統計的機械学習論
8. ベイズ推定
〜データを元に「確信」を高める手法〜

夏季集中講義(2019/09/02-13)

@九州工業大学

(飯塚キャンパス 大学院セミナー室７F)

このクラスのねらいと達成目標

ねらい

ベイズの定理の考え方を理解する。

達成目標

- 一般的な確率計算での確率の記述の仕方に慣れよう
- ベイズの定理を理解しよう
- 正規分布のパラメータをベイズ推定する過程を知ろう
- 正規分布のパラメータが１つに定まらない時の観測値の求め方を知ろう

クラスの進行

8 ベイズ推定

パラメトリックモデルの３つのステップ

① パラメータを含む数式モデルを設定する

② パラメータを評価する基準を定める

③ 最良の評価を与えるパラメータを決定する

②の評価基準では「誤差を最小にするパラメータを決定する方法」、「データセットが得られる確率を最大にするパラメータを決定する方法」が今まで使われた。

ここでは「パラメータがそれぞれの値をとる確率」という視点からアプローチしていく。

8.1 ベイズ推定モデルとベイズの定理

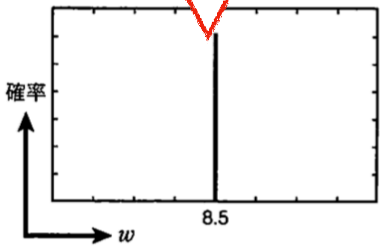
8.1.1 ベイズ推定の考え方

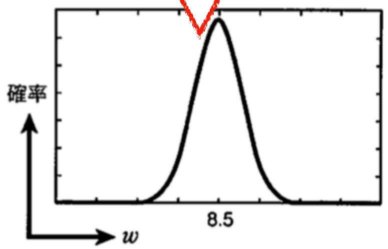
最尤推定法とベイズ推定の考え方の違いを知ろう

パラメータが確率的に予測されること

最尤推定法の結論は
1つ

ベイズ推定の結論は
確率的



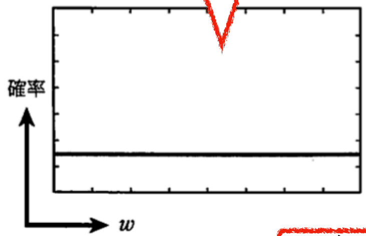


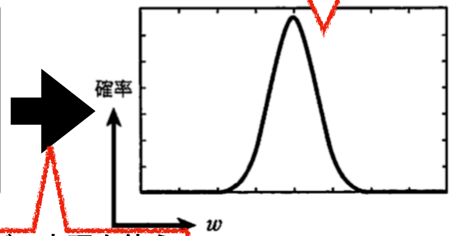
8.1.1 ベイズ推定の考え方

最尤推定法とベイズ推定の考え方の違いを知ろう

学習前：wの値はよくわか
らないので、全て同じ確率

学習後：wの値はこのあた
りの確率が高そう





ベイズの定理を使う

8.1.2 ベイズの定理入門

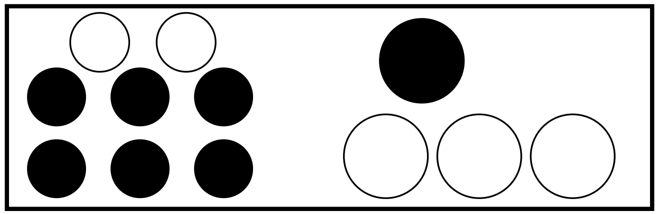
一般的な確率の計算方法について復習しよう

前提条件がない状況で事象Yが起こる確率$P(Y)$

ある条件Xを前提として事象Yが起こる確率$P(Y \mid X)$

例題

ある箱の中に大小・白黒のボールが入っている

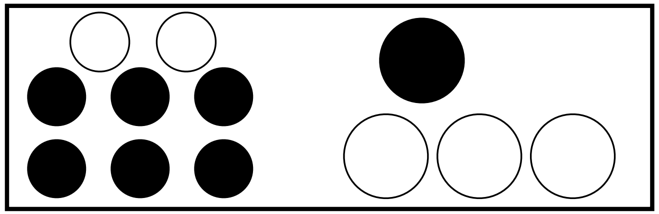


8.1.2 ベイズの定理入門

例題

大小・白黒のボールが入っている箱からボールを取り出す

1. 出たボールが「黒」である確率
2. 出たボールが「大きい」と分かっている時にそのボールが「黒」である確率
3. 出たボールが「大きい黒」である確率



8.1.2 ベイズの定理入門

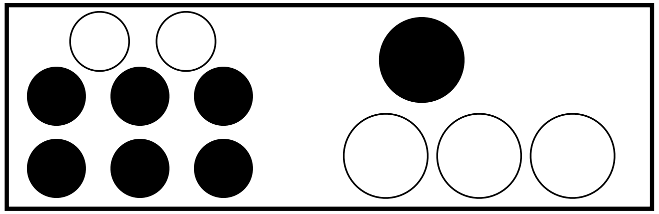
