

Введение в статистическую теорию распознавания образов. Байесовский подход.

Тверская Е. С.
e_tverskaya@bmstu.ru

Московский государственный технический университет им. Н. Э. Баумана
(национальный исследовательский университет)

Москва, 2025



ИУ-6
Компьютерные
системы и сети

Байесовские методы.

Введение. Рассуждения и пояснения.



Требования к реализуемому набору исходов

Вероятность – степень нашей разумной уверенности в некотором предположении.

Байесовские методы.

Введение. Рассуждения и пояснения.



Требования к реализуемому набору исходов

Вероятность – степень нашей разумной уверенности в некотором предположении.

НО!!!

Данная вероятность всегда является **условной**! То есть зависит от состояния информации в данный конкретный момент времени!



Байесовские методы.

Введение. Рассуждения и пояснения.

Требования к реализуемому набору исходов

Вероятность – степень нашей разумной уверенности в некотором предположении.

НО!!!

Данная вероятность всегда является **условной**! То есть зависит от состояния информации в данный конкретный момент времени!

Произошло изменение информации относительно какого-либо конкретного высказывания



Пересмотр вероятности высказывания



Байесовские методы.

Введение. Рассуждения и пояснения. Примеры.

Пример № 1.

Цена букинистического издания книги *«Прикладной регрессионный анализ»*, 1973г., Н. Дрейк, Г. Смит в «отличном» и «хорошем» качестве от 200 руб. до 1000 руб., то есть

$$M[\xi] = 600(\text{руб.})$$

При покупке выясняется, что состояние книги «хорошее». Следовательно, произошло уточнение цены: от 300 руб. до 600 руб., то есть

$$M[\xi] = 450(\text{руб.})$$



Байесовские методы.

Введение. Рассуждения и пояснения. Примеры.

Пример № 1.

Цена букинистического издания книги *«Прикладной регрессионный анализ», 1973г., Н. Дрейк, Г. Смит* в «отличном» и «хорошем» качестве от 200 руб. до 1000 руб., то есть

$$M[\xi] = 600(\text{руб.})$$

При покупке выясняется, что состояние книги «хорошее». Следовательно, произошло уточнение цены: от 300 руб. до 600 руб., то есть

$$M[\xi] = 450(\text{руб.})$$

Пример № 2.

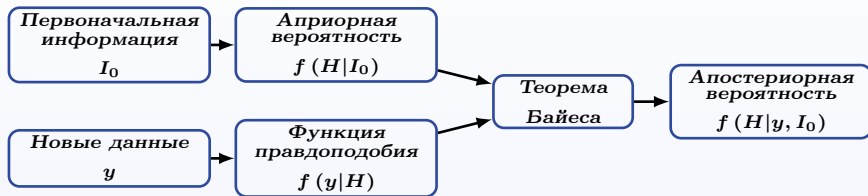
Автобус на остановку должен прийти в 10 ч. 30 мин. Из-за транспортной обстановки в городе он может опоздать (или прийти раньше на 7 мин.).

$$D[\xi] = 7 \text{ мин.}, \quad M[\xi] = 10 \text{ ч. } 30 \text{ мин.}$$

Пользуясь Yandex можно отследить автобус на карте и сделать вывод о его реальном времени прибытия ($M[\xi]$ и $D[\xi]$).

Байесовские методы.

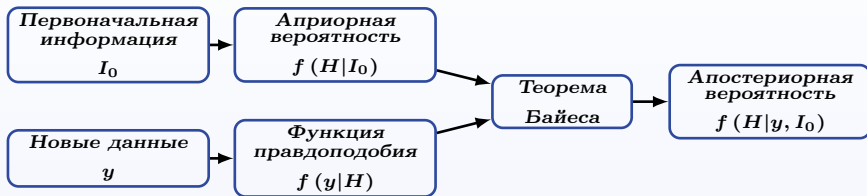
Введение. Рассуждения и пояснения. Схема процесса пересмотра вероятностей.





Байесовские методы.

Введение. Рассуждения и пояснения. Схема процесса пересмотра вероятностей.

