第八章

ベイズ推定:データを元に「確信」を高める手法

パラメトリックモデルの3つのステップ

- (1)パラメーターを含むモデル(数式）を設定する
- (2)パラメーターを評価する基準を定める：
    1，誤差を定義して誤差を最小にするようにパラメーターを決める
    2，最尤推定法：「トレーニングセットが得られる確率」である尤度[ゆうど]関数を定義して、これを最大にするようにパラメーターを決める
- (3)最良の評価を与えるパラメーターを決定する

最尤推定法とベイズ推定の違い

最尤推定法:

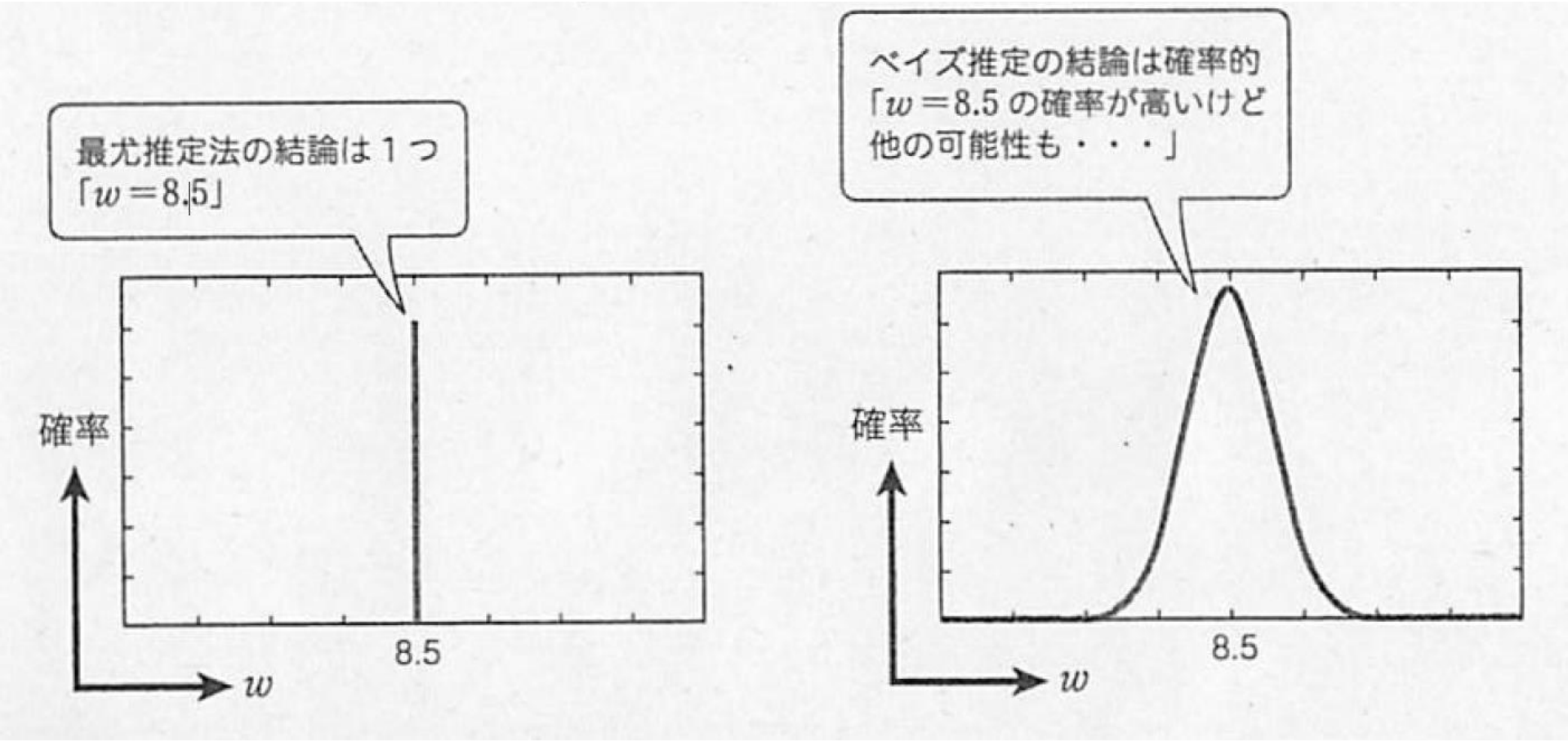
- あるデータが得られる確率P(x)を表す数式を用意する
- 確率P(x)に基づいて「トレーニングセットとして与えられたデータが得られる確率」を計算した上で、これを最大にするという条件でwの値を決定します
- パラメーターwの値は1つ