例題

ある箱の中に大小・白黒のボールが入っている

1. 出たボールが「黒」である確率

   まず、全体として12個あり、黒いボールは7個。

$$P(\text{Black}) = \frac{7}{12}$$



8.1.2 ベイズの定理入門

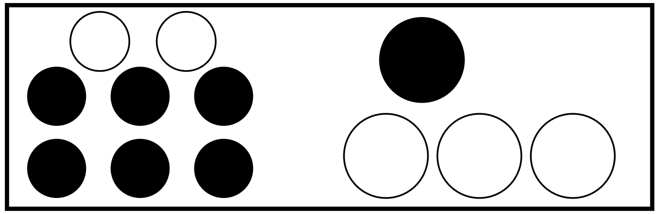
例題

ある箱の中に大小・白黒のボールが入っている

2. 出たボールが「大きい」と分かっている時にそのボールが「黒」である確率

   大きいボールは全部で4個、黒いボールはそのうち1個

条件付き確率 $P(\text{Black} \mid \text{Large}) = \frac{1}{4}$



8.1.2 ベイズの定理入門

例題

ある箱の中に大小・白黒のボールが入っている

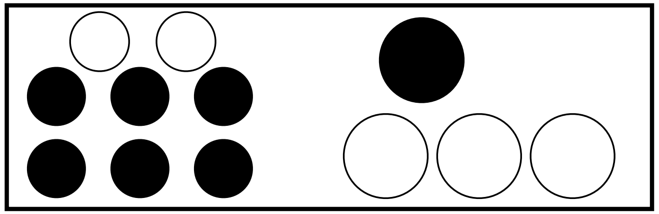
3. 出たボールが「大きい黒」である確率

   「大きく」かつ「黒い」ボールは1個

条件付き確率ではない

同時確率

$$P(\text{Black}\,,\,\text{Large}) = \frac{1}{12}$$



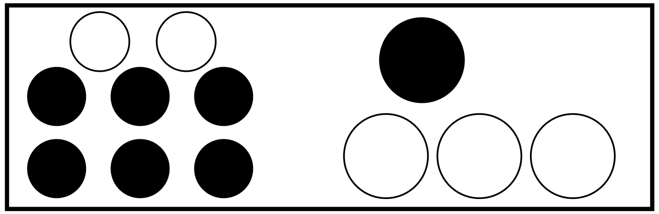
8.1.2 ベイズの定理入門

$P(\text{Black} \mid \text{Large}) =$ **「大の中で黒」の数÷「大」の数 =1/4**

$P(\text{Black}\,,\,\text{Large}) =$ **「黒かつ大」の数÷「全体の数」=1/12**

$\frac{P(\text{Black}\,,\,\text{Large})}{P(\text{Black} \mid \text{Large})} =$ **「大の数」÷「全体の数」** $= P(\text{Large})$

$P(\text{Black}\,,\,\text{Large}) = P(\text{Black} \mid \text{Large})P(\text{Large})$=(1/4)*(4/12)



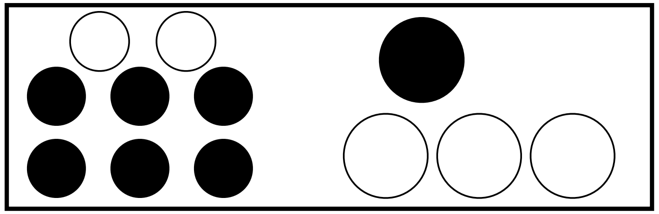
8.1.2 ベイズの定理入門

$P(\mathrm{Black}\,,\,\mathrm{Large}) = P(\mathrm{Black}\,|\,\mathrm{Large})P(\mathrm{Large})$=(1/4)*(4/12)

$P(\mathrm{Black}\,,\,\mathrm{Large}) = P(\mathrm{Large}\,|\,\mathrm{Black})P(\mathrm{Black})$=(1/7) * (7/12)