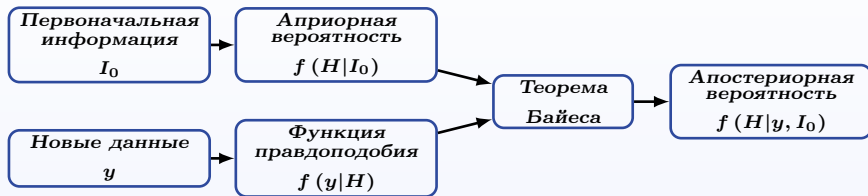


- I_0 – первоначальная информация. Представляет собой комбинацию информации, полученной от предыдущих исследований, теоретических соображений и случайных наблюдений.



Байесовские методы.

Введение. Рассуждения и пояснения. Схема процесса пересмотра вероятностей.

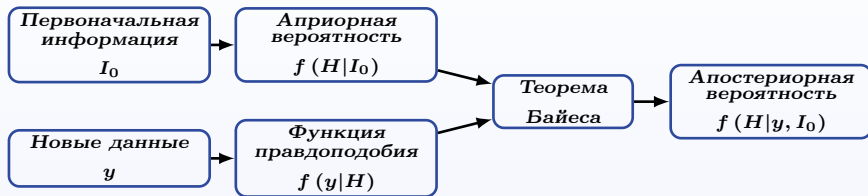


- I_0 – первоначальная информация. Представляет собой комбинацию информации, полученной от предыдущих исследований, теоретических соображений и случайных наблюдений.
- H – некоторое конкретное предположение.



Байесовские методы.

Введение. Рассуждения и пояснения. Схема процесса пересмотра вероятностей.

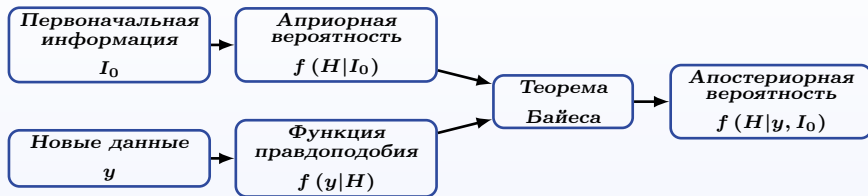


- I_0 – первоначальная информация. Представляет собой комбинацию информации, полученной от предыдущих исследований, теоретических соображений и случайных наблюдений.
- H – некоторое конкретное предположение.
- Первоначальные или априорные вероятности, характеризуемые $f(H|I_0)$, базируются на первоначальной информации I_0 .



Байесовские методы.

Введение. Рассуждения и пояснения. Схема процесса пересмотра вероятностей.

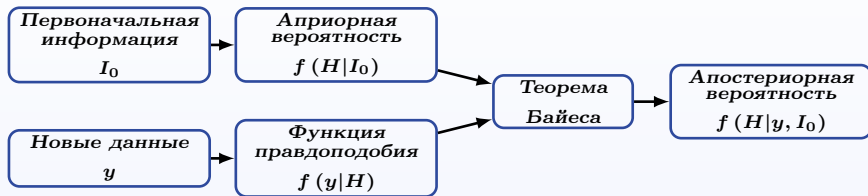


- I_0 – первоначальная информация. Представляет собой комбинацию информации, полученной от предыдущих исследований, теоретических соображений и случайных наблюдений.
- H – некоторое конкретное предположение.
- Первоначальные или априорные вероятности, характеризуемые $f(H|I_0)$, базируются на первоначальной информации I_0 .
- ФПВ $f(y|H)$ для новых наблюдений y при определенном условии H , т. е. заданном высказывании – **функция правдоподобия**.



Байесовские методы.

Введение. Рассуждения и пояснения. Схема процесса пересмотра вероятностей.

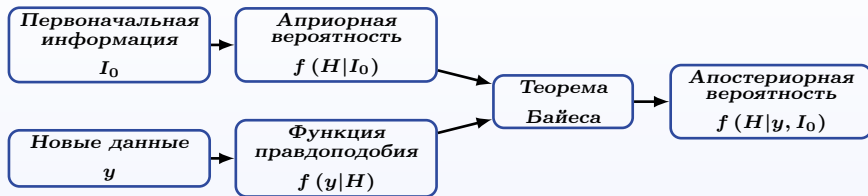


- I_0 – первоначальная информация. Представляет собой комбинацию информации, полученной от предыдущих исследований, теоретических соображений и случайных наблюдений.
- H – некоторое конкретное предположение.
- Первоначальные или априорные вероятности, характеризуемые $f(H|I_0)$, базируются на первоначальной информации I_0 .
- ФПВ $f(y|H)$ для новых наблюдений y при определенном условии H , т. е. заданном высказывании – **функция правдоподобия**.
- Объединяем априорную вероятность $f(H|I_0)$ с функцией правдоподобия $f(y|H)$ с помощью теоремы Байеса и получаем апостериорную вероятность $f(H|y, I_0)$.



