

Ⅱ．統計・確率の基礎知識

リスク計量化の前提となる統計・確率の基礎知識について整理、復習します。

図解中心の説明ですので、統計・確率は苦手だと感じている方も理解度アップに繋がります。

目 次

1. 基本統計量(1変量)
2. 基本統計量(2変量)
3. 確率変数と確率分布
4. 推定と検定

1. 基本統計量(1変量)

(1) 平 均

(2) 分 散

(3) 標準偏差

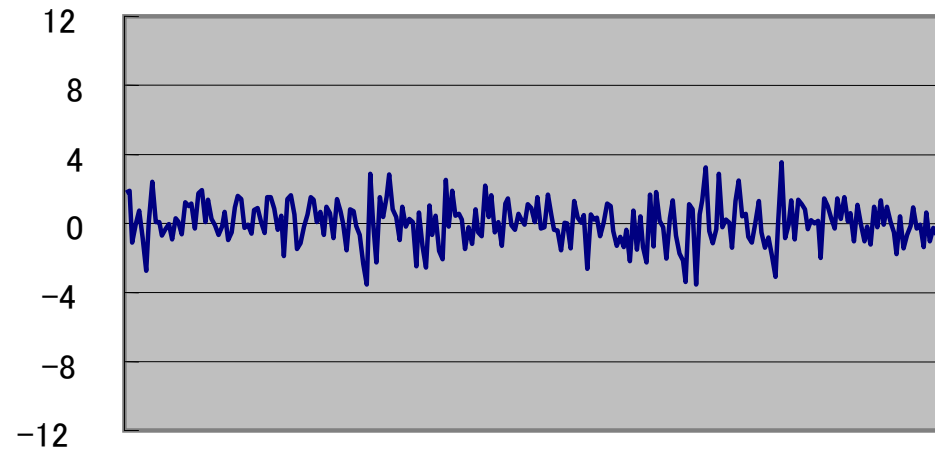
(4) パーセント点

講義の中では、以下の観測データを使います。

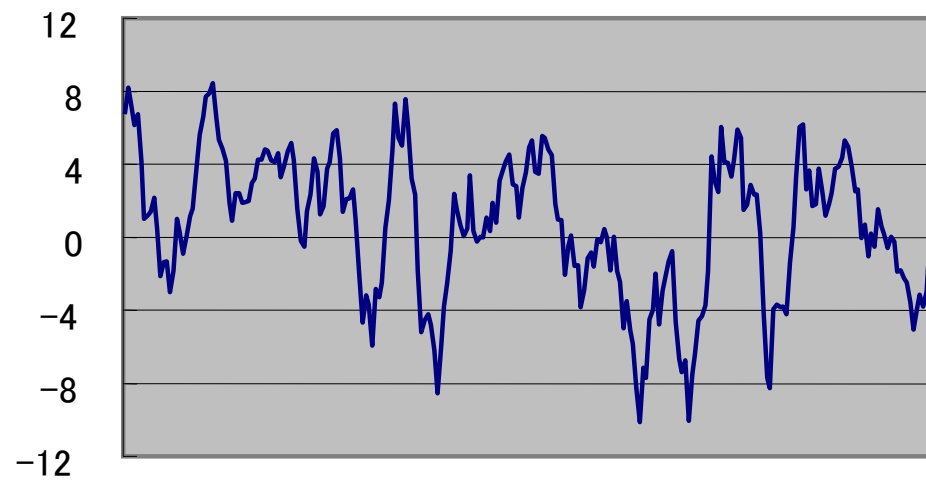
(例) 東証TOPIX・日次変化率 250個
東証TOPIX・10日間変化率 250個

	東証 指	日次変化	10日 変化
200X/9/29	1610.73	0.508	0.785
200X/9/28	1602.57	0.722	1.194
200X/9/27	1591.04	2.651	0.319
200X/9/26	1549.41	-0.667	-2.994
200X/9/25	1559.78	-0.245	-3.783
200X/9/22	1563.60	-1.048	-3.139
200X/9/21	1580.08	0.629	-3.894
200X/9/20	1570.18	-1.379	-5.040
200X/9/19	1591.98	-0.091	-3.538
200X/9/15	1593.43	-0.295	-2.474
200X/9/14	1591.04	2.651	0.319
⋮	⋮	⋮	⋮

東証TOPIX日次変化率の推移



東証TOPIX10日間変化率の推移



基本統計	Excel関数	日次変化率	10日間 変化率
データ	COUNT	250	250
平	AVERAGE	0.063	0.656
分	VARA	1.540	14.966
標準偏	STDEVA	1.241	3.869

(設問)

グラフと基本統計量を見て、どんなことに気づきましたか？

(ヒント)

気付いて欲しいことは4つあります。

答えは、講義の中で...

(1) 平均

- 平均は、観測データセットの「中心の位置」を示す指標の1つ。

$$\begin{aligned}\bar{X} &= \frac{\text{データの和}}{\text{データの数}} \\ &= \frac{X_1 + X_2 + \cdots + X_N}{N}\end{aligned}$$

- Excelでは、関数AVERAGE(データ範囲)を使って求める。

(2- i) 分 散 (記述統計の立場で定義)

- 分散は、観測データセットの「バラツキ」を示す指標の1つ。
 - ー データの「偏差平方和」(平均との差を2乗して合計)を求めて
「データの数」で割る。
 - ー 分散の「単位」は、データの持つ「単位」の2乗。

$$\begin{aligned} V_p = \sigma^2 &= \frac{\text{データの偏差平方和}}{\text{データの}} \\ &= \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \cdots + (X_N - \bar{X})^2}{N} \end{aligned}$$

- Excelでは、関数VARP(データ範囲)を使って求める。 8

記述統計： 中学・高校で学習する平均と分散

(例)

平均: 中心の位

観測データ	3	4	5	6	7	
						
偏差 (平均との差)	-2	-1	0	1	2	合計すると ゼロ

偏差平方	$(-2)^2$	$(-1)^2$	0^2	1^2	2^2	合計すると
						偏差平方和 10

- 観測データがバラつく(平均から離れる)と偏差平方和は増える。
- しかし、観測データ数が増えても偏差平方和は増えてしまう。

$$\text{分散} = \frac{\text{偏差平方和 } 10}{\text{観測データ数 } 5}$$

(参考)記述統計の考え方

- 観測データを母集団全体と考えると、統計量の算定を行い、観測データが持つ特性を分析・記述する。

(例)ある特定の集団(N人)の身長の平均と分散を計算する。

平均

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_N}{N}$$

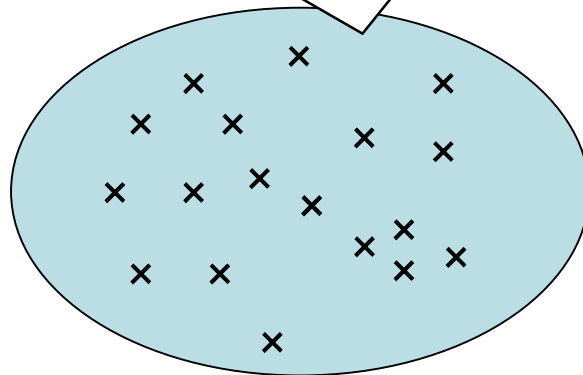
分散

$$V_p = \frac{(X_1 - \overline{X})^2 + (X_2 - \overline{X})^2 + \cdots + (X_N - \overline{X})^2}{N}$$

母集団＝標本の特性値を調べる

平均 μ
分散 V
標準偏差 σ
VaR など.

計測可



母集団＝標本

(2-ii) 分散(推測統計の立場で定義)

- 分散は、観測データセットの「バラツキ」を示す指標の1つ。
 - データの「偏差平方和」(平均との差を2乗して合計)を求めて「**データの数-1**」で割る。
 - 分散の「単位」は、データの持つ「単位」の2乗。

$$\begin{aligned} \text{Va} = \sigma^2 &= \frac{\text{データの偏差平方和}}{\text{データ数}-1} \\ &= \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_N - \bar{X})^2}{N-1} \end{aligned}$$

- Excelでは、関数VARA(データ範囲)を使って求める。₁₂

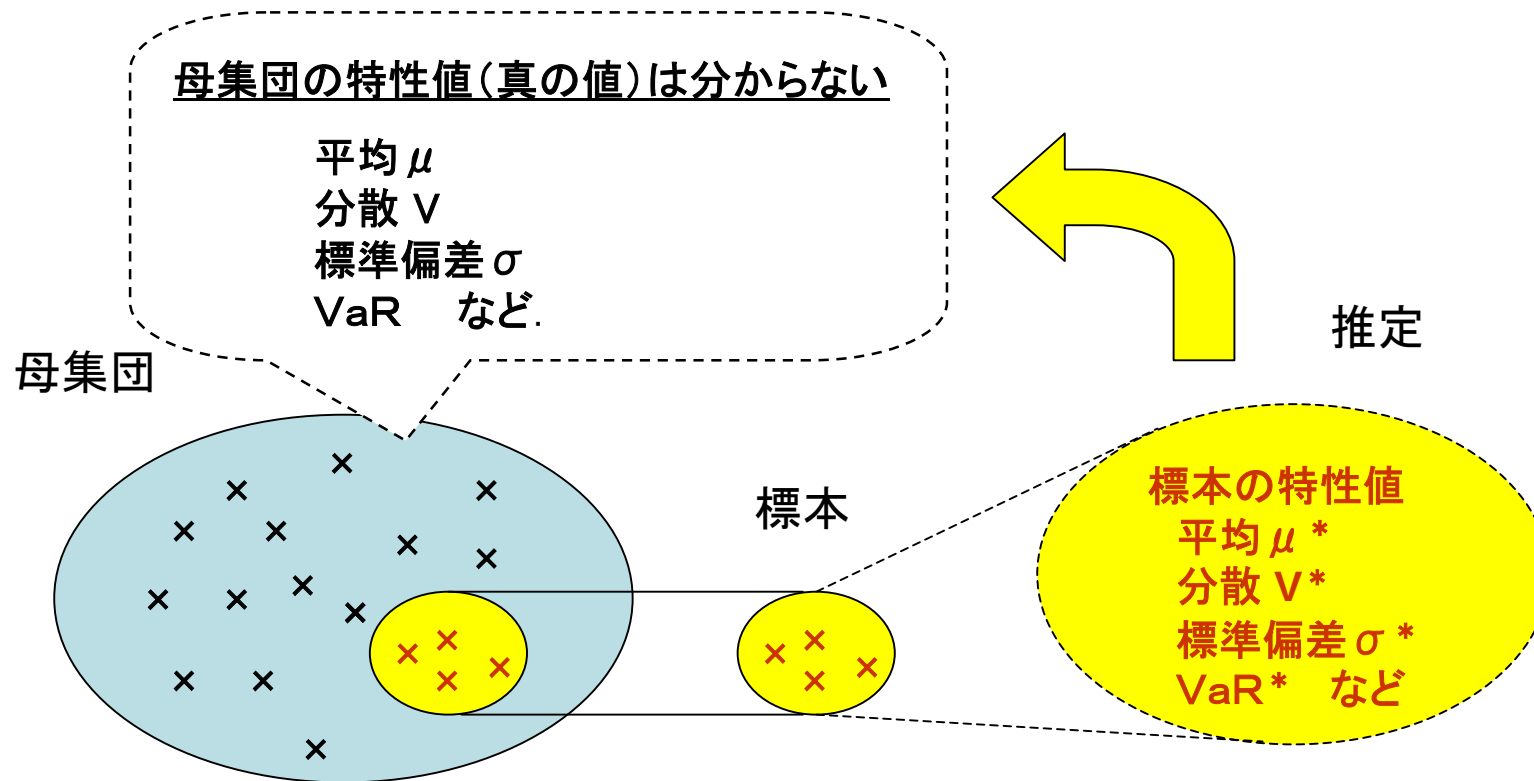
(参考) 推測統計の考え方

- 観測データを、母集団から抽出した標本(サンプル)と考えて、統計量の算定を行い、母集団の特性を推測し、検証する。

(例) 任意に抽出したN人(標本)の身長を計測して、日本人全体(母集団)の身長の平均と分散を推定する。

$$\begin{array}{l} \text{平均} \\ \overline{X} \end{array} = \frac{X_1 + X_2 + \cdots + X_N}{N}$$

$$\begin{array}{l} \text{分散(不偏標本分散)} \\ Va \end{array} = \frac{\overline{(X_1 - X)^2} + \overline{(X_2 - X)^2} + \cdots + \overline{(X_N - X)^2}}{N-1}$$



N-1で割った「標本分散」の特徴

- ・ 母集団の「真の分散」を、統計的手法で「推定」するときにN-1で割った「標本分散」を使うのは、以下のような特徴があるため。

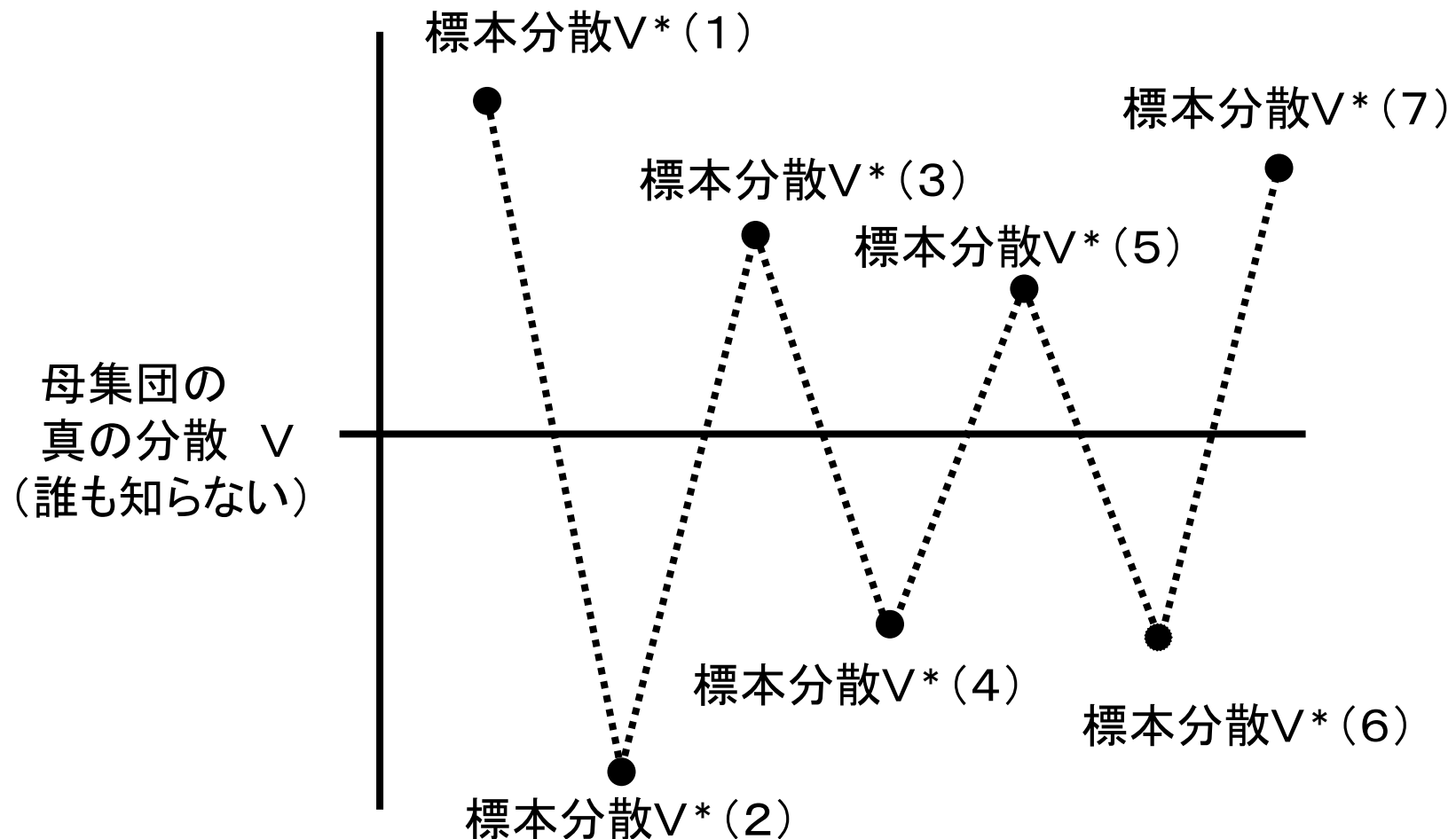
(一貫性)

- ・ 「標本分散」は、Nが大きくなると、母集団の「真の分散」に限りなく近づく

(不偏性)

- ・ 「標本分散」は、母集団の「真の分散」の偏りのない推定値となることが知られている

標本分散(V^*)を、標本を変えて繰り返し計算すると、
真の分散を中心にして偏りなく分布する(不偏性)





