

確率について

はじめに

統計モデリングとベイズ推論を行う上で、理解したい三つの概念

- ・ モデル
- ・ パラメータ
- ・ データ

モデルを立て、与えられたデータを用いて、パラメータを推定し、未知のデータを予測したり現象の起こる原理を説明する。

データ

データはある種のランダムさを持って得られたものであると考える。

データの発生の仕方をうまく説明したい。 また推定されたパラメータに基づいて、将来的に得られるであろうデータを予測したい。

モデル

データの発生の仕方を説明するための大枠。 または分析する前提としての枠組み。

基本的にはモデルをどう設定するか、よいモデルを作れるかが分析者のやること。

パラメータ

モデルの細かな設定を決めるための数値などをいう。

確率

乱数やサンプリングという考え方について、Rやstanを用いながら親しみを持ってもらおう。

例題

まずは上の統計モデリングの三つの概念について、特にパラメータの推定について理解するため、 次のような簡単な問題を考えよう。

手元にコインがある。これを投げて表が出るか、裏が出るかを予測したい。すでに10回投げて7回表が出るということを観測している。

- ・ モデルとして、コインはある一定の確率 p で表が出て、 $1 - p$ で裏が出るとする。
- ・ この p がパラメータである。
- ・ また10回投げて7回表が出た、というのが現在のデータである。

このデータを用いて p を推定する、というのがパラメータの推定。

推定されたパラメータを用いて、次に表が出るか裏が出るかを予測することができる。

どのようにパラメータ p を推定するか？

- ・ 10回中7回なので $p = 0.7$ と推定
- ・ コインだから表裏半々ぐらい出るはずなので、ちょっと表が出やすく $p = 0.6$ と推定

など、色々考え方があある。

一つ目の推測は、最尤法という方法に基づいた結論と同じものになる。最尤法の考え方を理解するために、次のような実験を行うことにしよう。

```
x <- rbinom(n = 1, size = 10, prob = 0.7)
x

## [1] 7

y <- rbinom(n = 1, size = 10, prob = 0.6)
y

## [1] 7

z <- rbinom(n = 1, size = 10, prob = 0.5)
z

## [1] 7

x1 <- rbinom(n = 100, size = 10, prob = 0.8)
x1

## [1] 8 8 8 8 9 7 10 7 9 8 9 7 7 7 9 10 7 10 9 8 8 8 6
## [24] 7 7 8 9 8 8 10 10 8 9 7 9 8 9 9 8 7 7 10 10 8 7 7
## [47] 6 9 7 7 6 6 7 10 8 6 8 10 7 7 8 9 10 7 9 7 8 8 7
## [70] 7 10 9 8 9 8 9 9 8 10 8 6 10 10 10 8 9 9 8 8 8 6 8
## [93] 8 9 8 8 6 7 9 8

x2 <- rbinom(n = 100, size = 10, prob = 0.7)
x2

## [1] 6 7 7 7 4 9 6 7 6 8 8 7 7 4 7 10 6 7 7 7 8 6 6
## [24] 7 8 7 4 8 5 6 5 8 8 6 8 5 7 4 7 8 8 6 7 5 5 10
## [47] 9 7 6 6 6 8 7 7 6 6 8 7 10 7 6 8 8 5 8 4 9 7 8
## [70] 6 9 5 9 7 7 6 7 8 9 6 8 8 7 7 7 7 6 6 8 8 5 5
## [93] 5 4 8 7 9 6 7 8

x3 <- rbinom(n = 100, size = 10, prob = 0.6)
x3

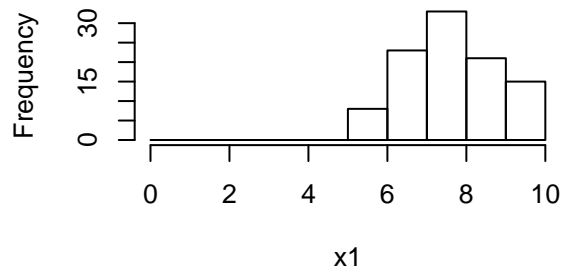
## [1] 8 7 4 4 6 4 5 7 7 6 6 6 8 5 7 5 8 8 8 5 6 7 5 7 6 6 4 6 7 7 6 6 7 8 4
## [36] 7 8 5 8 9 4 7 6 7 4 5 7 4 6 8 5 7 7 8 4 4 6 6 4 6 7 5 8 4 4 7 8 8 9 8
## [71] 3 6 7 6 6 7 6 5 5 3 7 8 7 5 5 7 5 7 7 7 7 6 8 5 8 5 6 6 6 6

x4 <- rbinom(n = 100, size = 10, prob = 0.5)
x4

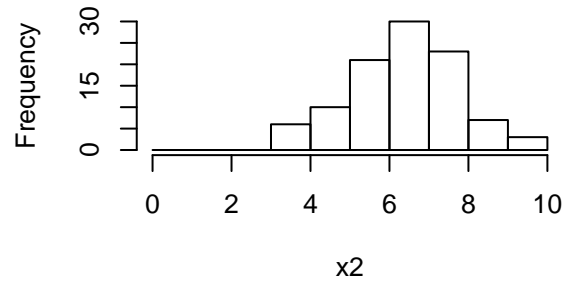
## [1] 4 2 6 7 6 2 5 6 4 6 8 3 4 4 4 5 5 5 3 7 5 5 5 4 7 5 4 5 5 6 8 5 2 5 4
## [36] 6 2 6 7 5 4 4 6 3 6 3 4 4 8 4 4 5 4 4 3 2 4 5 6 6 3 5 7 7 6 4 6 5 6 6
## [71] 5 5 5 4 6 5 5 8 6 5 4 6 2 3 8 7 4 5 4 6 7 2 5 4 5 6 2 6 5 7

par(mfrow=c(2,2))
hist(x1, breaks = 0:10)
hist(x2, breaks = 0:10)
hist(x3, breaks = 0:10)
hist(x4, breaks = 0:10)
```

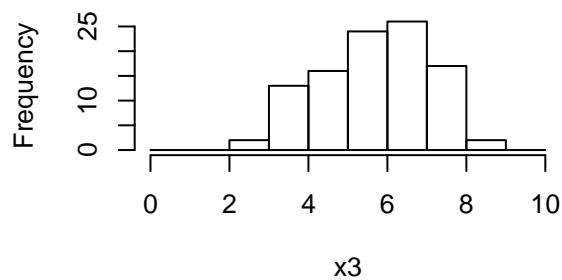
Histogram of x1



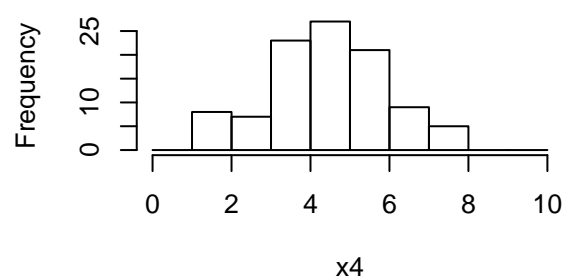
Histogram of x2



Histogram of x3



Histogram of x4



このようにパラメータ p の値によって、表の出る回数にばらつきがある。この表の出る回数の割合をパラメータ p に対する尤度という。

今回のデータでは、表の出た回数が7回なので、パラメータ p ごとに表が7回出る確率を計算し、それを $L(p)$ と表す。これを尤度関数という。

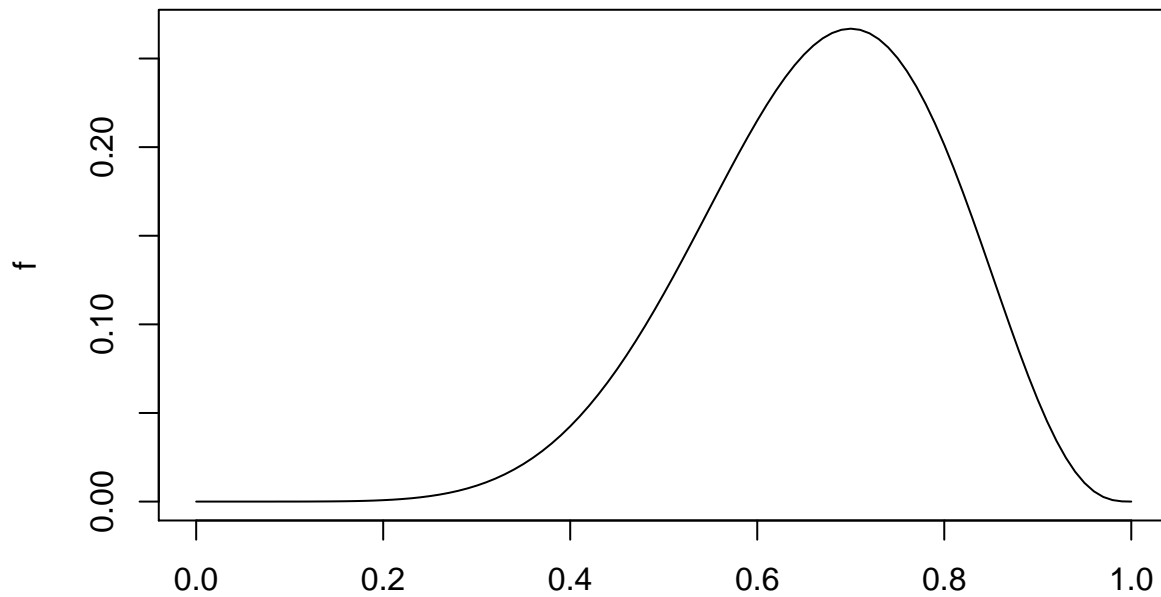
今回のデータに基づいて尤度関数を計算すると

$$L(p) = 120 p^7 (1 - p)^3$$

となる。

最尤法では尤度関数が最大となる p をパラメータの推定値とする。このグラフをかくと、

```
f <- function(p){120*p^7*(1-p)^3}
plot(f, 0, 1)
```