Байесовские методы.

Введение. Рассуждения и пояснения. Схема процесса пересмотра вероятностей.



- I_0 – первоначальная информация. Представляет собой комбинацию информации, полученной от предыдущих исследований, теоретических соображений и случайных наблюдений.
- H – некоторое конкретное предположение.
- Первоначальные или априорные вероятности, характеризуемые $f(H|I_0)$, базируются на первоначальной информации I_0 .
- ФПВ $f(y|H)$ для новых наблюдений y при определенном условии H , т. е. заданном высказывании – **функция правдоподобия**.
- Объединяем априорную вероятность $f(H|I_0)$ с функцией правдоподобия $f(y|H)$ с помощью теоремы Байеса и получаем апостериорную вероятность $f(H|y, I_0)$.

Пересмотр первоначальной вероятности $f(H|I_0) \rightarrow f(H|y, I_0)$

Байесовские методы.

Введение. Рассуждения и пояснения. Схема процесса пересмотра вероятностей.

ИУ-6
Компьютерные
системы и сети

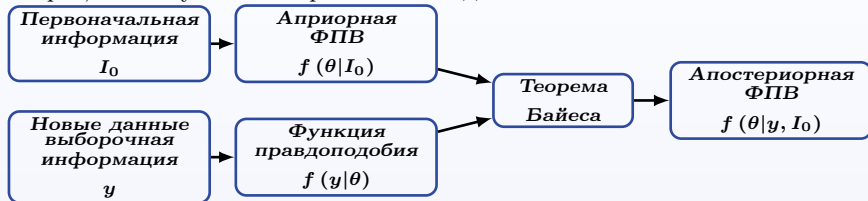
Если в задаче рассматривается не предположение H , а некоторый параметр θ , то схему можно переписать в виде

Байесовские методы.

Введение. Рассуждения и пояснения. Схема процесса пересмотра вероятностей.



Если в задаче рассматривается не предположение H , а некоторый параметр θ , то схему можно переписать в виде

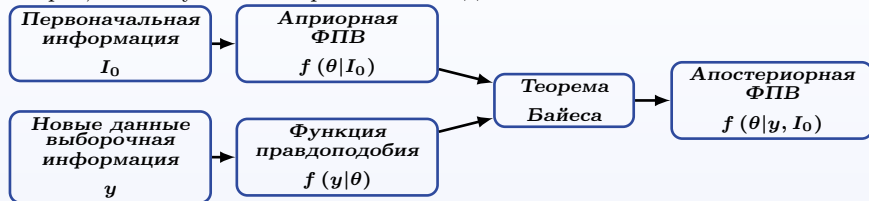




Байесовские методы.

Введение. Рассуждения и пояснения. Схема процесса пересмотра вероятностей.

Если в задаче рассматривается не предположение H , а некоторый параметр θ , то схему можно переписать в виде



Замечание

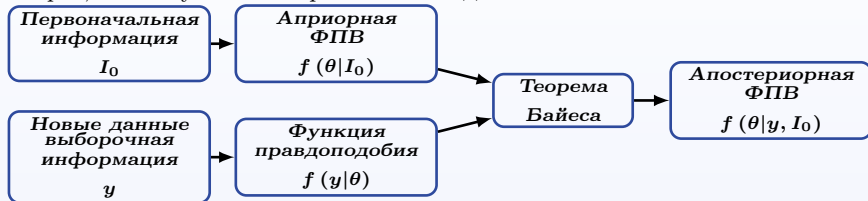
- По мере накопления выборочной информации, она начинает все более преобладать в апостериорной ФПВ, которая все более концентрируется вокруг истинного значения параметра.



Байесовские методы.

Введение. Рассуждения и пояснения. Схема процесса пересмотра вероятностей.