これらの式を一般化すると、

$$P(X\,,\,Y) = P(X\,|\,Y)P(Y) = P(Y\,|\,X)P(X)$$



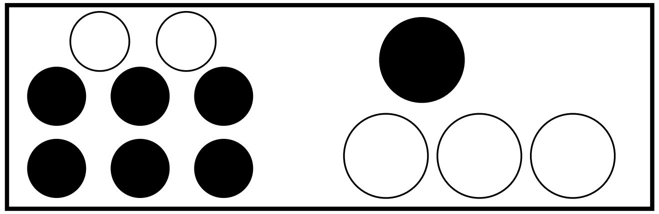
8.1.2 ベイズの定理入門

$$P(X\,,\,Y) = P(X\,|\,Y)P(Y) = P(Y\,|\,X)P(X)$$

⬇

$$P(Y\,|\,X) = \frac{P(X\,|\,Y)P(Y)}{P(X)}$$

「ベイズの定理」



8.1.2 ベイズの定理入門

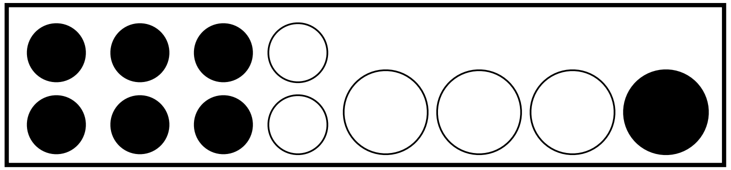
もう１つの公式

「黒かつ大」の確率＋「黒かつ小」の確率＝「黒」の確率

$$P(\mathrm{Black}\,,\,\mathrm{Large}) + P(\mathrm{Black}\,,\,\mathrm{Small}) = P(\mathrm{Black})$$

⬇

$$P(X) = \sum_{Y} P(X, Y)$$



8.1.2 ベイズの定理入門

$$P(X) = \sum_{Y} P(X, Y)$$

$$P(X\,,\,Y) = P(X\,|\,Y)P(Y) = P(Y\,|\,X)P(X)$$

⬇

$$P(X) = \sum_{Y} P(X\,|\,Y)P(Y)$$

「周辺確率の公式」

ベイズの定理に周辺確率の公式を代入すると

$$P(Y\,|\,X) = \frac{P(X\,|\,Y)}{\sum_{Y'} P(X\,|\,Y')P(Y')}P(Y)$$

8.1.2 ベイズの定理入門

Xである時のYの確率

Xである時のYの確率

$$P(Y\,|\,X) = \frac{P(X\,|\,Y)}{\sum_{Y'} P(X\,|\,Y')P(Y')}P(Y)$$

「条件と結果」を入れ替えた関係を計算するのが、
ベイズの定理の特徴になる。

8.1.2 ベイズの定理入門

例題からベイズの定理について理解する

「太郎さんはピロリ菌に感染している恐れがあるので、ピロリ菌検査を受けました。一般に、太郎さんの年代の人がピロリ菌に感染している確率（割合）は1%です。また、ピロリ菌検査の精度は95%です。

そして、太郎さんの検査結果は「陽性反応」でした。ただし、検査結果が間違っている可能性もあります。太郎さんが本当にピロリ菌に感染している確率を求めなさい。」

8.1.2 ベイズの定理入門

例題からベイズの定理について理解する

X：検査の結果陽性反応がでる,　Y：ピロリ菌に感染している

P(陽性反応|感染)：
感染している人に「陽性反応」がでる確率95%

P(陰性反応|感染)：
感染している人に「陰性反応」がでる確率は5%

P(陰性反応|非感染)：
感染していない人に「陰性反応」がでる確率は95%

P(陽性反応|非感染)：
感染していない人に「陽性反応」がでる確率は5%

8.1.2 ベイズの定理入門

例題からベイズの定理について理解する

太郎さんがピロリ菌に感染している確率について

検査を受ける前なら：1%

P(感染)=0.01

P(非感染)=0.99

検査の結果陽性反応が出た後の感染確率は

$$\mathrm{P}(感染|陽性反応) = \frac{\mathrm{P}(陽性反応|感染)}{(\mathrm{P}(陽性反応|感染)\mathrm{P}(感染)+\mathrm{P}(陽性反応|非感染)\mathrm{P}(非感染))}\mathrm{P}(感染)$$

$$P(Y\,|\,X) = \frac{P(X\,|\,Y)P(Y)}{\sum_{Y'} P(X\,|\,Y')P(Y')}P(Y)$$

8.1.2 ベイズの定理入門

例題からベイズの定理について理解する

$$P(\text{感染}\,|\,\text{陽性反応}) = \frac{P(\text{陽性反応}\,|\,\text{感染})}{(P(\text{陽性反応}\,|\,\text{感染})P(\text{感染})+P(\text{陽性反応}\,|\,\text{非感染})P(\text{非感染}))}P(\text{感染})$$