x

であり、 $p =$

0.7の時に最大となることがわかる。

今回はパラメータ p を持つ二項分布モデルに基づいて、現象を説明している。

ではベイズ推論の枠組みでは、上の問題に対してどのような回答をするか？

パラメータ p をある決まった値として推定するのではなく、ある程度幅を持ったなぜ分布として推定する。

分布にすると何が違うかを説明する。例えば与えられたデータが

- ・ 2回投げて1回表
- ・ 100回投げて50回表

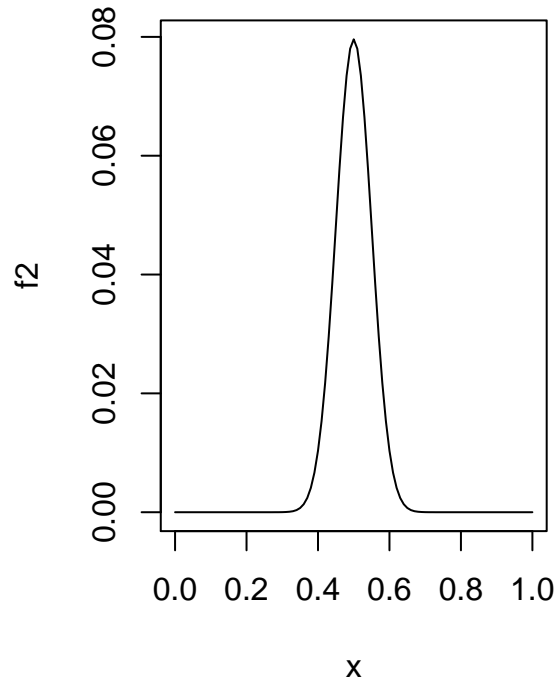
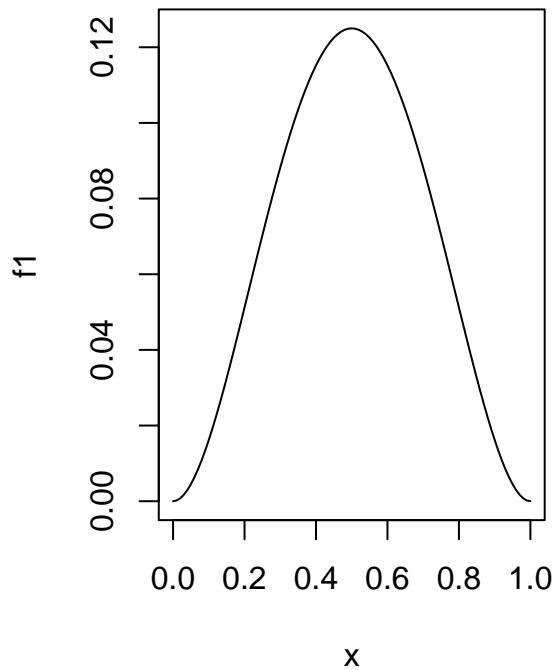
というふた通りの可能性を考えよう。この時、最尤法でパラメータを推定するとどちらも $p = 0.5$ となる。しかし、後者の方が信頼できそうだろう。

このような信頼度の違いを分布の形状として表現することができる。

例えば p の分布が以下の二つでどのように違うか検討せよ。

```
f1 <- function(p){choose(2,1) * p^2 * (1-p)^2}
f2 <- function(p){choose(100,50) * p^50 * (1-p)^50}
```

```
par(mfrow=c(1,2))
plot(f1, 0, 1)
plot(f2, 0, 1)
```



これがパラメータに分布を持たせることの一つの利点である。

分散の大きい分布であれば情報が少なく、分散の小さな分布であれば情報が多、と考えることができる。

ベイズ推論の考え方

パラメータはある決まった真の値があるのではなく、何らかの確率分布にしたがっているとして推測を行う。

- ・ 尤度関数
- ・ 事前分布
- ・ 事後分布

の三つの概念について理解する。

尤度関数

上ですでに出てきたように、パラメータを入力すると現在観測しているデータが発生する確率を計算する関数。

モデルとデータが与えられると関数が定まる。

事前分布と事後分布

いずれもパラメータの確率分布。データを観測する前の分布のことを事前分布、データを観測した後の分布を事後分布という。

観測を通して事前分布を事後分布に取り替えることを、パラメータ更新、ベイズ更新などという。

パラメータの分布

ベイズ推論ではデータの発生を確率現象と捉えるのみでなく、パラメータも確率的な現象と考える。

確率分布とベイズの定理

ここからは、ベイズ推論の枠組みを理解するために以下の設定を通して、確率分布、特にパラメータ分布の更新について理解しよう。

設定

外からは見分けのつかない二つの箱A, Bの中に赤いボールと白いボールがいくつか入っている。

以下の手順に従ってボールを取り出し赤いボールなら賞金1万円、白いボールなら罰金1万円もらえるというゲームを考えよう。

- ・ まずはじめにどちらかの箱を選ぶ。
- ・ 次に観察期間として選んだ箱から規定の回数、ボールを取り出し色を確認して戻すという操作を行う。
- ・ そののち、実際にゲームに挑戦するか決める。
- ・ 挑戦する場合は先ほど選んだ箱の中からボールを一つとり、色に応じて賞金を得る。

さて、どのような戦略を立てるべきか。

例題1

- ・ 箱Aには赤いボールが5個、白いボールが0個
- ・ 箱Bには赤いボールが0個、白いボールが5個

入っているということがあらかじめわかっているとしよう。

この場合には、観察期間として1回ボールを確認すれば、選んだ箱がAなのかBなのか完全に決定することができる。

従ってそれに応じて最終的に挑戦するかどうか決定すればよい。

例題2

- ・ 箱Aには赤いボールが3個、白いボールが1個
- ・ 箱Bには赤いボールが1個、白いボールが3個

入っていることがわかっているとしよう。

また観察期間として、一度だけボールの色を確認できるとする。

Rの乱数を利用して実験してみよう。

```
rbinom(n = 10, size = 1, prob = 0.75)
```

```
## [1] 1 0 1 1 1 1 1 1 1 1
```

```
rbinom(n = 10, size = 1, prob = 0.25)
```

```
## [1] 0 0 0 0 0 0 0 0 1 0
```

赤いボールが出た場合には箱Aと推定し、白いボールが出た場合には箱Bと推定するのがもっともらしいであろう。

さらにそれを踏まえた上で、最終的に挑戦するかを決定できる。

このような判断を数学的に考えていこう。 まずは

- ・ 確率分布
- ・ 条件付き確率
- ・ 同時確率
- ・ 周辺確率
- ・ 尤度関数
- ・ ベイズの定理

という概念を順番に理解していこう。

確率分布

X という出来事が起こる確率を $P(X)$ とかく。例えば

$$P(\text{コインを投げて表が出る}) = \frac{1}{2}$$

のように書く。これを省略して

$$P(\text{表}) = \frac{1}{2}$$

などとも書く。

今回の問題では箱がAなのかBなのか、ボールが赤なのか白なのか、といった出来事について確率を考える。

$$P(A) = \frac{1}{2}$$

$$P(B) = \frac{1}{2}$$

と設定しよう。二つの箱からランダムに選ぶのでこのように設定するのが妥当だろう。

確率変数と確率分布

上のようにある状況で起こりうる出来事全体が X_0, X_1, \dots, X_n である時、 $P(X_1)$ から $P(X_n)$ にたいして0以上1以下の数を与える。

このようなものを（離散的な）確率分布という。この時、合計が1になるように決めなければならない。

例1

「コインを1回投げる」という操作に注目すると、起こりうる結果は表が出るか裏が出るかどちらか。この結果を表す確率変数 X を導入する。

$P(\text{表})$ と $P(\text{裏})$ を決めればこの現象を記述できる、と考える。この二つの値が0以上1以下で、合計が1になるように配分してやる。

例えば

$$P(X = \text{表}) = \frac{1}{2}, P(X = \text{裏}) = \frac{1}{2}$$

としてもよいし、

$$P(X = \text{表}) = \frac{1}{3}, P(X = \text{裏}) = \frac{2}{3}$$

としてもよい。極端な場合には

$$P(X = \text{表}) = 1, P(X = \text{裏}) = 0$$

としてもよい。

例2

「サイコロを1回投げる」という操作に注目すると、起こりうる結果は1,2,3,4,5,6のどれかが出るということ。サイコロの目を表す確率変数を X とする。

この場合は

$$P(X = 1), P(X = 2), P(X = 3), P(X = 4), P(X = 5), P(X = 6)$$

を決めてやる。例えば

$$P(X = 1) = \frac{1}{6}, P(X = 2) = \frac{1}{6}, P(X = 3) = \frac{1}{6}, P(X = 4) = \frac{1}{6}, P(X = 5) = \frac{1}{6}, P(X = 6) = \frac{1}{6}$$