講義の中で、VaRを計測する際に使う
分散、標準偏差は、推測統計の立場

から定義したもの($N-1$ で割ったもの)
です。

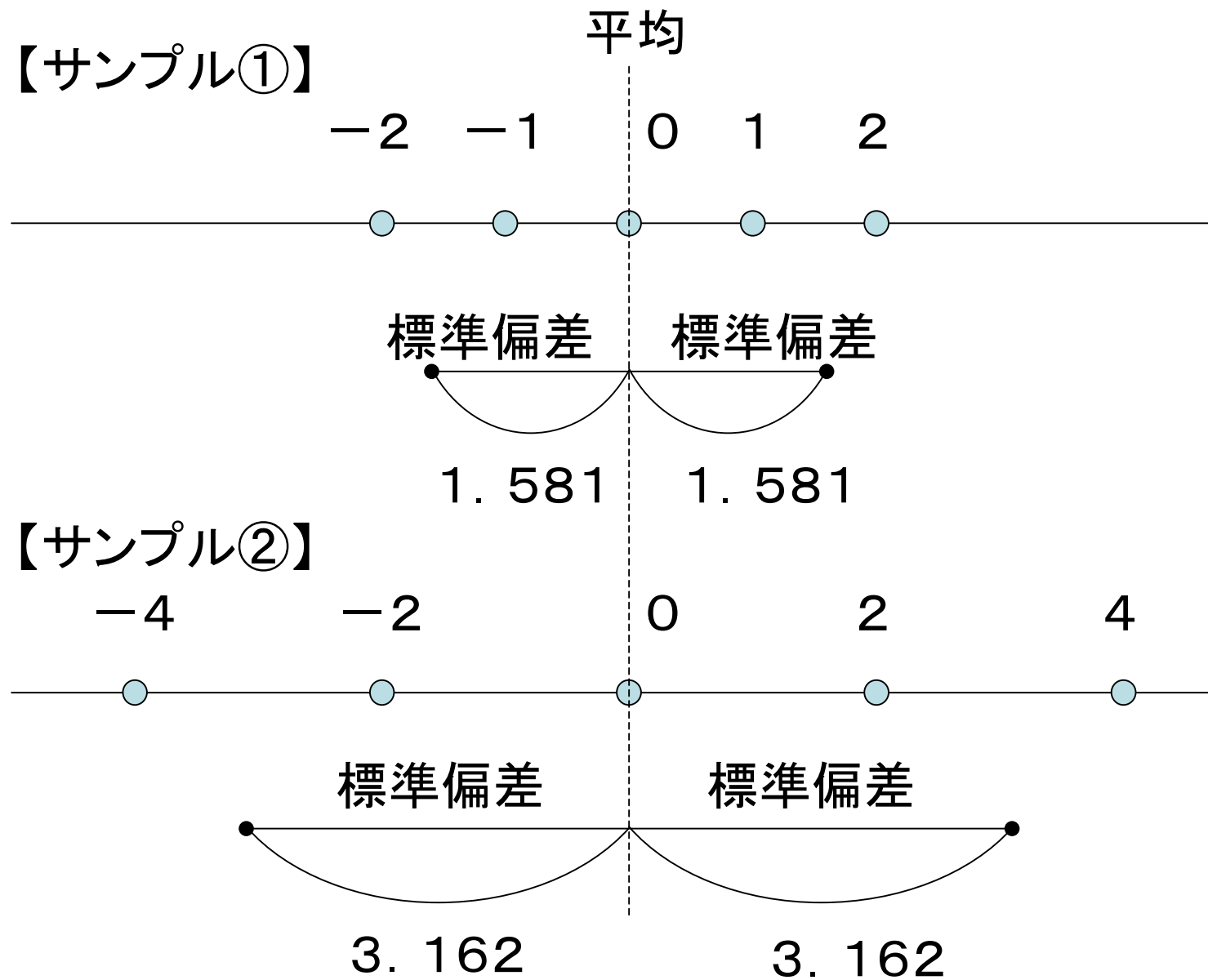
(3) 標準偏差(推測統計の立場で記載)

- 標準偏差は、観測データセットの「バラツキ」を示す指標の1つ。分散の平方根(ルート)をとって定義する。

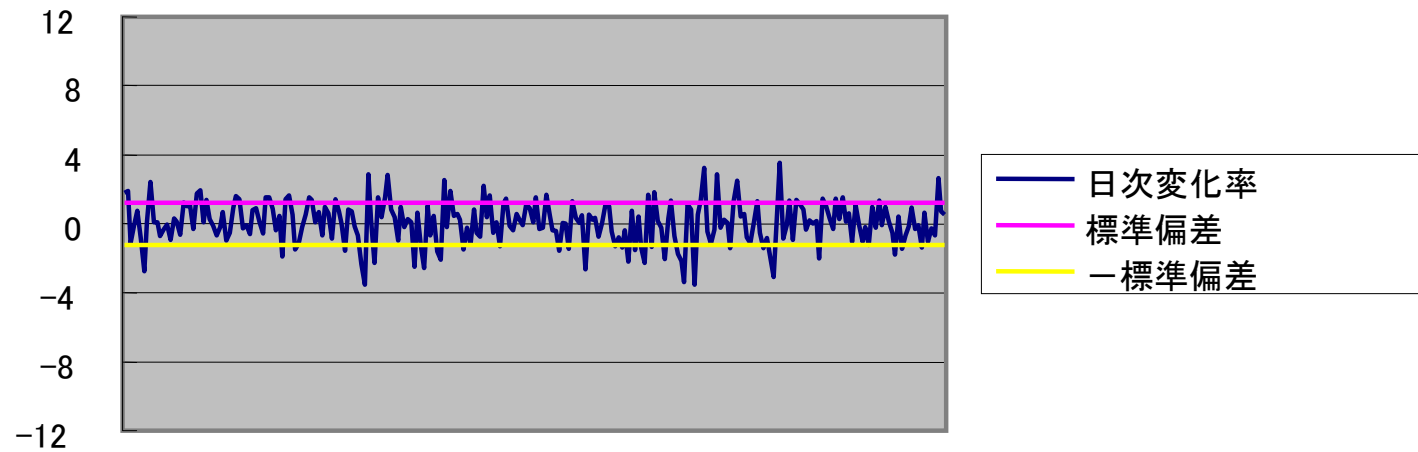
ー 標準偏差の「単位」は、データの持つ「単位」と同じ。

$$\begin{aligned}\sigma &= \sqrt{\frac{\text{データの偏差平方和}}{\text{データ数}-1}} \\ &= \sqrt{\frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_N - \bar{X})^2}{N-1}}\end{aligned}$$

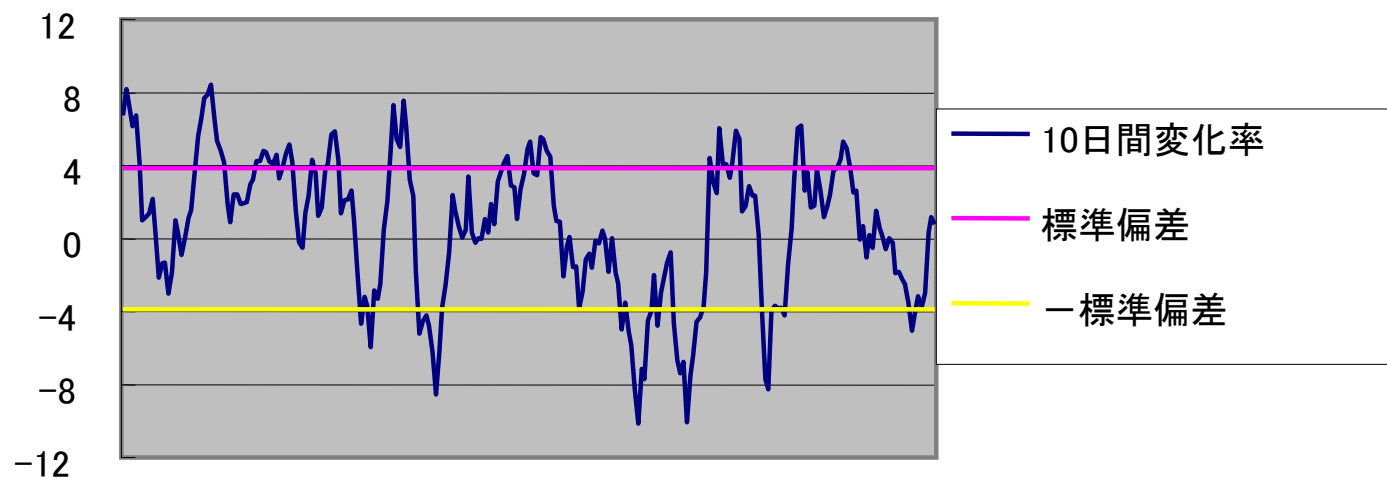
- Excelでは、関数STDEVA(データ範囲)を使って求める。



東証TOPIX日次変化率の推移



東証TOPIX10日間変化率の推移



基本統計	Excel関	日次変化	10日 変化
データ	COUN	250	250
平	AVERAG	0.063	0.656
分	VAR	1.540	14.966
標準偏	STDEV	1.241	3.869

- 平均をみると、日次変化率、10日間変化率とも概ねゼロとなっている。
- 分散をみると、10日間変化率の分散は、日次変化率の分散の概ね10倍となっている。
- 標準偏差をみると、10日間変化率の標準偏差は、日次変化率の標準偏差の概ね $\sqrt{10}$ 倍(=3.162倍)となっている。



株価・金利・為替等の変化率に関して

- ① その平均をゼロと仮定したり、
- ② T日間変化率の標準偏差は、日次変化率の標準偏差の $\sqrt{T}$ 倍と仮定して
市場VaRを計測することがある。

(4) パーセント点

- パーセント点とは、観測データを小さい順に並べたときに、その値よりも小さな値の割合が指定された割合（百分率）になるデータの値として定義される。
- 例えば、99パーセント点というのは、その値より小さなデータの割合が99%となるデータの値のことを指す。
 - 50パーセント点のことを中央値（メジアン）と呼ぶ。
 - 25パーセント点を第1四分位点、75パーセント点を第3四分位点と呼ぶ。
- Excelでは、関数PERCENTILE（データ範囲, 率）を使って求める。

(例) 1000個の損失データが観測されている場合、
99%点というのは、損失額を小さい順に並べて
990番目になるデータ値のこと。

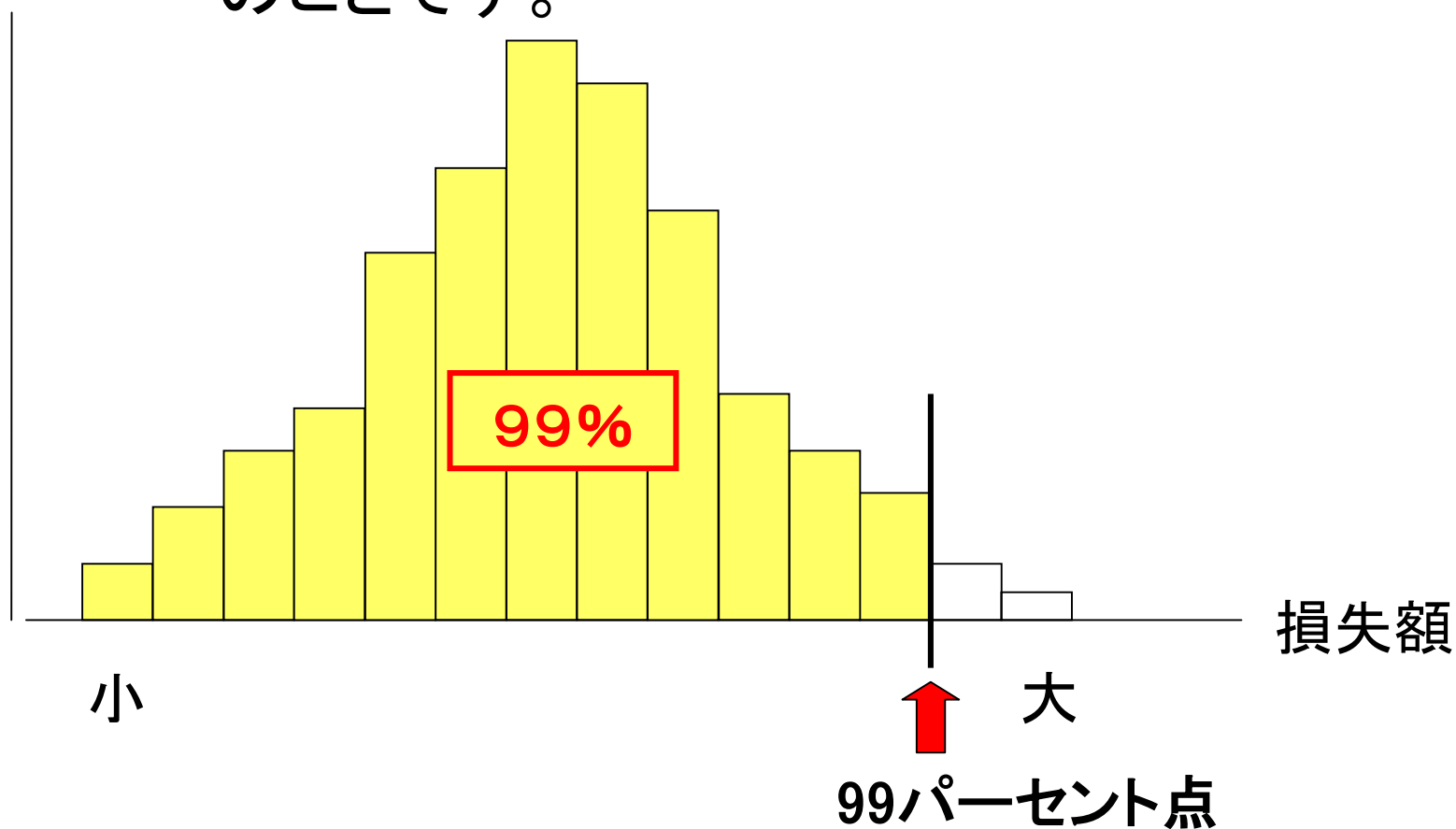
順位	百分位	損失額
985 番 目	98.5%	529
986 番 目	98.6%	558
987 番 目	98.7%	589
988 番 目	98.8%	618
989 番 目	98.9%	621
990 番 目	99.0%	632
991 番 目	99.1%	654
992 番 目	99.2%	671
993 番 目	99.3%	698
994 番 目	99.4%	703
995 番 目	99.5%	712
996 番 目	99.6%	776
997 番 目	99.7%	794
998 番 目	99.8%	810
999 番 目	99.9%	831
1000 番 目	100.0%	869





99%VaRは、文字通り、99パーセント点

のことです。



2. 基本統計量(2変量)

(1) 散布図

(2) 共分散

(3) 相関係数

(4) 相関行列

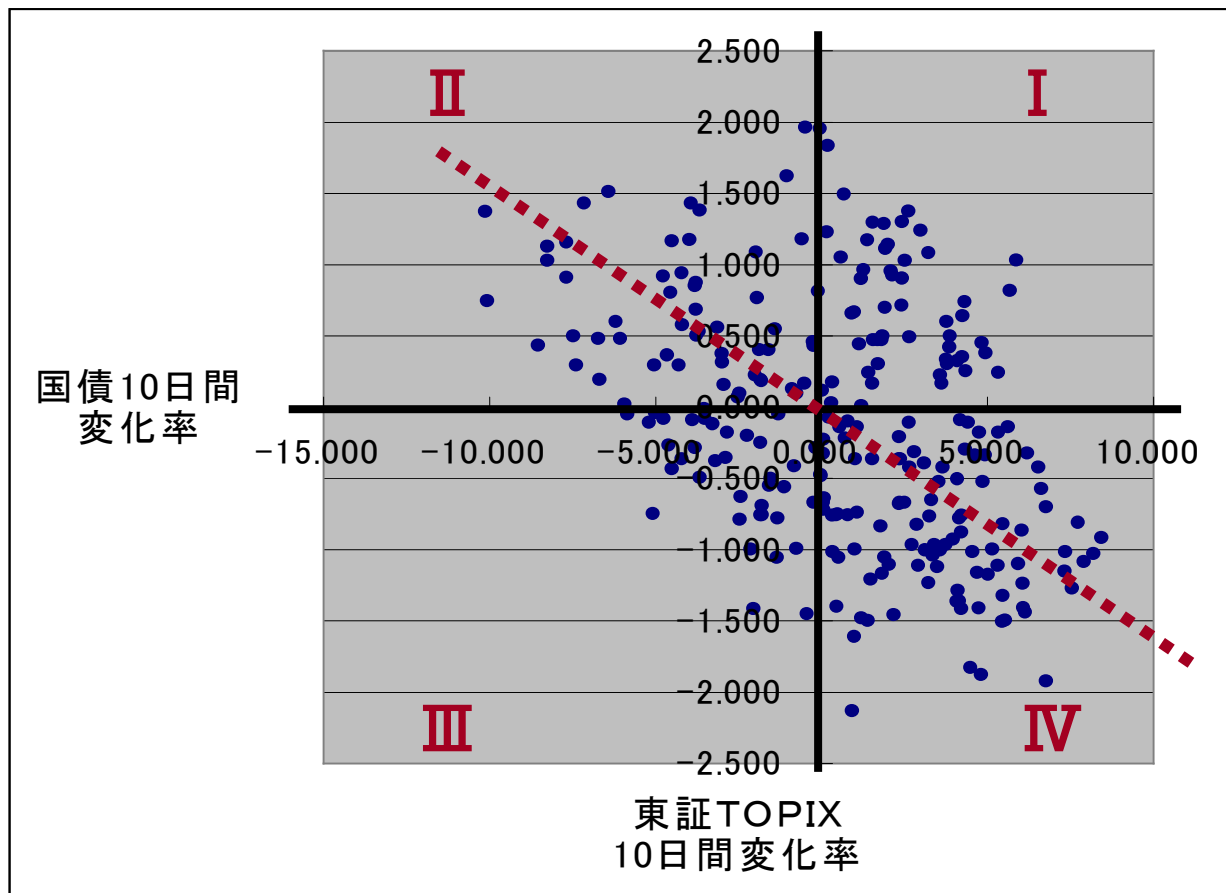
(1) 散布図

- 以下のような2変量の関係を調べるためには、散布図を書くのが直感的に理解しやすい。

	東証TOPIX 10日間変化率 (X)	10年割引国債 10日間変化率 (Y)
200X/9/29	0.785	-0.098
200X/9/28	1.194	0.010
200X/9/27	0.319	0.177
200X/9/26	-2.994	0.315
200X/9/25	-3.783	0.688
200X/9/22	-3.139	0.560
200X/9/21	-3.894	-0.088
200X/9/20	-5.040	0.295
200X/9/19	-3.538	-0.010
200X/9/15	-2.474	0.098
⋮	⋮	⋮