最尤推定法とベイズ推定の違い

ベイズ推定：

- パラメーターWの値を1つではない
- パラメーターWは、さまざまな値をとる可能性があると考えて、それぞれの値をとる確率を計算します。

最尤推定法とベイズ推定の違い

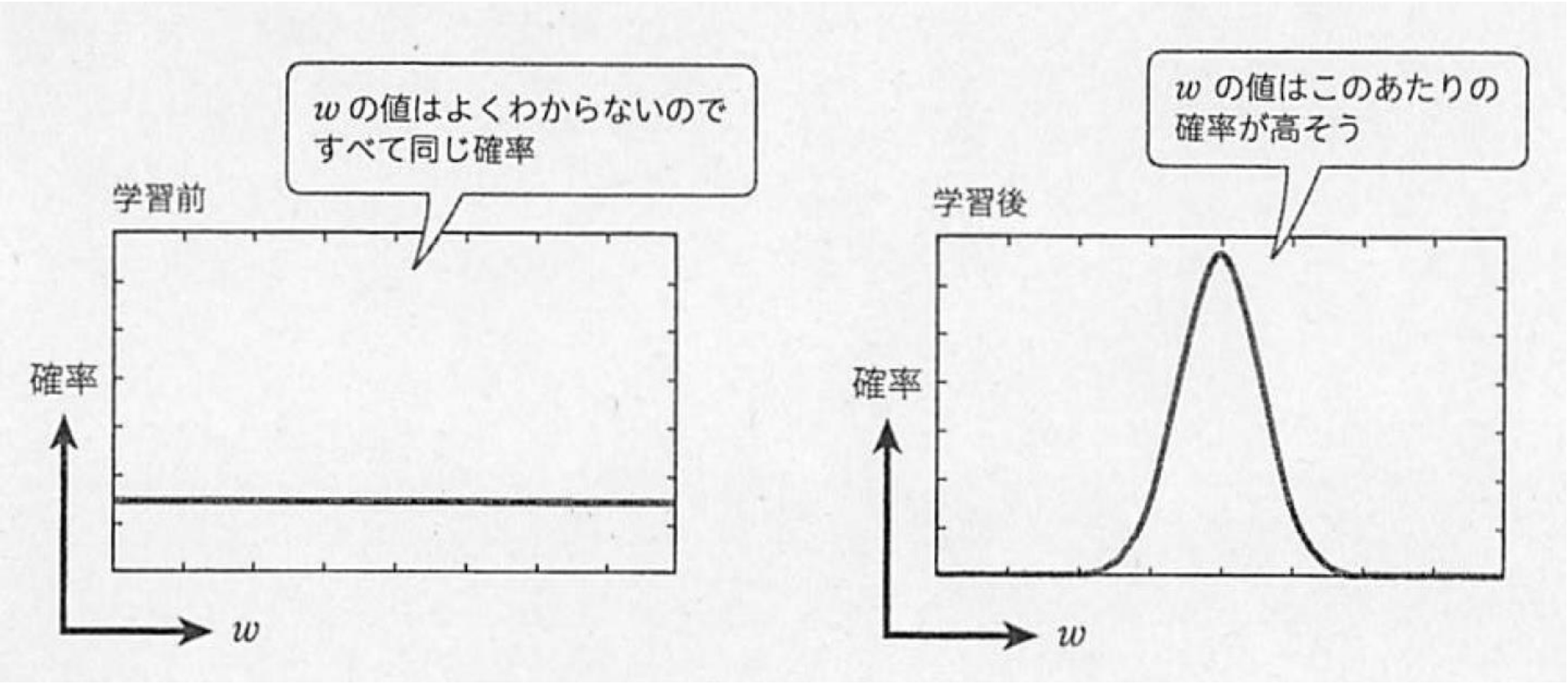


最尤推定法とベイズ推定の違い

ベイズ推定:

- 機械学習を行う前は、パラメーターW の値はいくらなのか、まったく見当がつかないので、すべての値が同じ確率になっています。
- そして、トレーニングセットして与えられたデータに基づいて、機械学習を実施すて、新たに更新された確率が得られます。

最尤推定法とベイズ推定の違い



ベイズの定理入門

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

- P(B) = 事象 A が起きる前の、事象 B の確率（事前確率, prior probability）
- P(B|A) = 事象 A が起きた後での、事象 B の確率（事後確率，条件付き確率, posterior probability, conditional probability）