としてもよいし、

$$P(X = 1) = \frac{1}{10}, P(X = 2) = \frac{2}{10}, P(X = 3) = \frac{3}{10}, P(X = 4) = \frac{1}{10}, P(X = 5) = \frac{2}{10}, P(X = 6) = \frac{1}{10}$$

としてもよい。

条件付き確率と同時分布

次にボールを一つ取り出した時、赤であるか白であるかの確率を計算したいが、このままでは難しい。

なぜならどちらの箱を選んだかによって赤いボールの出やすさが異なるからである。

このように箱の種類とボールの色といういくつかの条件を組み合わせる場合には、条件付き確率や同時分布を用いて記述する。

例えば選んだ箱がAだという条件のもと、赤いボールを取り出す確率を表現したい。

Y という条件のもとでの X の確率を

$$P(X | Y)$$

と表し、これを条件付き確率という。

箱の種類を表す確率変数を θ とし、取り出したボールの色を表す変数を X とする。

Aの箱という条件のもとで赤いボールを取り出す確率は

$$P(X = \text{赤} | \theta = A) = \frac{3}{4}$$

であり、Aの箱という条件のもとで白いボールを取り出す確率は

$$P(X = \text{白} | \theta = A) = \frac{1}{4}$$

となる。この二つの合計は1になっている。よってこれは確率分布を与えていることに注意しよう。

同様に箱Bを選んだという条件のもとでのボールの色に関する条件付き確率を表現すると、

$$P(X = \text{赤} | \theta = B) = \frac{1}{4}, P(X = \text{白} | \theta = B) = \frac{3}{4}$$

と書くことができる。これも確率分布を与えている。

同時分布と周辺分布

箱を選びボールを取り出す、という二つの条件をまとめて考えると、起こりうる出来事としては

- ・ 箱Aから赤いボール
- ・ 箱Aから白いボール
- ・ 箱Bから赤いボール
- ・ 箱Bから白いボール

の4パターンがある。このように二つ以上の条件をまとめて考えた確率を同時確率という。

これらの確率を計算すると、

$$P(X = \text{赤} \cap \theta = A) = \frac{3}{4} \times \frac{1}{2} = \frac{3}{8}$$

$$P(X = \text{白} \cap \theta = A) = \frac{1}{4} \times \frac{1}{2} = \frac{1}{8}$$

$$P(X = \text{赤} \cap \theta = B) = \frac{1}{4} \times \frac{1}{2} = \frac{1}{8}$$

$$P(X = \text{白} \cap \theta = B) = \frac{3}{4} \times \frac{1}{2} = \frac{3}{8}$$

と計算できる。4つを合計すると1になるので、これも確率分布である。

また $P(A)$, $P(B)$ や $P(\text{赤})$, $P(\text{白})$ と行ったように、どちらかの条件のみを考えた確率分布を周辺分布という。この場合には

$$P(X = \text{赤}) = P(X = \text{赤} \cap \theta = A) + P(X = \text{赤} \cap \theta = B) = \frac{1}{2}$$

$$P(X = \text{白}) = P(X = \text{白} \cap \theta = A) + P(X = \text{白} \cap \theta = B) = \frac{1}{2}$$

として計算できる。クロス集計表を思い浮かべるとよい。

ベイズの定理

クロス集計表をじっと見ていると、次のような関係が成り立つことがわかる。

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

これを式変形して得られる次の公式をベイズの定理という。

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

これを用いることで、 $P(A | \text{赤})$ と $P(B | \text{赤})$ のいずれが大きいかを計算することができる。

言い換えると、赤いボールを一個とったあとで、目の前の箱がAであるのかBであるのか、どちらの確率が高いかを計算できる。

例題1の場合で言えば、Aの箱を選んだという条件のもとで赤いボールが出る確率は

$$P(X = \text{赤} | \theta = A) = 1$$

であり、Aの箱を選んだという条件のもとで白いボールが出る確率は

$$P(X = \text{白} | \theta = A) = 0$$

となる。またBの箱から取ったという条件のもとでの赤白のそれぞれの確率は

$$P(X = \text{赤} | \theta = B) = 0, P(X = \text{白} | \theta = B) = 1$$

となる。

同時分布は

$$P(\theta = A \cap X = \text{白}) = 0, P(\theta = A \cap X = \text{赤}) = \frac{1}{2}, P(\theta = B \cap X = \text{白}) = \frac{1}{2}, P(\theta = B \cap X = \text{赤}) = 0$$

となり、さらに

$$P(X = \text{赤}) = P(\theta = A \cap X = \text{赤}) + P(\theta = B \cap X = \text{赤}) = \frac{1}{2}$$

$$P(X = \text{白}) = P(\theta = A \cap X = \text{白}) + P(\theta = B \cap X = \text{白}) = \frac{1}{2}$$

であることがわかる。

さて、赤いボールを取り出したとき、選んだ箱がどちらの箱の確率が高いだろうか？

$P(\theta = A | X = \text{赤})$ と $P(\theta = B | X = \text{赤})$ の大小を比較しよう。

ベイズの定理を使えば、

$$P(\theta = A | \text{赤}) = \frac{P(\text{赤} | \theta = A)P(\theta = A)}{P(\text{赤})} = 1 \times \frac{1}{2} \div \frac{1}{2} = 1$$

$$P(\theta = B | \text{赤}) = \frac{P(\text{赤} | \theta = B)P(\theta = B)}{P(\text{赤})} = 0 \times \frac{1}{2} \div \frac{1}{2} = 0$$

と計算できる。

例題2の場合を考えよう。

1回ボールを取り出すと赤いボールがでた、という事実を観測データとしよう。

すると、赤いボールが出る確率をパラメータの関数として表現すると以下のようなになる。

$$L(A) = P(\text{赤} | \theta = A) = \frac{3}{4}$$

$$L(B) = P(\text{赤} \mid \theta = B) = \frac{1}{4}$$

この L が尤度関数である。

ベイズの定理を使えば、

$$P(\theta = A \mid \text{赤}) = \frac{P(\text{赤} \mid \theta = A)P(\theta = A)}{P(\text{赤})} = \frac{L(A)P(\theta = A)}{P(\text{赤})} = \frac{3}{4} \times \frac{1}{2} \div \frac{1}{2} = \frac{3}{4}$$

$$P(\theta = B \mid \text{赤}) = \frac{P(\text{赤} \mid \theta = B)P(\theta = B)}{P(\text{赤})} = \frac{L(B)P(\theta = B)}{P(\text{赤})} = \frac{1}{4} \times \frac{1}{2} \div \frac{1}{2} = \frac{1}{4}$$

と計算できる。

これが事後分布。

この事後分布を用いて、次のボールが赤である確率を計算すると、

$$P(\text{赤}) = P(\text{赤} \mid \theta = A)P(\theta = A \mid \text{赤}) + P(\text{赤} \mid \theta = B)P(\theta = B \mid \text{赤}) = \frac{3}{4} \times \frac{3}{4} + \frac{1}{4} \times \frac{1}{4} = \frac{10}{16}$$

となる。

さてここまでの言葉遣いを用いて、ベイズ推測の枠組みを改めて説明すると、以下のようになる。

$P(\theta = A), P(\theta = B)$ という確率分布を事前分布、 $P(\theta = A \mid \text{白}), P(\theta = B \mid \text{白})$ という確率分布を事後分布という。

- ・ モデルは確率分布 $P(\text{赤} \mid \theta = A), P(\text{白} \mid \theta = A)$ という確率分布から決まるもの
- ・ パラメータは箱の種類AまたはB
- ・ データは観察期間にでたボールの記録

ということになる。

例題3

次に箱の設定は前と同じで観察できる回数が5回になったとしよう。

前回とはデータとして想定する出来事の起こり方が変わってくる。

- ・ 例題2ではボールを取り出して赤が出るか白が出るか
- ・ 例題3ではボールを5回取り出して何回赤が出るか

という点が異なる。

このような出来事の確率分布を記述するため

- ・ 二項分布

について理解する。

```
rbinom(n = 1, size = 5, prob = 0.75)
```

```
## [1] 4
```

```
rbinom(n = 1, size = 5, prob = 0.25)
```

```
## [1] 2
```

今回の問題設定は、このような数を見たときにprobの値がどのようにになっているか推測する、ということができる。

二項分布

表と裏がそれぞれ $\frac{1}{2}$ の確率で出るコイン、つまり $P(\text{表}) = P(\text{裏}) = \frac{1}{2}$ であるようなコインを N 回投げる、という操作を考えよう。

$N = 1$ のとき、表が出るか裏が出るかどうかで、表が1回出る確率、0回出る確率はどちらも $\frac{1}{2}$ である。

$N = 2$ のとき、表表、表裏、裏表、裏裏の4パターンあり、

$$P(\text{表表}) = P(\text{表裏}) = P(\text{裏表}) = P(\text{裏裏}) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

と全て等しい。これは1回目の確率と2回目の確率をかけて計算できる。

出方のパターンではなく、表の出る回数 X の分布を考えよう。 $N = 2$ の時は表の出る回数は0,1,2のいずれかなので $P(X = 0)$, $P(X = 1)$, $P(X = 2)$ という確率分布を与える。上で計算したことから

$$P(X = 0) = \frac{1}{4}, P(X = 1) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}, P(X = 2) = \frac{1}{4}$$