株価変化率と国債価格変化率との関係

- **Ⅱ、Ⅳ**のエリアに分布が多い。
- 株価変化率がプラス(マイナス)のとき、国債価格変化率はマイナス(プラス)となる傾向がある。



(2) 共分散 (推測統計の立場で記載)

- 共分散は、2つの変量(X、Y)の間の「直線的な比例関係の強さ」を示す指標。

- データの「偏差積和」を求めて、「**データ数－1**」で割る。
- 共分散の「単位」は、Xの持つ「単位」 掛ける Yの持つ「単位」。

$$\text{COV}(X, Y) = \frac{\text{データの偏差積和}}{\text{データ数} - 1}$$
$$= \frac{(X_1 - \bar{X})(Y_1 - \bar{Y}) + (X_2 - \bar{X})(Y_2 - \bar{Y}) + \dots + (X_N - \bar{X})(Y_N - \bar{Y})}{N - 1}$$

- Excelでは、関数COVAR(データ範囲(X)、データ範囲(Y))を使って求める。

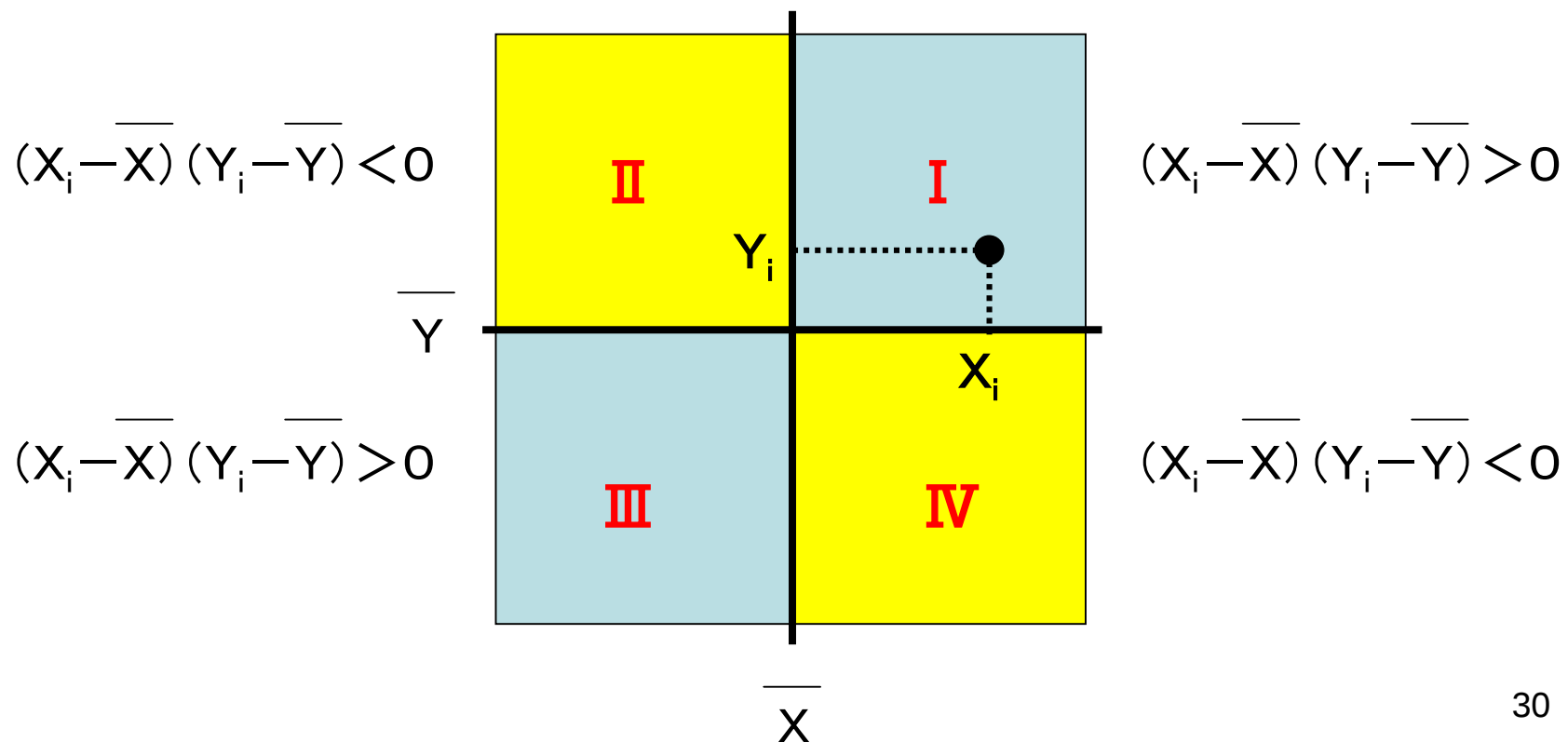
(注) Excelでは、データの偏差積和をN－1ではなく、Nで割って共分散を定義している(記述統計の立場で定義している)ため、別途、調整を行う必要がある。

偏差積和

$$= (X_1 - \bar{X})(Y_1 - \bar{Y}) + (X_2 - \bar{X})(Y_2 - \bar{Y}) + \cdots + (X_N - \bar{X})(Y_N - \bar{Y})$$

I、Ⅲのエリアに多く分布 $\Rightarrow$ 偏差積和 > 0 : 正の相関

Ⅱ、Ⅳのエリアに多く分布 $\Rightarrow$ 偏差積和 < 0 : 負の相関



(3) 相関係数

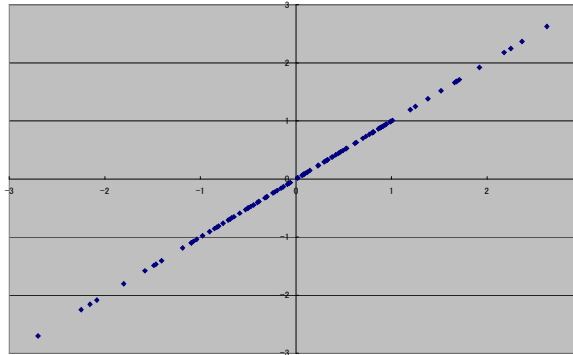
- 相関係数は、2つの変量(X、Y)間の「直線的な比例関係の強さ」を示す指標。
- 共分散を、2つの標準偏差の積で割って定義する。
 - 相関係数は-1～+1までの値をとる。「単位」を持たない無名数。
 - 相関係数の定義には、データ数Nが含まれていない(定義は1通りのみ)。

$$\rho(X, Y) = \frac{\text{COV}(X, Y)}{\sigma(X) \sigma(Y)}$$
$$= \frac{(X_1 - \bar{X})(Y_1 - \bar{Y}) + \cdots + (X_N - \bar{X})(Y_N - \bar{Y})}{\sqrt{(X_1 - \bar{X})^2 + \cdots + (X_N - \bar{X})^2} \sqrt{(Y_1 - \bar{Y})^2 + \cdots + (Y_N - \bar{Y})^2}}$$

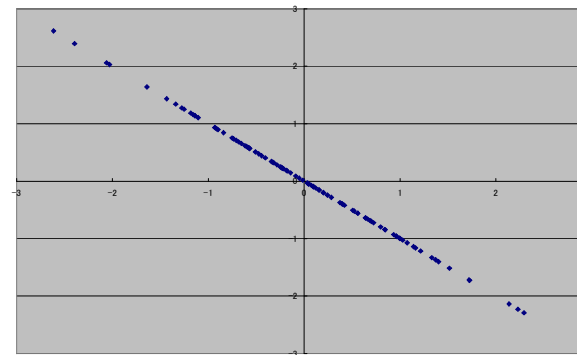
- Excelでは、関数CORREL(データ範囲(X)、データ範囲(Y))を使って求める。

相関係数と散布図

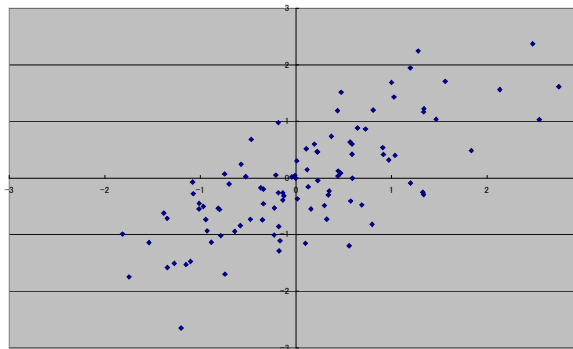
$\rho = 1.0$
(正の完全相関)



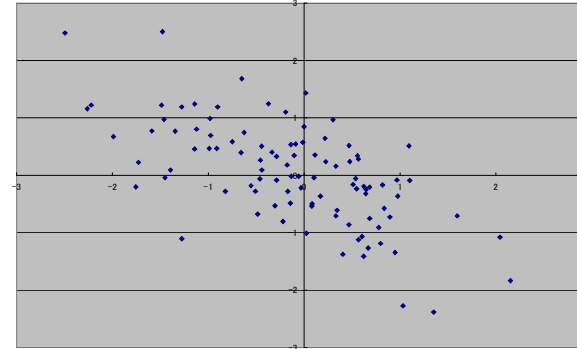
$\rho = -1.0$
(負の完全相関)



$\rho = 0.7$



$\rho = -0.7$



相関係数の定義

$$\rho_{xy} = \text{COV}(X,Y) / \sigma_x \sigma_y$$

$\text{COV}(X,Y)$: X, Y の共分散 $= (1/N-1) * \sum (X_t - EX)(Y_t - EY)$

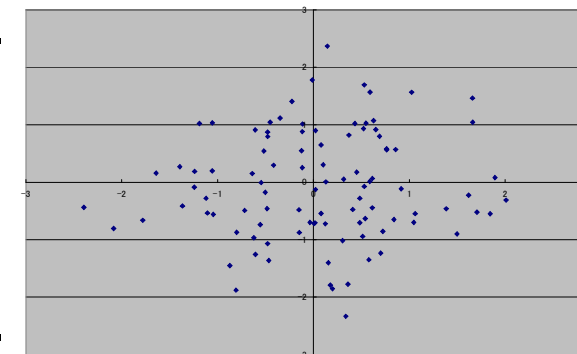
σ_x : X の標準偏差

EX : X の平均値

σ_y : Y の標準偏差

EY : Y の平均値

$\rho = 0$
(無相関)



(4) 相関行列と分散共分散行列

太枠内が相関行列

	X_1	X_2	X_3	$\dots$	X_N
X_1	1	$\rho(X_1, X_2)$	$\rho(X_1, X_3)$	$\dots$	$\rho(X_1, X_N)$
X_2	$\rho(X_2, X_1)$	1	$\rho(X_2, X_3)$	$\dots$	$\rho(X_2, X_N)$
X_3	$\rho(X_3, X_1)$	$\rho(X_3, X_2)$	1	$\dots$	$\rho(X_3, X_N)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
X_N	$\rho(X_N, X_1)$	$\rho(X_N, X_2)$	$\rho(X_N, X_3)$	$\dots$	1

$\rho(X_i, X_i) = 1$: 同じ変量(X_{ii})同士の相関は1

$\rho(X_i, X_j) = \rho(X_j, X_i)$: 2つの変量(X_i, X_j)の順序を変えて計算しても相関係数の値は同じ。

太枠内が分散共分散行列

	X_1	X_2	X_3	$\dots$	X_N
X_1	V_{X1}	$\text{COV}(X_1, X_2)$	$\text{COV}(X_1, X_3)$	$\dots$	$\text{COV}(X_1, X_N)$
X_2	$\text{COV}(X_2, X_1)$	V_{X2}	$\text{COV}(X_2, X_3)$	$\dots$	$\text{COV}(X_2, X_N)$
X_3	$\text{COV}(X_3, X_1)$	$\text{COV}(X_3, X_2)$	V_{X3}	$\dots$	$\text{COV}(X_3, X_N)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
X_N	$\text{COV}(X_N, X_1)$	$\text{COV}(X_N, X_2)$	$\text{COV}(X_N, X_3)$	$\dots$	V_{XN}



VaRの計測手法として、分散共分散法の説明をします。

VaRの計測において、分散共分散行列、相関行列が重要な働きをします。

Ⅲ. VaRの計測と検証より

分散共分散法(デルタ法)による計算例②

ー リスクファクターが2つのケース

VaRの計算シート 分散共分散法(デルタ法)

【ポートフォリオ】

株式投信	100	億円
10年割引国債	100	億円

保有期間	10	日
信頼水準	99.00	%

観測データ	250	日
-------	-----	---

	東証TOPIX 10日間変化率	10年割引国債 10日間変化率
2006/9/29	0.785	-0.098
2006/9/28	1.194	0.010
2006/9/27	0.319	0.177
2006/9/26	-2.994	0.315
2006/9/25	-3.783	0.688
2006/9/22	-3.139	0.560
2006/9/21	-3.894	-0.088
2006/9/20	-5.040	0.295
2006/9/19	-3.538	-0.010
2006/9/15	-2.474	0.098
2006/9/14	-2.248	-0.197
2006/9/13	-1.822	0.187
2006/9/12	-1.875	0.403
2006/9/11	-0.235	0.433
2006/9/8	0.007	0.118
2006/9/7	-0.591	1.179
2006/9/6	0.155	1.228
2006/9/5	0.582	1.051
2006/9/4	1.534	1.296
2006/9/1	-0.495	1.964
2006/8/31	0.184	1.837

単独VaR	標準偏差	× 信頼係数	× 感応度
株式投信 9.00	3.8686	2.33	100
割引国債 1.99	0.8568	2.33	100

ポートVaR	①	②
単純合算 10.99	①	②
相関考慮後 8.35	① > ②: ポートフォリオ効果	

投信VaR	国債VaR	相関行列	投信VaR	国債VaR				
9.00	1.99	<table><tr><td>1</td><td>-0.4233</td></tr><tr><td>-0.4233</td><td>1</td></tr></table>	1	-0.4233	-0.4233	1	9.00	1.99
1	-0.4233							
-0.4233	1							
		↓						
		行列計算式						
		<table><tr><td>8.1560</td><td>-1.8162</td></tr></table>	8.1560	-1.8162	<table><tr><td>9.00</td></tr><tr><td>1.99</td></tr></table>	9.00	1.99	
8.1560	-1.8162							
9.00								
1.99								
			↓					
		VaR^2 :	69.78	行列計算式				
		VaR :	8.35					
投信感応度	国債感応度	分散共分散行列	投信感応度	国債感応度				
100.00	100.00	<table><tr><td>14.96626</td><td>-1.3938</td></tr><tr><td>-1.3938</td><td>0.7364709</td></tr></table>	14.96626	-1.3938	-1.3938	0.7364709	100.00	100.00
14.96626	-1.3938							
-1.3938	0.7364709							
		↓						
		行列計算式						
		<table><tr><td>1357.2481</td><td>-65.7303</td></tr></table>	1357.2481	-65.7303	<table><tr><td>100.00</td></tr><tr><td>100.00</td></tr></table>	100.00	100.00	
1357.2481	-65.7303							
100.00								
100.00								
			↓					
		ポート分散 :	12.92	行列計算式				
		ポート標準偏差 :	3.59	(単位調整)				
		信頼係数	2.33					
		ポートVaR	8.36					

行列計算式

行列計算式
(単位調整)

