二項分布

正田 備也

masada@rikkyo.ac.jp

Contents

確率分布とは

- 確率分布とは、確率変数について、各々の値をとる確率を表したもの
- 離散確率分布とは、確率変数が離散的な値をとる場合の確率分布
- 連続確率分布とは、確率変数が連続的な値をとる場合の確率分布

離散確率分布

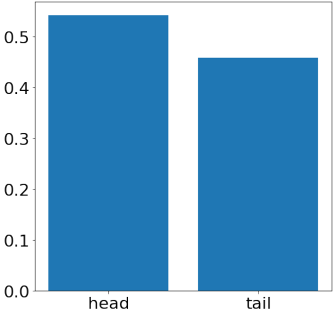
- 離散確率分布は、以下の三つにより特徴付けられる

1. パラメータ
   - 確率分布を特徴付ける数
2. 台
   - 確率変数のとる値の集合
3. 確率質量関数
   - 離散確率変数に、その値をとる確率を対応させる関数

ベルヌーイ分布 Bernoulli distribution

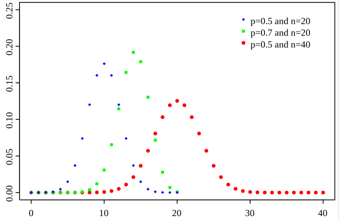
- $V = \{\mathrm{v}_1, \mathrm{v}_2\}$ という 2 種類のアイテムの集合上に定義される確率分布
  - 例えば、コインの表が v_1 でコインの裏が v_2 など
- パラメータは $\boldsymbol{\phi} = (\phi_1, \phi_2)$
  - アイテム v_1 が出現する確率 ϕ_1
  - アイテム v_2 が出現する確率 ϕ_2
  - $\phi_1 + \phi_2 = 1$ が成り立つので、自由度は 1

（右図は確率質量関数の例）



二項分布 binomial distribution

- 複数回のコイン投げのモデリングには二項分布を使う
  - ベルヌーイ分布は 1 回のコイン投げのモデリングに使う
- 試行回数を n として、“n 回のうち v_1 が k 回出現する確率がこれこれ” というふうに、$k=0$ から $k=n$ までの、ありうる出現回数すべてに、確率を割り振る確率分布
- パラメータは n と $\boldsymbol{\phi}=(\phi_1,\phi_2)$
  - 試行の回数 n（これは観測データから決まる）
  - アイテム v_1 が出現する確率 ϕ_1
  - アイテム v_2 が出現する確率 ϕ_2
  - $\phi_1+\phi_2=1$ が成り立つので、自由度は 1



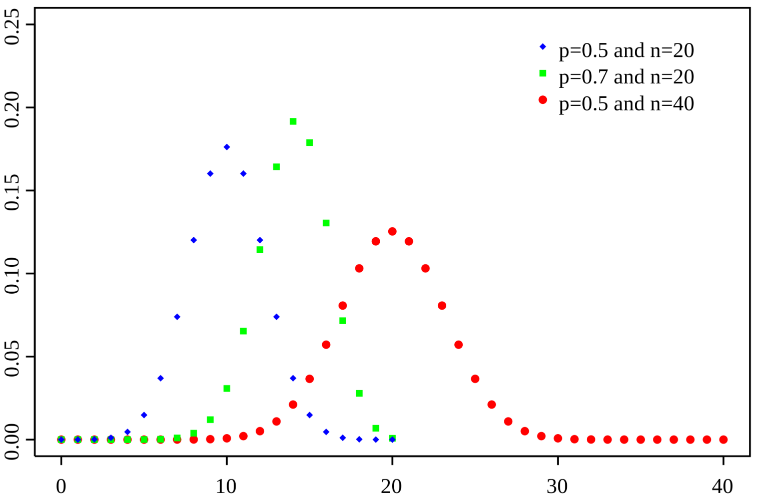


Figure: 二項分布の確率質量関数の例

例. 5回のコイン投げ (1/3)

- 5回のコイン投げを二項分布でモデル化する
- コインの表と裏が出る確率は同じ 1/2 だとする
- 二項分布は以下の 6 種類の事象の集合の上に定義される
  - 5 回のうち表が 0 回出る
  - 5 回のうち表が 1 回出る
  - 5 回のうち表が 2 回出る
  - 5 回のうち表が 3 回出る
  - 5 回のうち表が 4 回出る
  - 5 回のうち表が 5 回出る