となることがわかる。 $X = 1$ となるのは2通りあって、それらの確率の和を考えればよい。

$N = 3$ ならば、表裏のでる出方は表表表、表表裏、表裏表、表裏裏、裏表表、裏表裏、裏裏表、裏裏裏、と八通りあり、いずれも確率 $\frac{1}{8}$ である。従って

$$P(X = 0) = \frac{1}{8}, P(X = 1) = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8}, P(X = 2) = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8}, P(X = 3) = \frac{1}{8}$$

と計算できる。

$N = 4$ の場合、表裏の出るパターンを全て列挙し、表の出る回数の確率分布を計算しよう。

一般に N 回コインを投げた時、表が k 回出るパターンは

$${}_N C_k = \frac{N!}{k!(N-k)!}$$

通りある。

例えば

$${}_2 C_1 = \frac{2}{1} = 2, {}_4 C_2 = \frac{4!}{2! \times 2!} = 6$$

などとなる。

したがって N 回コインを投げた時の、表が出る回数の確率分布は

$$P(X = k) = {}_N C_k \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{N-k}$$

と計算できる。

より一般に1回コインを投げた時に表の出る確率 p とする。つまり

$$P(\text{表}) = p, P(\text{裏}) = 1 - p$$

とする。この時 N 回コインを投げた時の、表が出る回数の確率分布は

$$P(X = k) = {}_N C_k p^k (1 - p)^{N-k}$$

となる。これが二項分布である。

例題4

箱の個数がA, Bで異なるとしよう。例えばAの箱が3個、Bの箱が2個あるとする。

この時は、パラメータの事前分布を

$$P(\theta = A) = \frac{3}{5}, P(\theta = B) = \frac{2}{5}$$

と設定する。

5回ボールを取り出した結果、赤いボールが4個、白いボールが1個、というデータが与えられたとしよう。

確率変数 X を赤いボールの個数とする。

このとき尤度関数は、

$$L(A) = P(X = 4 | \theta = A) = {}_5C_4 \left(\frac{3}{4}\right)^4 \left(\frac{1}{4}\right)^1 = \frac{5 \times 3^4}{4^5}$$

$$L(B) = P(X = 4 | \theta = B) = {}_5C_4 \left(\frac{3}{4}\right)^1 \left(\frac{1}{4}\right)^4 = \frac{5 \times 3}{4^5}$$

であり、

$$P(X = 4) = P(X = 4 | \theta = A)P(\theta = A) + P(X = 4 | \theta = B)P(\theta = B) = \frac{5 \times 3^4}{4^5} \times \frac{3}{5} + \frac{5 \times 3}{4^5} \times \frac{2}{5} = \frac{3^5 + 6}{4^5}$$

となる。

したがってベイズの定理を使えば、事後分布は

$$P(\theta = A | X = 4) = \frac{P(X = 4 | \theta = A)P(\theta = A)}{P(X = 4)} = \frac{L(A)P(\theta = A)}{P(X = 4)} = \frac{5 \times 3^4}{4^5} \times \frac{3}{5} \div \frac{3^5 + 6}{4^5} = \frac{3^5}{3^5 + 6}$$

$$P(\theta = B | X = 4) = \frac{P(X = 4 | \theta = B)P(\theta = B)}{P(X = 4)} = \frac{L(B)P(\theta = B)}{P(X = 4)} = \frac{5 \times 3}{4^5} \times \frac{2}{5} \div \frac{3^5 + 6}{4^5} = \frac{6}{3^5 + 6}$$

と計算できる。

この事後分布を用いて、次のボールが赤である確率を計算すると、

$$P(\text{赤}) = P(\text{赤} | \theta = A)P(\theta = A | X = 4) + P(\text{赤} | \theta = B)P(\theta = B | X = 4) = \frac{3}{4} \times \frac{3^5}{3^5 + 6} + \frac{1}{4} \times \frac{6}{3^5 + 6}$$

となる。

例題5

次に箱の種類がA, B, C, Dと4種類あり、それぞれ赤白のボールの内訳が、

- ・ 箱Aは赤いボールが1個、白いボールが4個
- ・ 箱Bは赤いボールが2個、白いボールが3個
- ・ 箱Cは赤いボールが3個、白いボールが2個
- ・ 箱Dは赤いボールが4個、白いボールが1個

となっているとしよう。またこの箱が全て一つずつあるとする。

パラメータの事前分布としては

$$P(A) = P(B) = P(C) = P(D) = \frac{1}{4}$$

とすればよい。

5個ボールを取り出して赤いボールを3個白いボールを3個取り出した、というデータが与えられたとしよう。

確率変数 X を赤いボールの個数とする。

このとき尤度関数は、

$$L(A) = P(X = 3 | A) = {}_5C_3 \left(\frac{1}{5}\right)^3 \left(\frac{4}{5}\right)^2$$

$$L(B) = P(X = 3 | \theta = B) = {}_5C_3 \left(\frac{2}{5}\right)^3 \left(\frac{3}{5}\right)^2$$

$$L(C) = P(X = 3 | C) = {}_5C_3 \left(\frac{3}{5}\right)^3 \left(\frac{2}{5}\right)^2$$

$$L(D) = P(X = 3 | D) = {}_5C_3 \left(\frac{4}{5}\right)^3 \left(\frac{1}{5}\right)^2$$

であり、

$$P(X = 3) = P(X = 3 | A)P(A) + P(X = 3 | \theta = B)P(\theta = B) + P(X = 3 | C)P(C) + P(X = 3 | D)P(D) = \frac{1}{4} \times 10 \times \frac{1}{5^5} \times (4^2 + 2^3 + 3^3 + 4^3)$$

となる。

したがってベイズの定理を使えば事後分布を、

$$P(A | X = 3) = \frac{P(X = 3 | A)P(A)}{P(X = 3)} = \frac{L(A)P(A)}{P(X = 3)} = 10 \times \frac{4^2}{5^5} \times \frac{1}{4} \div \frac{26}{5^3} = \frac{4}{65}$$

$$P(\theta = B | X = 3) = \frac{P(X = 3 | \theta = B)P(\theta = B)}{P(X = 3)} = \frac{L(B)P(\theta = B)}{P(X = 3)} = 10 \times \frac{2^3 \times 3^2}{5^5} \times \frac{1}{4} \div \frac{26}{5^3} = \frac{18}{65}$$

$$P(C | X = 3) = \frac{P(X = 3 | C)P(C)}{P(X = 3)} = \frac{L(C)P(C)}{P(X = 3)} = 10 \times \frac{3^3 \times 2^2}{5^5} \times \frac{1}{4} \div \frac{26}{5^3} = \frac{27}{65}$$

$$P(D | X = 3) = \frac{P(X = 3 | D)P(D)}{P(X = 3)} = \frac{L(D)P(D)}{P(X = 3)} = 10 \times \frac{4^3}{5^5} \times \frac{1}{4} \div \frac{26}{5^3} = \frac{16}{65}$$

と計算できる。

この事後分布を用いて、次のボールが赤である確率を計算すると、

$$P(\text{赤}) = P(\text{赤} | A)P(A | X = 4) + P(\text{赤} | \theta = B)P(\theta = B | X = 4) = \frac{3}{4} \times \frac{3}{4} + \frac{1}{4} \times \frac{1}{4} = \frac{10}{16}$$

