学習エンジンとしてのベイズ確率

M.Shimura
JCD02773@nifty.ne.jp

2008 年 5 月 21 日

目次

1 **ベイズの定理** 2

2 **ベイズと信頼性の計測** 2

3 **ベルヌイ試行とベイズ確率** 5

4 **ポートフォリオの構築/正規分布** 12

5 **ポートフォリオ** 16

6 Reference 21

付録 A **ベイズ確率** 22

付録 B **ベイズに馴染もう** 23

付録 C **ベイズ確率の事例抄** 25

付録 D calc memo 26

概要

250 年前のベイズ確率が注目され、スパムフィルタや金融工学などで活躍している。中妻照雄「入門　ベイス統計学」の comparison としてベイズ理論にベルヌイ分布を用いて信頼性の計測を、正規分布を用いてポートフォリオの作成を実際に行ってみた

1　ベイズの定理

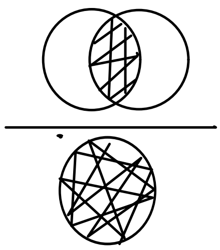
■Thomas Bayes　(1702-1761)

ロンドン生まれ。父も牧師。エディンバラ大学で宗教学を学ぶ（数学も学んだようだ）[*1]イングランドの長老派の牧師で数学者。国教徒ではなかった。

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A|B) \times P(B)}{P(A|B) \times P(B) + P(A|B^c) \times P(B^c)}$$

$P(A|B)P(B) = P(A|B^c)P(B^c) = P(A \cap B) + P(A \cap B^c) = P(A)$

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



CAP

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A|B) \times P(B)}{P(A|B) \times P(B) + P(A|B^c) \times P(B^c)}$$

$\frac{P(A \cap B)}{P(A)}$

cap（共通部分）を全体で割ったもの

2　ベイズと信頼性の計測

2.1　仏の顔も 3 度まで

この格言は江戸期の大阪カルタで用いられ、近松門左衛門「冥土の飛脚」にもでてくる話が起源のようだ。建築設計、建築資材、伊勢名物や、思いも寄らない年賀葉書の古紙やプラスチックのリサイクル表示に広がり、超高級料亭がとどめを刺したようだ。

次の EXAMPLE は納品の遅れがどのように信用に影響するかの計算方法の例である。大型地震で、ある地方の部品産業が損壊すると各地で生産が止まる様子が報道され、僅少在庫ストックの実体がが垣間見られる。

[*1] 国教徒ではないので $Oxford$ や $Cambridge$ では学べなかった。

ジャストインタイムに過度に依存する程工場は早く止まる。

この信頼性に関するベイズ理論はスパムフィルタにも応用されて活躍している。

■Woeked Example(1)　issue:Nakatsuma p36

この例題は中妻による。

2.1.1　事後確率

事象 A：納期に遅れる

事象 B：X 社が信頼できる企業とわかる

ここでは信頼できる企業でも $\frac{1}{20}$ の割合で納期に遅れる。

X 社の信頼度の事前確率は最初だから $\frac{1}{2}$ とする。

	納期に遅れる A	納期を守る A^c
信頼できる (B)	$P(A\|B)$	$P(A^c\|B)$
信頼できない (B^c)	$P(B^c)$	$P(B)$

信頼できる企業の納期に関する条件付き確率の例	納期に遅れる A	納期を守る A^c
信頼できる (B)	1/20	19/20
信頼できない (B^c)	1/2	1/2

必要なパラメータは $P(A|B)$ と $P(A|B^c)$ と $P(B)$

信頼度は何となく $\dfrac{1}{2}$ からスタートした場合。

$$
\begin{array}{r|lcl}
1 & P(B|A) & = \dfrac{P(A|B)P(B)}{P(A|B)P(B)+P(A|B^c)P(B^c)} & = \dfrac{(1/20)(1/2)}{(1/20)(1/2)+(1/2)(1/2)} = \dfrac{1}{11} = P(B|A_1) \\
2 & P(B|A_1 \cap A_2) & = \dfrac{P(A_2|B)P(B|A_1)}{P(A_2|B)P(B|A_1)+P(A_2|B^c)P(B^c|A_1)} & = \dfrac{(1/20)(1/11)}{(1/20)(1/11)+(1/2)(10/11)} = \dfrac{1}{101} \\
3 & P(B|A_1 \cap A_2 \cap A_3) & = \dfrac{P(A_3|B)P(B|A_1 \cap A_2)}{P(A_3|B)P(B|A_1 \cap A_2)+P(A_3|B^c)P(B^c|A_1 \cap A_2)} & = \dfrac{(1/20)(1/101)}{(1/20)(1/101)+(1/2)(100/101)} = \dfrac{1}{1001}
\end{array}
$$

信頼を一回失うと事後確率は著しく低下する。

事後確率の変化	(事前確率)	遅れ 1 の事後確率	遅れ 2 の事後確率	遅れ 3 の事後確率
信頼できる (B)	1/2	1/11	1/101	1/1001
信頼できない (B^c)	1/2	10/11	100/101	1000/1001

2.1.2　目標未達成度

この部品製作会社の事後分布を使って、今後の取引の方向を定めたい。

目標未達成度によってリスクの比較を行う。

$$
\text{目標未達成度} = \begin{cases} max(50 - {}_{-}50, 0) = 100 & \text{(納期に遅れる)} \\ max(0 - 100, 0) = 0 & \text{（納期を守る）} \end{cases}
$$

部品製作会社 X 社が納期を守ると A 社は 100 万円の利益。遅れると A 社は 50 万円の損失。

2.1.3　予測分布

リスクの予測分布を求める。事前分布では納期リスクは広く $\frac{1}{2}$ と取っておいた。事後分布のデータにより、予測分布を求める。

$$
\begin{array}{r|ll}
(0) & P(A_1) = P(A_1|B)P(B) + P(A_1|B^c)P(B^c) & = 1/20 \times 1/2 + 1/2 \times 1/2 = 11/40 \\
(1) & P(A_2|A_1) = P(A_2|B)P(B|A_1) + P(A_2|B^c)P(B^c|A_1) & = 1/20 \times 1/11 + 1/2 \times 10/11 = 101/220 \\
(2) & P(A_3|A_1 \cap A_2) = P(A_3|B)P(B|A_1 \cap A_2) + P(A_3|B^c)P(B^c|A_1 \cap A_2) & = 1/20 \times 1/101 + 1/2 \times 100/101 = 1001/2020
\end{array}
$$