Если в задаче рассматривается не предположение H , а некоторый параметр θ , то схему можно переписать в виде



Замечание

- По мере накопления выборочной информации, она начинает все более преобладать в апостериорной ФПВ, которая все более концентрируется вокруг истинного значения параметра.
- Если два исследователя первоначально располагали различными априорными ФПВ (располагали различной начальной информацией), то их апостериорные ФПВ будут сближаться по мере присоединения дополнительных общих данных.



Байесовские методы.

Введение. Рассуждения и пояснения. Функция правдоподобия.

Определение

Совокупность независимых случайных величин $\{\xi_k(\omega)\}_{k=1}^n$, каждая из которых имеет тот же закон распределения, что и наблюдаемая случайная величина $\xi(\omega)$, называют **случайной выборкой объема n для $\xi(\omega)$** и обозначают $\overrightarrow{\xi_n(\omega)}$ или $\overrightarrow{X_n(\omega)}$.



Байесовские методы.

Введение. Рассуждения и пояснения. Функция правдоподобия.

Определение

Совокупность независимых случайных величин $\{\xi_k(\omega)\}_{k=1}^n$, каждая из которых имеет тот же закон распределения, что и наблюдаемая случайная величина $\xi(\omega)$, называют **случайной выборкой объема n для $\xi(\omega)$** и обозначают $\overrightarrow{\xi_n(\omega)}$ или $\overrightarrow{X_n(\omega)}$.

Определение

Любую возможную **реализацию** $\overrightarrow{x} = [x_1, \dots, x_n]^T$ случайного вектора $\overrightarrow{\xi_n(\omega)}$ называют **выборкой объема n для случайной величины $\xi(\omega)$** , а x_k – k -ый элемент выборки \overrightarrow{x} , $\forall k = \overline{1, n}$.



Байесовские методы.

Введение. Рассуждения и пояснения. Функция правдоподобия.

Замечание

1. Случайная величина $\xi(\omega)$ может быть как скалярной так и векторной.
2. Для обозначения случайной выборки используют следующую форму записи $\overrightarrow{\xi_n(\omega)} = [\xi_1(\omega), \dots, \xi_n(\omega)]^T$ или $\overrightarrow{X_n(\omega)} = [X_1(\omega), \dots, X_n(\omega)]^T$
 $\xi_k(\omega)$ – k -ый элемент случайной выборки, $k = \overline{1, n}$.
 $\overrightarrow{\xi_n(\omega)}$ – n -мерный случайный вектор, с независимыми компонентами, каждая из которых имеет тот же закон распределения, что и случайная величина $\xi(\omega)$.
3. Если $F_\xi(x)$ – функция распределения вероятностей случайной величины $\xi(\omega)$, то функция распределения вероятностей случайной выборки $\overrightarrow{\xi_n(\omega)}$ имеет вид:

$$\begin{aligned}
 F_{\overrightarrow{\xi_n}}(X) &= F_{\overrightarrow{\xi_n}}(x_1, \dots, x_n) = P \{ \xi_k(\omega) < x_k, \forall k = \overline{1, n} \} = \\
 &= \prod_{k=1}^n P \{ \xi_k(\omega) < x_k \} = \prod_{k=1}^n F_\xi(x_k)
 \end{aligned}$$



Байесовские методы.

Введение. Рассуждения и пояснения. Функция правдоподобия.

Пусть $\mathcal{F} = \{F(x; \theta) \mid \theta \in \Theta\}$ – класс допустимых распределений наблюдаемой случайной величины ξ .

$f(x; \theta)$ – плотность наблюдаемой случайной величины ξ .

Пусть $\overrightarrow{X_n(\omega)} = [X_1(\omega), \dots, X_n(\omega)]^T$ – случайная выборка из $\mathcal{L}(\xi) \in \mathcal{F}$ и

$\overrightarrow{x} = [x_1, \dots, x_n]^T$ – реализация этой случайной выборки $\overrightarrow{X_n(\omega)}$.

$L(\overrightarrow{x}; \theta) = f(x_1; \theta) \dots f(x_n; \theta)$ – плотность распределения случайного вектора $\overrightarrow{X_n(\omega)}$.



Байесовские методы.

Введение. Рассуждения и пояснения. Функция правдоподобия.

Пусть $\mathcal{F} = \{F(x; \theta) \mid \theta \in \Theta\}$ – класс допустимых распределений наблюдаемой случайной величины ξ .