となる。

例題6

箱AからDの中身は前と同じだが、今回は箱の個数がどのようになっているかわからないとする。

特に何も情報がない場合、全ての箱の割合が均等である一様分布を設定するべきであろう。

また、事前に情報がある場合や、例えば賞金や罰金の額が均等でないといった場合、何らかの情報を持った事前分布を設定することもありうる。

このようにパラメータの事前分布と呼ばれる確率分布を自分で設定する必要がある。多くの場合、事前分布は一様分布を設定する。

これを無情報事前分布といたりする。

例題7

さらに箱の種類が101種類、型番0番から100番まであり、赤いボールの割合 p が型番と一致するとする。

さらに、各番号の箱の個数が一定であるかはわからない状況であるとする。

この場合、モデルが二項分布であることは同じであるが、事前分布の設定を変えることになる。

まず尤度関数を計算しよう。

型番 p の箱を選んでいとしたとき、 n 回ボールを取り出したうち k 回赤いボールになる確率は

$$L(p) = P(X = k \mid p) = {}_nC_k p^k (1-p)^{n-k}$$

と計算できる。

例えば、5回中4回赤いボールというデータが与えられたとすると、

$$L(p) = P(X = 4 \mid p) = {}_5C_4 p^4 (1-p)^1$$

さて、事前分布は各 p ごとに p 番の箱がいくつあるか？を事前に設定しておき、各 p ごとに値を定めておくものだった。

今回の設定では箱の種類が多いので、連続的な分布を用いて近似的に考えることにしよう。

連続的な分布の場合は、 p の取りうる値が無限にあるので、各 p の箱の割合を与えるのではなく、 p 付近の密度を与える関数を考える。それを確率密度関数という。

例えば事前になんの情報もない場合には $p = 0$ から $p = 1$ までが均等にあることを表す一様分布を事前分布として用いる。一様分布の確率密度関数は

$$f(p) = 1 \quad (0 \leq p \leq 1)$$

である。

上と同じようにベイズの定理を用いて事後分布を計算すると、

$$f(p \mid X = 4) = \frac{L(p)f(p)}{P(X = 4)} = \frac{5}{P(X = 4)} p^4 (1-p)^1$$

となる。

ここで $P(X = 4)$ を計算するためには積分の計算が必要になる。

例題8

箱に関する設定は上の例題と同じとしよう。

なんらかの情報がある場合には事前分布を自由に設定できる。

ここではベータ分布と呼ばれる確率分布を使うことにしよう。ベータ分布の確率密度関数は

$$f(p) = \frac{1}{C} p^{\alpha-1} (1-p)^{\beta-1}$$

である。ここで α, β はベータ分布のパラメータである。

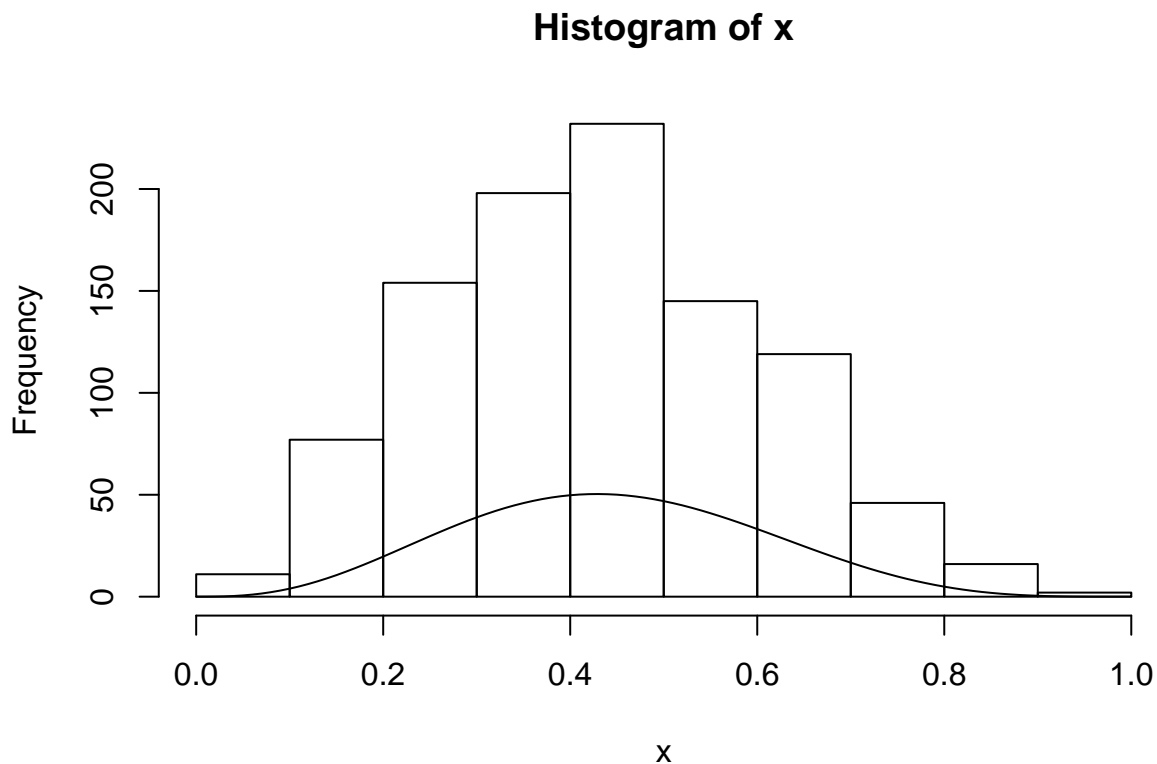
また分母の C は面積を1にするため（確率分布の条件）の定数なので、今回はあまり気にしないでよい。

この x の関数を密度関数に持つ確率分布をベータ分布という。例えば $Be(1, 1)$ は一様分布である。

ベータ分布からサンプリングしてみよう。また α , β を変えると分布の様子がどのように変わるか観察してみよう。

```
n <- 1000
alpha <- 3
beta <- 4
x<-rbeta(n = n, shape1 = alpha, shape2 = beta)
```

```
f <- function(x){x^alpha*(1-x)^beta*n*0.1/beta(alpha, beta)}
hist(x)
plot(f,0,1, add=TRUE)
```



事前分布をベータ分布とした時、事後分布は

$$f(p \mid X = k) = L(p) \times f(p) = p^{\alpha-1}(1-p)^{\beta-1} \times p^k(1-p)^{N-k} = p^{\alpha+k-1}(1-p)^{\beta+N-k-1}$$

となるので、またベータ分布になる。

連続型確率分布

改めて連続型確率分布の考え方について整理しよう。

データやパラメータなどの値が有限でない、あるいはそのように近似できる場合を考える。例えば、色々なものの大きさや量を測る場合を考えよう。このようなものを連続型の確率変数という。

それに対して数え上げられるだけの可能性しかとらないものを離散型の確率変数という。

上の例題で言えば箱の種類であったり、ある出来事が何回起こるかといった整数の値で表されるものなどである。

離散型の確率変数の場合、確率分布とは起こりうる結果が $X = x_1, x_2, \dots, x_n$ に対して、

$$P(X = x_1), P(X = x_2), \dots, P(X = x_n)$$

に0以上の数を割り当て、それらの和が1となるように分布させたものだった。

一方で連続型の確率変数の場合、飛び飛びの値ではなくぎっしり詰まった（ある範囲の）数すべてを取るようになる。したがって、ある決まった値をとる k は0でないとおかしなことになってしまう。

連続分布の場合、 $P(a \leq X \leq b)$ のように範囲を指定して確率を計算する。

一般に、確率変数 X が連続分布に従うとき、密度関数と呼ばれる関数 $f(x)$ があって $P(a \leq X \leq b)$ は $f(x)$ と x 軸で囲まれた部分の面積として与えられ、この面積を表す記号として