ベイズの定理入門

$$P(A,B) = P(B|A)P(A) = P(A|B)P(B)$$

$$P(A) = \sum_{B} P(A,B)$$

$$P(A) = \sum_{B} P(A|B)P(B)$$

ベイズの定理入門

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$$= \frac{P(B|A)P(A)}{\sum_A P(B|A)P(A)}$$

ベイズ推定による正規分布の決定：パラメーター推定

正規分布:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (x \in \mathbb{R})$$

N個の観測値$\{t_n\}_{n=1}^{N}$,をトレーニングセットとしま

$$\mathcal{N}(t_n \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(t_n-\mu)^2}$$

ベイズ推定による正規分布の決定：パラメーター推定

$$P(t|\mu) = N(t_1|\mu,\sigma^2) \times \cdots \times N(t_N|\mu,\sigma^2)$$

$$= \prod_{N=1}^{N} N(t_n|\mu,\sigma^2)$$

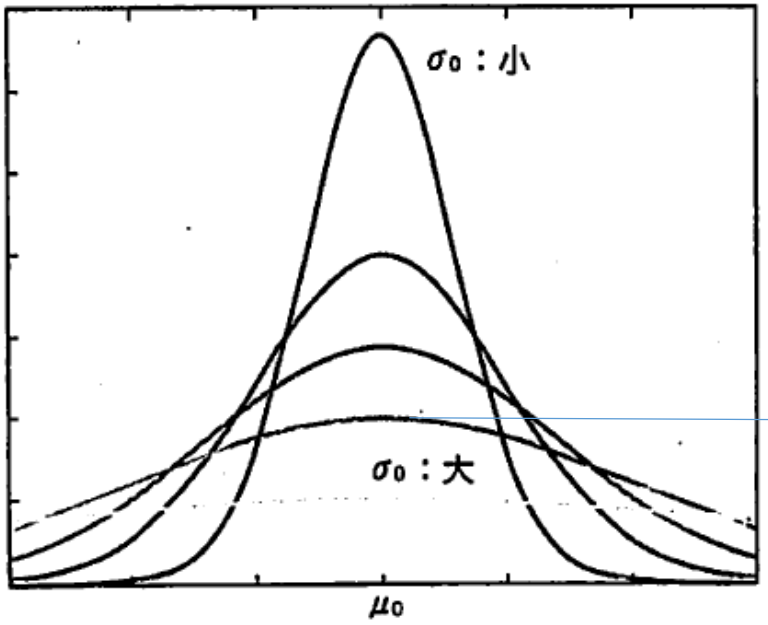
- 観測データtに基づいて、パラメーター狸の確率をP(μ)からP(μ|t)に更新している