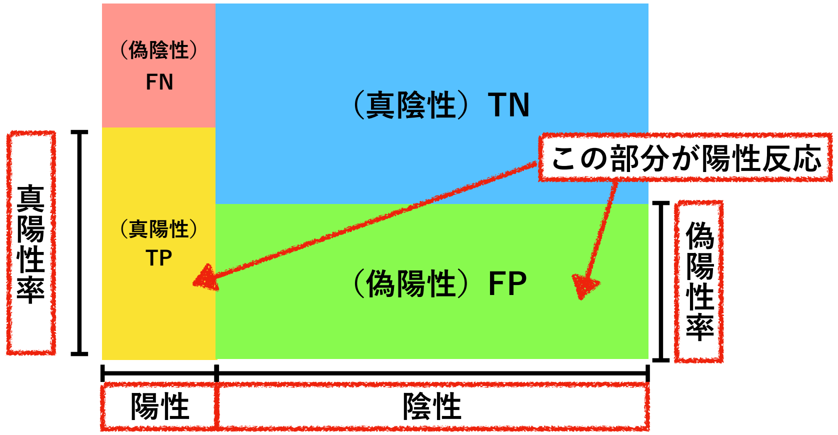
$$\begin{aligned} P(Y\,|\,X) &= \frac{0.95}{0.95\times 0.01 + 0.05\times 0.99}\times 0.01 \\ &\fallingdotseq 0.16 \end{aligned}$$

太郎さんが感染している確率は16%

8.1.2 ベイズの定理入門

例題からベイズの定理について理解する

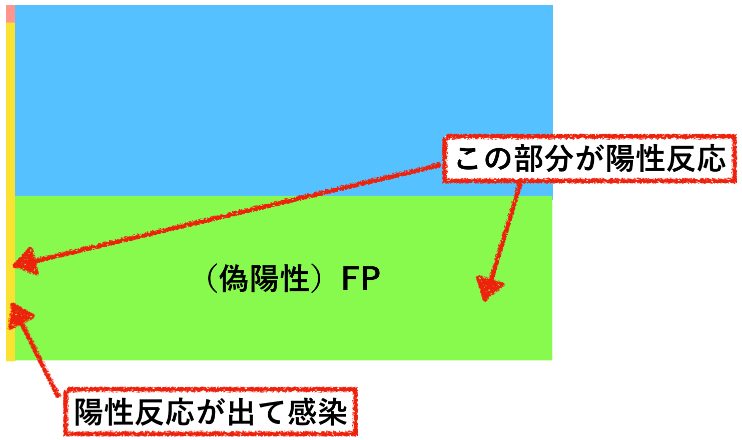
太郎さんが感染している確率は16%→検査精度95%なのに低い？



8.1.2 ベイズの定理入門

例題からベイズの定理について理解する

太郎さんが陽性である確率＝真陽性÷（真陽性＋偽陽性）



8.1.3 ベイズ推定による正規分布の決定: パラメータ推定

正規分布の平均の推定にベイズを適用する。

$\{t_n\}_{n=1}^{N}$：トレーニングセット→ベクトル $\mathbf{t}$

仮に平均が分かっている場合、ある特定のt_n が得られる確率は

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

トレーニングセット全体が観測される確率は

$$\begin{aligned} P(\mathbf{t}\,|\,\mu) &= N(t_1\,|\,\mu,\sigma^2)\times \ldots \times N(t_N\,|\,\mu,\sigma^2) \\ &= \prod_{n=1}^{N} N(t_n\,|\,\mu,\sigma^2) \end{aligned}$$

8.1.3 ベイズ推定による正規分布の決定: パラメータ推定

$$P(\mathbf{t}\,|\,\mu) = N(t_1\,|\,\mu,\sigma^2)\times \ldots \times N(t_N\,|\,\mu,\sigma^2) = \prod_{n=1}^{N} N(t_n\,|\,\mu,\sigma^2)$$