$$\int_a^b f(x)dx$$

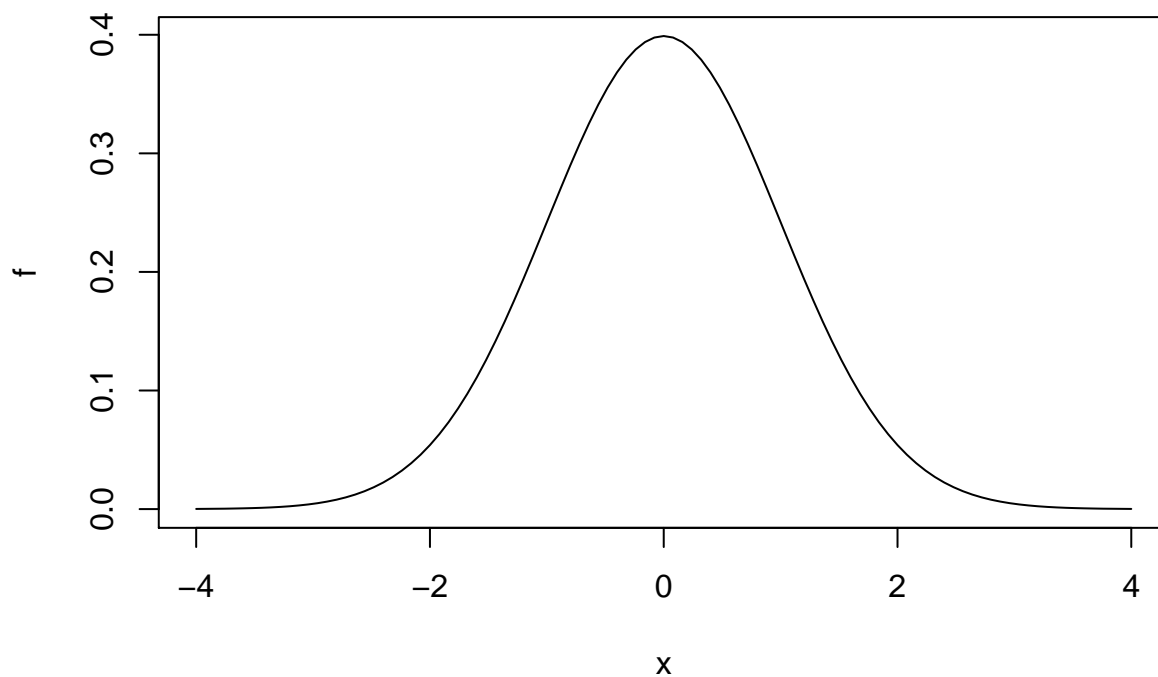
を定める。

例えばあとで見る標準正規分布の場合、その密度関数は

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

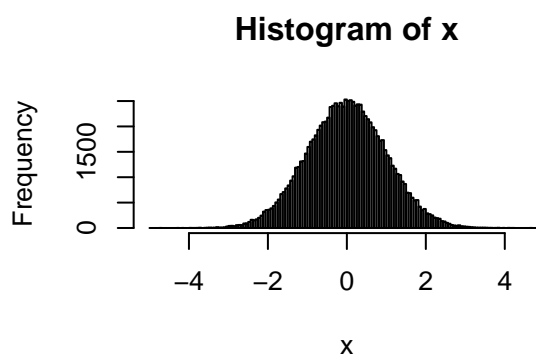
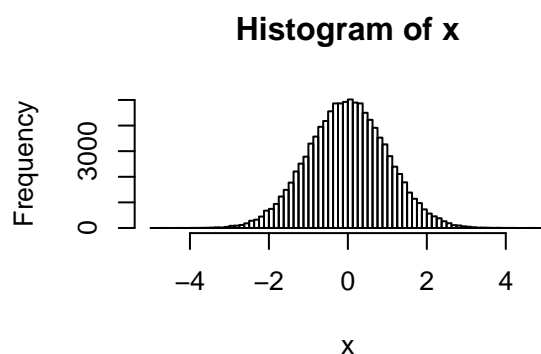
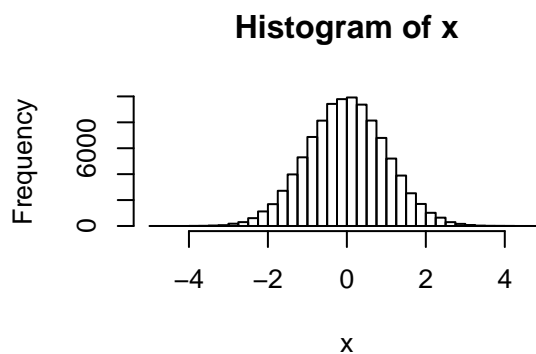
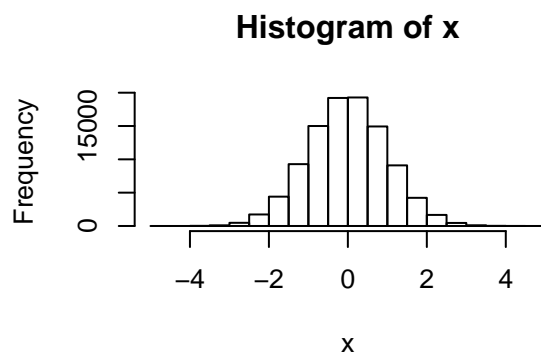
で与えられ、そのグラフは

```
mu <- 0
sigma <- 1
f <- function(x){(1/sqrt(2*pi)*sigma)*exp(-(x-mu)^2/(2*sigma^2))}
plot(f,-4,4)
```



ここから乱数をサンプリングし、ヒストグラムをかくと、

```
N <- 100000
x <- rnorm(N)
par(mfrow=c(2,2))
hist(x,breaks = (0:20)/2-5)
hist(x,breaks = (0:40)/4-5)
hist(x,breaks = (0:80)/8-5)
hist(x,breaks = (0:160)/16-5)
```

となる。

幅を細くしていけば、密度関数の形に近づいていくことが見える。

ヒストグラムでは高さが x がある範囲にあるサンプルの個数で、全体の個数に対する割合が確率であると考えられる。

このように見れば、確率密度関数とその面積を用いて確率を表現できることが納得できる。

連続型確率分布に対するベイズの定理

ベイズ推論に置いて扱う事象は、

- ・ パラメータがどのような条件をみたすか
- ・ 得られたデータがどのようなものであるか

の大きく二つがある。

例えば先ほどの例題で言えば

- ・ 箱の名前、種類を表す変数 X
- ・ 赤いボールを何個取り出したかを表す変数 θ

の二つがあった。

このように二つの連続型確率変数 X, Y を同時に考えたい場合、二変数の密度関数 $f(x, y)$ を与えることで確率分布を記述する。このとき

$$P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f(x, y) dx dy$$

として確率分布が与えられる。

$$f(x) = \int_{-\infty}^{\infty} f(x, y) dy$$

とすれば X の密度関数が得られる。これは離散分布の場合に

$$P(X = x) = \sum_i P(X = x, Y = y_i)$$

としたのと同じこと。

またこの場合に条件付き分布は密度関数

$$f(y | x) = \frac{f(x, y)}{f(x)}$$

により与えられる。

連続分布に対してもベイズの定理が成り立つ。

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

連続型確率分布の例

以下ではいくつかの連続型確率分布の例とそれを用いたベイズ推定の例題を紹介する。

- ・ 一様分布
- ・ 正規分布
- ・ ガンマ分布

一様分布

ある範囲の実数が一様に現れるような確率分布を一様分布という。

例えば0以上1以下の実数が一様に現れるような確率分布であれば、その確率密度関数は

$$f(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & x < 0, 1 < x \end{cases}$$

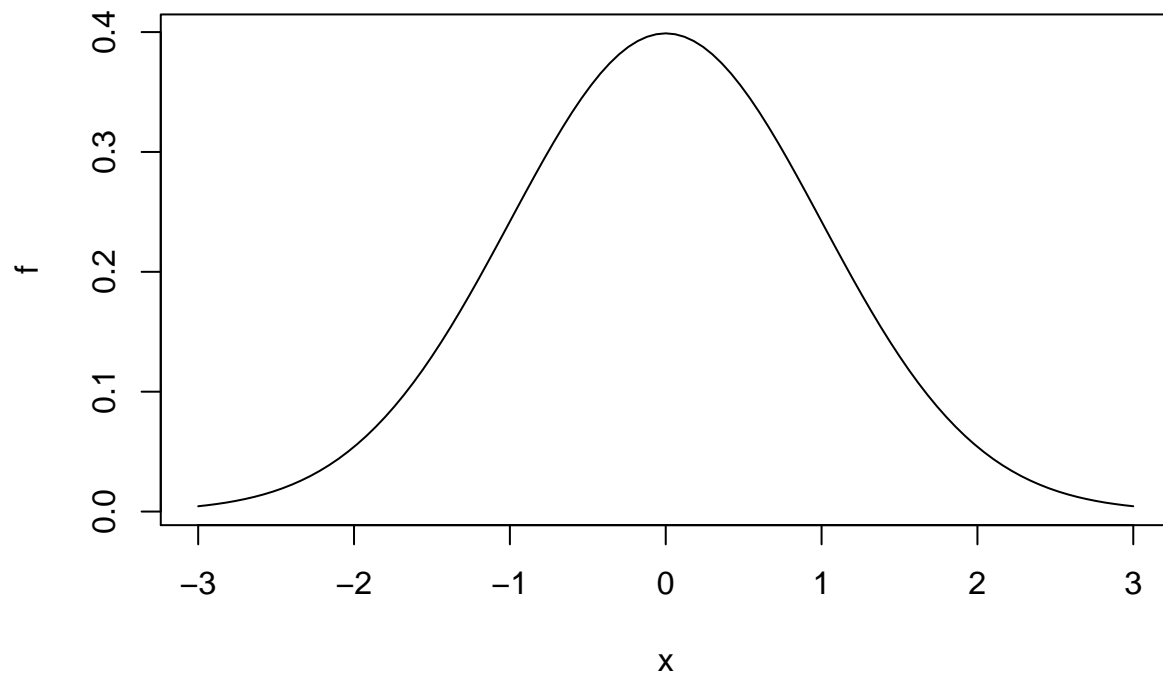
となる。このとき、例えば $P(0.5 \leq x \leq 0.6) = 0.1$ と計算できる。

正規分布

正規分布は次の密度関数 $f(x)$ で定まる確率分布。

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

```
mu <- 0
sigma <- 1
f <- function(x){(1/sqrt(2*pi)*sigma)*exp(-(x-mu)^2/(2*sigma^2))}
plot(f,-3,3)
```



確率はRで計算できる

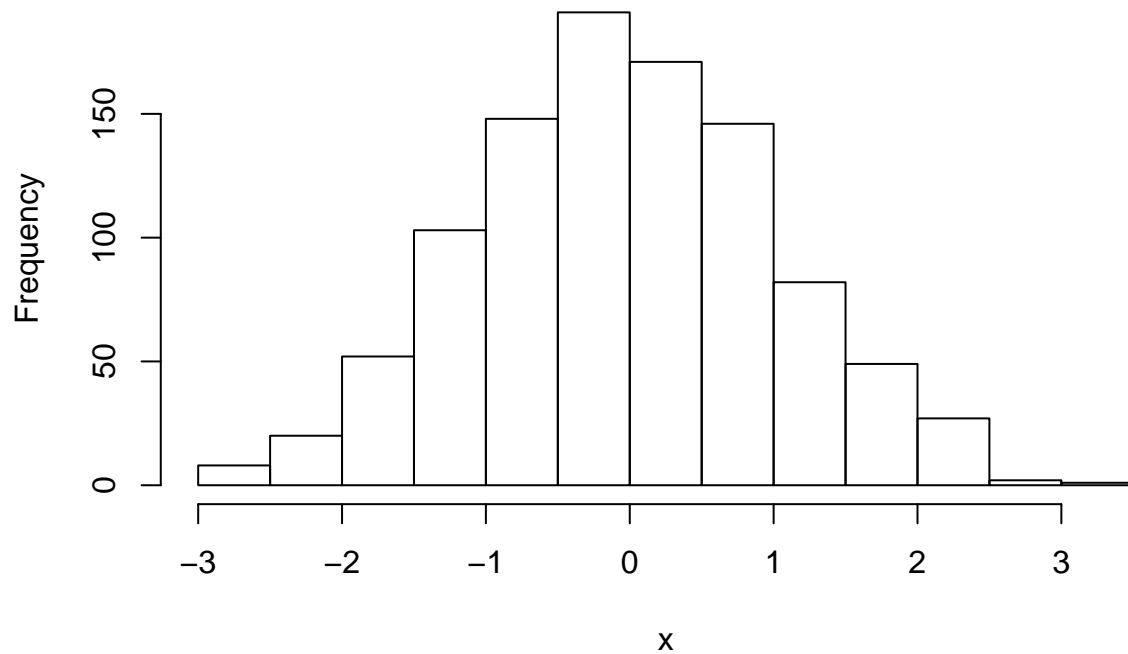
```
pnorm(1)
```

```
## [1] 0.8413447
```