例. 5回のコイン投げ (2/3)

- 例えば、5回のうち表が2回出る場合は何通りあるか？

11000　10001　01001　00011

10100　01100　00110

10010　01010　00101

- 答えは、$\frac{5!}{2!3!} = 10$ 通り
- そして、個々の列が生起する確率は $\phi_1^2 \phi_2^3 = 1/32$
- よって、5回のうち表が2回出る確率は $10 \times 1/32 = 10/32$

例. 5回のコイン投げ (3/3)

- コインの表裏は同確率で出ると仮定している
- よって、どんな裏表の列も、同じ $1/32$ の確率で生起する
- すると、6種類の事象のそれぞれの生起確率は、以下のようになる
  - 5回のうち表が0回出る確率は $\frac{1}{32}$
  - 5回のうち表が1回出る確率は $\frac{5}{32}$
  - 5回のうち表が2回出る確率は $\frac{10}{32}$ ←前のスライドを参照
  - 5回のうち表が3回出る確率は $\frac{10}{32}$
  - 5回のうち表が4回出る確率は $\frac{5}{32}$
  - 5回のうち表が5回出る確率は $\frac{1}{32}$

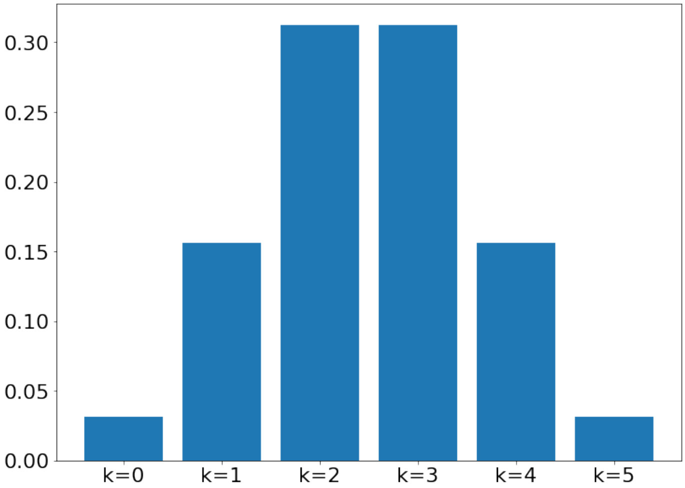


Figure: 5 回のコイン投げに対応する二項分布の確率質量関数（表裏は同確率）

二項分布の確率質量関数

- 確率質量関数 probability mass function; pmf
  - 離散確率変数に、その値をとる確率を対応させる関数
  - 二項分布の場合は、表が出る回数にその確率を対応させる
- 二項分布の確率質量関数
  - n 回の試行のうち v_1 が k 回出現する確率は：

$$p(k; \boldsymbol{\phi}, n) = \frac{n!}{k!(n-k)!}\phi_1^k \phi_2^{n-k} \tag{1}$$

  - 「;」は、その右側にある ϕ と n が自由パラメータ（我々が値を定める必要があるパラメータ）であることを意味する

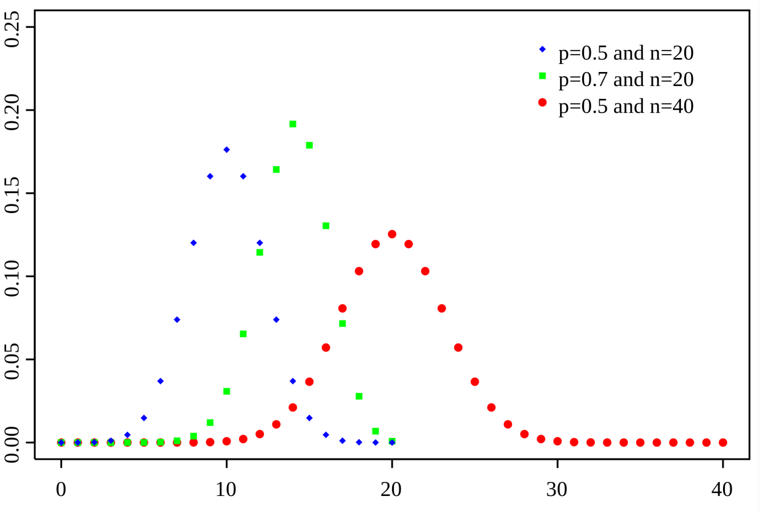


Figure: 二項分布の確率質量関数の例

独立同分布 (i.i.d.; independent and identically distributed)