ベイズの定理に当てはめると、

$$P(Y\,|\,X) = \frac{P(X\,|\,Y)}{\sum_{Y'} P(X\,|\,Y')P(Y')} P(Y)$$

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

μは連続的に変化しているため、積分を用いる

8.1.3 ベイズ推定による正規分布の決定: パラメータ推定

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

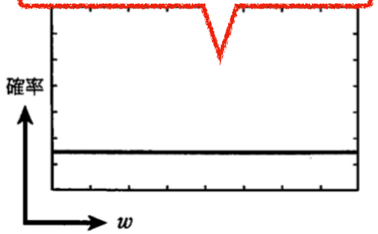
連続変数μに対して一旦平均と分散を仮定する

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

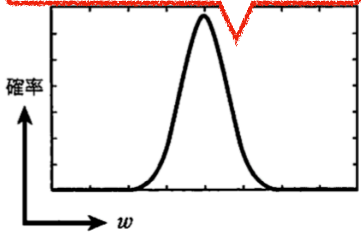
平均を中心として標準偏差の幅に広がる確率

8.1.3 ベイズ推定による正規分布の決定

学習前：wの値はよくわからないので、全て同じ確率

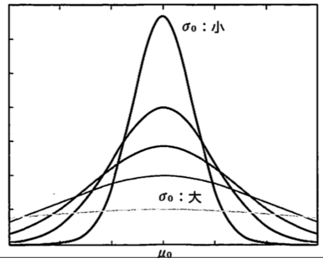


学習後：wの値はこのあたりの確率が高そう



平均を中心として標準偏差の幅に広がる確率

$\sigma_0\to\infty$の極限をとり初期状態のフラットな確率分布を再現



8.1.3 ベイズ推定による正規分布の決定: パラメータ推定

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

事後分布　**事前分布**

事前分布：トレーニングセットを観測する前の分布

事後分布：トレーニングセットを観測した後の分布

8.1.3 ベイズ推定による正規分布の決定: パラメータ推定

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

ここで、複数の正規分布を掛け合わせたものは再び正規分布になる性質を利用すると

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

$$\beta = \frac{1}{\sigma^2}$$

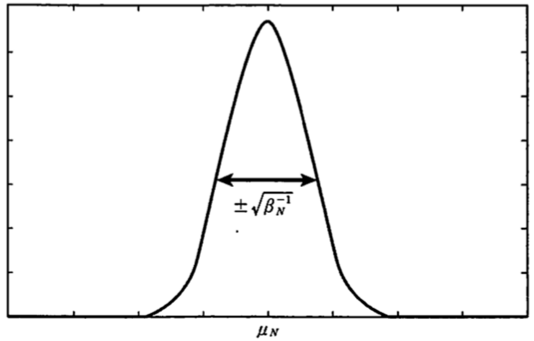
$$\beta_0 = \frac{1}{\sigma_0^2}$$

$$\beta_N = N\beta + \beta_0$$

$$\mu_N = \frac{\beta \sum_{n=1}^{N} t_n + \beta_0\mu_0}{N\beta + \beta_0}$$

8.1.3 ベイズ推定による正規分布の決定: パラメータ推定

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



事後分布

8.1.3 ベイズ推定による正規分布の決定

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

$\sigma_0 \to \infty$より、$\beta_0 \to 0$の極限に対応

$$\mu_N = \frac{\beta \sum_{n=1}^{N} t_n + \beta_0\mu_0}{N\beta + \beta_0}$$

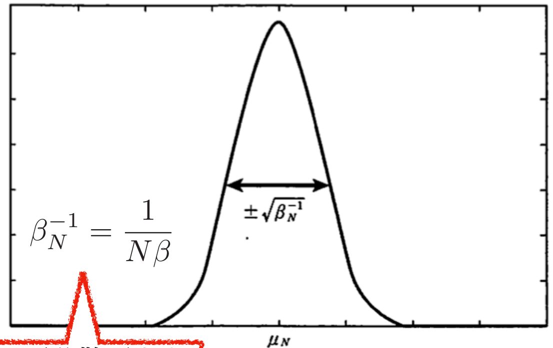
について極限$\beta_0 \to 0$をとると、

$$\mu_N = \frac{1}{N}\sum_{n=1}^{N} t_n$$