正規分布からのサンプリングは

```
mu <- 0
sigma <- 1
N <- 1000
x <- rnorm(n = N, mean = mu, sd = sigma)
hist(x)
```

Histogram of x



ここで μ と σ は正規分布のパラメータ。

確率密度関数のグラフは $x = \mu$ がピークで左右対称。この幅の広がり具合を決めるのが σ であり、

パラメータ μ, σ の正規分布に従う確率変数は

- ・ 平均 μ
- ・ 分散 σ^2

である。

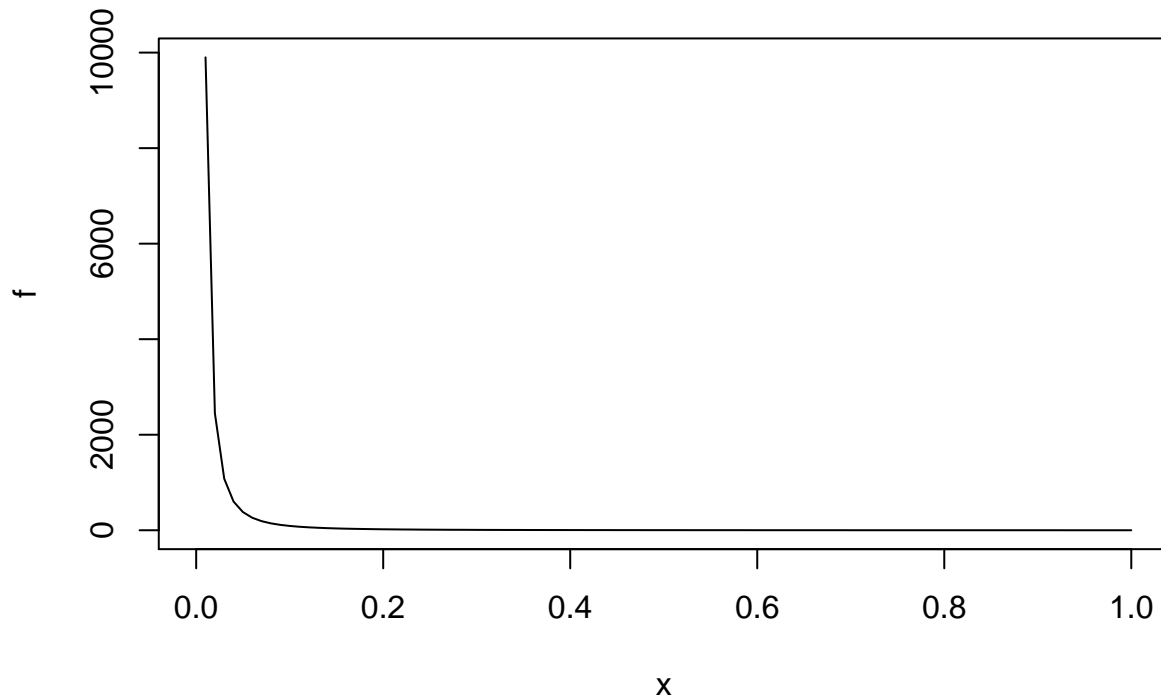
正規分布は誤差の分布として使われる。色々和良好的性質を持ち、もっとも基本的な連続分布。

ガンマ分布

ガンマ分布は次のような密度関数を持つ。

$$f(x) = \frac{\theta^k}{\Gamma(k)} e^{-\theta x} x^{k-1}$$

```
theta <- 2
k <- 2
f <- function(x){exp(-x)*x^{theta}}
plot(f)
```



ガンマ分布のパラメータは k と

連続型確率分布を用いたベイズ推定

それでは改めて連続型確率分布を用いたベイズ推定の問題を考えてみよう。

例題：ポアソン分布のパラメータ推定

ある地域での一日の事故件数 X を予測したい。

これをポアソン分布モデルを用いて予測してみよう。データとして、これまで10日間の事故件数データが、

```
x<-c(8,4,5,5,7,9,7,5,5,4)
```

と与えられているとしよう。

ポアソン分布はパラメータ λ をもつ離散型確率分布である。パラメータ λ のポアソン分布に従う確率変数 X は

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

と確率を計算でき、

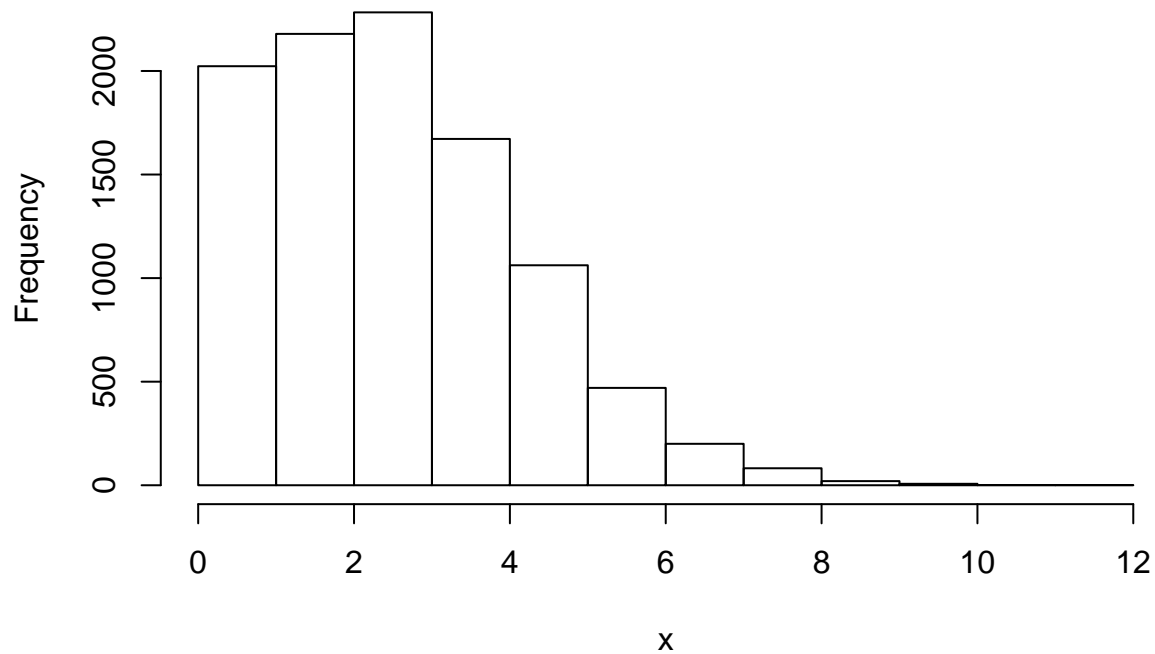
- ・ 平均 λ
- ・ 分散 λ

である。

ポアソン分布からサンプリングした変数のヒストグラムを、 λ の値を変えて観察してみよう。

```
N <- 10000
lambda <- 3
x <- rpois(n = N, lambda = lambda)
hist(x,breaks=0:max(x))
```

Histogram of x



今回のデータから尤度関数を計算すると、

$$L(\lambda) = e^{-\lambda} \frac{\lambda^{x_1}}{x_1!} \times \dots \times e^{-\lambda} \frac{\lambda^{x_n}}{x_n!} = \prod_{i=1}^N e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} = e^{-10\lambda} \frac{\lambda^{59}}{x_1! \dots x_n!}$$

パラメータ λ の事前分布として、ガンマ分布を用いることにする。

事前分布をパラメータ k, θ のガンマ分布としよう。

$$f(\lambda) = \frac{\theta^k}{\Gamma(k)} e^{-\theta\lambda} \lambda^{k-1}$$

すると、事後分布は

$$e^{-(10+\theta)\lambda} \lambda^{59+k-1}$$

により決まるガンマ分布になる。つまりパラメータ k, θ のガンマ分布である。

予測は

例題

上の例題ではパラメータの分布に連続型確率分布を用いたが、データの分布にも連続型分布を用いることができる。そのような例題として、正規分布に

正規分布の例題

工業製品の規格を調査する。ある工場で作られた製品の大きさが、以下のようであった。

```
x <- c(102.3, 100.4, 99.3, 100.2, 100.3, 99.4, 101.5, 99.3, 98.9, 100.1)
```

製品が規格通り製造されているか調査したい。

正規分布モデルを用いる。パラメータは平均 μ と標準偏差 σ で、これに対して上のデータから尤度関数を計算すると

$$L(\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_1 - \mu)^2}{2\sigma^2}\right) \times \cdots \times \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_N - \mu)^2}{2\sigma^2}\right) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right)$$

である。

```
x <- c(102.3,100.4,99.3,100.2,100.3,99.4,101.5,99.3,98.9,100.1)
L <- function(mu, sigma){
  L <- 1
  for(i in 1:10){
    L <- L*dnorm(x[i],mu,sigma)
  }
}
```