$f(x; \theta)$ – плотность наблюдаемой случайной величины ξ .

Пусть $\overrightarrow{X_n(\omega)} = [X_1(\omega), \dots, X_n(\omega)]^T$ – случайная выборка из $\mathcal{L}(\xi) \in \mathcal{F}$ и

$\overrightarrow{x} = [x_1, \dots, x_n]^T$ – реализация этой случайной выборки $\overrightarrow{X_n(\omega)}$.

$L(\overrightarrow{x}; \theta) = f(x_1; \theta) \dots f(x_n; \theta)$ – плотность распределения случайного вектора $\overrightarrow{X_n(\omega)}$.

Определение

Рассматриваемая функция $L(\overrightarrow{x}; \theta)$ при фиксированном \overrightarrow{x} как функция параметра θ называется **функцией правдоподобия**.



Байесовские методы.

Введение. Рассуждения и пояснения. Функция правдоподобия.

Пусть $\mathcal{F} = \{F(x; \theta) \mid \theta \in \Theta\}$ – класс допустимых распределений наблюдаемой случайной величины ξ .

$f(x; \theta)$ – плотность наблюдаемой случайной величины ξ .

Пусть $\overrightarrow{X_n(\omega)} = [X_1(\omega), \dots, X_n(\omega)]^T$ – случайная выборка из $\mathcal{L}(\xi) \in \mathcal{F}$ и

$\overrightarrow{x} = [x_1, \dots, x_n]^T$ – реализация этой случайной выборки $\overrightarrow{X_n(\omega)}$.

$L(\overrightarrow{x}; \theta) = f(x_1; \theta) \dots f(x_n; \theta)$ – плотность распределения случайного вектора $\overrightarrow{X_n(\omega)}$.

Определение

Рассматриваемая функция $L(\overrightarrow{x}; \theta)$ при фиксированном \overrightarrow{x} как функция параметра θ называется **функцией правдоподобия**.

Иногда обозначают $L(\overrightarrow{x}; \theta) = L(\overrightarrow{x} | \theta)$.



Байесовские методы.

Введение. Рассуждения и пояснения. Функция правдоподобия.

Пусть $\mathcal{F} = \{F(x; \theta) \mid \theta \in \Theta\}$ – класс допустимых распределений наблюдаемой случайной величины ξ .

$f(x; \theta)$ – плотность наблюдаемой случайной величины ξ .

Пусть $\overrightarrow{X_n(\omega)} = [X_1(\omega), \dots, X_n(\omega)]^T$ – случайная выборка из $\mathcal{L}(\xi) \in \mathcal{F}$ и

$\overrightarrow{x} = [x_1, \dots, x_n]^T$ – реализация этой случайной выборки $\overrightarrow{X_n(\omega)}$.

$L(\overrightarrow{x}; \theta) = f(x_1; \theta) \dots f(x_n; \theta)$ – плотность распределения случайного вектора $\overrightarrow{X_n(\omega)}$.

Определение

Рассматриваемая функция $L(\overrightarrow{x}; \theta)$ при фиксированном \overrightarrow{x} как функция параметра θ называется **функцией правдоподобия**.

Иногда обозначают $L(\overrightarrow{x}; \theta) = L(\overrightarrow{x} | \theta)$.

Замечание

В ряде задач предположения о независимости и одинаковости распределенности компонент $X_k(\omega)$ не выполняются.

Байесовские методы.

Условная вероятность и теорема Байеса.



Определение

Пусть (Ω, \mathcal{A}, P) – вероятностное пространство некоторого случайного испытания, $A_k \in \mathcal{A}$ – характеристическое множество случайного события α_k , $k = \overline{1, n}$ и $\{\alpha_k\}_{k=1}^n$ – полная группа событий. Если $B \in \mathcal{A}$ – характеристическое множество случайного события β , то имеет место *формула полной вероятности*:

$$P(\beta) = \sum_{k=1}^n P(\beta|\alpha_k)P(\alpha_k).$$



Байесовские методы.

Условная вероятность и теорема Байеса.