	目標に至らない	事前予測 (0)	事後予測 (1)	事後予測 (2)
納期に遅れる	100	11/40	101/220	1001/2020
納期を守る	0	29/40	119/220	1019/2020
risk		11/40 = 27.5	101/220 = 45.1	1001/2020 = 49.55

事後分布によりリスクの予測分布が徐々に固まってくる。

2.1.4　計算結果

```
   (100;0) p_pred 1r20;1r2;1r2
+-+----------+---------+-------+
| |believable|pred/delay|risk-% |
+-+----------+---------+-------+
|0|1r2       |1r20     |-      |
+-+----------+---------+-------+
|1|1r11      |11r40    |27.5   |
+-+----------+---------+-------+
|2|1r101     |101r220  |45.9091|
+-+----------+---------+-------+
|3|1r1001    |1001r2020|49.5545|
+-+----------+---------+-------+
```

1 回の信頼失墜は $\frac{1}{2}$から$\frac{1}{11}$ へ
2 回目は相当厳しい。

```
   (100;0) p_pred 1r20;1r2;9r10
+-+----------+---------+-------+
| |believable|pred/delay|risk-% |
+-+----------+---------+-------+
|0|9r10      |1r20     |-      |
+-+----------+---------+-------+
|1|9r19      |19r200   |9.5    |
+-+----------+---------+-------+
|2|9r109     |109r380  |28.6842|
```

```
+-+----------+---------+-------+
|3|9r1009    |1009r2180 |46.2844|
+-+----------+---------+-------+
```

当初の信頼が 90 %の場合。もう 1 回信頼回復のチャンスがありそうだ。
$\frac{9}{10} \to \frac{9}{19} \to \frac{9}{109}$

3　ベルヌイ試行とベイズ確率

ベルヌイ試行はコインの表裏、相撲の勝ち負け、納期に遅れる · 守るなどの　ビットで言えば 0/1 の結果が出るトライアルである。

3.1　成功確率の事後分布

3.1.1　ベルヌイ分布

$X_1 = \begin{cases} 1 & (\text{遅れる}) \\ 0 & (\text{守る}) \end{cases}$	$Pr(X_1 = x_i) = \begin{cases} \pi & (x_i = 1) \\ 1 - \pi & (x_i = 0) \end{cases}$
ベルヌイ分布	$p(x_i \vert \pi) = Pr(X_i = x_i) = \pi^{x_i}(1-\pi)^{1-x_i}, \ (x_i = 0, 1)$

[*2]

ベルヌイ分布により推計されたデータから未知のパラメータ π を推計する。

納期遅延のサンプル（X 社）

	1	2	3	4	5
納期	×	○	×	○	○
A, A^c	A	A^c	A	A^c	A^c
x_i（ビット）	1	0	1	0	0

ベイズの定理による納期の事後分布 (X 社が納期に遅れる確率 π の事後分布 $p(\pi \vert x_1)$	$p(\pi \vert x_i) = \dfrac{p(x_i \vert \pi)p(\pi)}{\int_0^1 p(x_i \vert \pi)p(\pi)d\pi}$ この式の分母は π に依存していないので次のように表すことができる $p(\pi \vert x_i) \propto p(x_1 \vert \pi)p(\pi)$

[*2] π は円周率ではなく、確率を表す単なる記号

■Worked Example　計算例

pi の取りうる値 (X 社) は $\frac{1}{20}$ と $\frac{1}{2}$ とする。(事前分布)

$$p(\pi) = \begin{cases} \frac{1}{2} & (\pi = \frac{1}{20}) \\ \frac{1}{2} & (\pi = \frac{1}{2}) \\ 0 & (\text{それ以外}) \end{cases}$$

$$p(\pi = \frac{1}{20}|x_1 = 1) = \frac{(1/20)(1/2)}{(1/20)(1/2) + (1/2)(1/2)} = \frac{1}{11}$$

π の取りうる値が J 通り $(\pi_1, \cdots, \pi_J)$ の場合の π の事後分布	$p(\pi_j \mid x_i = 1) = \frac{p(x_i = 1 \mid \pi_j)p(\pi_j)}{\sum_{j=1}^{J} p(x_i = 1 \mid \pi_j)p(\pi_j)}, \ (j = 1, \cdots J$

3.2　事後情報による逐次更新

3.2.1　β　分布

pi の事後分布のカーネルを求めるには積分を求めなければならないが、ベータ関数を用いると簡単に導出できる。

ベルヌイ分布　2 項分布の $P = 1.q = 1 - p = 0$ の場合である。

β　分布　$B(x, y) = \int_0^1 u^{x-1}(1-u)^{y-1}du$　　$x > 0, y > 0$

$B(m, n) = \frac{(m-1)!(n-1)!}{(m+n-1)!}$　　m, n は自然数

3.2.2　β 関数から規準化定数を求める

$$\int_0^1 \pi d\pi = B(2,1) = \frac{(2-1)!(1-1)!}{(2+1-1)!} = \frac{1}{2}$$
$$\int_0^1 \pi(1-\pi)d\pi = B(2,2) = \frac{(2-1)!(2-1)!}{(2+2-1)!} = \frac{1}{6}$$
$$\int_0^1 \pi^2(1-\pi)d\pi = B(3,2) = \frac{(3-1)!(2-1)!}{(3+2-1)!} = \frac{1}{12}$$
$$\int_0^1 \pi^2(1-\pi)^2d\pi = B(3,3) = \frac{(3-1)!(3-1)!}{(3+3-1)!} = \frac{1}{30}$$
$$\int_0^1 \pi^2(1-\pi)^3d\pi = B(3,4) = \frac{(3-1)!(4-1)!}{(3+4-1)!} = \frac{1}{60}$$