$$p(k;\boldsymbol{\phi},n) = \frac{n!}{k!(n-k)!}\phi_1^k \phi_2^{n-k} \tag{2}$$

- 二項分布において、n 回の試行のうちコインの表が k 回出る確率を上の式で表すとき、以下のことを仮定している

1. 表が出る確率は履歴に左右されない
   - 試行の独立性の仮定
2. 表が出る確率は同じままである
   - 確率分布（この場合はベルヌーイ分布）の同一性の仮定

- 二項分布で 2 種類の事象が生起する列をモデル化するとき、この仮定を置いてもいいかどうか、考える必要がある

問題2-1

- 表が出る確率が 0.6 であるコインを 3 回投げたとき
  - 表が 1 回も出ない確率
  - 表が 1 回だけ出る確率
  - 表がちょうど 2 回出る確率
  - 3 回とも表が出る確率
- を、それぞれ求めよ。ただし i.i.d. を仮定する。

統計モデリングの問い

観測されたデータを

どのモデルが

一番うまく説明してくれるか？

- 「どのモデル」
  - モデルを選ぶ範囲をどう設定するか？
- 「一番うまく」
  - どういう基準でモデルを選ぶか？

二項分布の場合の統計モデリングの問い

観測されたデータを

パラメータ ϕ_1 の値がいくらの二項分布が

一番うまく説明してくれるか？

- 「どのモデル」
  - n が観測データの個数である二項分布のなかから選ぶ
- 「一番うまく」
  - まだどういう基準でモデルを選ぶかは述べられていない

パラメータを使ってデータの確率を書く

- 「コインを n 回投げて表が k 回出る確率はいくら？」
- この問いに、二項分布によるモデリングで答えると・・・

$$p(k; \boldsymbol{\phi}, n) = \frac{n!}{k!(n-k)!}\phi_1^k \phi_2^{n-k} \tag{3}$$

- パラメータを使えば、観測データが生成される確率を表せる
- しかしパラメータは未知数
- パラメータの値を定めること＝モデルを選ぶこと

観測データからパラメータを推定する

- 目の前にあるこのコインを投げて表が出る確率がいくらか、分からない
- 観測できるのは、実際にコインを投げて得られる表裏の列
- そこで、実際にコインを投げて表が出た回数から、表が出る確率を推定 estimate する
- そのとき、観測データを「一番うまく説明してくれる」値を推定値とする
- 「一番うまく」とは？

問題2-2

- コインを100回投げたら、表が52回出たとする
- このコインの表が出る確率はいくら？

問題2-2の解答例

- 100回のうち表が52回出たのだから・・・

$$\phi_1 = \frac{52}{100} \tag{4}$$

- なぜこの計算でいいと考えるのか？
  - 普通はこう計算すると思うが、なぜこう計算するのでいいと考えるのか？

解答例への理屈づけ

- この 100 回のコイン投げを二項分布でモデリングする
  - 二項分布でモデリングすること自体は正しいのか？
- すると、100 回のうち表が 52 回出る、という事象の確率は

$$p(k=52;\boldsymbol{\phi},n=100)=\frac{100!}{52!48!}\phi_1^{52}\phi_2^{48} \tag{5}$$

- この式を ϕ_1 の関数とみなし、この関数を最大化する ϕ_1 の値を求めてみる
  - ϕ_2 は $1-\phi_1$ で置き換える

$$f(\phi_1) = \frac{100!}{52!48!}\phi_1^{52}(1-\phi_1)^{48} \tag{6}$$

$$\begin{aligned}
\frac{df(\phi_1)}{d\phi_1} &= \frac{100!}{52!48!}\{52\phi_1^{51}(1-\phi_1)^{48} - 48\phi_1^{52}(1-\phi_1)^{47}\} \\
&= \frac{100!}{52!48!}\phi_1^{51}(1-\phi_1)^{47}\{52(1-\phi_1) - 48\phi_1\} \\
&= \frac{100!}{52!48!}\phi_1^{51}(1-\phi_1)^{47}(52 - 100\phi_1)
\end{aligned} \tag{7}$$