トレーニングセットの標本平均と等価

8.1.3 ベイズ推定による正規分布の決定: パラメータ推定

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



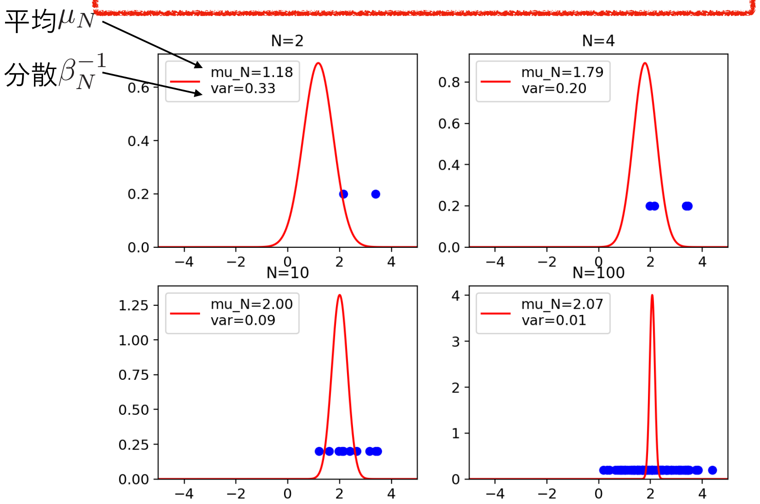
$$\beta_N^{-1} = \frac{1}{N\beta}$$

Nが増えるほど分散が小さくなる

事後分布

8.1.3 サンプルコードによる確認

平均μの事後分布P(μ | t)のグラフ



8.1.4 ベイズ推定による正規分布の決定: 観測値の分布の推定

本当に知りたいのは次に得られる観測値tの値

平均と分散が分かっている場合、観測値t_n が得られる確率

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

今の状況では平均μは１つに定まっていないので

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

8.1.4 ベイズ推定による正規分布の決定: 観測値の分布の推定

次に観測されるデータがtである確率

$$P(t) = \int_{-\infty}^{\infty} P(\mu \,|\, \mathbf{t}) N(t \,|\, \mu, \sigma^2) d\mu$$

ここで

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

を利用すると

$$P(t) = \int_{-\infty}^{\infty} N(\mu \,|\, \mu_N, \beta_N^{-1}) N(t \,|\, \mu, \beta^{-1}) d\mu$$

8.1.4 ベイズ推定による正規分布の決定: 観測値の分布の推定

$$P(t) = \int_{-\infty}^{\infty} N(\mu \,|\, \mu_N, \beta_N^{-1}) N(t \,|\, \mu, \beta^{-1}) d\mu$$

積分が実行されると、結果は再び正規分布になる。

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

元々の分散はβ^{-1}のはずだが事後分布の分散β_N^{-1}だけ
広がった分散で推定している。

8.1.4 ベイズ推定による正規分布の決定: 観測値の分布の推定

$$\beta_N^{-1} = \frac{1}{N\beta}$$

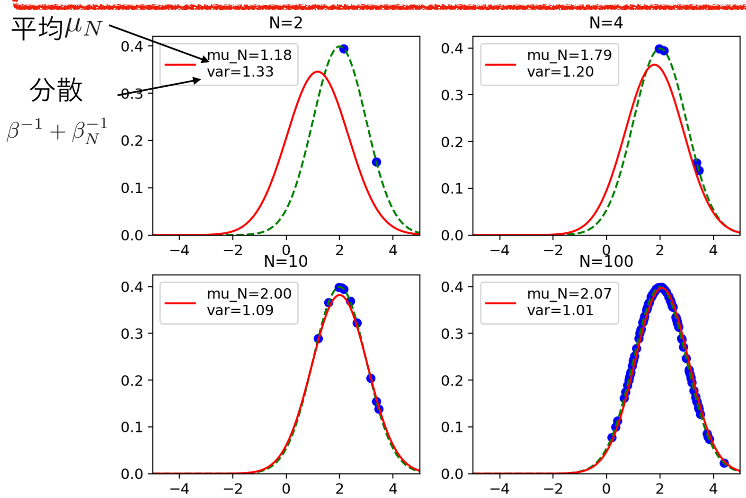
より、分散はNが増えるほど小さくなっていく。
N→∞の極限をとると

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

トレーニングセットのデータ数Nを大きくしていくと「確信度」が高くなりμ_Nが真の平均μと同じであると考えることができる。

8.1.5 サンプルコードによる確認

事後分布P(μ | t)に基づいて次に得られる観測値の確率P(t)



まとめ

- **一般的な確率計算での例題を元にベイズを理解した**
- **モデルのパラメータをベイズ推定する手続きを知った**
- **正規分布のパラメータが１つに定まらない時に観測値を求めるための手続きを知った**

ミニッツペーパー

- **身近な話題の中でベイズの定理を適用できる問題があれば紹介してください。**