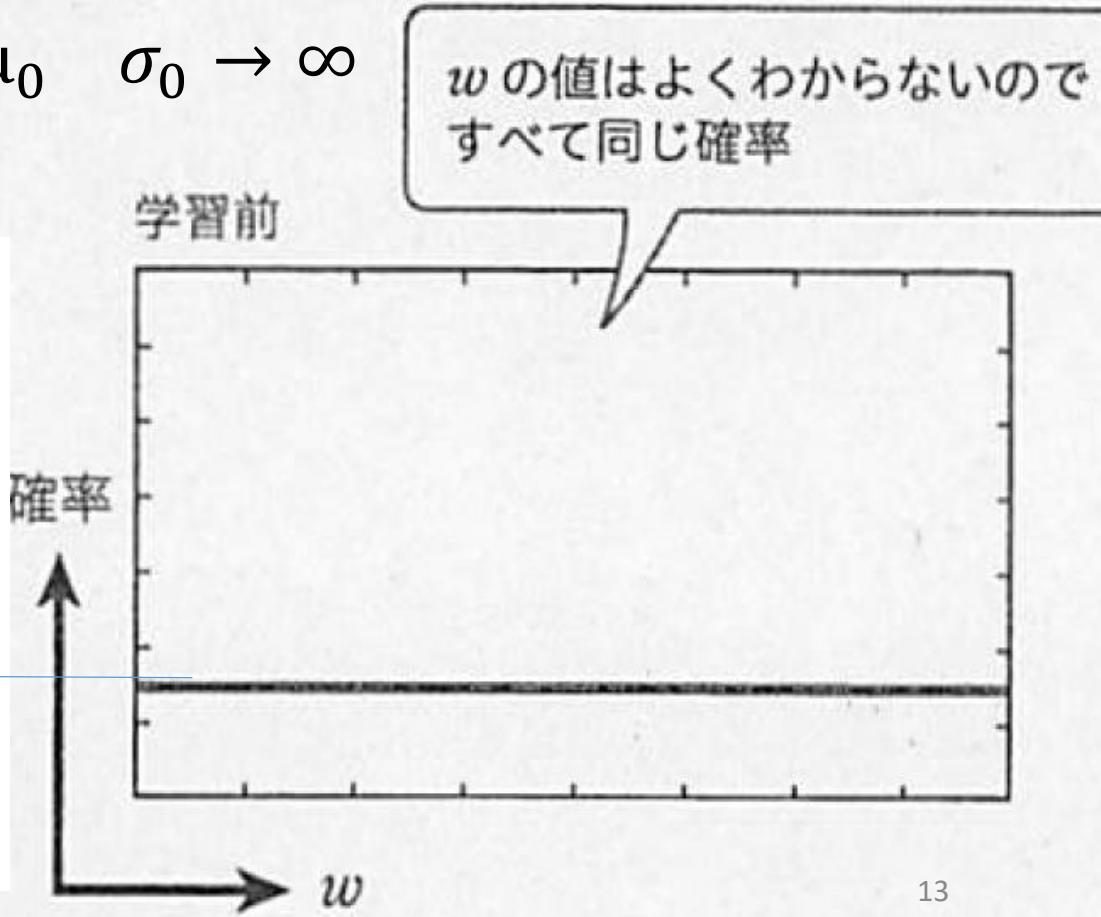
$$P(\mu|\mathrm{t}) = \frac{P(\mathrm{t}|\mu)}{\int_{-\infty}^{\infty} P(\mathrm{t}|\mu')P(\mu')d\mu'} P(\mu)$$