$\frac{df(\phi_1)}{d\phi_1} = 0$ とおくと、$\phi_1 = 0$、$1 - \phi_1 = 0$、$52 - 100\phi_1 = 0$
このうち最大値が得られるのは、$\phi_1 = \frac{52}{100}$ のとき。

尤度 likelihood

$$p(k=52;\boldsymbol{\phi},n=100)=\frac{100!}{52!48!}\phi_1^{52}\phi_2^{48} \tag{8}$$

- 上の例のように、観測されたデータの確率を、未知数であるパラメータの関数として表したものを、尤度と呼ぶ
- 尤度はパラメータの関数である
  - コイン投げ 100 回の結果という観測データは、既知
  - つまり、観測データは定数であって、尤度はパラメータの関数
    - パラメータの値を動かすと、尤度も動く

最尤推定 maximum likelihood estimation

- 式 (7) のように、与えられた観測データについて尤度を最大化するパラメータの値をもってパラメータの推定値とすることを、最尤推定という
- 尤度はパラメータの関数
  - パラメータの値を変えると尤度が変わる、ということ
  - 観測データは既知、つまり固定された値

Contents

問題2-2の解答例（再び）

- 100回のうち表が52回出たのだから・・・

$$\phi_1 = \frac{52}{100} \tag{9}$$

- しかし…この求め方で本当にいいのか？
- ダメだとすればその理由は？

- 表が出る確率を、その100回のコイン投げだけで決めてしまっていいのか？

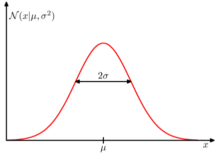
ベイズ的な考え方

- 表が出る確率を決めること自体に不確かさがある、と考える
- つまり、ϕ_1 が 0 から 1 のあいだのどの値をとるかも確率的に決まる、と考える
- 言い換えれば、パラメータもまた確率変数だと考える
  - これがベイズ的な考え方
- つまり、ベイズ的なデータのモデル化は二段構え
  1. コイン投げで表が何回出るかは確率的に決まる
  2. それだけでなく、表が出る確率も確率的に決まる

パラメータの値が確率的に決まるとは？

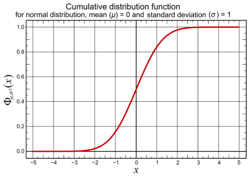
- 表が出る確率 ϕ_1 は様々でありうる
  - 0.01, 0.1, 0.3, 0.6, 0.98 等々・・・0 以上 1 以下なら何でも OK
- とびとびの値ではなく連続的な値
  - 表が出る確率の値が何通りあるかなんて言えない
    - 強いて言えば、無限通りある
- つまり、0 から 1 までの連続的な値それぞれをとる確率を考える必要がある
- 言い換えれば、区間 $[0, 1]$ 上に定義される連続確率分布を考える必要がある

連続確率分布



- 連続分布は確率密度関数 probability density function (pdf) で表すことが多い（例. 右図は正規分布の場合）
- 確率密度関数は、定義域全域で積分すると 1 になる
- 連続分布を累積分布関数 cumulative distribution function (cdf) で表すこともある
  - ざっくり言えば、pdf を積分すると cdf

$$F(x) = \int_{-\infty}^{x} f(t)dt$$



一様分布

- 連続値に確率を定める確率分布のひとつの例
- 定義域はいろいろでありうる
  - コインの表が出る確率 ϕ_1 の場合は $[0, 1]$ の範囲

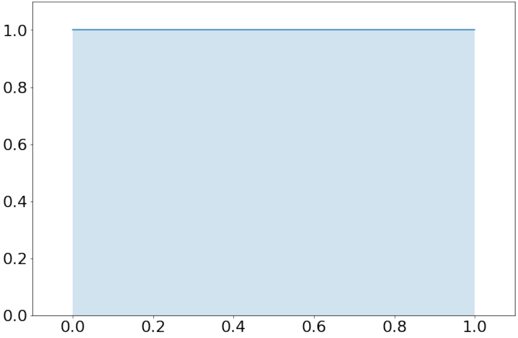


Figure: 区間 $[0, 1]$ 上の一様分布

復習：ベイズ則

$$p(z|x) \propto p(x|z)p(z)$$

- ある仮説が成り立つ確率 $p(z)$ は・・・
- その仮説が成り立つときに尤もらしさ $p(x|z)$ が大きいデータ x が観測されると・・・
- 高い値の確率 $p(z|x)$ になる