3. 確率変数と確率分布

(1) 確率変数

(2) 確率分布

— 確率密度関数、分布関数

(3) 様々な確率分布

— 一様分布、正規分布、対数正規分布
ポワソン分布、2項分布

(4) 確率変数の独立

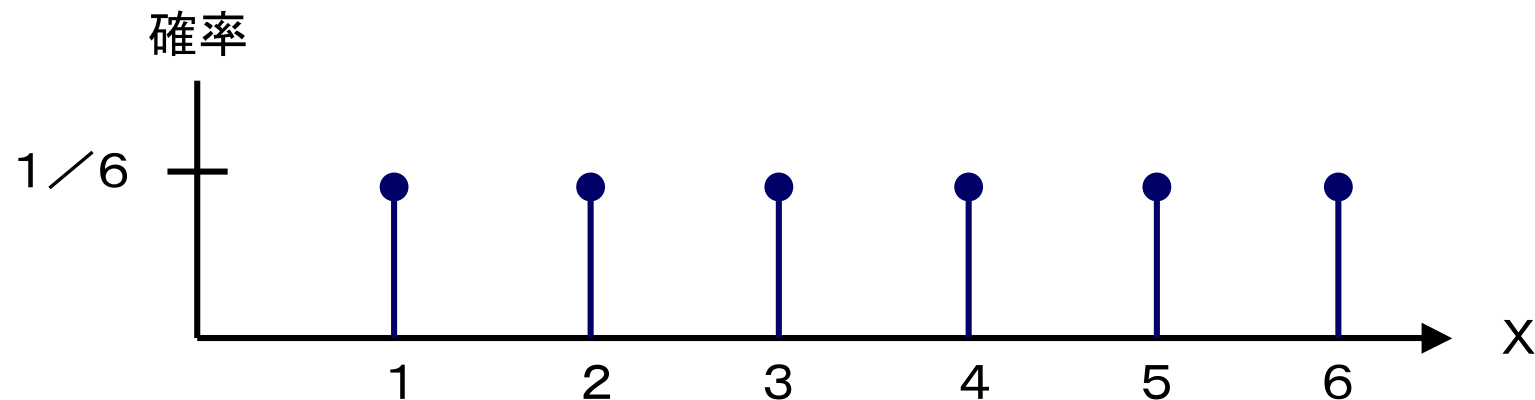
(1) 確率変数

- 予め定まった確率にしたがって値が変動する数のことを「確率変数」という

(例) サイコロを振ったときに出る目の数

— 離散的な確率変数

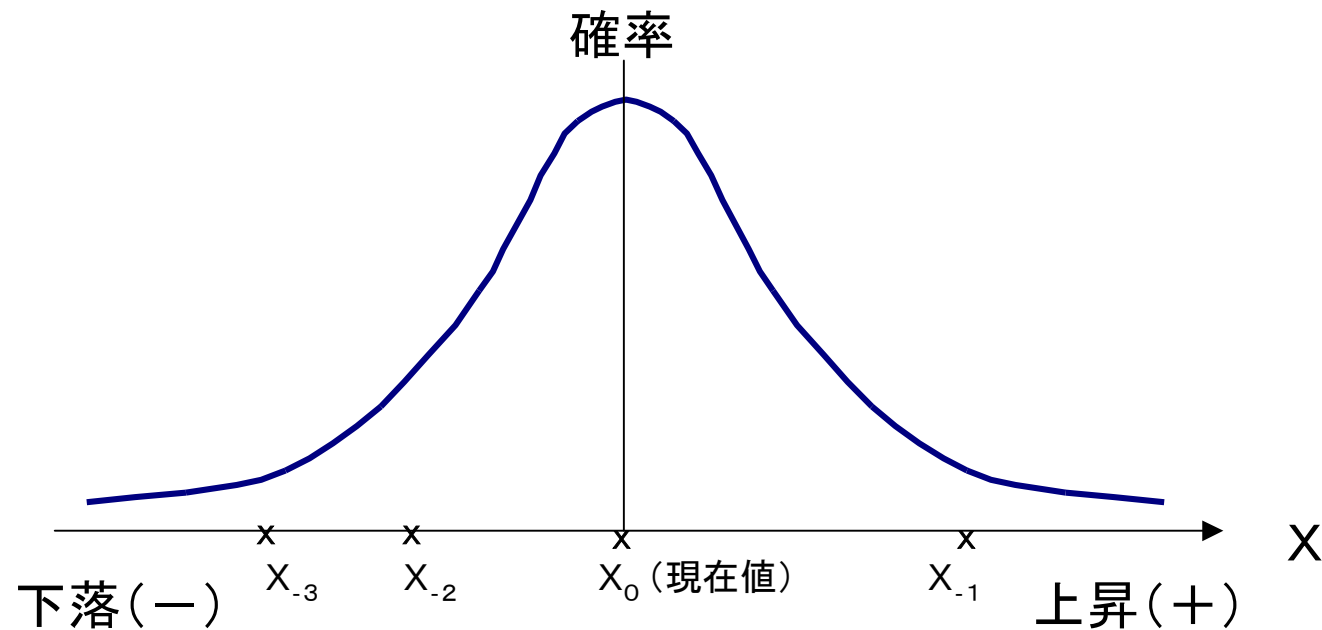
サイコロの目 (X)	1	2	3	4	5	6
確 率	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$





株価、金利、為替等の変化率を、
「確率変数」として捉えることも可能。

— 連続的な確率変数





その他の確率変数

- VaRを250回計測して、VaRを超える損失が発生する回数
- 事件・事故発生に伴う損失の発生額(1回当たり)
- 事件・事故の年間発生件数
- 個別企業の信用状態

(2) 確率分布

- ・ 確率分布を表わすとき、2種類の関数がある。

① 確率密度関数

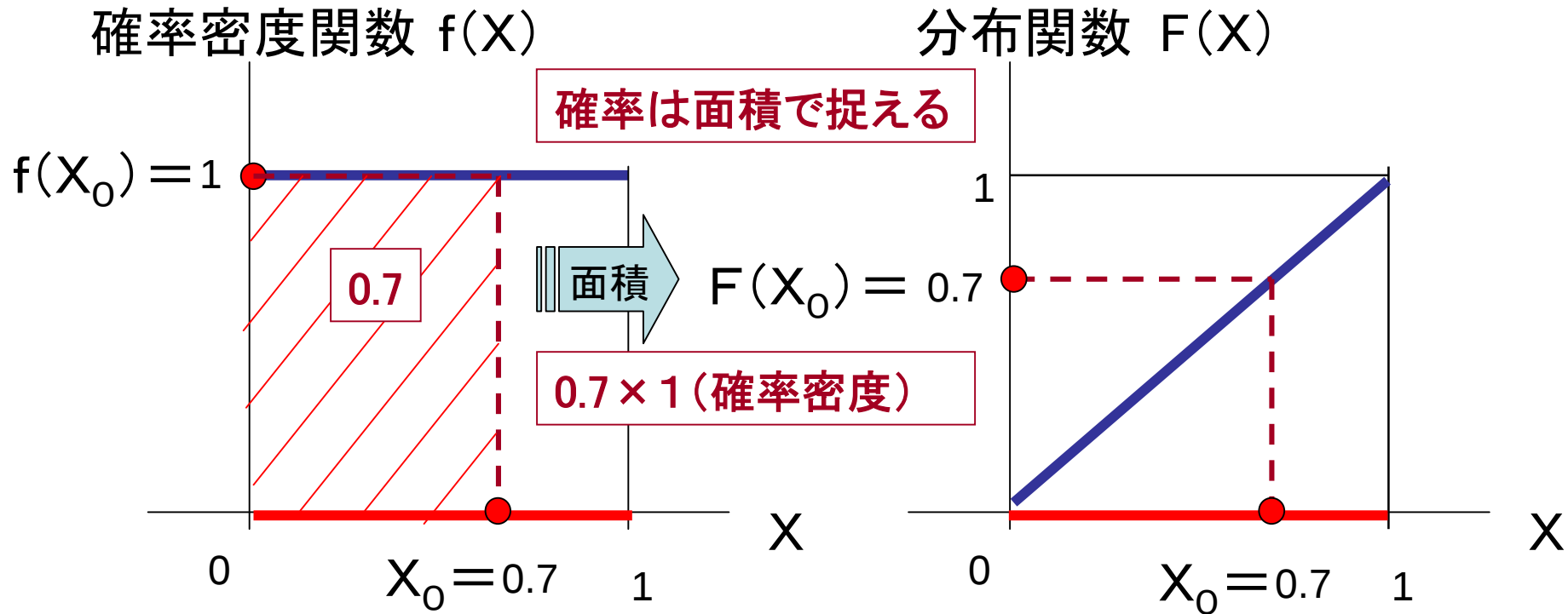
確率変数(X)が「ある値」をとる確率(確率密度)
を表わす関数

② 分布関数(累積確率密度関数)

確率変数(X)が「ある値以下」になる確率を表わ
す関数

(例) 数直線上で、0から1までの値をランダムにとる
確率変数(X)を考える。

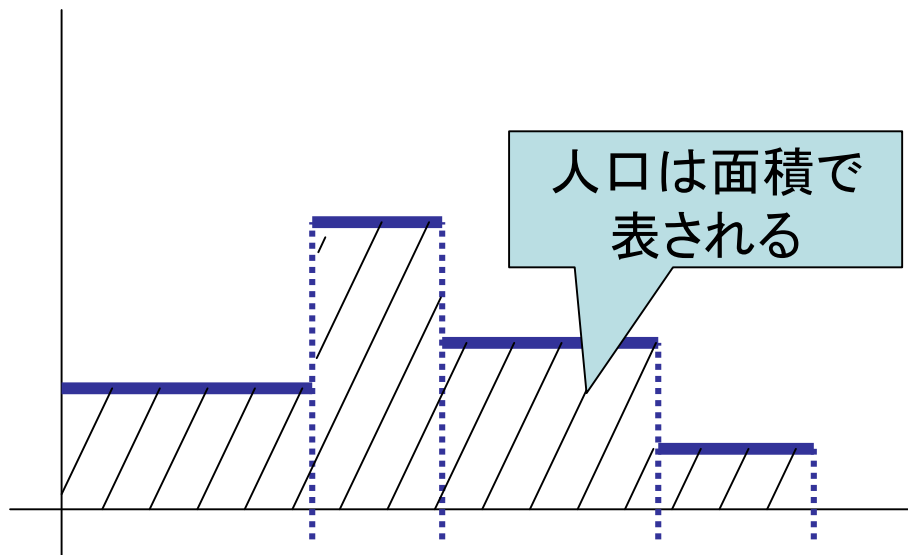
X は 0～1の間で無限の値をとる可能性がある
 X が 0.7の値をとる確率はゼロ
 X が 0.7以下の値をとる確率は 0.7(斜線部の面積)



X が 0.7の値をとる「確率密度」は 1

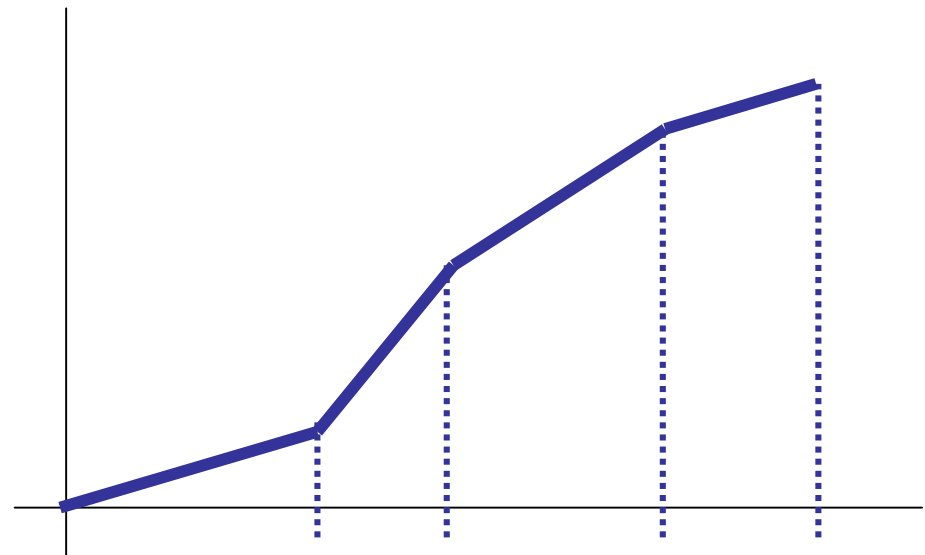
(参考)人口と人口密度

$f(X)$: 人口密度 (万人/ km^2)



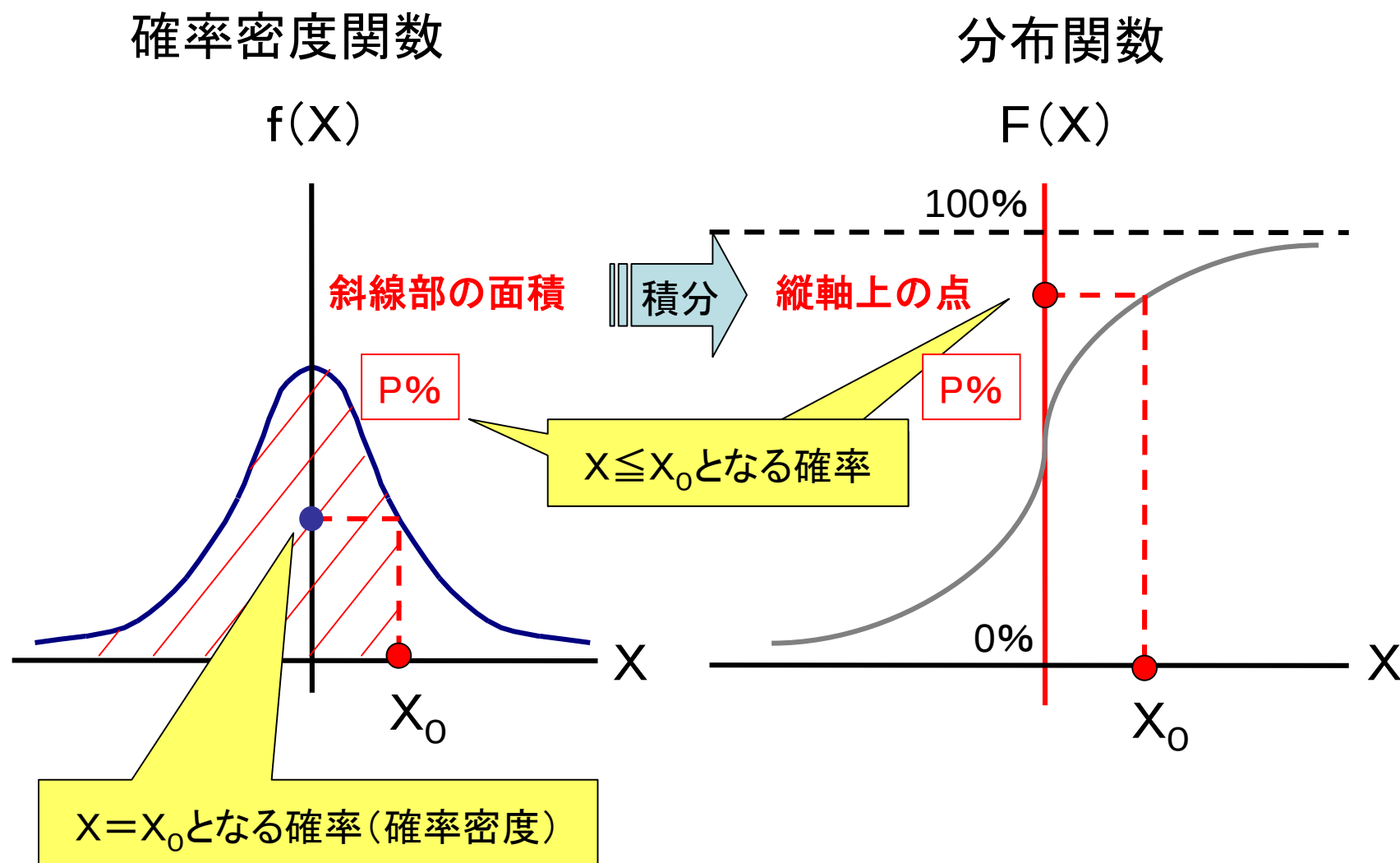
X : 各地域の広さ (km^2)

$F(X)$: 人口 (万人)



X : 各地域の広さ (km^2)

- より一般的に概念図で示すと

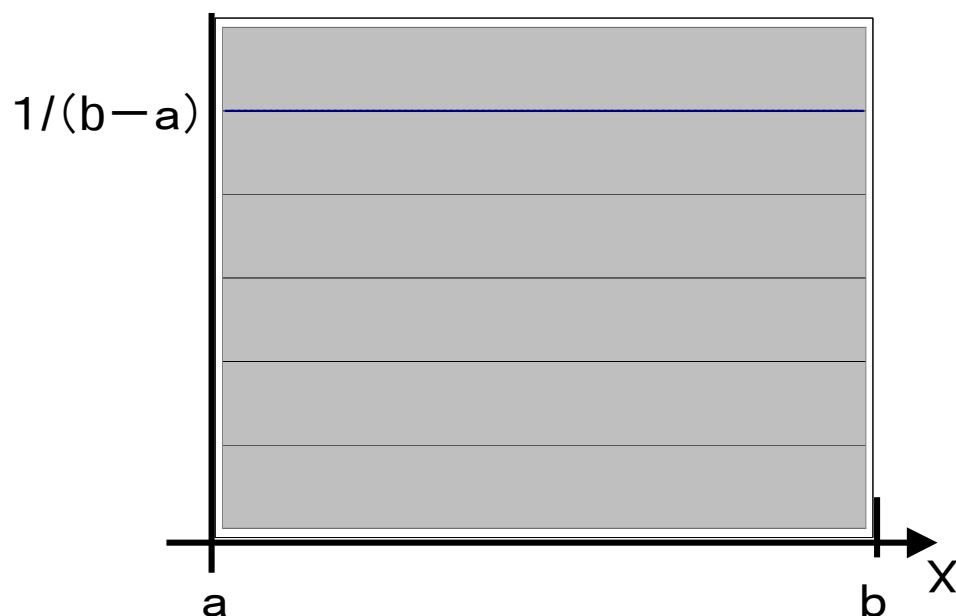


(3) 様々な確率分布

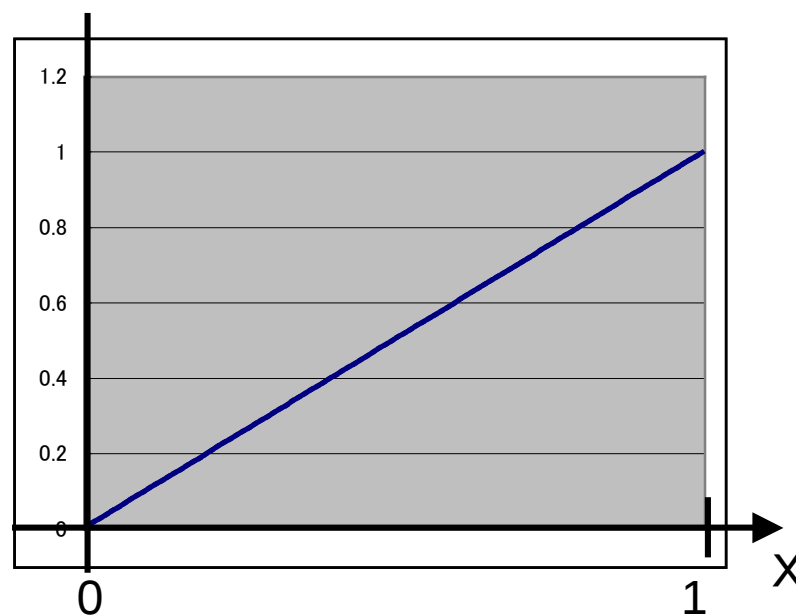


一様分布: ある区間の中の値が同じ確率で生起する分布。

$f(X)$ 確率密度関数



$F(X)$ 分布関数



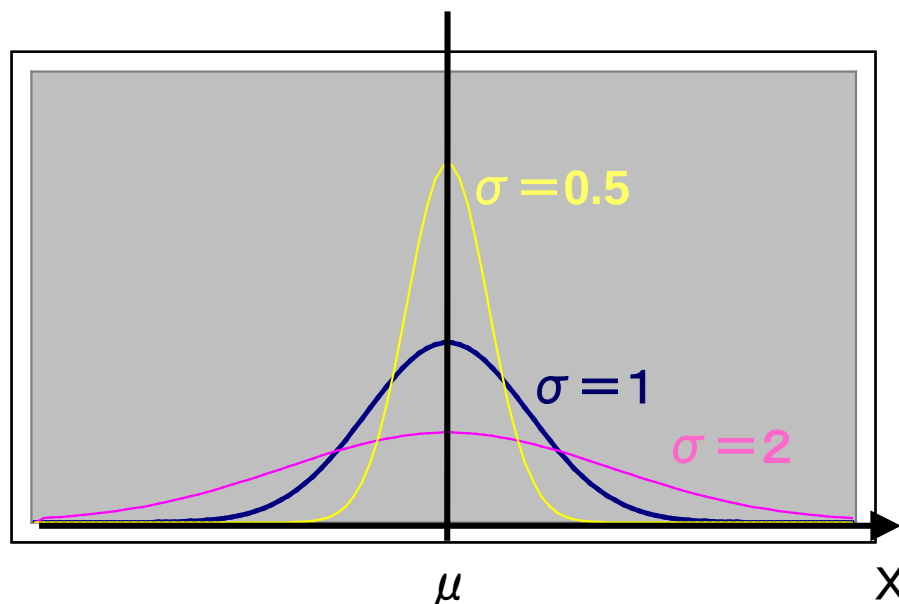
- 一様分布にしたがう乱数(一様乱数)は、Excel関数RAND()を使って生成することができる。