ベータ分布　$B(m, n) = \frac{(m-1)!(n-1)!}{(m+n-1)!}$ を J で計算する。

```
 a,:  calc_beta_sub (L:0) a=. 2 1; 2 2;3 2;3 3;3 4
+---+--------+---------+---------+---------+
|2 1|2 2     |3 2      |3 3      |3 4      |NB. B(2,1)..
+---+--------+---------+---------+---------+
|0.5|0.166667|0.0833333|0.0333333|0.0166667|
```

```
+---+--------+---------+---------+---------+
1r2  1r6     1r12      1r30      1r60   NB.
```

3.2.3 π の事後分布

[*3]

$p(\pi|(x_1 = 1) \propto \pi$
$p(\pi|(x_1, x_2) = (1, 0)) \propto \pi(1 - \pi)$
$p(\pi|(x_1, x_2, x_3) = (1, 0, 1)) \propto \pi^2(1 - \pi)$
$p(\pi|(x_1, x_2, x_3, x_4) = (1, 0, 1, 0)) \propto \pi^2(1 - \pi)^2$
$p(\pi|(x_1, x_2, x_3, x_4, x_5) = (1, 0, 1, 0, 0)) \propto \pi^2(1 - \pi)^3$

$p(\pi|(x_1 = 1) = 2\pi\mathbf{1}(0, 1)(\pi)$
$p(\pi|(x_1, x_2) = (1, 0)) = 6\pi(1 - \pi)\mathbf{1}(0, 1)(\pi)$
$p(\pi|(x_1, x_2, x_3) = (1, 0, 1)) = 12\pi^2(1 - \pi)\mathbf{1}(0, 1)(\pi)$
$p(\pi|(x_1, x_2, x_3, x_4) = (1, 0, 1, 0)) = 30\pi^2(1 - \pi)^2\mathbf{1}(0, 1)(\pi)$
$p(\pi|(x_1, x_2, x_3, x_4, x_5) = (1, 0, 1, 0, 0)) = 60\pi^2(1 - \pi)^3\mathbf{1}(0, 1)(\pi)$

[*3] 指示関数の表記 $p(\pi|x_1 = 1) = \begin{cases} 2\pi & (a < x < b) \\ 0 & (x \leq a, b \leq x) \end{cases}$
は
$p(\pi|x_1 = 1) = 2\pi\mathbf{1}_{(0,1)}(\pi)$
とあらわすと $select - case$ の表記が省略できる

事前分布=1

事前分布は全て 1 の横線

1 回目

$\int_0^1 \pi d\pi = B(2,1) = \frac{(2-1)!(1-1)!}{(2+1-1)!} = \frac{1}{2}$

$p(\pi|(x_1 = 1) = 2\pi 1_{0.1}(\pi)$

plot_beyes0 2 1

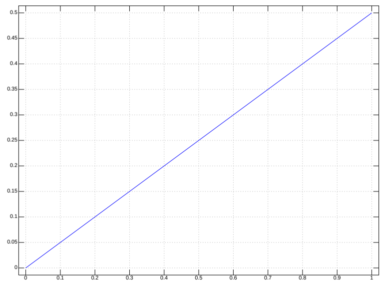


図 1　B(2,1)

2 回目

$\int_0^1 \pi(1-\pi) d\pi = B(2,2) = \frac{(2-1)!(2-1)!}{(2+2-1)!} = \frac{1}{6}$

$p(\pi|(x_1, x_2) = (1,0)) = 6\pi^2(1-\pi)1_{0.1}(\pi)$

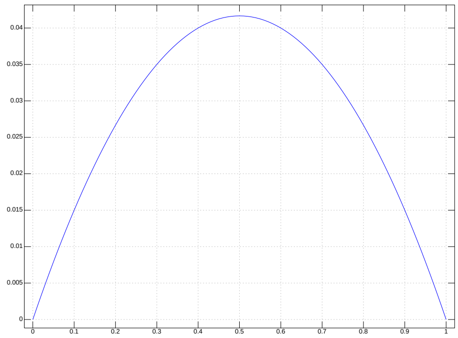


図 2　B(2,2)

3 回目

$\int_0^1 \pi^2(1-\pi) d\pi = B(3,2) = \frac{(3-1)!(2-1)!}{(3+2-1)!} = \frac{1}{12}$

$p(\pi|(x_1, x_2, x_3) = (1,0,1)) = 12\pi^2(1-\pi)1_{0.1}(\pi)$

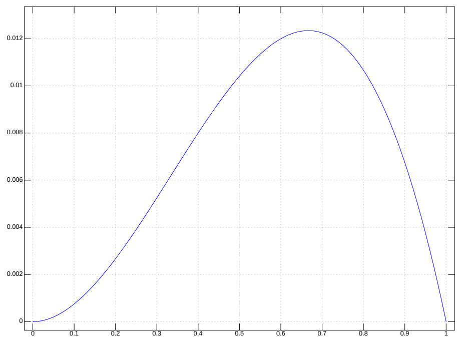


図 3　B(3,2)

3.3　一括資料で事後分布を求める

逐次で求めても一括で求めても計算方法は同じである。

$p(D|\pi) = \Pi_{i=1}^n p(x_i|\pi) = \Pi_{i=1}^n \pi^{x_i}(1-\pi)^{1-x_i}$

$= \pi^{\sum_{i=1}^n x_i}(1-\pi)^{n-\sum_{i=1}^n x_i}$

$p(D|\pi) = \pi^{y_n}(1-\pi)^{n-y_n}$

$p(\pi|D) = \frac{\pi^{y_n}(1-\pi)^{n-y_n}}{B(y_n+1, n-y_n+1)} 1_{0,1}(\pi)$

4 回目

$\int_0^1 \pi^2(1-\pi)^2 d\pi = B(3,3) = \frac{(3-1)!(3-1)!}{(3+3-1)!} = \frac{1}{30}$