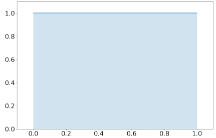
ベイズ則をベイズ的なモデリングで使う

$$p(\theta|x) \propto p(x|\theta)p(\theta) \tag{11}$$

- モデルのパラメータ θ がある値をとる確率 $p(\theta)$ は・・・
- パラメータがその値をとるときに尤度 $p(x|\theta)$ が大きいデータ x が観測されると・・・
- 高い値の確率 $p(\theta|x)$ になる

Contents

一様分布からスタートする



- コインの表が出る確率がいくらかなんて全く分からないという状態からスタートすることにする
  - 常識的には「ほぼ 1/2」と仮定するのが自然だが、今回はこうする
- そこで、表が出る確率 ϕ_1 が区間 $[0, 1]$ のどの値をとるかについては、一様分布を仮定することにする
- つまり、$p(\phi_1) = 1$ と仮定する
  - これは、ϕ_1 がどの値をとることも全く同じようにありえそう、という意味

1回コインを投げたら表だったとする

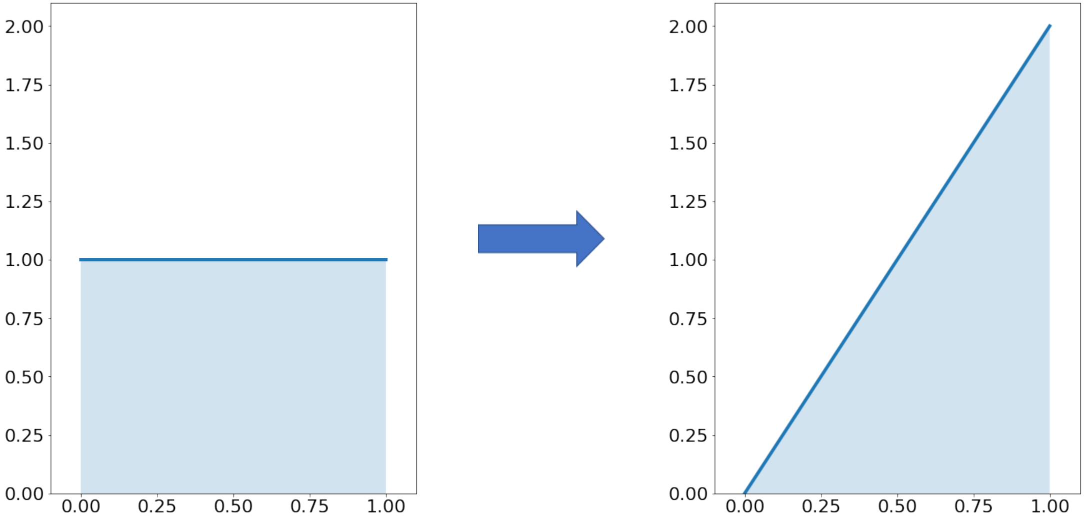
$$p(\phi_1|x_1 = \mathsf{H}) \propto p(x_1 = \mathsf{H}|\phi_1)p(\phi_1)$$

$p(\phi_1)$ は一様分布だと仮定したので、$p(\phi_1) = 1$。また、尤度 $p(x_1 = \mathsf{H}|\phi_1)$ は ϕ_1 に等しい。よって、

$$p(\phi_1|x_1 = \mathsf{H}) \propto \phi_1 \times 1 = \phi_1 \tag{12}$$

右辺を規格化する。つまり、$[0, 1]$ で積分して 1 になるようにする。$\int_0^1 \phi_1 d\phi_1 = \left[\frac{\phi_1^2}{2}\right]_0^1 = \frac{1}{2}$ より、右辺を $\frac{1}{2}$ で割って

