 - 2\mu n\overline{\mathbf{x}} + \sum_{i=1}^{n} x_i^2\right]\right)p(\mu,\tau)$$

Зато подходит следующее априорное:

$$p(\mu,\tau) = \frac{\beta^\alpha}{\Gamma(\alpha)}\tau^{\alpha-1}\exp\left(-\beta\tau\right)\cdot\sqrt{\frac{\lambda\tau}{2\pi}}\exp\left(-\frac{\lambda\tau}{2}(\mu-a_0)^2\right).$$

Такое распределение называется *гамма-нормальным* (link) с параметрами $a_0 \in \mathbb{R}$, $\lambda > 0$ и $\alpha, \beta > 0$:

$$(\mu,\tau) \sim \mathrm{NormalGamma}\left(a_0,\lambda,\alpha,\beta\right),$$

Его можно задать и проще, заметив что

$$\tau \sim \Gamma(\alpha,\beta)$$

$$(\mu|\tau) \sim \mathcal{N}\left(a_0,(\lambda\tau)^{-1}\right)$$

Нормальная модель, 3

Тогда при гамма-нормальном априорном апостериорное будет иметь вид:

$$(\mu, \tau|\mathbf{x}) \sim \mathrm{NormalGamma}\left(\frac{\lambda a_0 + n\overline{\mathbf{x}}}{\lambda + n}, \lambda + n, \alpha + \frac{n}{2}, \beta + \frac{1}{2}\left(ns^2 + \frac{\lambda n(\overline{\mathbf{x}} - a_0)^2}{\lambda + n}\right)\right),$$

где, как и ранее, $\overline{\mathbf{x}}$ и s^2 — выборочные среднее и дисперсия.

Итак, семейство гамма-нормальных распределений является сопряжённым к нормальному распределению с неизвестными мат. ожиданием и дисперсией.

ДЗ: сделать байесовский вывод для наблюдения x_{new}, то есть посчитать $p(x_{\text{new}}|\mathbf{x})$.

Экспоненциальное семейство распределений

$$p(x_1|\theta) = h(x_1)\exp\left(\sum_{j=1}^{s} D_j(x_1)\eta_j(\theta) - a(\theta)\right), \quad x_1 \in \mathbb{R}.$$

Правдоподобие:

$$p(\mathbf{x}|\theta) = H(\mathbf{x})\exp\left(\sum_{j=1}^{s} T_j(\mathbf{x})\eta_j(\theta) - A(\theta)\right), \quad \mathbf{x} \in \mathbb{R}^n,$$

где $H(\mathbf{x}) = h(x_1)\ldots h(x_n)$, $\quad T_j(\mathbf{x}) = \sum_{i=1}^{n} D_j(x_i)$, $\quad A(\theta) = na(\theta)$.
Кстати, $\{T_j(\mathbf{x})\}_{j=1}^{s}$ — набор достаточных статистик в этой модели.
Априорное:

$$p(\theta) \ \propto \ \exp\left(\sum_{j=1}^{s} \chi_j\eta_j(\theta) - \nu a(\theta)\right).$$

Апостериорное: тот же вид с $\chi_j' = \chi_j + T_j(\mathbf{x})$ и $\nu' = \nu + n$.

Сопряжённые распределения

Список сопряжённых семейств распределений: wikipedia