$p(\pi|(x_1,x_2,x_3,x_4)=(1,0,1,0)) = 30\pi^2(1-\pi)^2 1_{0.1}(\pi)$

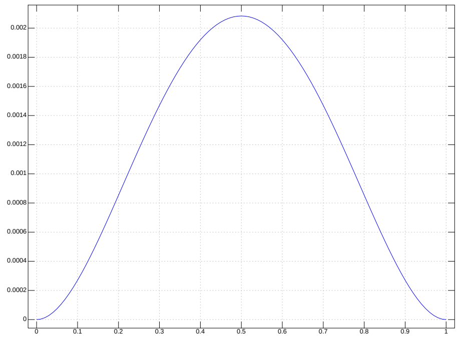


図 4　B(3,3)

5 回目

$\int_0^1 \pi^2(1-\pi)^3 d\pi = B(3,4) = \frac{(3-1)!(4-1)!}{(3+4-1)!} = \frac{1}{60}$

$p(\pi|(x_1,x_2,x_3,x_4,x_5) = (1,0,1,0,0)) = 60\pi^2(1-\pi)^3 1_{0.1}(\pi)$

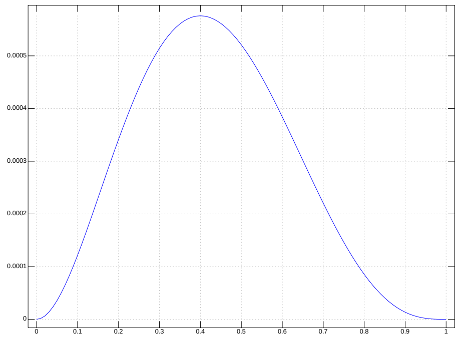


図 5　B(3,4)

3.4　予測分布とベイズ推論

予測分布 $p(\tilde{x}|D) = \frac{p(\tilde{x},D)}{p(D)}$

$$p(\tilde{x}|D) = \frac{p(\tilde{x},D)}{p(D)} = \frac{\int_0^1 p(\tilde{x},D|\pi)p(\pi)d\pi}{\int_0^1 p(D|\pi)p(\pi)d\pi}$$

$$p(\tilde{x},D)|\pi)p(\pi) = p(\tilde{x}|\pi)p(D|\pi)p(\pi)$$

予測分布 $p(\tilde{x}|D) = \int_0^1 p(\tilde{x},\pi)p(\pi|D)d\pi$

企業が破綻$\tilde{x}=1$　$p(\tilde{x}=1|D) = \frac{B(y+2.n-y+1)}{B(y+1.n-y+1)}$

ベータ関数から　$B(y+2.n-y+1)B(y+1.n-y+1) = \frac{(y+1)y!(n-y)!}{(n+2)!}$

$$= \frac{(y+1)!(n-y)!}{(n+2)(n+1)!} = \frac{y+1}{n+2}B(y+1,n-y+1)$$

企業が破綻する予測確率　$p(\tilde{x}=1|D) = \frac{y+1}{n+2}$

企業が破綻しない予測確率　$p(\tilde{x}=0|D) = \frac{n-y+1}{n+2}$

企業破綻の予測分布　$p(\tilde{x}|D) = \left(\frac{n+1}{n+2}\right)^{\tilde{x}}\left(\frac{1}{n+2}\right)^{1-\tilde{x}}$

Worked Example

$$\text{目標達成度} = \begin{cases} 100 & \times \\ 0 & ◯ \end{cases} \quad \text{(100 とするとリスク評価の値が 100 になる)}$$

$$\text{リスク} = \frac{n+1}{n+2} \times + \frac{1}{n+2} \times 0 = \frac{n+1}{n+2} \times 100$$

100：0 ならば手計算でできる。こちらの方が理解しやすい。

	リスク (×100 = %)
1 回目	2/3
2 回目	3/4
3 回目	4/5
10 回目	11/12
100 回目	101/102

```
   1.0  risk_prob L:0 {@> 1 2 3 10 100
+--------+----+---+--------+--------+
|       1|   2|  3|      10|     100|
|0.666667|0.75|0.8|0.916667|0.990196|
+--------+----+---+--------+--------+
```

3.5 Script

calc_beta_sub 規準化定数を求める。 $\int_0^1 \pi^2(1-\pi)^3 d\pi = B(3,4) = \frac{(3-1)!(4-1)!}{(3+4-1)!} = \frac{1}{60}$

*calc_beyes*0 0 から 1 の間を 100 等分して、各点での値を求める $p(\pi|(x_1, x_2, x_3, x_4, x_5) = (1,0,1,0,0)) = 60\pi^2(1-\pi)^3 1_{0.1}(\pi)$

$safe = 0/out = 1$

```
NB. ==============================
NB. *calc_beta_sub 3 2
NB. *plot_beyse0 3 2
NB. 3 is safe=0/ 2 is out=1
NB. ------------------------------
calc_beta_sub=: 3 : 0
NB. e.g. B(3,2) Nakazuma p54
NB. Usage: u 3 2//sum of safe=0/out=1
'M0 N0'=. y
(*/(! <: M0),! <: N0) % ! <: +/ M0,N0
)

calc_beyes0=: 3 : 0
```

```
NB.  plot calc_beyes0 3 2
'P0 P1'=. <: y
BP=. calc_beta_sub y
STEP=. steps 0 1 100
STEP;BP* (STEP^P0)*(1-STEP)^P1
)

plot_beyes0=: 3 : 'plot calc_beyes0 y'
NB. ----------------------
risk_prob=: 4 : 0
NB. x is p // y is n
NB.    1.0 risk_prob  L:0 {@>1 2 3 10 100
TMP0=. x * (>: y) % 2 + y
TMP1=.(1-x)* % 2 + y
,. y , TMP0 + TMP1
)
```