$$p(\phi_1|x_1 = \mathsf{H}) = 2\phi_1 \tag{13}$$


2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00

もう1回コインを投げたら裏だったとする

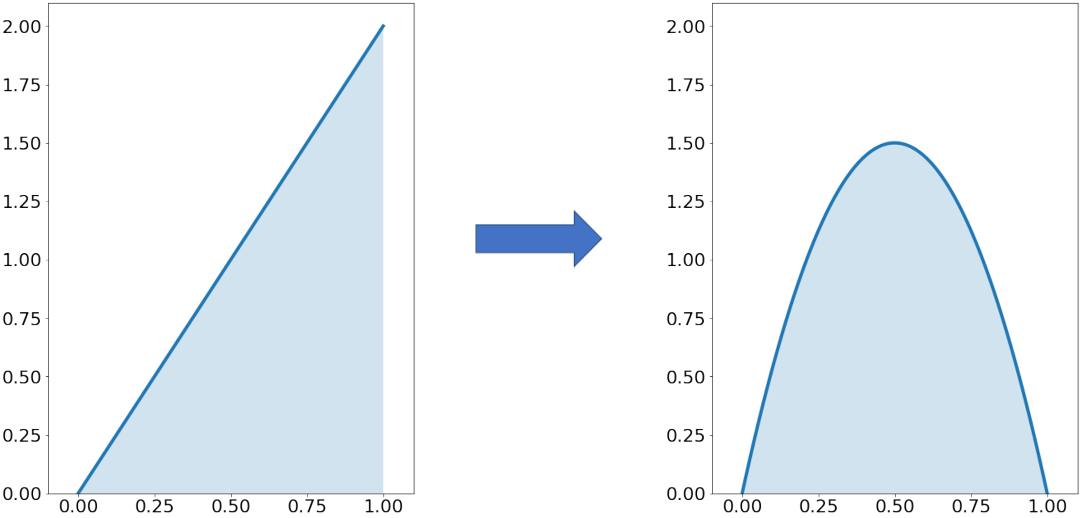
$$p(\phi_1|x_1 = \mathsf{H}, x_2 = \mathsf{T}) \propto p(x_1 = \mathsf{H}, x_2 = \mathsf{T}|\phi_1)p(\phi_1)$$

$p(\phi_1)$ は一様分布だと仮定したので、$p(\phi_1) = 1$。また、尤度 $p(x_1 = \mathsf{H}, x_2 = \mathsf{T}|\phi_1)$ は $\phi_1(1 - \phi_1)$ に等しい。よって、

$$p(\phi_1|x_1 = \mathsf{H}, x_2 = \mathsf{T}) \propto \phi_1(1 - \phi_1) \tag{14}$$

右辺を規格化する。つまり、$[0, 1]$ で積分して1になるようにする。$\int_0^1 \phi_1(1 - \phi_1)d\phi_1 = \left[\frac{\phi_1^2}{2} - \frac{\phi_1^3}{3}\right]_0^1 = \frac{1}{6}$ より、右辺を $\frac{1}{6}$ で割って

$$p(\phi_1|x_1 = \mathsf{H}, x_2 = \mathsf{T}) = 6\phi_1(1 - \phi_1) \tag{15}$$


2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00

もう1回コインを投げたら表だったとする

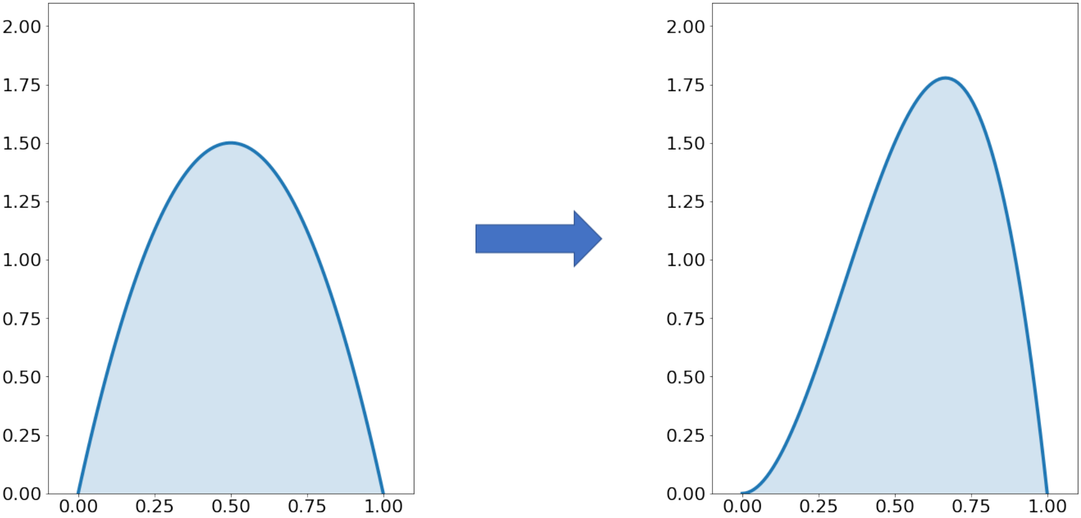
$$p(\phi_1|x_1=\mathsf{H},x_2=\mathsf{T},x_3=\mathsf{H}) \propto p(x_1=\mathsf{H},x_2=\mathsf{T},x_3=\mathsf{H}|\phi_1)p(\phi_1)$$