さてここでは

- ・ 平均の事前分布として、正規分布
- ・ 分散の事前分布として、逆ガンマ分布

を用いることにしよう。

すると事後分布は

となる。

これは σ, μ について、二変数の確率分布となっているが、 σ に適当な値を入れれば正規分布、 μ に適当な値を入れれば逆ガンマ分布になる。

この事後分布を用いての予測をする。例えば規格外の製品が製造される確率は、

と計算できる。

自然な共役分布

一般的にはパラメータの事後分布の密度関数を正確に計算したり、それを用いてパラメータ推定値の平均や分散、あるいは予測分布を計算するのは困難で

これを克服するための一つの方法は、事前分布として自然共役分布と呼ばれる特定の分布を用いることである。ベイズ推論の枠組みで、尤度関数はモデル

例えば

- ・ 二項分布のパラメータ推定にベータ分布
- ・ 正規分布のパラメータ推定に正規分布、逆ガンマ分布
- ・ ポアソン分布のパラメータ推定にガンマ分布

などがある。これらを使えば、正確に事後分布を決定することができる。

階層モデル

ある学校で問題数10問のテストを20人の生徒に行ったところ、以下のような結果をえた。生徒が問題に正解する能力を正解率 p として、これをパラメータとする二項分布モデルで考えてみよう。

すると、二項分布で想定される分布の様子、あるいは分散の大きさと異なってしまう。

このような場合、正解率に個人差があると考える。ただし個々の正解率をパラメータ p_1, p_2, \dots, p_{20} とするのではなく、正解率が別の分布に従って決ま

このようなモデルを階層モデルという。

マルコフ連鎖モンテカルロ法(MCMC)

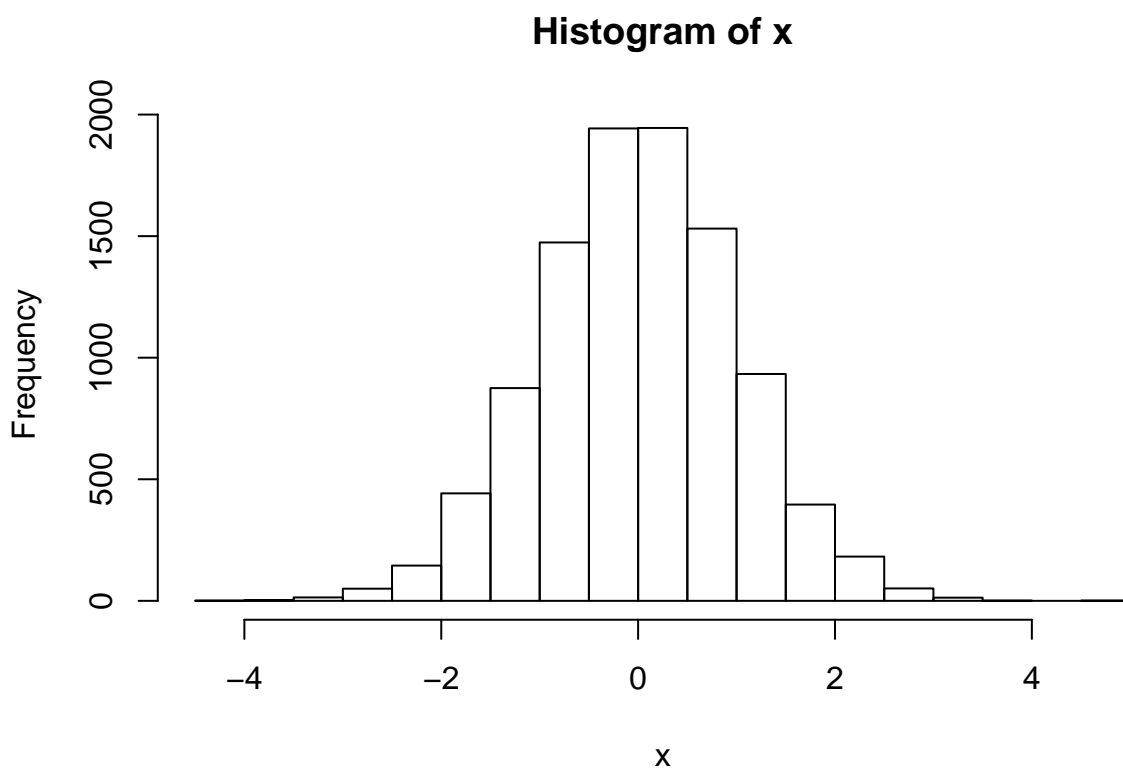
例えば階層モデルなどより複雑なモデルなどの場合にも、コンピュータの助けを借りることでパラメータの推定やデータの予測が可能になる。これを実現するための一つの方法がMCMCを用いたランダムサンプリングである。

ただし、この場合には正確な分布を決定するわけではなく、事後分布に従ったランダムサンプリングが近似的にできるのみであることに注意しよう。問題設定によっては収束しにくい分布などもあるので、上の自然な共役分布を使った場合と違ってより慎重な運用が必要になる。

MCMCの概要

確率密度関数が与えられたとき、その密度関数に従ってサンプリングを行いたい。よく知られている確率密度関数であれば、以下のようにRでも簡単にサ

```
N<-10000
x<-rnorm(N)
hist(x)
```



しかし、より複雑なモデリングを考える場合には、あらかじめ用意されている関数では乱数をサンプリングできない。

モンテカルロ法の例

0または1の値をとり、1をとる確率が円の面積の1/4になるような乱数を生成しよう。一様分布に従う乱数はあらかじめ用意されているとする。

```
sample <- function(n){
  x <- c()
  for(i in 1:n){
    r <- runif(2)
    x[i] <- ifelse(r[1]^2+r[2]^2<1, 1, 0)
  }
  return(x)
}
```



```
N <- 1000
4*sum(sample(N))/N
```

```
## [1] 3.152
```

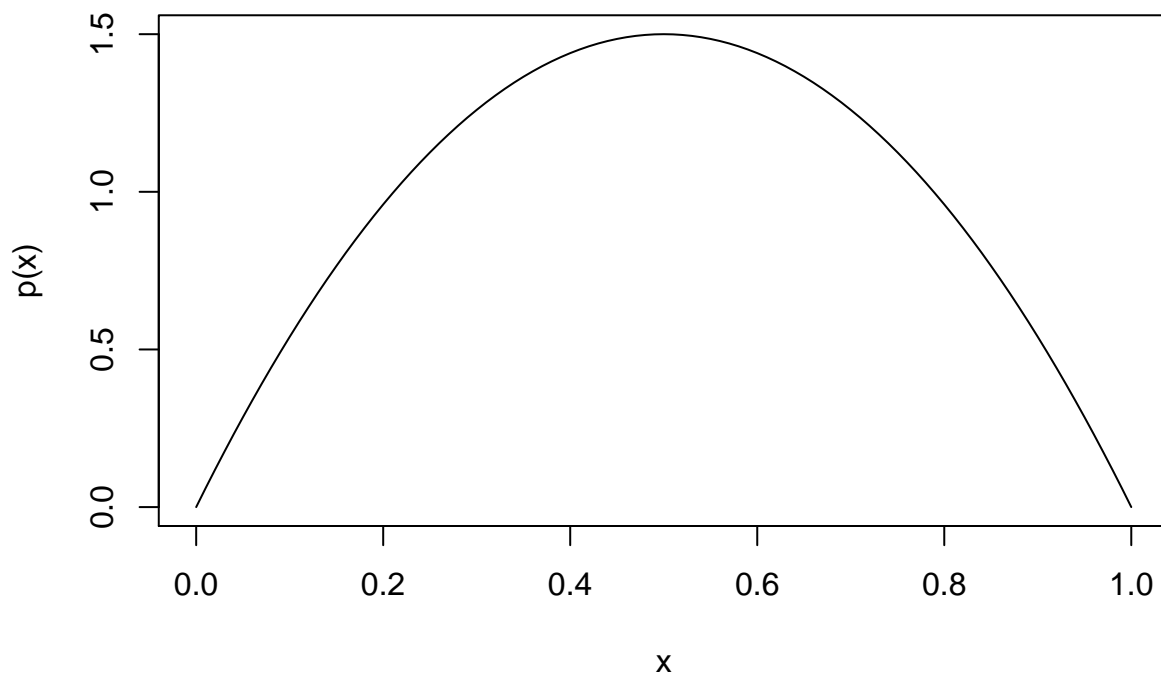
メトロポリス法

一様乱数があらかじめ与えられている時、確率密度関数 $f(x)$ に従う乱数 $\{x_1, x_2, \dots\}$ を以下の手順でサンプリングする。

x_i から x_{i+1} を作る方法。

- ・ まず適当な一様乱数から ϵ を生成し、次ステップの候補を $x' = x_i + \epsilon$ とする。
- ・ 次のような条件で実際に $x_{i+1} = x'$ とするか、あるいは $x_{i+1} = x_i$ のままにするかを決める。
- ・ $r = f(x_{i+1})/f(x_i)$ とする。
- ・ r が1より大きい時、 $x_{i+1} = x'$ とする。
- ・ r が1より小さい場合、確率 r で $x_{i+1} = x'$ にする。
- ・ 実際には、0から1の間の一様乱数 q を発生させ、 $r > q$ なら $x_{i+1} = x'$ とし $r < q$ なら $x_{i+1} = x_i$ とする。

```
p <- function(x){
  return(-6*x*(x-1))
}
x<-0:100/100
plot(x,p(x),'l')
```



```
step <- function(x){
  x_next <- x + runif(1, min=-1, max=1)#
  r <- p(x_next)/p(x)
  if(r > 1){
    return(x_next)
  }
  else{
    q <- runif(1)
    if(r>q) return(x_next)
  }
}
```

```
    else return(x)
  }
}
sample <- c()
sample[1] <- runif(1)
N <- 1000
for(i in 1:(N-1)){
  sample[i+1] <- step(sample[i])
}
hist(sample)
```

Histogram of sample