4　ポートフォリオの構築/正規分布

4.1　正規分布

幾つかの正規分布の関数をレビューしてみよう。

4.1.1　鈴木のスクリプト

Ｊのライブラリに入っている（予定）鈴木の正規分布の関数

標準正規分布の確率密度関数の 0, 0.5, 1 での値	`   ndens 0 0.5 1`
	`0.398942 0.352065 0.241971`
	
	`   ndens i:3`
	`0.00443185 0.053991 0.241971 0.398942 0.241971`
	`0.053991 0.00443185`

x で与えた平均と分散をもつ正規分布の確率密度を与える	`   0 1 nden 0 0.5 1`
	`0.398942 0.352065 0.241971`
	
	`   0 1 nden i:3`
	`0.00443185 0.053991 0.241971 0.398942 0.241971`
	`0.053991 0.00443185`

0 から y の絶対値までの標準正規分布の積分を与える	`   ndfs 1`
	`0.341344`
	
	`   ndfs _1`
	`0.341344`

密度関数の y までの積分の値	`ndf1 1.96`
	`0.975002`
	
	`ndf1 0.3685 NB. BS Model`
	`0.64375`
	
	`ndf1 0.2133 NB. BS Model`
	`0.584454`

x から y までの標準正規分布の積分の値	`1 ndf2 2` `0.135905`

4.1.2　J **のライブラリの関数**

```
   normalprob 0 1  0   NB. mean var point
0.5

   normalprob 0 1  1.96
0.0249978

  1- normalprob 0 1  1.96
0.975002
```

4.1.3　E.Show

```
 plot tmp;n01pdf tmp=. steps _3 3 100
   pd 'eps /temp/normal_01.eps'



   ndens i:3
0.00443185 0.053991 0.241971 0.398942 0.241971 0.053991 0.00443185

   n01pdf i:3
0.00443185 0.053991 0.241971 0.398942 0.241971 0.053991 0.00443185

   n01cdf i:3
0.0013499 0.0227501 0.158655 0.5 0.841345 0.97725 0.99865
```

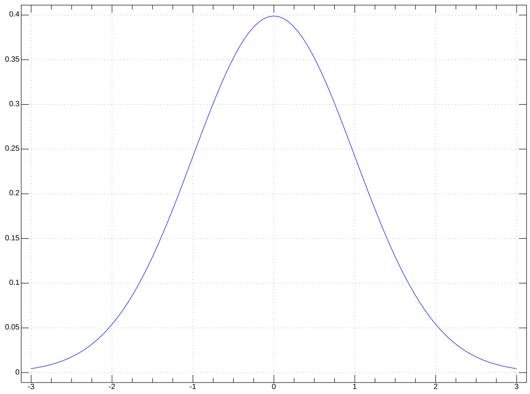


図 6　normal distribution

4.1.4　正規分布のロケーションとスケール

正規分布のパラメータの意味と機能を確認しておく。

1. 正規分布:$N(\mu, \sigma^2)$
2. μ は分布の山の位置を決定する (Location parameter)
3. σ^2 は分布の広がりを決定する Scale Parameter である。
4. この分散（σ^2）は金融ではボラティリティと呼ばれる。

```
'key 0 1' plot tmp; ;("1) ,.
(0 1;1 1) nden L:0 tmp=. steps _3 3 100

 (ndens i:3),.  0 1 nden i:3
0.00443185 0.00443185
  0.053991   0.053991
  0.241971   0.241971
  0.398942   0.398942
  0.241971   0.241971
  0.053991   0.053991
0.00443185 0.00443185
```

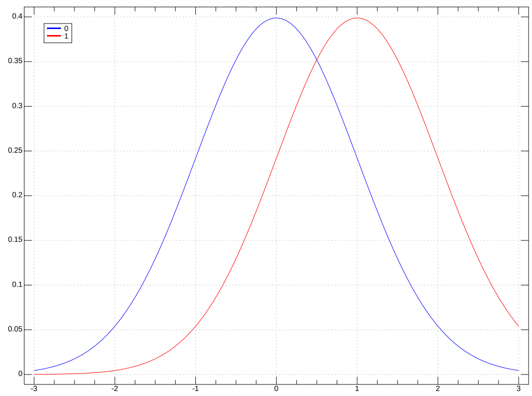


図 7　$\mu = 0, 1$

Scale

```
'key 0 1' plot tmp; ;("1) ,.
(0 1;0 2) nden L:0 tmp=. steps _3 3 100
```

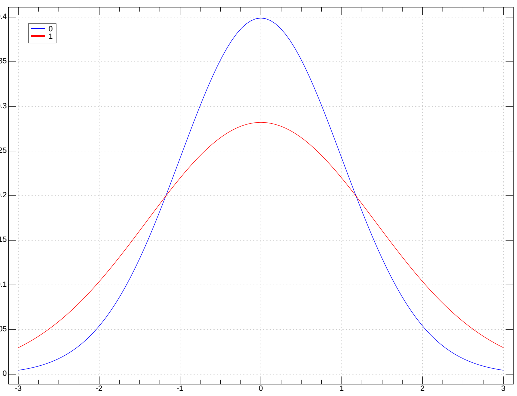


図 8　$\sigma^2 = 1, 2$

5 ポートフォリオ

5.1 期待収益率

Example:日経平均の年次収益率と定期預金金利（一年、年利）は次のようである。
(issue:Nakatsuma)

```
  D0
       NK   IR
2000  4.59 0.17
2001  4.36 0.06
2002  2.07 0.04
2003  8.67 0.04
2004 16.78 0.04
2005 17.48 0.07
```

5.2 期待収益率の事前分布

ここに 3 人の投資かがいる。強気の A 氏。中立の B 氏。慎重な C 氏。3 人の事前予測は

事前予測	A	B	C
予測レンジ	$0 \leftrightarrow 60$	$-30 \leftrightarrow 30$	$-60 \leftrightarrow 0$
中心値	30	0	-30
幅 $\times \frac{1}{2}$	30	30	30
事前分布	$N(30, 10^2)$	$N(30, 10^2)$	$N(30, 10^2)$

ここで

1. 期待収益率の事後分布
2. 将来の収益率の予測分布

を求める。
（単位は全て % である。）

μ : 期待収益率

ボラティリティ（一定とする）:

$\mu_0 - 3\tau_0 \leq \mu \leq \mu_0 + 3\tau_0$は 3σを取っているので、99.7% がこの範囲に収まる。

τ_0 : 3σと中心値の$\frac{1}{3}$

μ_0 : 予測レンジの中心値

投資家

$A : N(30, 10^2) : Bull$

$B : N(0, 10^2)$

$C : N(-30, 10^2) : Bear$

入力単位は %

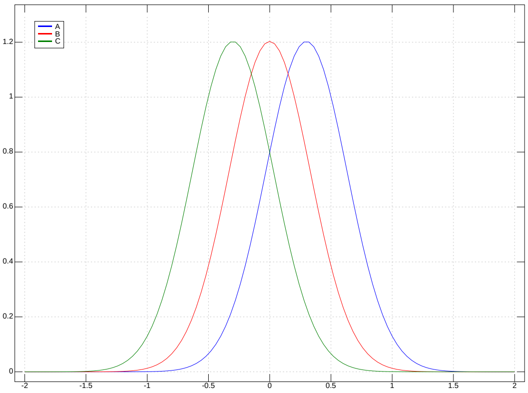


図 9　$\mu = 0.3, 0, -0.3$

5.3　期待収益率の事後分布

正規分布の確率密度関数　$p(x|\mu) = \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{(x-\mu)^2}{2\sigma^2})$

前年度の収益率 x_1　$x = x_1 \rightarrow p(x_1|\mu)$ はμの尤度になる

ベイズの定理　$p(\mu|x_1) = \frac{p(x_1|\mu)p(\mu)}{\int_{\infty}^{\infty} p(x_1|\mu)p(\mu)d\mu}$　: μの事後分布

μの事後分布

$$p(\mu|x_1) = \frac{1}{\sqrt{2\pi\tau_1^2}} exp(-\frac{(\mu-\mu_1)^2}{2\tau_1^2})$$

$$\mu_1 = \frac{\sigma^{-2}x_1 + \tau_0^{-2}\mu_0}{\sigma^{-2} + \tau_0^{-2}}$$

$$\tau_1^2 = \frac{1}{\sigma^{-2} + \tau_0^{-2}}$$

$$p(\mu|x_1) = \frac{1}{\sqrt{2\pi\tau_1^2}} exp(-\frac{(\mu-\mu_1)^2}{2\tau_1^2})$$

$N(\mu_1, \tau_1^2)$ の確率密度関数となる

逐次更新の場合の期待収益率の事後分布は次の式を計算すれば求まる。

$$\mu_1 = \frac{\sigma^{-2}x_1 + \tau_0^{-2}\mu_0}{\sigma^{-2} + \tau_0^{-2}}$$

$$\tau_1^2 = \frac{1}{\sigma^{-2} + \tau_0^{-2}}$$

A	B	C
$(N(30, 10^2)$	$N(0, 10^2)$	$N(-30, 10^2)$

`(30 10 20) bn  D0`
`NB.yearNikei int mu_n  tau_n`
`2000  4.59 0.17  24.918 8.94427`
`2001  4.36 0.06 21.4917 8.16497`
`2002  2.07 0.04 18.7171 7.55929`
`2003  8.67 0.04 17.4612 7.07107`
`2004 16.78 0.04 17.3856 6.66667`
`2005 17.48 0.07  17.395 6.32456`
	`(0 10 20) bn  D0`
	`NB. mu_n    tau_n`
	`  0.918 8.94427`
	`1.49167 8.16497`
	`1.57429 7.55929`
	`2.46125 7.07107`
	`4.05222 6.66667`
	`  5.395 6.32456`
		`(_30 10 20) bn D0`
		`NB. mu_n    tau_n`
		` _23.082 8.94427`
		`_18.5083 8.16497`
		`_15.5686 7.55929`
		`_12.5387 7.07107`
		`_9.28111 6.66667`
		`  _6.605 6.32456`

事後分布の推移

```
30 10 20  plot_bn D0
 pd 'eps /temp/bn_a0.eps'

 0 10 20  plot_bn D0
 pd 'eps /temp/bn_b0.eps'

 _30 10 20  plot_bn D0
 pd 'eps /temp/bn_c0.eps'
```