Определение

Пусть (Ω, \mathcal{A}, P) – вероятностное пространство некоторого случайного испытания, $A_k \in \mathcal{A}$ – характеристическое множество случайного события α_k , $k = \overline{1, n}$ и $\{\alpha_k\}_{k=1}^n$ – полная группа событий. Если $B \in \mathcal{A}$ – характеристическое множество случайного события β , то имеет место **формула полной вероятности**:

$$P(\beta) = \sum_{k=1}^n P(\beta|\alpha_k)P(\alpha_k).$$

Теорема

Если случайное событие β имеет ненулевую вероятность, то имеет место **формула Байеса**:

$$P(\alpha_k|\beta) = \frac{P(\beta|\alpha_k)P(\alpha_k)}{P(\beta)} = \frac{P(\beta|\alpha_k)P(\alpha_k)}{\sum_{k=1}^n P(\beta|\alpha_k)P(\alpha_k)}.$$

Байесовские методы.

Условная вероятность и теорема Байеса. Адаптация под схему пересмотра вероятностей.



Замечание

Теорема Байеса известна в литературе под названием **принцип обратной вероятности**: в таких задачах на основе данных, имеющихся у исследователя, то есть на основе содержащейся в этих данных информации, пытаются вывести генерировавший эти данные случайный процесс.



Байесовские методы.

Условная вероятность и теорема Байеса. Адаптация под схему пересмотра вероятностей.

Замечание

Теорема Байеса известна в литературе под названием **принцип обратной вероятности**: в таких задачах на основе данных, имеющихся у исследователя, то есть на основе содержащейся в этих данных информации, пытаются вывести генерировавший эти данные случайный процесс.

$$f(\theta|y) = \frac{f(\theta)f(y|\theta)}{f(y)}$$

При практической реализации.

$f(\theta|y) \sim f(\theta)f(y|\theta) \sim$ априорная ФПВ \times функция правдоподобия

1. $1/f(y)$ – нормирующая постоянная.

$$2. 1 \equiv \int_{-\infty}^{\infty} f(\theta|y) d\theta = \frac{1}{f(y)} \int_{-\infty}^{\infty} f(y|\theta)f(\theta) d\theta,$$

$$f(y) = \int_{-\infty}^{\infty} f(y|\theta)f(\theta) d\theta$$



Байесовские методы.

Условная вероятность и теорема Байеса. Пример.

Постановка задачи.

Пусть мы располагаем n независимыми наблюдениями $\vec{y} = [y_1, \dots, y_n]$, которые представляют собой выборку из нормально распределенной генеральной совокупности с неизвестным математическим ожиданием μ и известной дисперсией $\sigma^2 = \sigma_0^2$. **Получить** апостериорную ФПВ для μ .

Имеем

$$f(\mu|\vec{y}, \sigma_0^2) \sim f(\mu)f(\vec{y}|\mu, \sigma_0^2),$$

где $f(\mu|\vec{y}, \sigma_0^2)$ – апостериорная ФПВ для параметра μ при имеющейся выборочной информации \vec{y} и известной дисперсии σ_0^2 .

$f(\mu)$ – априорная ФПВ для μ .

$f(\vec{y}|\mu, \sigma_0^2)$ – функция правдоподобия. Тогда

$$f(\vec{y}|\mu, \sigma_0^2) = \prod_{i=1}^n f(y_i|\mu, \sigma_0^2) = \frac{1}{(2\pi\sigma_0^2)^{n/2}} \exp \left[-\frac{1}{2\sigma_0^2} \sum_{i=1}^n (y_i - \mu)^2 \right]$$



Байесовские методы.

Условная вероятность и теорема Байеса. Пример.

Проведя необходимые преобразования, получаем

$$f(\vec{y}|\mu, \sigma_0^2) = \prod_{i=1}^n f(y_i|\mu, \sigma_0^2) = \\ = \frac{1}{(2\pi\sigma_0^2)^{n/2}} \exp \left[-\frac{1}{2\sigma_0^2} \sum_{i=1}^n [\nu s^2 + n(\mu - \hat{\mu})^2] \right],$$

где $\nu = n - 1$, $\hat{\mu} = \frac{1}{n} \sum_{i=1}^n y_i$ – выборочное среднее, а

$s^2 = \frac{1}{\nu} \sum_{i=1}^n (y_i - \hat{\mu})^2$ – выборочная дисперсия.

Предположим, что

$$f(\mu) = \frac{1}{\sqrt{2\pi}\sigma_a} \exp \left[-\frac{1}{2\sigma_a^2} (\mu - \mu_a)^2 \right],$$

где μ_a – априорное математическое ожидание (параметр, устанавливаемый исследователем на основании первоначальной информации).