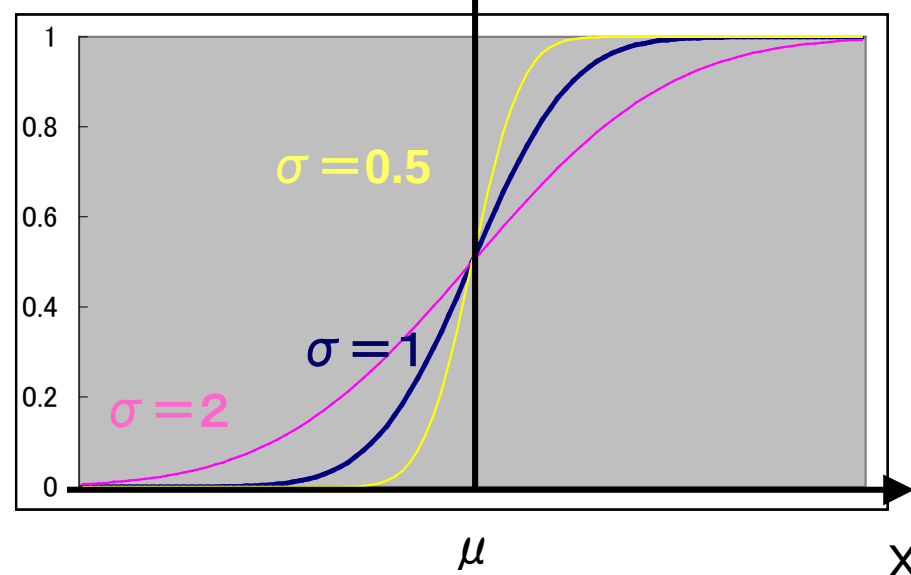
正規分布： 左右対称の釣鐘型をした確率分布。

平均(μ)、標準偏差(σ)を与えると分布の
EXCEL関数 ~~NORMDIST(X, μ , σ , 関数形式?)~~ **形状が決まる。** $\Rightarrow$ **関数形式?)**と表す。

f(X) 確率密度関数

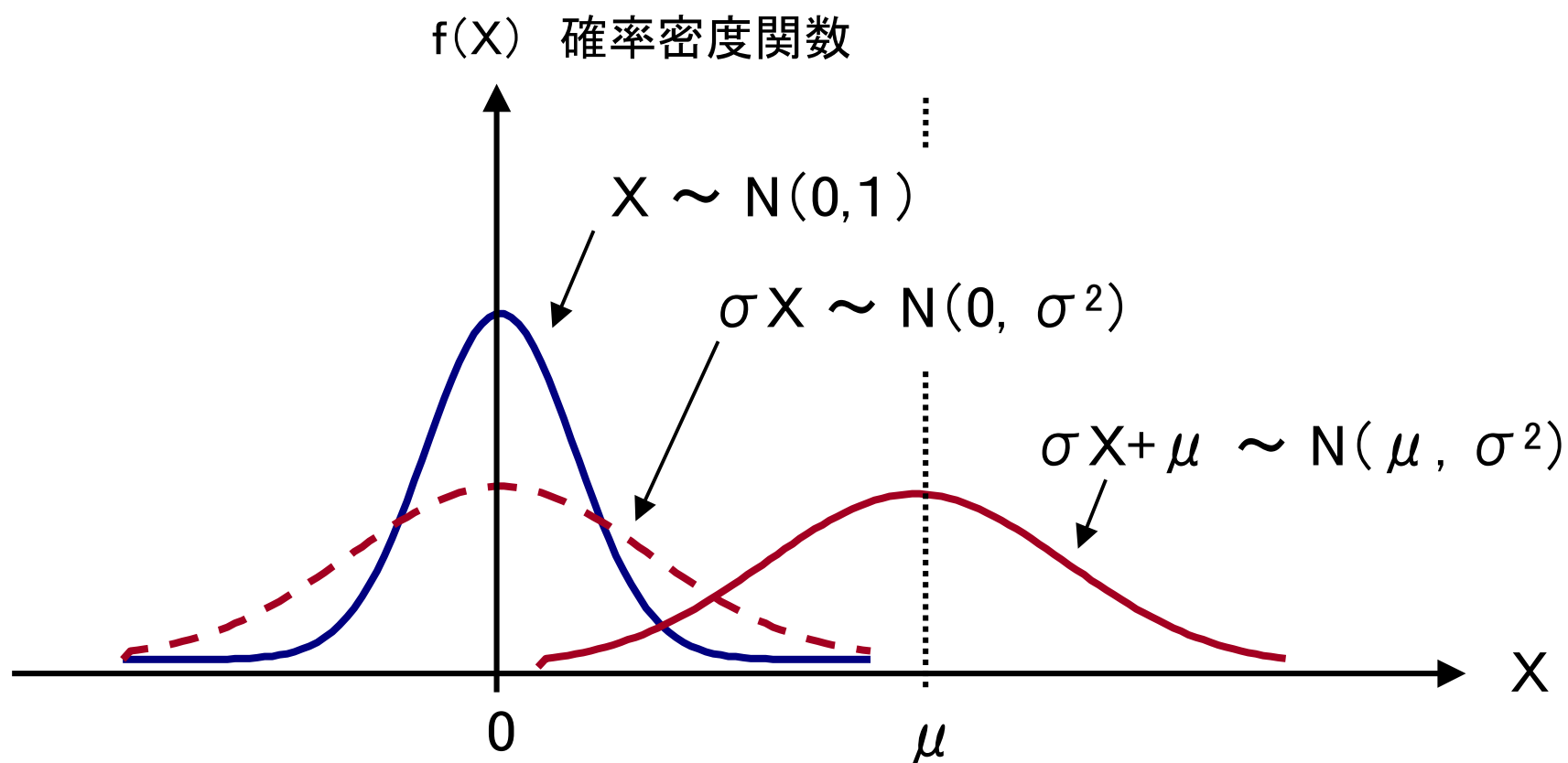


F(X) 分布関数



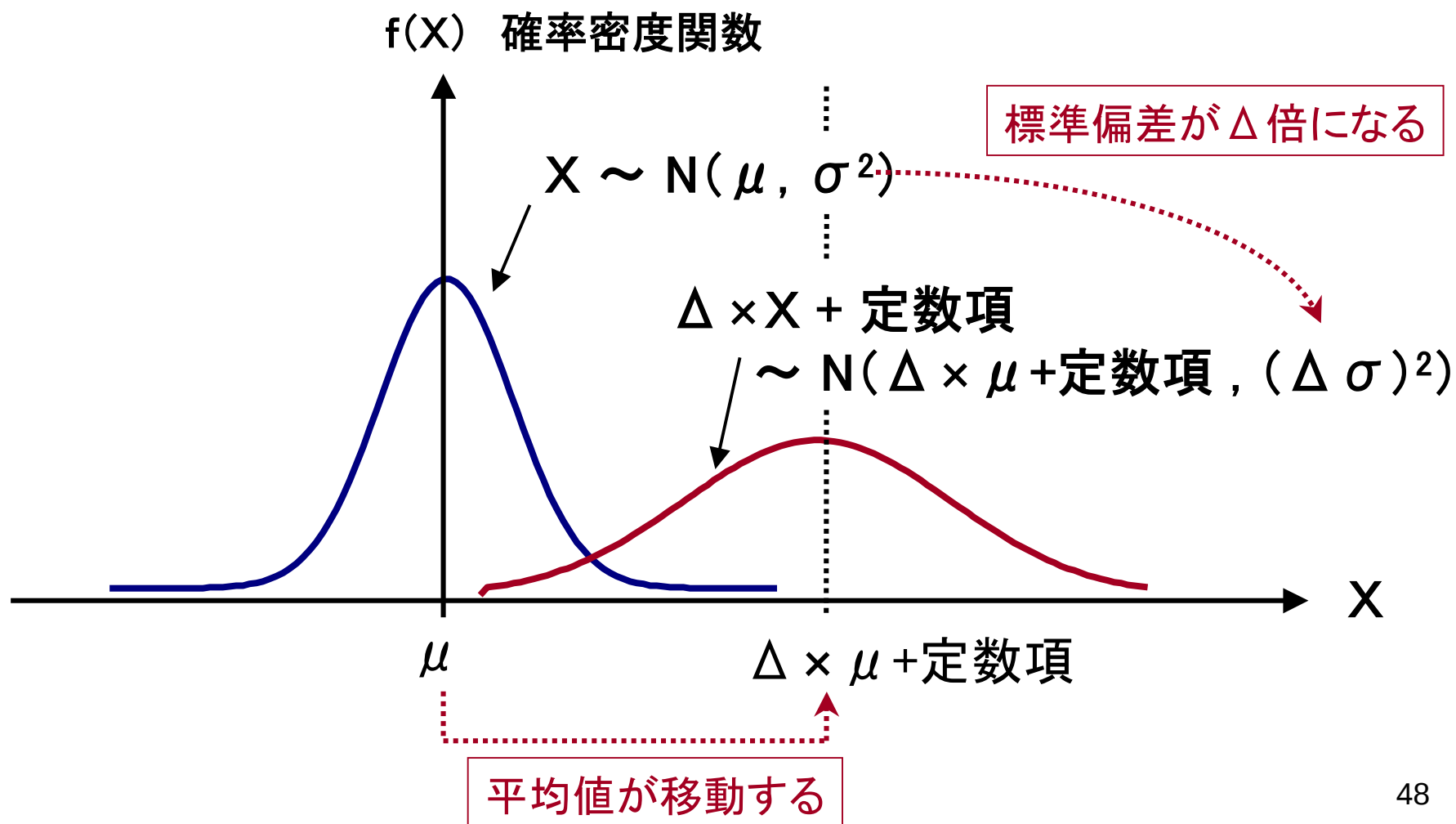
- 平均(μ)=0、標準偏差(σ)=1の正規分布を標準正規分布
と言い、 $N(0,1)$ と表す。

確率変数 X が 標準正規分布にしたがうとき
確率変数 $\sigma X + \mu$ は 正規分布にしたがう。



確率変数 X が 正規分布にしたがうとき

確率変数 $\Delta \times X + \text{定数項}$ は 正規分布にしたがう。





正規分布の特徴

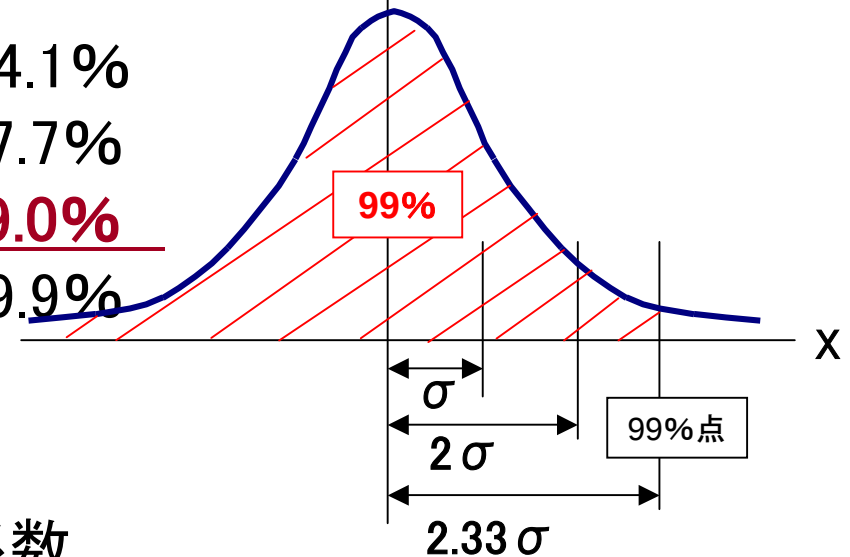
- 平均からどれだけ離れているか（標準偏差の何倍か）という情報から、 X 以下の値をとる確率が分かる。
- 例えば、 X が $N(0, \sigma^2)$ の正規分布にしたがって生起するとき

$X \leq \sigma$ となる確率は 84.1%

$X \leq 2\sigma$ となる確率は 97.7%

$X \leq 2.33\sigma$ となる確率は 99.0%

$X \leq 3\sigma$ となる確率は 99.9%



となることが知られている。

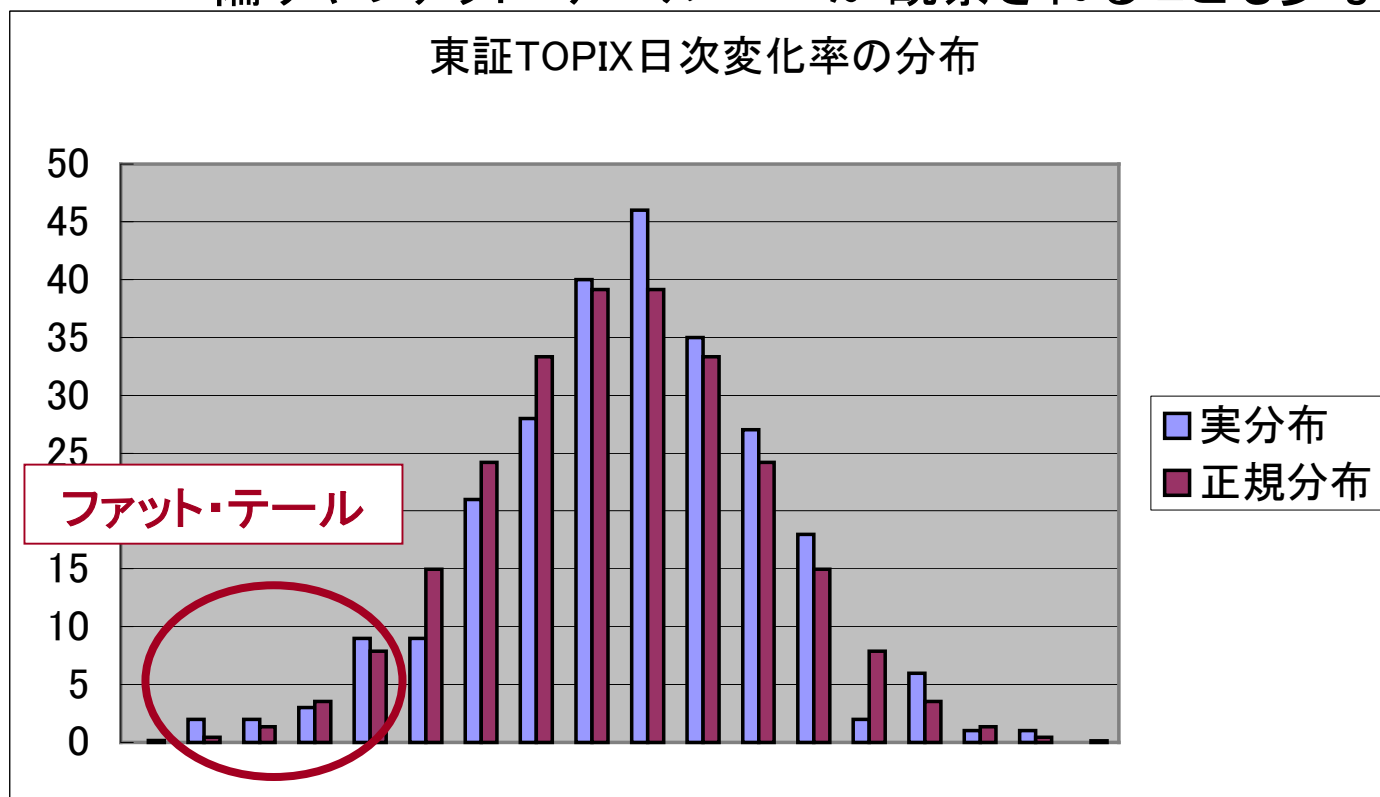
- このとき、 σ の前に付いている係数を「信頼係数」という。
- 正規分布は、 X が「信頼係数」 $\times \sigma$ 以下となる確率が分かる
便利な確率分布の1つ。