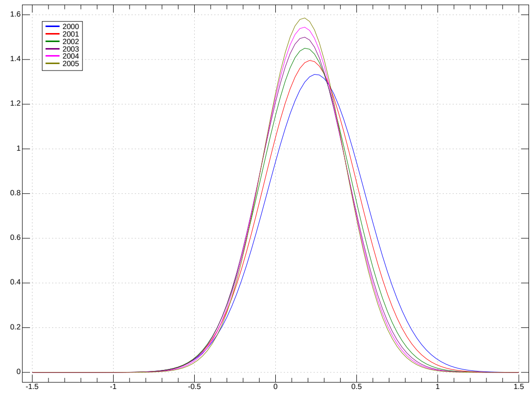


図 10　A

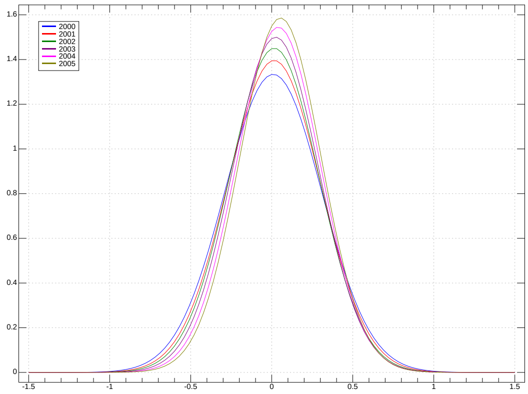


図 11　B

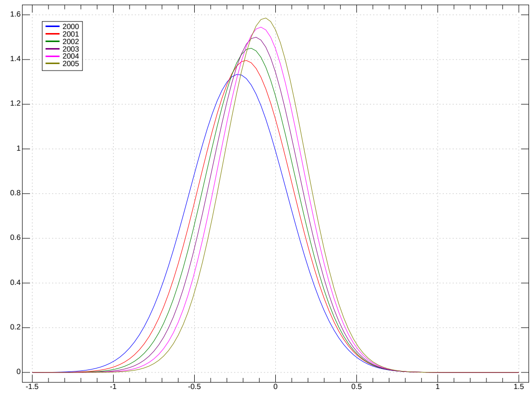


図 12　C

5.4 予測分布 $\tilde{x}$

事後分布	$p(\mu\vert x_1) = \frac{1}{\sqrt{2\pi\tau_1^2}} exp(-\frac{(\mu-\mu_1)^2}{2\tau_1^2})$ 正規分布 $N(\mu_1, \tau_1^2)$ の確率密度関数
新しい事後分布	$p(\mu\vert x_1, x_2) = \frac{1}{\sqrt{2\pi\tau_2^2}} exp(-\frac{(\mu-\mu_1)^2}{2\tau_2^2})$
μ_2	$\mu_2 = \frac{\sigma^{-2}x_2 + \tau_1^{-2}\mu_1}{\sigma^{-2} + \tau_1^{-2}}$
τ_2^2	$\tau_2^2 = \frac{1}{\sigma^{-2} + \tau_1^{-2}}$
	$N(\mu_2, \sigma^2 + \tau_2^2)$ の確率密度関数となる
予測分布	$p(\tilde{x}\vert x_1) = \frac{1}{\sqrt{2\pi(\sigma^2+\tau_2^2)}} exp(-\frac{(\tilde{x}-\mu_2)^2}{2(\sigma^2+\tau_2^2)})$ $N(\mu_1, \sigma^2 + \tau_1^2)$ の確率密度関数である

$N(\mu_1, \sigma^2 + \tau_1^2)$ は
平均は μ_1 で期待収益率の事後分布の平均と同じもの。
分散は収益率の分散 σ^2 と期待収益率 μ の事後分布の分散 τ_1^2 の和になっており、将来の収益率の不確実性を**収益率そのものの不確実性** σ^2 と**期待収益率の不確実性** τ_1^2 に分解されている。

予測分布は次の式を計算すればよい。

予測分布 $(\tilde{x}) = \sqrt{\sigma^2 + \tau_n^2}$

```
    (30 10 20) bn  D0
```

NB.	日経	定期金利	mu_1	tau_1	予測分布
2000	4.59	0.17	24.918	8.94427	21.9089
2001	4.36	0.06	21.4917	8.16497	21.6025
2002	2.07	0.04	18.7171	7.55929	21.3809
2003	8.67	0.04	17.4612	7.07107	21.2132
2004	16.78	0.04	17.3856	6.66667	21.0819
2005	17.48	0.07	17.395	6.32456	20.9762

この例題ではボラティリティを固定 (20%) にしており、各年末のインプライド · ボラティリティにはしていない。(Script も同様、少しの変更で対応できる。)

強気の A 氏の予測 $N(\mu, \tau^2) = N(30, 10^2)$ より幾分低めであるが、B,C 氏よりは高めのパフォーマンスがでている。

5.5 Portfolio

最小化問題	$min\frac{1}{2}\lambda a^2(\sigma^2 + \tau_n^2) - a(\mu_n - R_f)$
最少化問題の解 a^*	$a^* = \frac{\mu_n - R_f}{\lambda(\sigma^2 + \tau_n^2)}$

5.5.1 計算過程

λ を 10% に固定。他の変数は bn で求められている。

$a^* = \dfrac{\mu_n - R_f}{\lambda(\sigma^2 + \tau_n^2)}$

```
   (30 10 20) portfolio D0   NB. A
0.515583 0.45925 0.408563 0.387139 0.390275 0.39375


   (0 10 20 ) portfolio D0  NB. B
0.0155833 0.0306786 0.0335625 0.0538056 0.090275 0.121023


   (_30 10 20 ) portfolio D0 NB. C
_0.484417 _0.397893 _0.341438 _0.279528 _0.209725 _0.151705
```

市場の予測により A,B,C のポートフォリオは相当異なる。C なら先行の空売りが中心。

```
portfolio=: 4 : 0
NB. Usage: (30 10 20) portfolio  D0
NB. lambda is fixed 10(=0.10)
'RF M0 T0 SQRST'=: { 2}. |: x bn y
100* (M0 - RF) % 10 * *:SQRST
)
```

5.6 Script

```
bn=:bayse_normal=: 4 : 0
NB. Usage: (30 10 20) bn  D0
NB. x is mu_0, sd0, sigma(implied-volatility)
NB. y is x1 (nikkei intarest rate)
'M0 SD0 IV'=: x    NB. mu_0, tau_0,iv_0
NB. mu_0 is standard deviation(0) /not var
NK =:1{|:  y     NB. xn
ANS=. <''
for_ctr. i. # NK do.
TMP0=.(IV,SD0) bn_sub (ctr { NK),M0
ANS=. ANS,<TMP0
'M0 SD0'=. TMP0          NB. replace new mu_1 tau_1
end.
TMP1=.;("1) ,. }. ANS
PRED=. %: +/"1  ^&2 IV,.{:"1 TMP1 NB. sqrt sigma^2 + tau_2 ^2
y,.TMP1,.PRED
```

```
)

bn_sub=: 4 : 0
NB. calc mu_1 and tau_1^2
PAF0=. (^&_2) x            NB. IV,SD0  // ^&_2 is 1/a^2
T1=. % %: +/ PAF0          NB. Tau_1   //convert tau^2 -> tau
M0=. PAF0 * y              NB. NK,M0
M1=. (+/ M0) % +/ PAF0     NB. mu_1
M1, T1
)
```

6 Reference

中妻照雄　「入門ベイズ統計学」朝倉書店　2007

付録 A　ベイズ確率

A.1　cup cap

cup 茶碗　$A \cup B$
cap　帽子 $A \cap B$

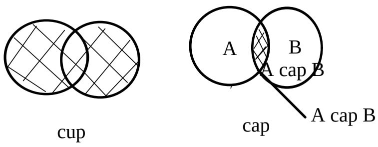


図 13　cup and cap

$$
\begin{aligned}
&P(A \cup B) = P(A) + P(B) - P(A \cap B) && \text{加法定理} \\
&P(A \cap B) = P(A)P(B) && \text{事象 } A \text{ と } B \text{ は独立} \\
&P(A \cup B) = P(A) + P(B) && \text{事象 } A \text{ と } B \text{ は背反}
\end{aligned}
$$