σ_a^2 – априорная дисперсия (параметр, устанавливаемый исследователем на основании первоначальной информации).



Байесовские методы.

Условная вероятность и теорема Байеса. Пример.

Применим теорему Байеса

$$f(\mu|\vec{y}, \sigma_0^2) \sim f(\mu)f(\vec{y}|\mu, \sigma_0^2) \sim \exp \left\{ -\frac{1}{2} \left[\frac{(\mu - \mu_a)^2}{\sigma_a^2} + \frac{n}{\sigma_0^2} (\mu - \hat{\mu})^2 \right] \right\} \sim$$

$$\sim \exp \left\{ -\left(\frac{\sigma_a^2 + \sigma_0^2/n}{2\sigma_a^2\sigma_0^2/n} \right) \left(\mu - \frac{\hat{\mu}\sigma_a^2 + \mu_a \frac{\sigma_0^2}{n}}{\sigma_a^2 + \sigma_0^2/n} \right)^2 \right\},$$

Вывод: μ распределено апостериорно нормально с математическим ожиданием и дисперсией равными соответственно:

$$M\mu = \frac{\hat{\mu}\sigma_a^2 + \mu_a \frac{\sigma_0^2}{n}}{\sigma_a^2 + \sigma_0^2/n} = \frac{\hat{\mu}(\sigma_0^2/n)^{-1} + \mu_a(\sigma_a^2)^{-1}}{(\sigma_0^2/n)^{-1} + (\sigma_a^2)^{-1}} = \frac{\hat{\mu}h_0 + \mu_a h_a}{h_0 + h_a}.$$

$$D\mu = \frac{\sigma_a^2\sigma_0^2/n}{\sigma_a^2 + \sigma_0^2/n} = \frac{1}{(\sigma_0^2/n)^{-1} + (\sigma_a^2)^{-1}} = \frac{1}{h_0 + h_a}$$

См. числовой пример в файле

Байесовские методы.

Условная вероятность и теорема Байеса. Несколько массивов данных.



Пусть $f(\theta)$ – априорная ФПВ для вектора параметров θ .

\vec{y}_1 – полученный массив данных с ФПВ $f(\vec{y}_1|\theta)$.

Тогда имеем апостериорную ФПВ:

$$f(\theta|\vec{y}_1) \sim f(\theta)f(\vec{y}_1|\theta).$$

Байесовские методы.

Условная вероятность и теорема Байеса. Несколько массивов данных.



Пусть $f(\theta)$ – априорная ФПВ для вектора параметров θ .

\vec{y}_1 – полученный массив данных с ФПВ $f(\vec{y}_1|\theta)$.

Тогда имеем апостериорную ФПВ:

$$f(\theta|\vec{y}_1) \sim f(\theta)f(\vec{y}_1|\theta).$$

Далее получаем новый массив данных \vec{y}_2 , который был сгенерирован независимо от первого с ФПВ $f(\vec{y}_2|\theta)$.



Байесовские методы.

Условная вероятность и теорема Байеса. Несколько массивов данных.

Пусть $f(\theta)$ – априорная ФПВ для вектора параметров θ .

\vec{y}_1 – полученный массив данных с ФПВ $f(\vec{y}_1|\theta)$.

Тогда имеем апостериорную ФПВ:

$$f(\theta|\vec{y}_1) \sim f(\theta)f(\vec{y}_1|\theta).$$

Далее получаем новый массив данных \vec{y}_2 , который был сгенерирован независимо от первого с ФПВ $f(\vec{y}_2|\theta)$.

Тогда $f(\theta|\vec{y}_1)$ – априорная ФПВ для анализа нового массива данных.

По теореме Байеса получим апостериорную ФПВ:

$$f(\theta|\vec{y}_1, \vec{y}_2) \sim f(\theta|\vec{y}_1)f(\vec{y}_2|\theta)$$

Окончательно

$$f(\theta|\vec{y}_1, \vec{y}_2) \sim f(\theta) \underbrace{f(\vec{y}_1|\theta)f(\vec{y}_2|\theta)}_{\substack{\text{функция} \\ \text{правдоподобия} \\ \text{для } \theta}}.$$