ベイズ推定による正規分布の決定：パラメーター推定

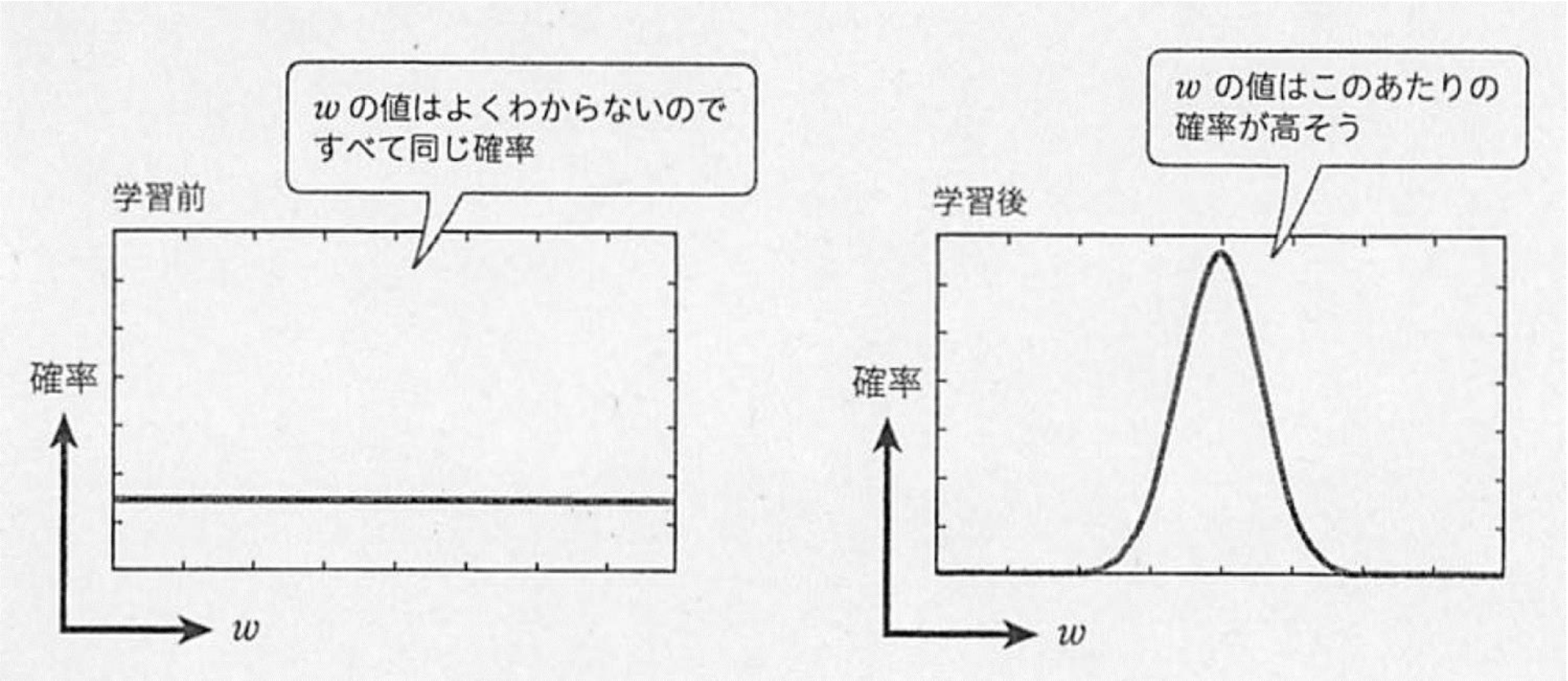
平均μ_0、分散σ_0^2の正規分布を仮定します。

$P(\mu|\mu_0, \sigma_0^2)$　　$\mu=\mu_0$　$\sigma_0 \to \infty$





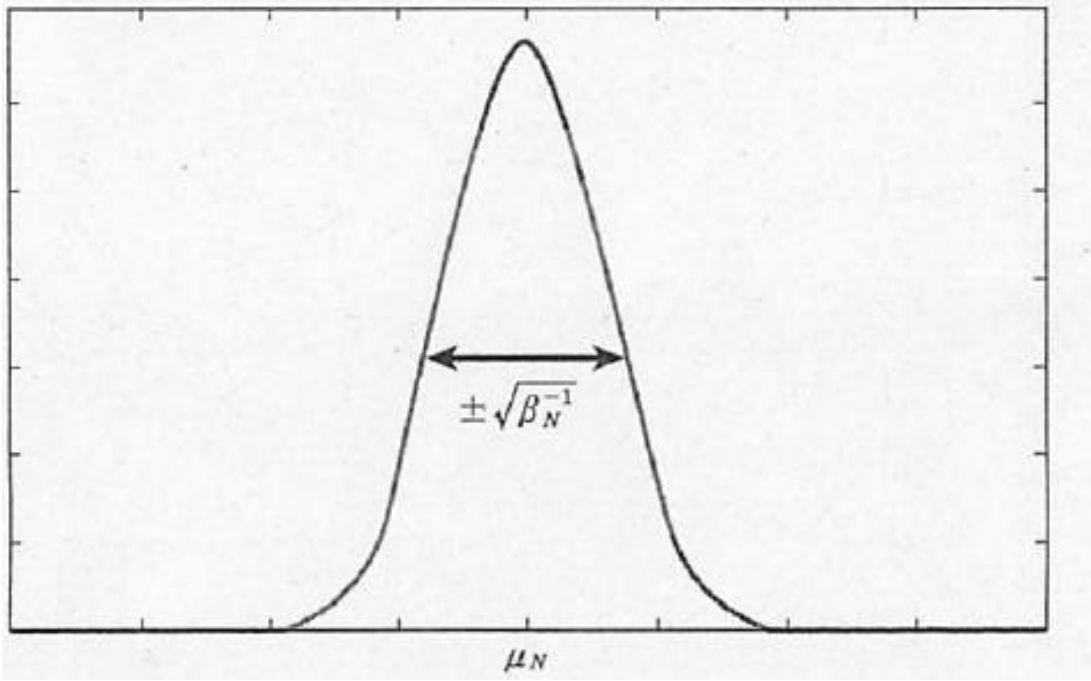
ベイズ推定による正規分布の決定:パラメーター推定



ベイズ推定による正規分布の決定：パラメーター推定

- $P(\mu)$：事前分布
- $P(\mu|t)$：事後分布
- $\beta = \frac{1}{\sigma^2}$ $\beta_0 = \frac{1}{\sigma_0^2}$
- $\beta_N = N\beta + \beta_0$
- $\mu_N = \frac{\beta \sum_{n=1}^{N} t_n + \beta_0 \mu_0}{N\beta + \beta_0}$
- $\mu_N = \frac{1}{N} \sum_{n=1}^{N} t_n$

