2.1.1　ベータ分布

茨城大学工学部情報工学科
05T4072R
相原　功昌

ベータ分布

- 式2.13で表される

$$Beta(\mu \mid a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \mu^{a-1}(1-\mu)^{b-1} \quad \cdots (2.13)$$

- 正規化されている→式2.14

$$\int_0^1 Beta(\mu \mid a,b)\, d\mu = 1 \quad \cdots (2.14)$$

ガンマ関数

- 式1.141で定義される

$$\Gamma(x) \equiv \int_0^{\infty} u^{x-1} e^{-u} du \qquad \cdots (1.141)$$

ベータ分布の必要性

- ベルヌーイ分布と二項分布のパラメータμの最尤推定（式2.8）

→データ集合が小さいと過学習してしまう

⇩

ベイズ主義的に扱う

⇩

μの事前分布p(μ)を導入

共役性

尤度関数: $\mu^x(1-\mu)^{1-x}$

事前分布が、μと$(1-\mu)$のべき乗に比例するように選択

事後分布は、事前分布と尤度関数の積に比例

→事後分布は事前分布と同じ関数形式

ベータ分布の平均と分散

- 平均→式2.15

$$E[\mu] = \frac{a}{a+b} \qquad \cdots (2.15)$$

- 分散→式2.16

$$\mathrm{var}[\mu] = \frac{ab}{(a+b)^2(a+b+1)} \qquad \cdots (2.16)$$

超パラメータ

- パラメータaとbはパラメータμの分布を決める

→超パラメータ（hyperparameter）と呼ばれることがある

ベータ分布のグラフ

- ベータ分布のグラフを図2.2に示す

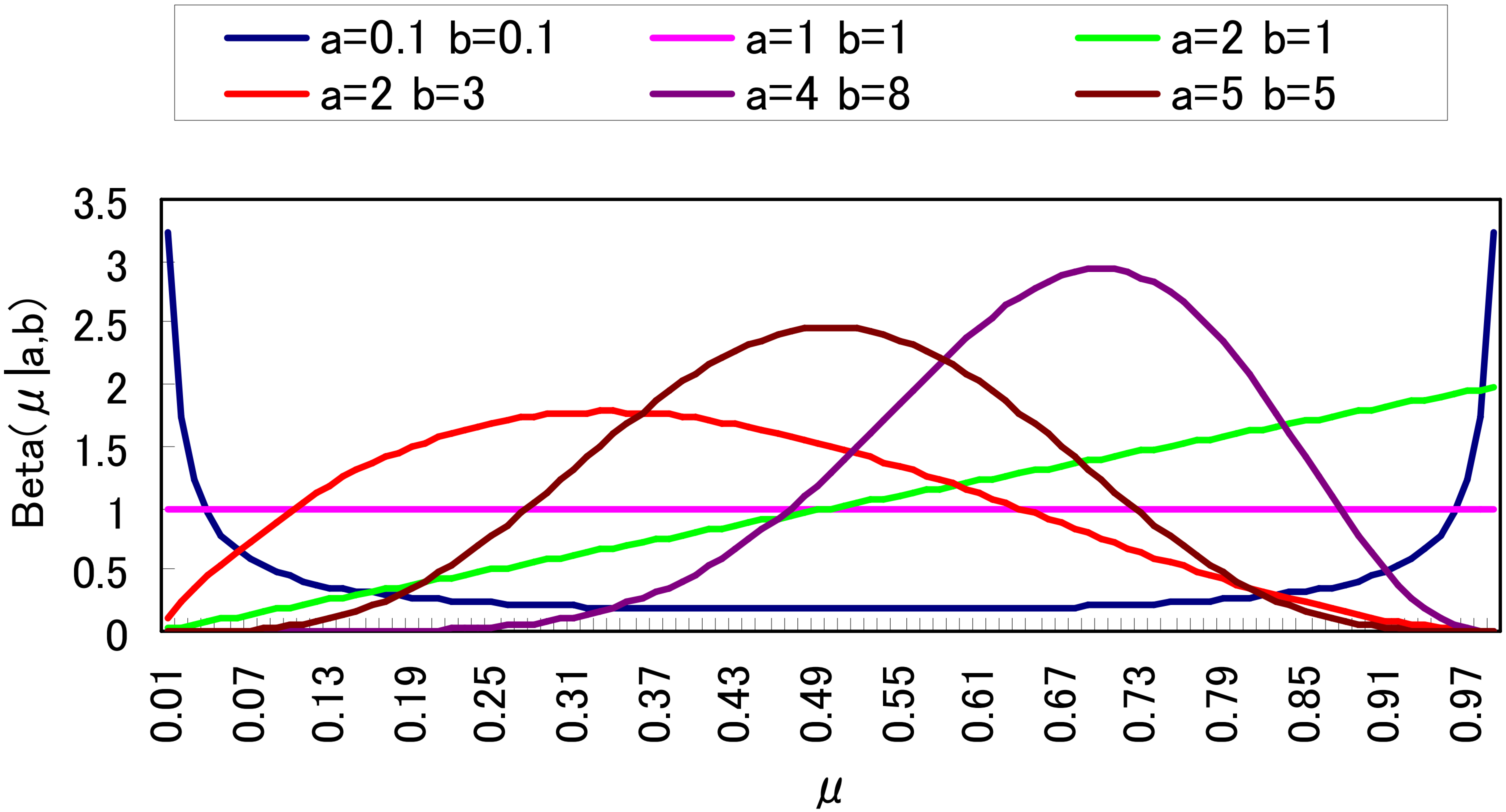

図2.2 ベータ分布

有効観測数(１)

- μ の事後分布

＝事前分布(式2.13)×二項尤度関数(式2.9)

- μ に依存する要素を取り出す

→式2.17

$$p(\mu \mid m, l, a, b) \propto \mu^{m+a-1}(1-\mu)^{l+b-1} \quad \cdots (2.17)$$