重複を含め全て列挙し、

cup 重複部分をシングル化

cap 重複した数のみを列挙

A.2　表記法

$P(B|A)$　A を事前確率とする事後確率が PB。[*4]

$P_B(A)$ と表記する方法もある。

A.3　計算法

$P(A|B)\dfrac{P(A \cap B)}{P(B)}$

事象 A, B が独立とは、$P(A \cap B) = P(A)P(B)$

P（$X = i$）> 0

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

[*4] ごぜんさま。御前　御膳　午前

$X \backslash Y$	0	3	
1	$1r2$	$1r3$	$5r6$
2	$1r12$	$1r12$	$1r6$
	$7r12$	$5r12$	

$$P(Y=0|X=1) = \frac{P(Y=0 \cap X=1)}{P(X=1)} = \frac{1r2}{5r6} = 3r5$$

$$P(X=2|Y=3) = \frac{P(X=2 \cap Y=3)}{P(Y=3)} = \frac{1r12}{5r12} = 1r5$$

A.4　with J

```
   A
2 4 6 8 10
   B
1 2 3 4 5

   A cop B
+---+---------------+---+---+
|CUP|1 2 3 4 5 6 8 10|CAP|2 4|
+---+---------------+---+---+

cop=: 4 : 0
NB. cup and cap
NB. usage: A cop B
CUP=.( ~: TMP)# TMP=. /:~ x,y
CAP=.(-.~: TMP) # TMP
('CUP');CUP;('CAP');CAP
)
```

付録 B　ベイズに馴染もう

神の国を数学で描くことを考えたかもしれないベイズである。
予見や経験を先験てき知識と考え知見を更新していくと、認識エンジンとして活用できる。

■Worked Example(1)　斑と伝染病, *Rich&Knight's example*

病気により斑点がでる	率	
$P(spots\|measles) =$	0.5	麻疹
$P(spots\|chickenpox) =$	0.6	水疱瘡
$P(spots\|lassa) =$	0.8	ラッサ熱
$P(measles) =$	0.01	
$P(chichenpox) =$	0.01	
$P(lassa) =$	0.0005	

[*5]

$$P(measles|spot) = \frac{0.5 \times 0.01}{(0.5 \times 0.01) + 0.6 \times 0.01 + 0.8 \times 0.0005} = 0.44$$

斑点がでて、麻疹の可能性は 44%

issue: $http : //www.macs.hw.ac.uk/ alison/ai3notes/$

■Worked Example　bag

bag A_1 mostly white balls....white is 3/4

bag A_2 mostly black balls....white is 1/4

$P(A_1) = P(A_2) = 1/2$

5 回のトライで白を 4 個、黒を 1 個選ぶ確率

2 項分布を用いる。

$$P(B|A_1) = \begin{pmatrix} 5 \\ 1 \end{pmatrix} \begin{pmatrix} 3 \\ 4 \end{pmatrix}^4 \begin{pmatrix} 1 \\ 4 \end{pmatrix}^1 = \frac{405}{1024}$$

$$P(B|A_2) = \begin{pmatrix} 5 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ 4 \end{pmatrix}^4 \begin{pmatrix} 3 \\ 4 \end{pmatrix}^1 = \frac{15}{1024}$$

$$P(A_1|B) = \frac{405/1024}{(405/1024) + 15/1024} = \frac{405}{420} = 0.964$$

```
  a1=. (1!5)*(3r4^4)*1r4
405r1024

  a2=. (1!5)*(1r4^4)*3r4
15r1024

   a1% +/ a1,a2
27r28

_1 x: 27r28
0.964286
```

issue: http://www.dcs.qmul.ac.uk/~norman/

[*5] spots 斑点

付録 C　ベイズ確率の事例抄

■Worked Example(2)　事後確率

症状			検査による判別度	
重症	$P(H_1)$	0.01	$P(A\|H_1)$	0.96
軽症傷	$P(H_2)$	0.05	$P(A\|H_2)$	0.1
健康	$P(H_3)$	0.94	$P(A\|H_3)$	0.05

$$P(A) = \sum_{i=1}^{3} P(H_i)P(A|H_i) = 0.0616$$
$$P(H_1|A) = \frac{P(A|H_1)P(H_1)}{P(A)} = \frac{0.0096}{0.0616} = 0.156$$

```
] a=. |:  2 3 $ 0.01 0.05 0.94 0.96 0.10 0.05

0.01 0.96
0.05  0.1
0.94 0.05

  +/ */"1 a
0.0096 0.005 0.047

  +/ */"1 a
0.0616
```

■Worked Example()　ある大学の学年別男女構成比。

1. A は女性
2. B_i は学年 ($i = 1, 2, 3, 4$)
3. $Pr(B_i|A)$ はランダムに抽出した学生が女子学生であった
4. その学生が B_i 学年である確率を求める

学年 (B)	男	女(A)
1	90	36
2	76	45
3	87	43
4	94	39

```
   SUM0
+-------+---------------+
|347 163|126 121 130 133|
+-------+---------------+
   SUM2
```

```
510

   beise0 DAT
NB.   1         2        3        4
0.220859 0.276074 0.263804 0.239264

beise0=: 3 : 0
SUM2=: +/;{. SUM0=:(+/ y); +/"1 y
TMP1=:((;{:SUM0)%SUM2)*({:"1 y)%;{:SUM0  NB. Pr(B)*Pr(A|B)
TMP2=:({:;{.SUM0)%SUM2                   NB. Pr(A)
TMP1%TMP2
)

DAT=:  4 2 $ 90 36 76 45 87 43 94 39
```

付録 D　calc memo

$2^{-2} = \frac{1}{2^2}$

$\frac{1}{2^{-2}} = \frac{1}{\frac{1}{2^2}} = 2^2$

```
   2^_2
0.25

  %. 2^_2
4
```