  （$\beta_0 \to 0$）



ベイズ推定による正規分布の決定；観測値の分布の推定

- さまざまなμに対する正規分布$N(t|\mu, \sigma^2)$をそれぞれの確率$P(\mu|t)$の重みで足し合わせる

- 前提条件を与えて考える範囲を制限すると、確率の値が変化します。この性質を利用して、確率の値を修正します

- $P(t) = N(t|\mu_N, \beta^{-1} + \beta_N^{-1})$

ベイズ推定による正規分布の決定;観測値の分布の推定

- $P(t) = N(t|\mu_N, \beta^{-1} + \beta_N^{-1})$
- 平均μ_Nが、真の平均μと同じである自信がない

- N→∞の極限をとると、分散はβ^{-1}になります
- $P(t) = N(t|\mu_N, \beta^{-1})$
- μ_Nが、真の平均μと同じであると自信を持って言えるので、もともとわかっている分散β-'で、次の観測データを予測できるようになります。

ベイズ推定の回帰分析への応用

- 「データの発生確率」の設定を用いて

$$\mathcal{N}(t \mid f(x_n), \beta^{-1}) = \sqrt{\frac{\beta}{2\pi}} e^{-\frac{\beta}{2}\{t-f(x_n)\}^2}$$

$$f(x) = \sum_{m=0}^{M} w_m x^m$$

- パラメーター　W=$(w_0, \cdots, w_M)^T$
- 観測値　　$t = (t_1, \cdots, t_N)^T$

ベイズ推定の回帰分析への応用

$$P(\mathbf{w}\,|\,\mathbf{t}) = \frac{P(\mathbf{t}\,|\,\mathbf{w})}{\int_{-\infty}^{\infty} P(\mathbf{t}\,|\,\mathbf{w}')P(\mathbf{w}')d\mathbf{w}'}P(\mathbf{w})$$

$$P(\mathbf{w}\,|\,\mathbf{t}) = \frac{1}{Z}P(\mathbf{t}\,|\,\mathbf{w})P(\mathbf{w})$$

$$P(\mathbf{w}\,|\,\mathbf{t}) = \mathrm{Const}\times\exp\left[-\frac{\beta}{2}\sum_{n=1}^{N}\{f(x_n)-t_n\}^2-\frac{\alpha}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w}\right]$$

$$E = \frac{\beta}{2}\sum_{n=1}^{N}\{f(x_n)-t_n\}^2+\frac{\alpha}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w}$$

ベイズ推定の回帰分析への応用

$$E=\frac{\beta}{2}\sum_{n=1}^{N}\{f(x_n)-t_n\}^2+\frac{\alpha}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w}$$

- 最小二乗法では多項式の次数が上がるとオーバーフイッティングが発生する
- 第2項はオーバーフイッティングを抑える
- 抑制の程度は、αの値の設定によって決まります

ベイズ推定の回帰分析への応用

- 事後分布全体の形：

$$P(\mathrm{w}\,|\,\mathrm{t}) = \mathcal{N}\left(\mathrm{w}\,\middle|\,\beta \mathrm{S}\sum_{n=1}^{N} t_n \phi(x_n), \mathrm{S}\right)$$

- 分散sは行列形式(分散共分散行列）になります：

$$\mathrm{S}^{-1} = \alpha \mathrm{I} + \beta \sum_{n=1}^{N} \phi(x_n)\phi(x_n)^{\mathrm{T}}$$

- $\phi(x)$は、xを$0 \sim M$乗した値を並べたベクトル

$$\phi(x) = \begin{pmatrix} x^0 \\ x^1 \\ \vdots \\ x^M \end{pmatrix}$$

観測値の分布の推定

- 「次に観測されるデータの確率」の計算
- これをさまざまなwについて、事後分布P(wlt)の重みで足し合わせます。

$$P(x,t)=\int_{-\infty}^{\infty}P(\mathrm{w}\mid \mathrm{t})\mathcal{N}(t\mid f(x),\beta^{-1})d\mathrm{w}$$

- その結果、P(x,t)は次の正規分布になります。

$$P(x,t)=\mathcal{N}(t\mid m(x),\ s(x))$$

観測値の分布の推定

- 平均m(x)と分散s(x)は、次式で与えられます。

$$m(x) = \beta \boldsymbol{\phi}(x)^{\mathrm{T}} \mathrm{S} \sum_{n=1}^{N} t_n \boldsymbol{\phi}(x_n)$$

$$s(x) = \beta^{-1} + \boldsymbol{\phi}(x)^{\mathrm{T}} \mathrm{S} \boldsymbol{\phi}(x)$$

- 結論：観測点xを決めると、その点の観測データは、平均m(x)、分散s(x)の正規分布に従う