株価、金利、為替等の変化率は、正規分布にしたがうと想定されることが多い。

ー しかし、実際の分布をみると、正規分布と比較して、歪み、
(注)

偏りやファット・テール (注)が観察されることも少なくない。





対数正規分布:

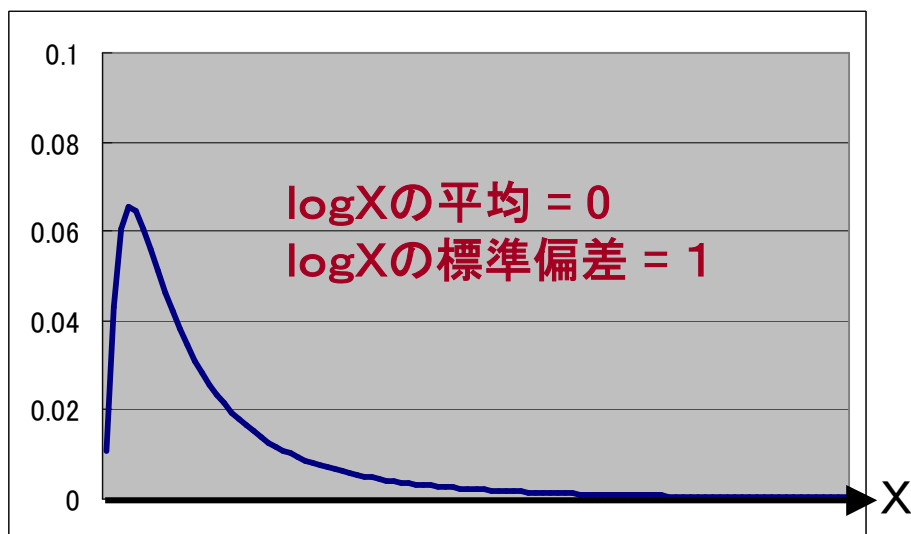
左右非対象、片側に裾野が長いファットテールな分布。

変数 X の対数値($\log X$)が正規分布にしたがうとき、
変数 X は対数正規分布にしたがう、という。

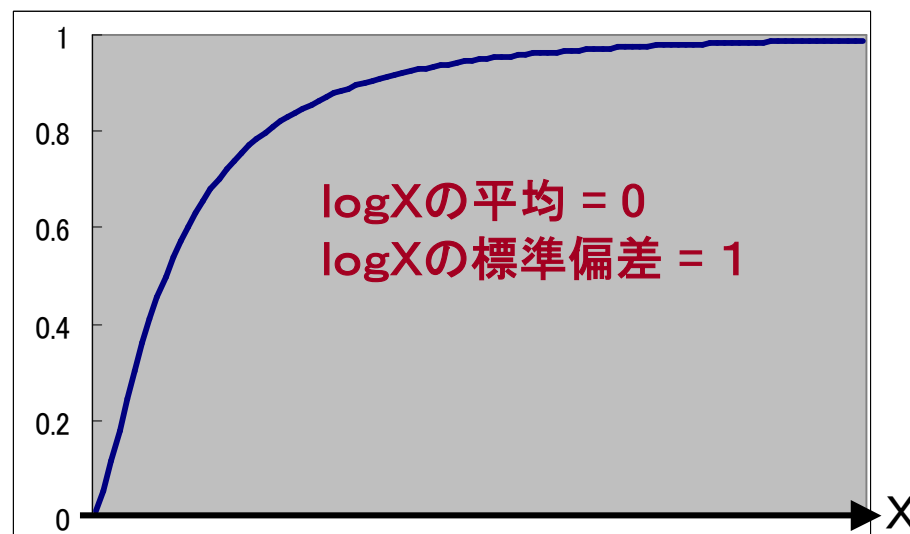
**$\log X$ の平均(μ)、 $\log X$ の標準偏差(σ)を与えると
分布の形状が決まる。**

EXCEL関数 LOGNORMDIST(X, μ, σ)

$f(X)$ 確率密度関数



$F(X)$ 分布関数





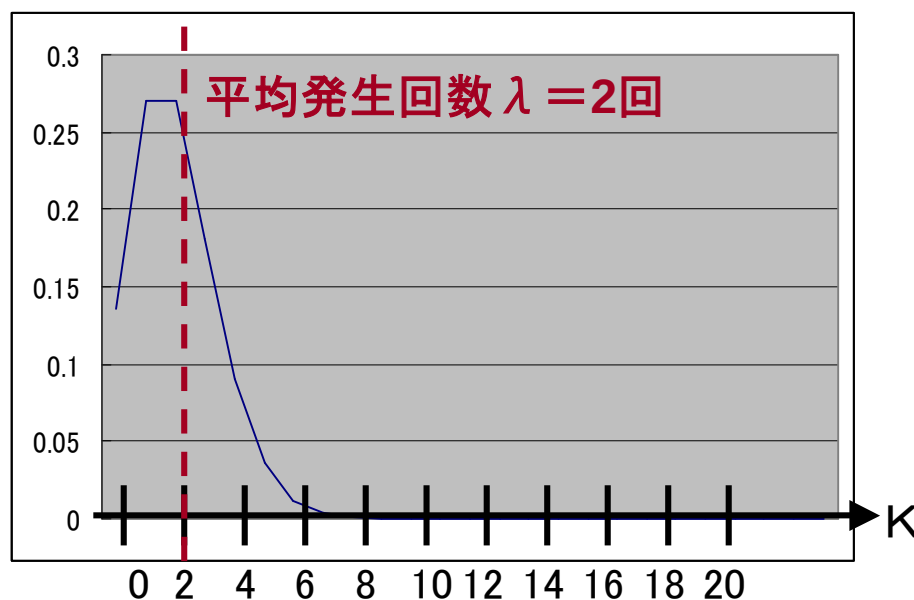
ポワソン分布:

所与の領域、あるいは、所与の時間内において、0回、1回、2回、3回・・・と発生する事象が、ちょうどK回発生する確率を示す。

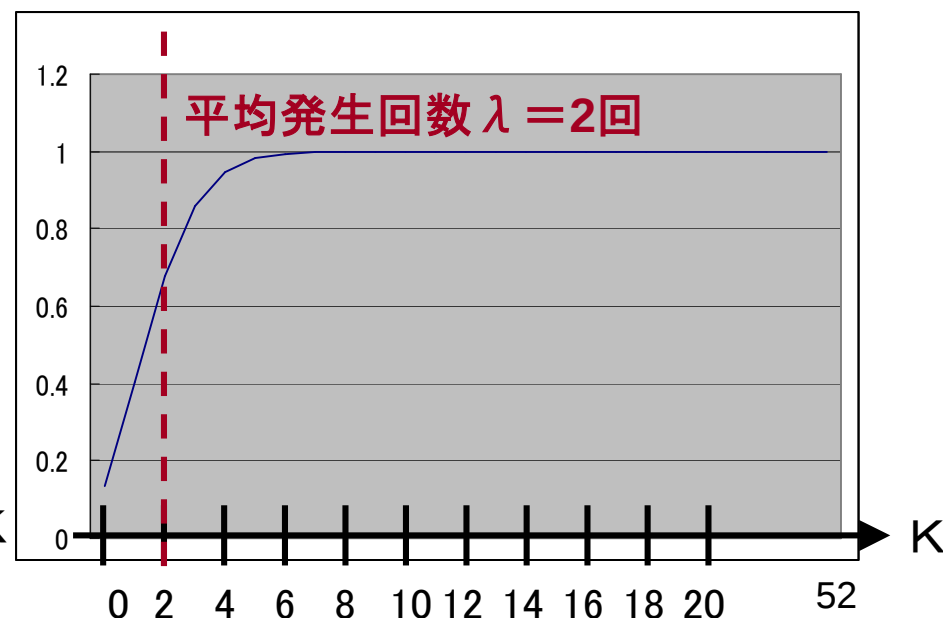
平均発生回数(λ 回)を与えると分布の形状が決まる。

EXCEL関数 POISSON(K, λ , 関数形式)

$f(K)$ 確率密度関数



$F(K)$ 分布関数





講義(Ⅲ.)の中で、

市場VaRを計測(分散共分散法)するとき
正規分布を利用する例をあげます。

信用VaRを計測(モンテカルロ・シミュレー
ション法)するとき、正規分布を利用する例
をあげます。

オペリスクVaRを計測(モンテカルロ・シミュレ
ーション法)するとき、対数正規分布とポワソン
分布を利用する例をあげます。

— 実務的には、フィットのよい別の確率分布
を利用することもあります。



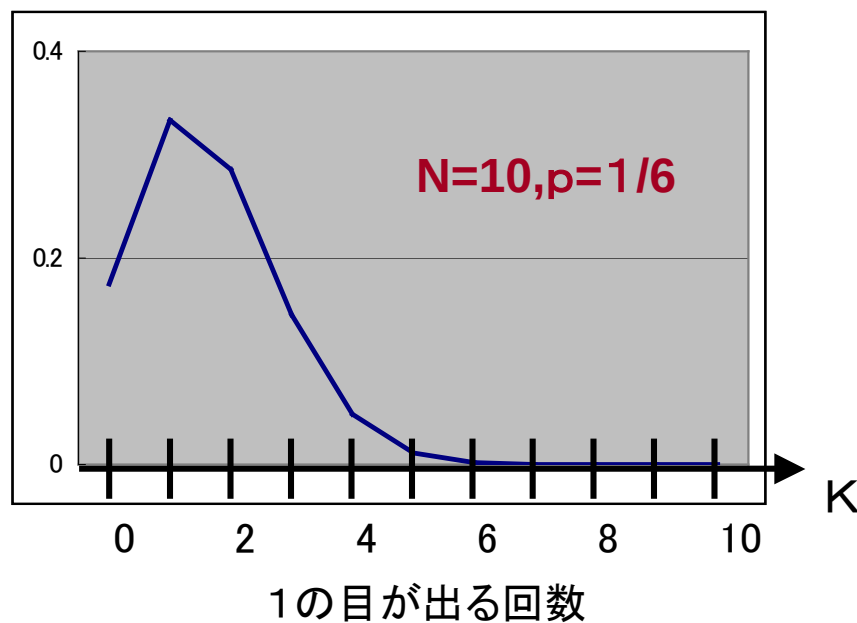
2項分布:

結果が2通りある試行(実験)を N 回繰り返したとき、片方の結果が起こる回数(K)の確率分布。

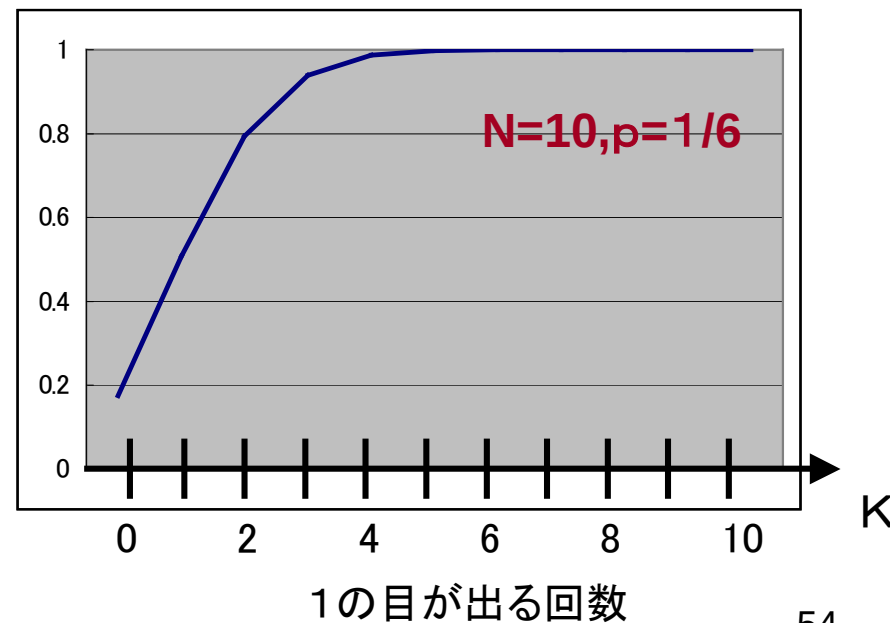
試行回数(N 回)と、片方の結果が起きる確率(p)を与えると分布の形状が決まる。

(例)サイコロを10回振って 1の目が出る回数(K)

$f(K)$ 確率密度関数



$F(K)$ 分布関数



ある事象が起きる確率は p 。
N回の試行のうち、K回は
ある事象が起きる。

ある事象が起きない確率は $1 - p$ 。
N回の試行のうち、 $N - K$ 回は
ある事象は起きない。

2項分布(Excel関数)

$$\text{BINOMDIST}(K, N, p, \text{false}) = {}_N C_K p^K (1 - p)^{N-K}$$

N回の試行の中から ある事象が起きるK回の試行を
取り出す組み合わせ

$${}_N C_K = \frac{N \times (N-1) \times \cdots \times (N-K+1)}{K \times (K-1) \times \cdots \times 2 \times 1}$$

(例)サイコロを10回振ったときに2回、1の目が出る確率

$$\text{BINOMDIST}(2, 10, 1/6, \text{false})$$

$$= {}_{10} C_2 (1/6)^2 (5/6)^{10-2} = \frac{10 \times 9}{2 \times 1} \times (1/6)^2 (5/6)^8$$



講義の中で、

VaR計測モデルのバックテストを行なう
とき、2項分布を利用します。

(4) 確率変数の独立

【定義】

- 確率変数 X_1 、 X_2 が互いに影響されず、それぞれの確率分布にしたがって値をとるとき、

確率変数 X_1 、 X_2 は、互いに「独立」であるという。

(例)サイコロを振ったときに出る目の数

1回目: $X_1 = 1$ 、2回目: $X_2 = 1$

3回目: $X_3 = ?$

サイコロの目 (X_3)	1	2	3	4	5	6
確 率	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$

- 2回続けて1の目が出ても、3回目の結果には影響を及ぼさない。
- 3回目は、いずれの目が出る確率も $1/6$ 。

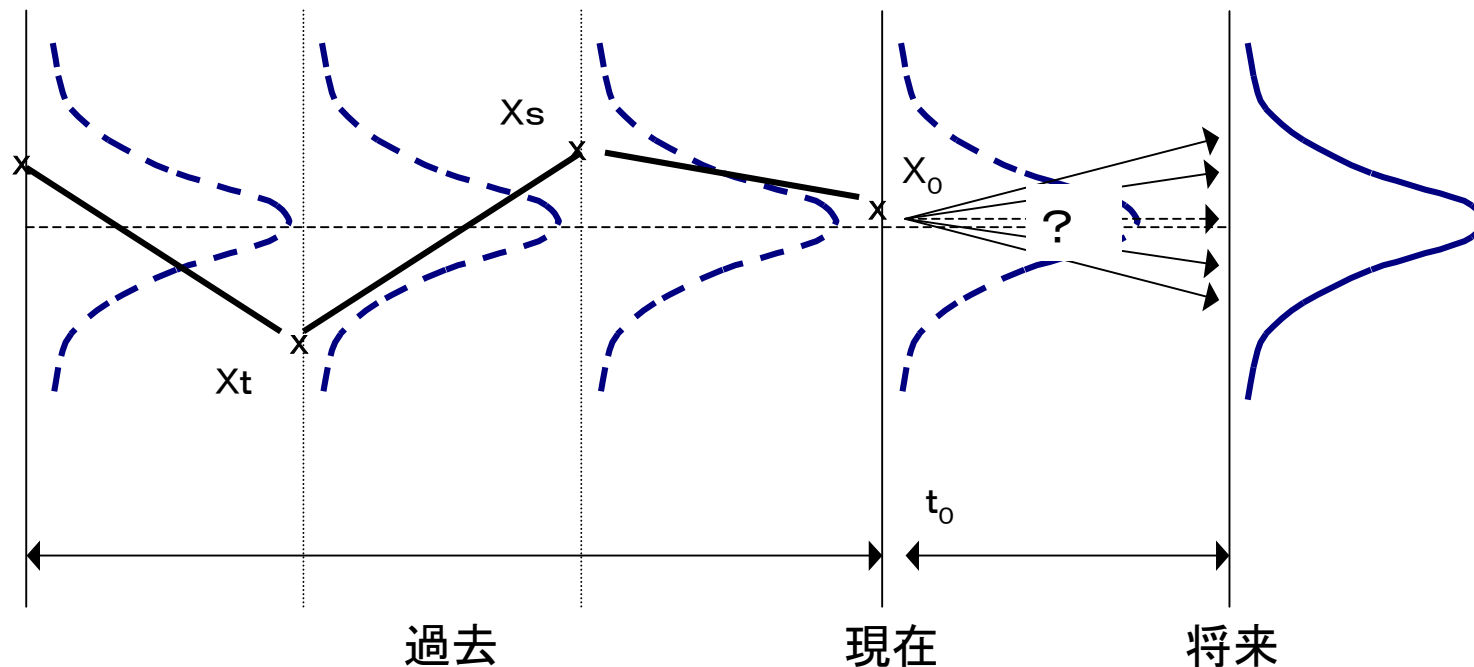


株価、金利、為替等の変化について

互いに独立かつ同一の確率分布にしたがって
変動している、と考えられることが多い。

⇒ i.i.d.の想定

確率変数 X の推移と、その確率分布



【独立の定義】

- 確率変数 X_t 、 X_s の確率関数に関して、以下の式が成り立つとき、確率変数 X_t 、 X_s は互いに「独立」と言う

$$P(X_t=a, X_s=b) = P(X_t=a)P(X_s=b)$$

【i. i. d. の定義】

- 確率変数 X_t 、 X_s について、以下の2つの条件を満たすとき、

確率変数 X_t 、 X_s は互いに「i. i. d. 」(注)であると言う。

(注)independently and identically distributed

- ① 確率変数 X_t 、 X_s は互いに独立である。
- ② 確率変数 X_t 、 X_s は同一の確率分布にしたがう。

【定理】

- 確率変数 X_1 、 X_2 が互いに「独立」のとき、以下のことが成り立つ。

- ① 確率変数 $X_1 X_2$ の期待値は、それぞれの確率変数の期待値の積になる。

$$E(X_1 X_2) = E(X_1) E(X_2)$$

- ② 確率変数 $X_1 + X_2$ の分散は、それぞれの確率変数の分散の和に等しい。

$$V(X_1 + X_2) = V(X_1) + V(X_2)$$

- ③ 確率変数 X_1 と X_2 は無相関である。

$$\rho(X_1, X_2) = 0$$

(証明省略)