- l=N-m はコイン投げの例の「裏」の数に相当

有効観測数(２)

- 式2.17をベータ分布として完結に表す

  →式2.18

$$p(\mu \mid m,l,a,b) = \frac{\Gamma(m+a+l+b)}{\Gamma(m+a)\Gamma(l+b)} \mu^{m+a-1}(1-\mu)^{l+b-1} \quad \cdots(2.18)$$

- x=1となる観測値をｍ個

  x=0となる観測値を l 個　となるデータ集合を考える

有効観測数(3)

- 事前分布から、この集合を観測した後の事後分布を求める

  →aの値をm、bの値をｌだけ増やせばいい
- 超パラメータaとb

  →x=1とx=0の有効観測数として解釈

逐次学習（1）

（１）事前分布を得る
（２）値を観測する
（３）事後分布を得る
（４）（３）で得た事後分布を事前分布とみなして、（２）～（４）を繰り返す

逐次学習（２）

- データが独立同分布に従えば成り立つ
- 全てのデータが届く前に予測をしなければならない実時間での学習などに利用
- 大規模データ集合についても有用

正確な予測（１）

- できるだけ正確に、次の試行の出力を予測する
→観測データ集合Dが与えられた時のxの予測分布を考える　→式2.19

$$
\begin{aligned}
p(x=1|D) &= \int_0^1 p(x=1|\mu)\,p(\mu|D)\,d\mu \\
&= \int_0^1 \mu\, p(\mu|D)\,d\mu \\
&= E[\mu|D] \qquad \cdots (2.19)
\end{aligned}
$$

正確な予測（２）

- 式2.18と式2.15から、式2.20を得る

$$p(x=1|D)=\frac{m+a}{m+a+l+b} \quad \cdots(2.20)$$

- 観測値の、x＝１に相当するものの、総数に対する割合
- m,l→∞の極限では、式2.20の結果は最尤推定の結果と等しくなる

ベイズ学習の特性

- 多くのデータを観測するほど、事後分布が示す不確実性は恒常的に減少する

→図2.2から、観測値の数が増えるに従って、事後分布のピークが鋭くなる