$p(\phi_1)$ は一様分布だと仮定したので、$p(\phi_1)=1$。また、尤度 $p(x_1=\mathsf{H},x_2=\mathsf{T},x_3=\mathsf{H}|\phi_1)$ は $\phi_1^2(1-\phi_1)$ に等しい。よって、

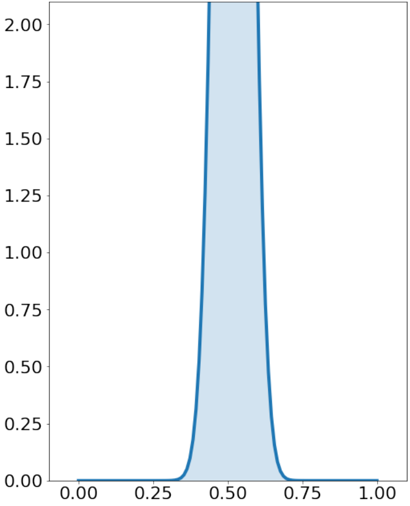
$$p(\phi_1|x_1=\mathsf{H},x_2=\mathsf{T},x_3=\mathsf{H}) \propto \phi_1^2(1-\phi_1) \tag{16}$$

右辺を規格化する。つまり、$[0,1]$ で積分して1になるようにする。$\int_0^1 \phi_1^2(1-\phi_1)d\phi_1 = \left[\frac{\phi_1^3}{3}-\frac{\phi_1^4}{4}\right]_0^1 = \frac{1}{12}$ より、右辺を $\frac{1}{12}$ で割って

$$p(\phi_1|x_1=\mathsf{H},x_2=\mathsf{T},x_3=\mathsf{H}) = 12\phi_1^2(1-\phi_1) \tag{17}$$

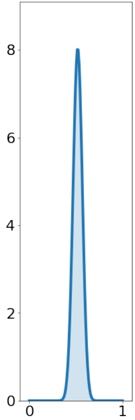

2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00

結局、合計100回投げたら表が52回出たとする



問題2-2の先ほどの解答例との違い

- 「合計100回投げて表が52回出た」という出題に、最初の解答例では52/100と答えを求めていた
- 今回は、連続的な関数（右図）で答えている
  - 表が出る確率として考えられる0から1の値について、それぞれがどのくらいありえそうかで、答えている
  - 答えをひとつの値に決めていない、とも言える



ベータ分布 beta distribution

- 上の説明に出てきた連続分布をベータ分布と呼ぶ
- ベータ分布は、2つのパラメータ a と b を持つ
- 確率密度関数は以下の式

$$p(x; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$$

  - $\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$ は、規格化定数

事前分布と事後分布

- 事前分布 prior distribution
  - 今日の話では、最初にそこからスタートしたところの一様分布が、事前分布
- 事後分布 posterior distribution
  - 今日の話でそのつど求めていた $p(\phi_1|\boldsymbol{x})$ が、事後分布
- ベイズ的な統計モデリングでは、観測データにもとづいて事後分布を求めることが、主な課題！

ベイズ的モデリングにおけるベイズ則

- ベイズ推論で使うベイズ則は、単に条件付き確率どうしの関係を表すだけの式ではない

$$（\text{事後分布}）\propto（\text{尤度}）\times（\text{事前分布}）$$

- 上の式が、ベイズ的な統計モデリング特有のベイズ則の使い方

課題2

- ϕ の関数 $f(\phi) = \phi^3(1-\phi)^2$ を $[0,1]$ の範囲で積分せよ。
- 積分した結果が 1 になるようにするためには、$f(\phi)$ にいくらを掛けておけばよかったか。