【ルートT倍ルール】

日次ベースの対数変化率 or 変化幅を $X_1, X_2, X_3, \dots, X_T$ とすると、

T日間の対数変化率 or 変化幅は $X_1 + X_2 + X_3 + \dots + X_T$ と表される。

各期のリスクファクター($X_1, X_2, X_3, \dots, X_T$)が、**互いに独立かつ同一の確率分布にしたがう**と想定する。 【i.i.dの定義】

日次ベースの対数変化率 or 変化幅 $X_1, X_2, X_3, \dots, X_T$ の分散を σ^2 とすると、
標準偏差を σ とすると、

T日間の対数変化率 or 変化幅 $X_1 + X_2 + X_3 + \dots + X_T$ の分散は $T \times \sigma^2$ とすると、
標準偏差は $\sqrt{T} \times \sigma$ となる。

(参考)対数変化率の定義

日次対数変化率

$$\log \frac{X_t}{X_{t-1}} \doteq \frac{X_t - X_{t-1}}{X_{t-1}} = \frac{X_t}{X_{t-1}} - 1$$

10日間対数変化率

$$\log \frac{X_t}{X_{t-10}} \doteq \frac{X_t - X_{t-10}}{X_{t-10}} = \frac{X_t}{X_{t-10}} - 1$$

- 対数変化率は、通常の変化率と近似的に等しいことが知られている。
- $\log$ (自然対数)は、Excelでは関数 $\text{LN}(\cdot)$ で与えられる。

対数変化率の特徴

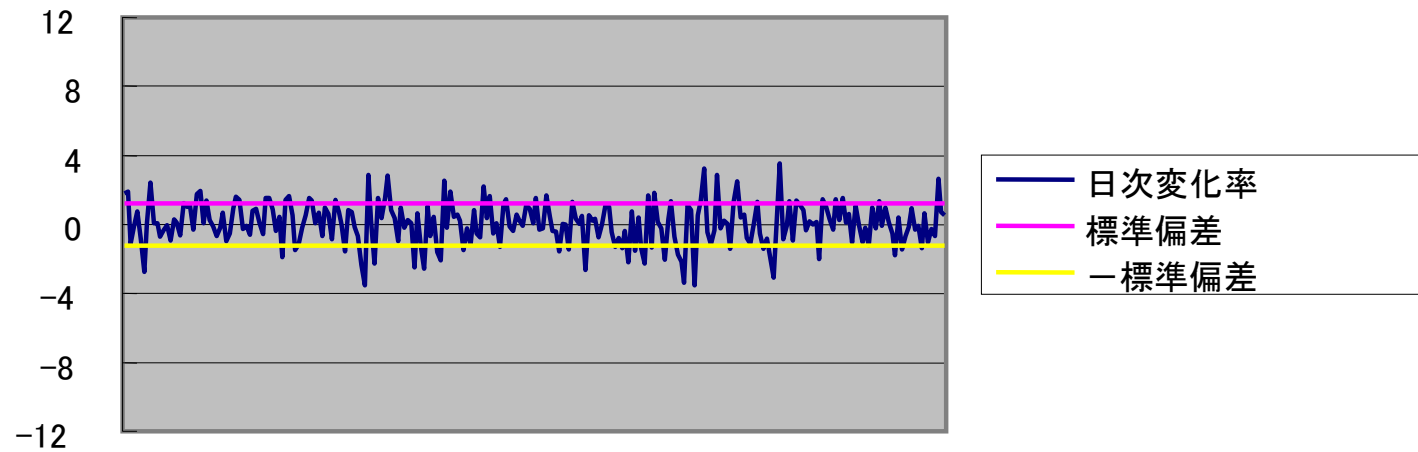
- 対数変化率は、同率の低下、上昇により、元の値に戻る。
- 10日間対数変化率は、日次対数変化率(10日分)の和となる。

	変化率(日次)	対数変化率(日次)
100	0.0101	0.0101
99	-0.0100	-0.0101
100	0.0526	0.0513
95	-0.0500	-0.0513
100	0.1111	0.1054
90	-0.1000	-0.1054
100	0.2500	0.2231
80	-0.2000	-0.2231
100	0.4286	0.3567
70	-0.3000	-0.3567
100	0.6667	0.5108
60	-0.4000	-0.5108
100	1.0000	0.6931
50	-0.5000	-0.6931
100	—	—

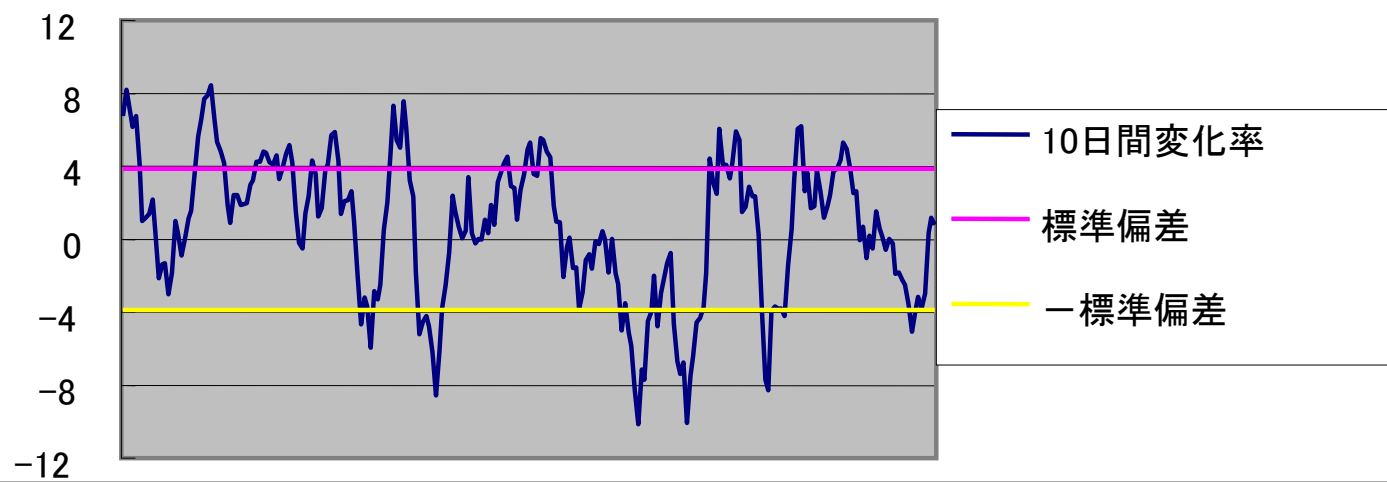
		対数変化率(日次)
X10	100	0.2877
X9	75	-0.4700
X8	120	1.3863
X7	30	-0.6931
X6	60	-0.9163
X5	150	0.5108
X4	90	1.0986
X3	30	-0.6931
X2	60	-0.2877
X1	80	-0.1178
X0	90	—
$\sum \log(X_t/X_{t-1})$		0.1054

	対数変化率(10日間)
$\log(X_{10}/X_0)$	0.1054

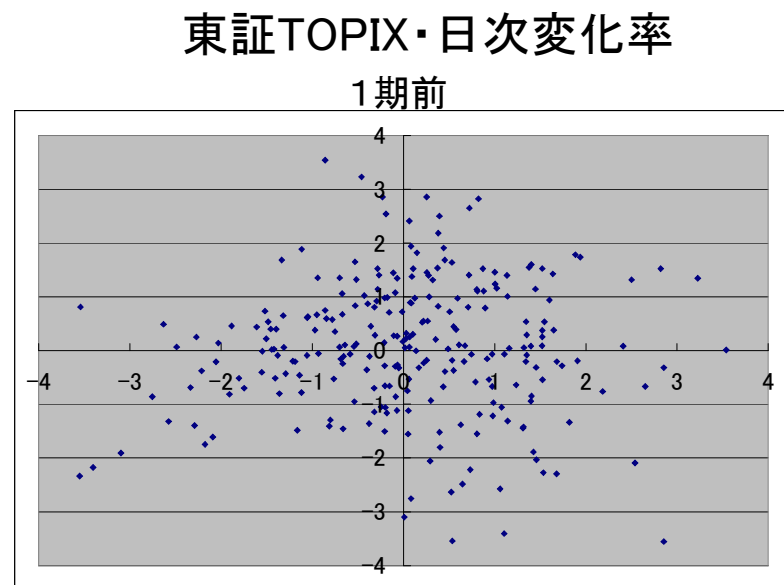
東証TOPIX日次変化率の推移



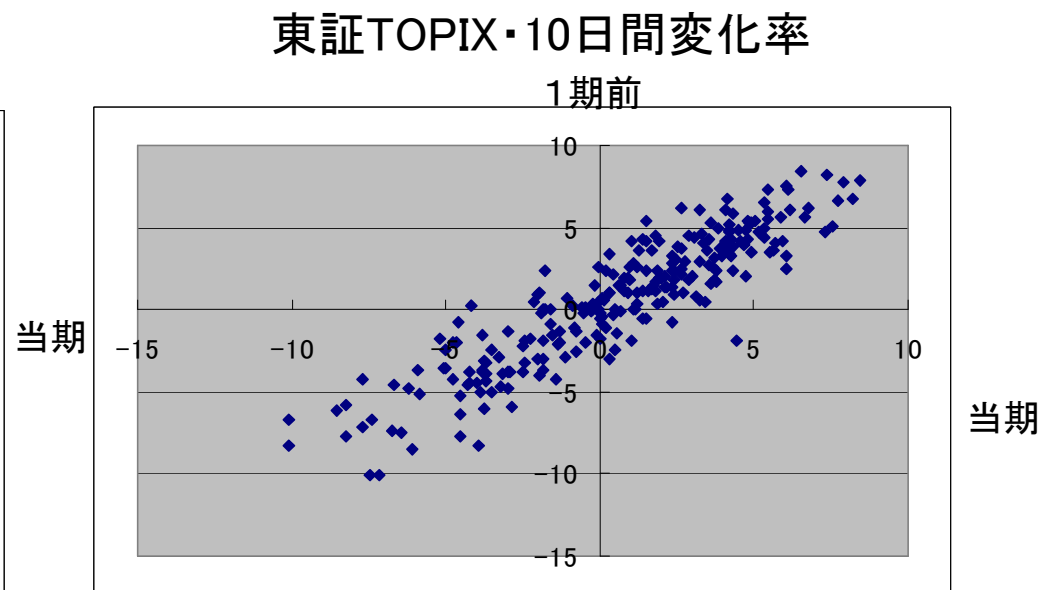
東証TOPIX10日間変化率の推移



- 下図は、過去1年間のデータをもとに、東証TOPIX・変化率と、1期前の変化率との相関関係（自己相関）をみたもの。



相関係数 $\rho = 0.037$



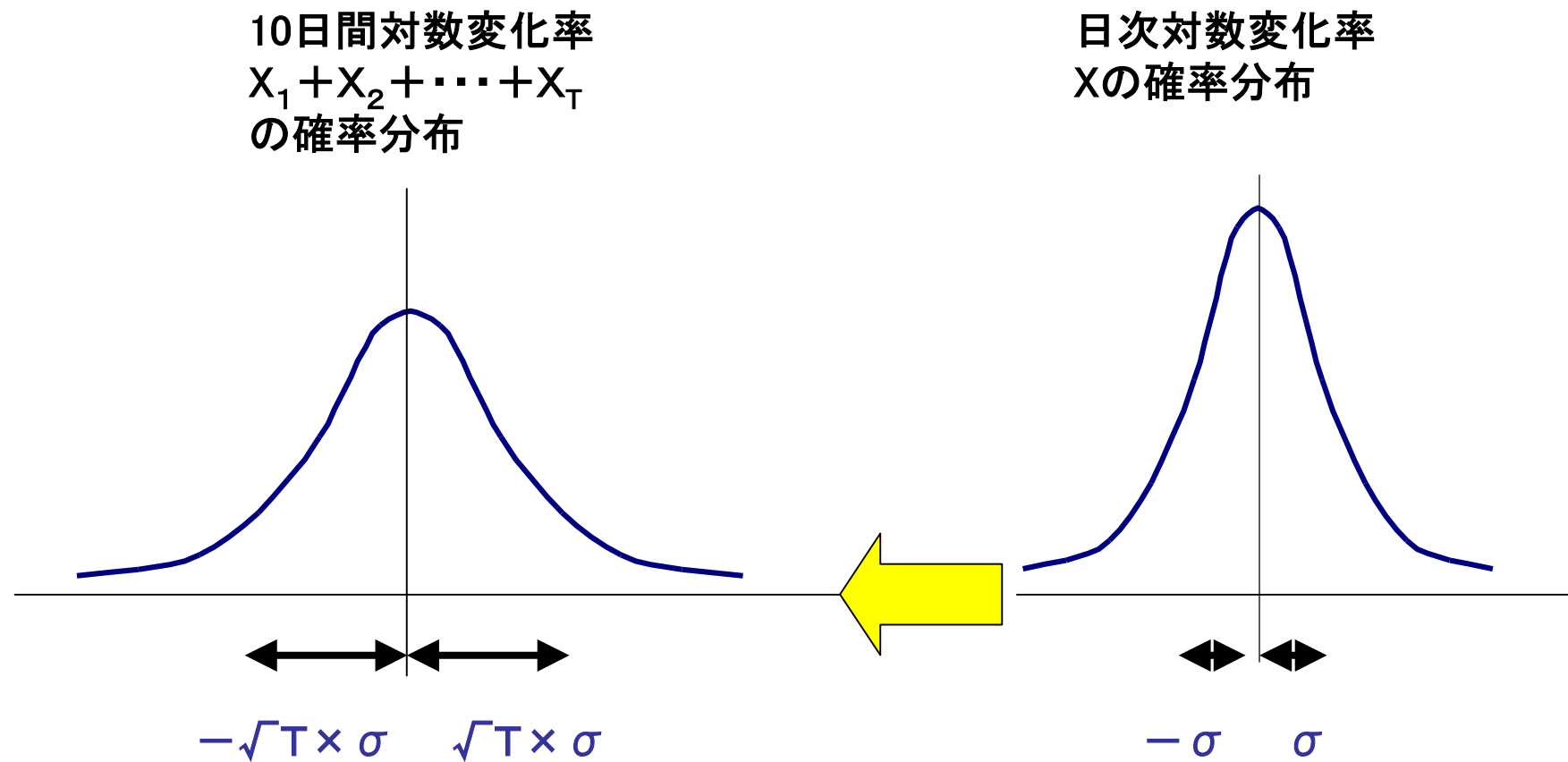
相関係数 $\rho = 0.905$

- 日次変化率の自己相関は弱い、10日間変化率の自己相関は強いことが観察される。
- 統計的に厳密に検証すると、多くの時系列データが（日次変化率でみても10日間変化率でみても）独立とは言えないことが多い。

基本統計量	Excel関数	日次 対数変化率	10日間 対数変化率
データ数	COUNT	250	250
平均	AVERAGE	0.063	0.656
分散	VARA	1.540	14.966
標準偏差	STDEVA	1.241	3.869

- 分散をみると、10日間対数変化率の分散は、日次対数変化率の分散の概ね10倍となっている。
- 標準偏差をみると、10日間対数変化率の標準偏差は、日次対数変化率の標準偏差の概ね $\sqrt{10}$ 倍(=3.162倍)となっている。

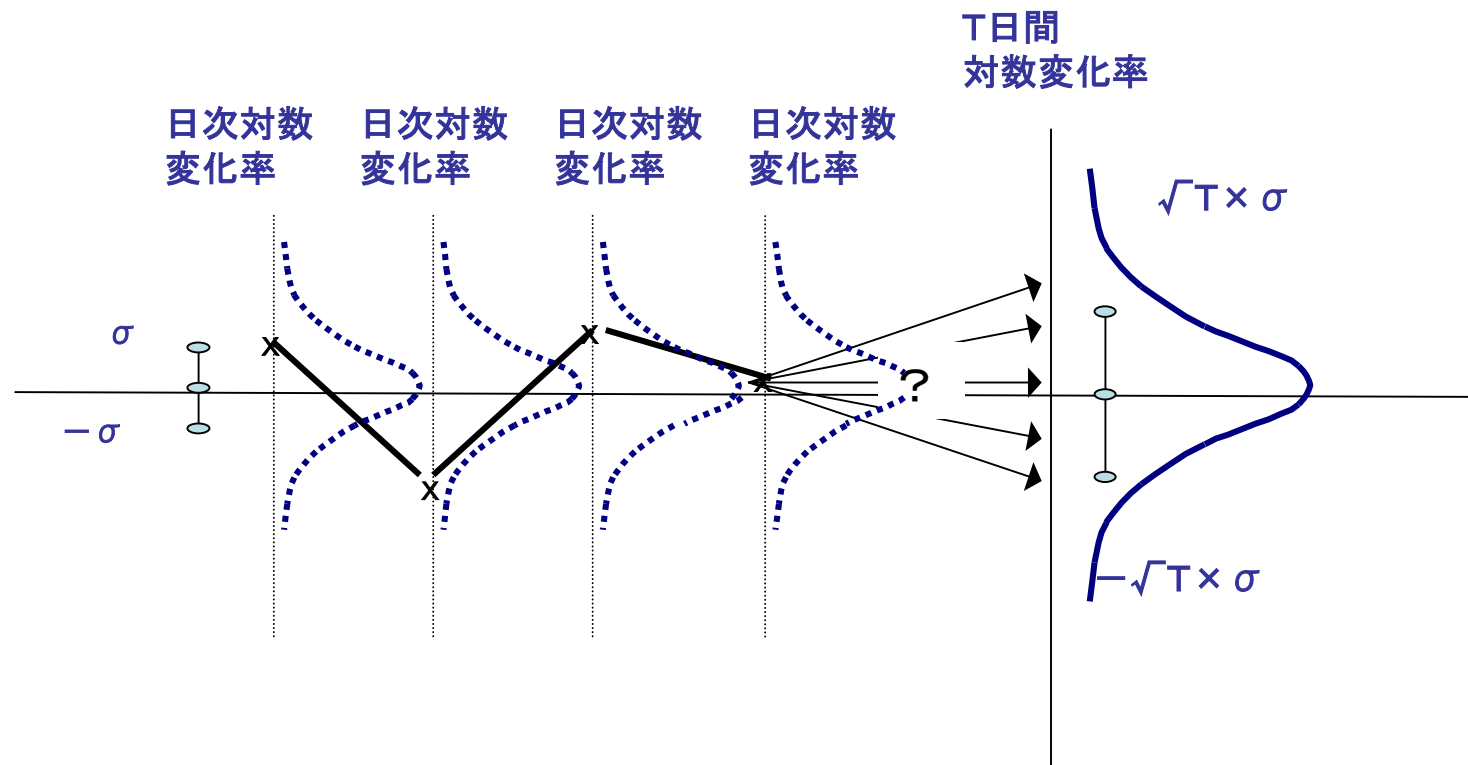
ルートT倍ルール



仮定

リスクファクターの確率分布は i. i. d.

ルートT倍ルール



仮定

リスクファクターの確率分布は i. i. d.

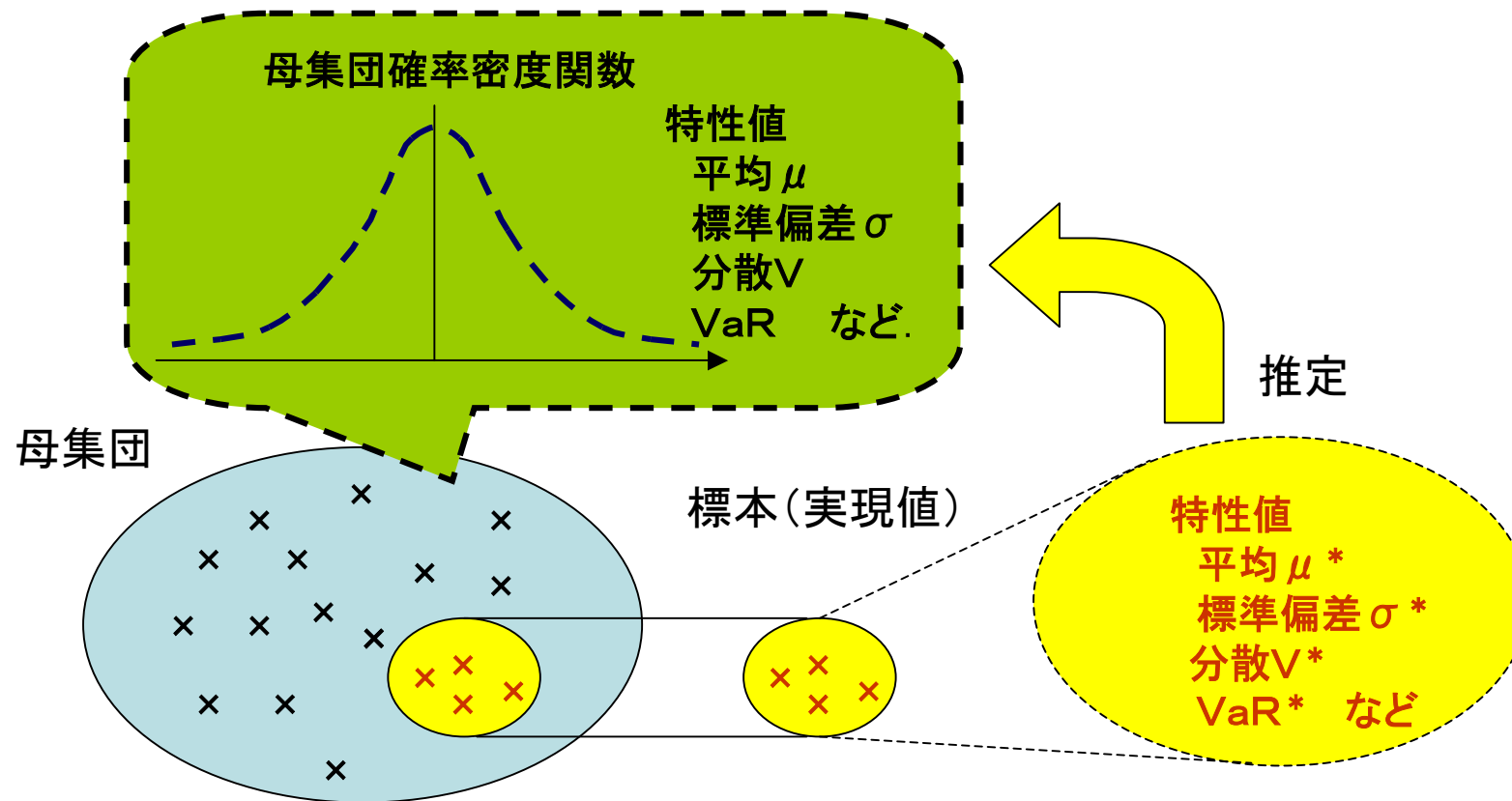
4. 推定と検定

(1) 推 定

(2) 検 定

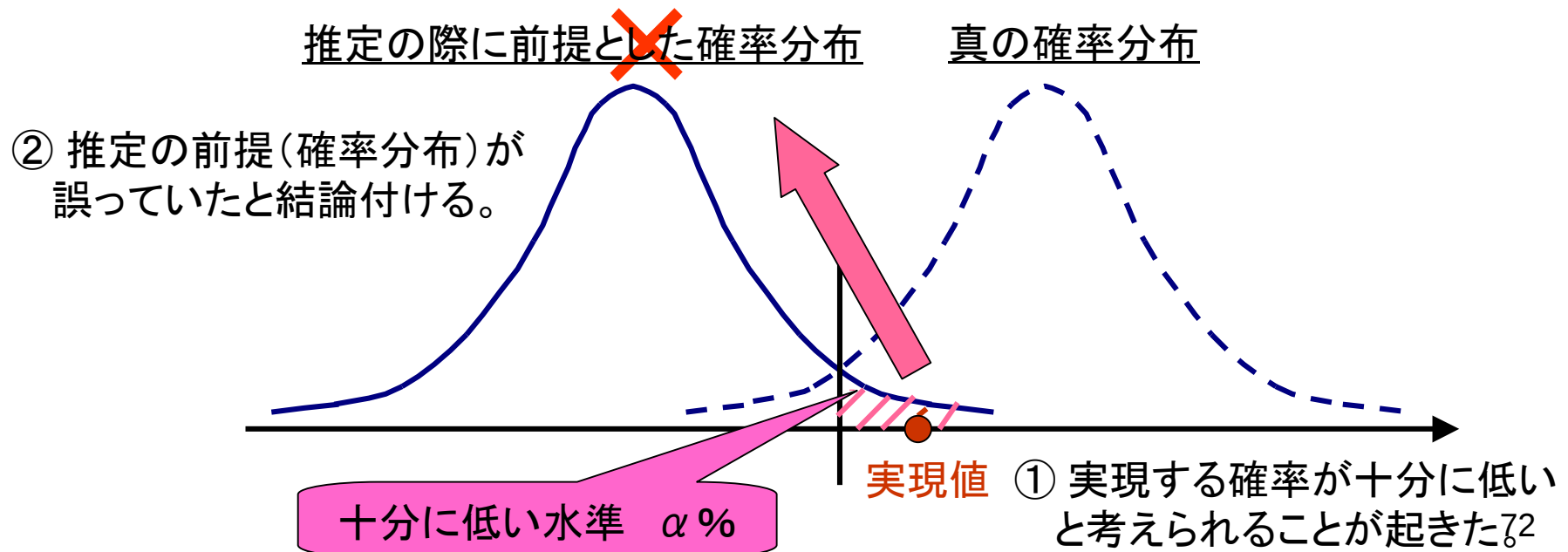
(1) 推 定

- 母集団の確率分布、特性値は、誰にも分からない。
- 標本の特性値から母集団の特性値を統計的に推測する。



(2) 検 定

- 一定の確率分布を前提にして推定した値について、その値をとる確率が十分に低いとき、「偶然、珍しいことが起きた」と考えるのではなく、「推定の際に置いた前提が誤っていた」と結論付ける。



(設問)

1の目がやすいサイコロがあります。
サイコロを割ったり、X線透視などをせず、
サイコロを振るだけで、このサイコロが
「イカサマ」かどうかを決めたいと思います。

あなたは、このサイコロを600回振って、
何回、1の目が出たら、「イカサマ」だと判断しますか？

120回で「イカサマ」だと判断しますか？

150回で「イカサマ」だと判断しますか？

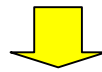
200回で「イカサマ」だと判断しますか？

300回で「イカサマ」だと判断しますか？

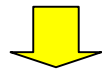
400回で「イカサマ」だと判断しますか？

(例) 1の目がやすい「イカサマ・サイコロ」の見付け方

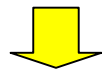
- このサイコロを振ったとき、1の目が出る確率は $1/6$ である。



- このサイコロを600回振ったとき、1の目が ? 回以上発生した。



- このサイコロを振ったとき、1の目が出る確率が $1/6$ だとすると、600回のうち ? 回以上、1の目が出る確率は十分に低い (例えば0.1%未満) ことが分かる。



- このサイコロを振ったとき、1の目が出る確率は $1/6$ とは言えない。

N回の観測で、K回、1の目が出る確率

2項分布 ${}_N C_K p^K (1-p)^{N-K}$

N=600回

p=1/6

1-p=5/6

K回	確率	確率	K回以上
0	0.000%	100.000%	0回以上
100	4.264%	60.278%	100回以上
110	2.904%	20.634%	110回以上
120	0.652%	3.051%	120回以上
130	0.052%	0.184%	130回以上
140	0.002%	0.004%	140回以上
150	0.000%	0.000%	150回以上
160	0.000%	0.000%	160回以上
170	0.000%	0.000%	170回以上
180	0.000%	0.000%	180回以上
190	0.000%	0.000%	190回以上
200	0.000%	0.000%	200回以上
300	0.000%	0.000%	300回以上
400	0.000%	0.000%	400回以上
500	0.000%	0.000%	500回以上
600	0.000%	0.000%	600回以上

検定の一般的手続き

- ①「帰無仮説」を立てる。
- ②「帰無仮説」が「真」(true)であるという仮定の下に「検定統計量」を決定する。
 - ただし「検定統計量の確率分布は既知とする。
- ③試行や標本(サンプル)の抽出により、「検定統計量」を計算する。
- ④「検定統計量」の実現値(計算値)がどの程度の確率でおき得ることかを確認する。
- ⑤「検定統計量」の実現値(計算値)が十分に低い確率(「有意水準」以下)でしか置きえないとき、「帰無仮説」を棄却する。

2種類の過誤

- 「検定」では、次の2通りの「過誤」(エラー)が起きる可能性がある。
- したがって、バックテストの結果も「過誤」(エラー)を伴っている可能性がある点、注意を要する。

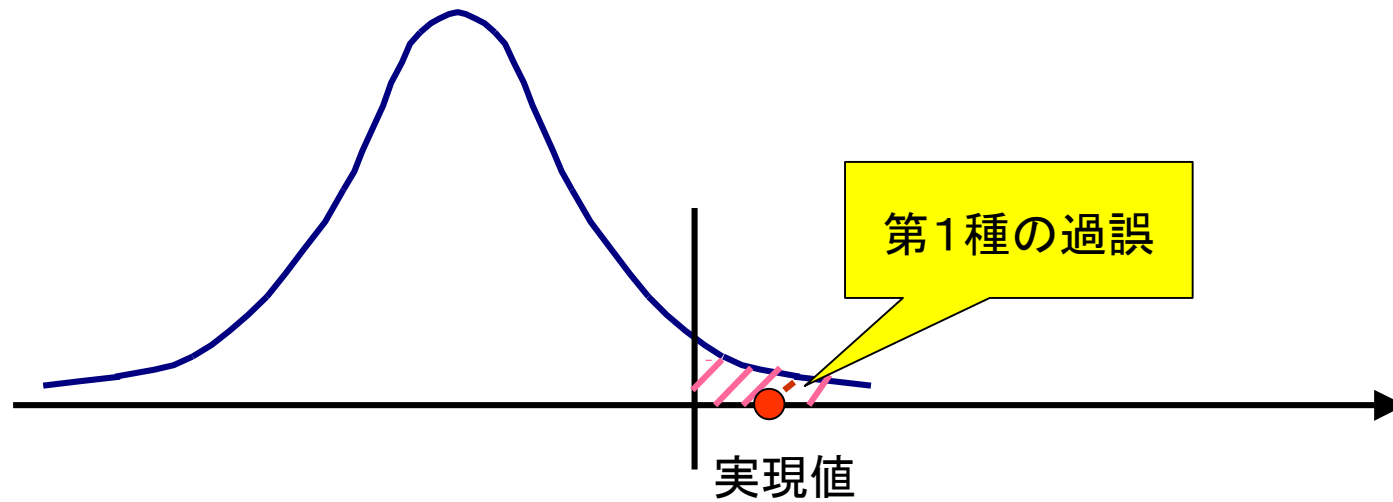
第1種の過誤(エラー)

本当は帰無仮説が正しいのに、
検定の結果、
帰無仮説が誤っていると結論付けてしまう。

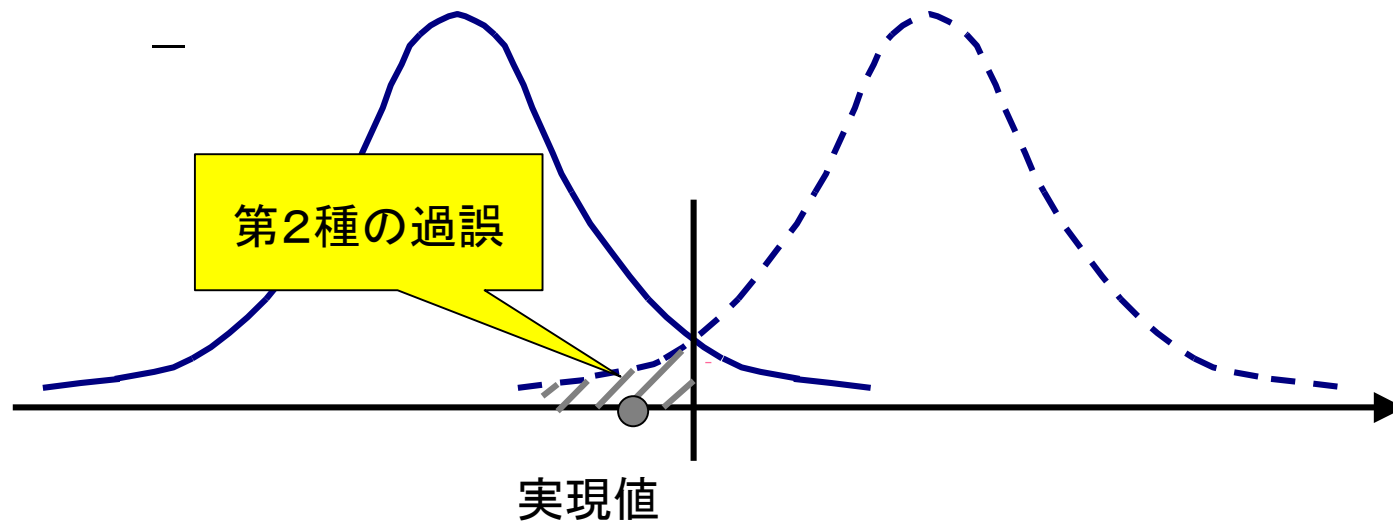
第2種の過誤(エラー)

本当は帰無仮説が正しくないのに、
検定の結果、
帰無仮説が正しいと結論付けてしまう。

推定に利用した確率分布 = 真の確率分布



推定に利用した確率分布 $\neq$ 真の確率分布





VaR計測モデルのバックテストは
「検定」の考え方に基づいて行います。

参考文献・資料

「イラスト・図解 確率・統計のしくみが分かる本」

長谷川勝也 著 技術評論社

「初等統計学」

P.G.ホーエル 著 浅井晃、村上正康 訳 培風館

日本銀行「市場リスク管理の基礎」セミナー

補足1 「確率・統計の基礎」

金融高度化センター 碓井茂樹