エクセルの準備

- NORM.S.DIST(z, TRUE): 標準正規分布のz値から累積確率p値を算出する
- 標本分散 VAR.P(データ配列)、不偏分散 VAR.S(データ配列)
→sum{(データ-データの平均)^2}/データ数
　sum{(データ-データの平均)^2}/(データ数-1)、にそれぞれ対応
- (次回使う)「データ」タブ→（分析ツールのチェックボックスにチェックを入れる）
→**データ分析**
もしない場合は、「ファイル」タブ→オプション→アドイン→Excelアドインの設定
→分析ツールを選択 (参考URL)
- (今回は使わない: t検定)TTEST(データ範囲1, データ範囲2, 1:片側or2:両側, 1:対応のあるデータの検定)：p値を返す

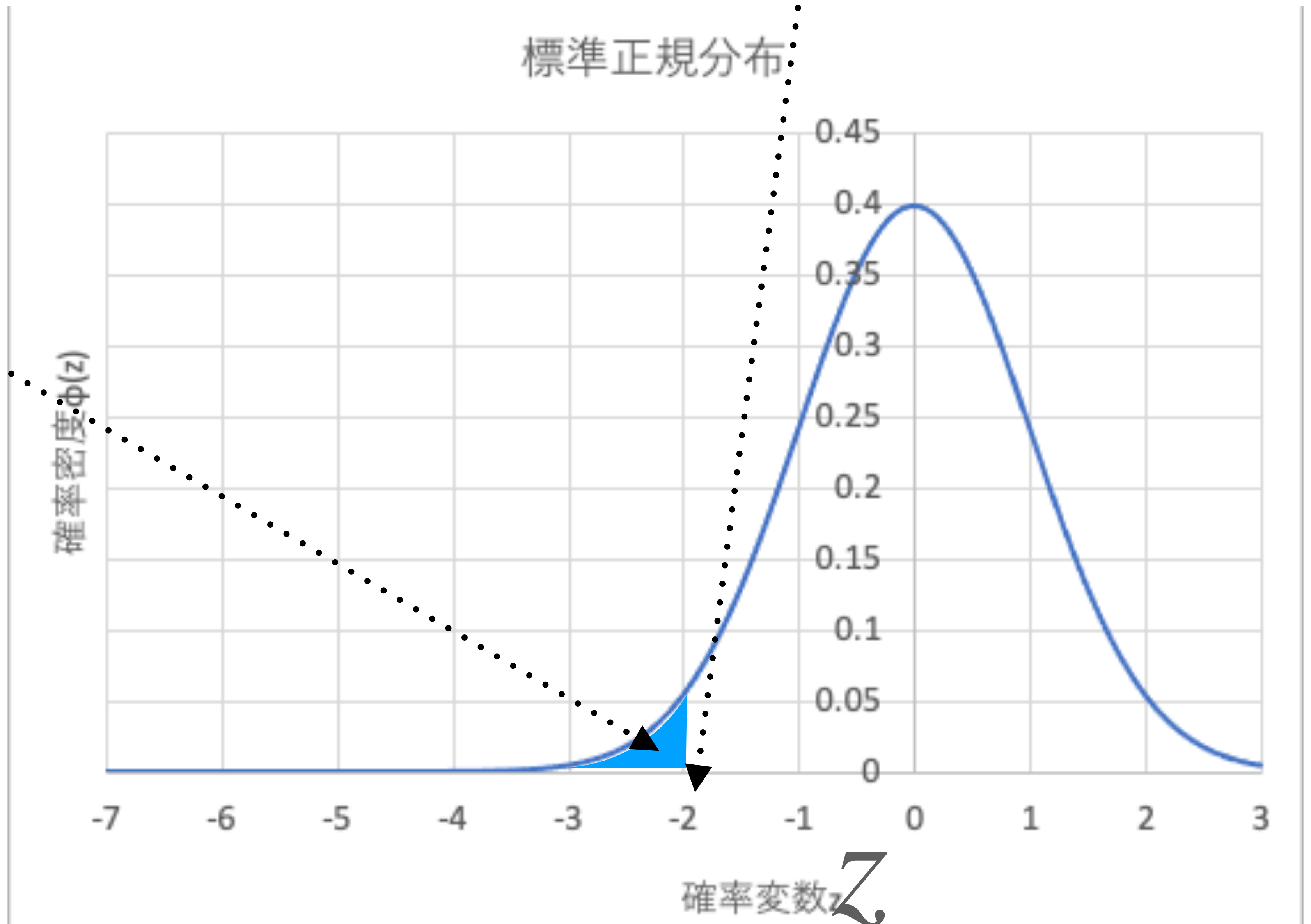
仮説・検定の手順

1. 帰無仮説を立てる
   (主張したい命題の否定を仮定する)

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

2. 確率を計算する　μ　$\bar{X}$
   z検定の場合は、母平均と標本平均の差を使って標準化変換→検定値zを求める
   →zから累積確率(p値:zが分布の端からどれだけの面積に位置するか)を求める

3. 帰無仮説を棄却する（危険率 or 有意水準 5%, 0.5%はよく使う）
   (危険率5%の時は、p値<0.05で仮定を棄却、p値＞0.05だと棄却できない)



例題　テキストp.119 & Excel sheet1

ある探検家が洞窟で太古の硬貨25枚を発見した。発見された硬貨の重さの標本平均は34.77(g)であった。一方、実際に太古の時代に流通していた効果の重さは、平均35.03(g)、標準偏差0.925(g)であった。(各硬貨の重さはp.118 orエクセル参照)
この時、探検家が発見した硬貨は、重さから本物と言えるか？危険率（有意水準）10%で検定せよ。

帰無仮説：洞窟で発見された硬貨の重さの平均=本物の硬貨の平均値
対立仮説：洞窟で発見された硬貨の重さの平均≠本物の硬貨の平均値
→もし洞窟で発見された硬貨の重さの平均が、本物の硬貨の平均よりも軽いor重いのどちらかを検定する場合は片側検定　(p値と危険率0.1を比較すれば良い)
→今回の帰無仮説は、洞窟で発見された硬貨の重さの平均が軽くても重くてもダメ→両側検定
→p値と、危険率0.1/2=0.05を比較する

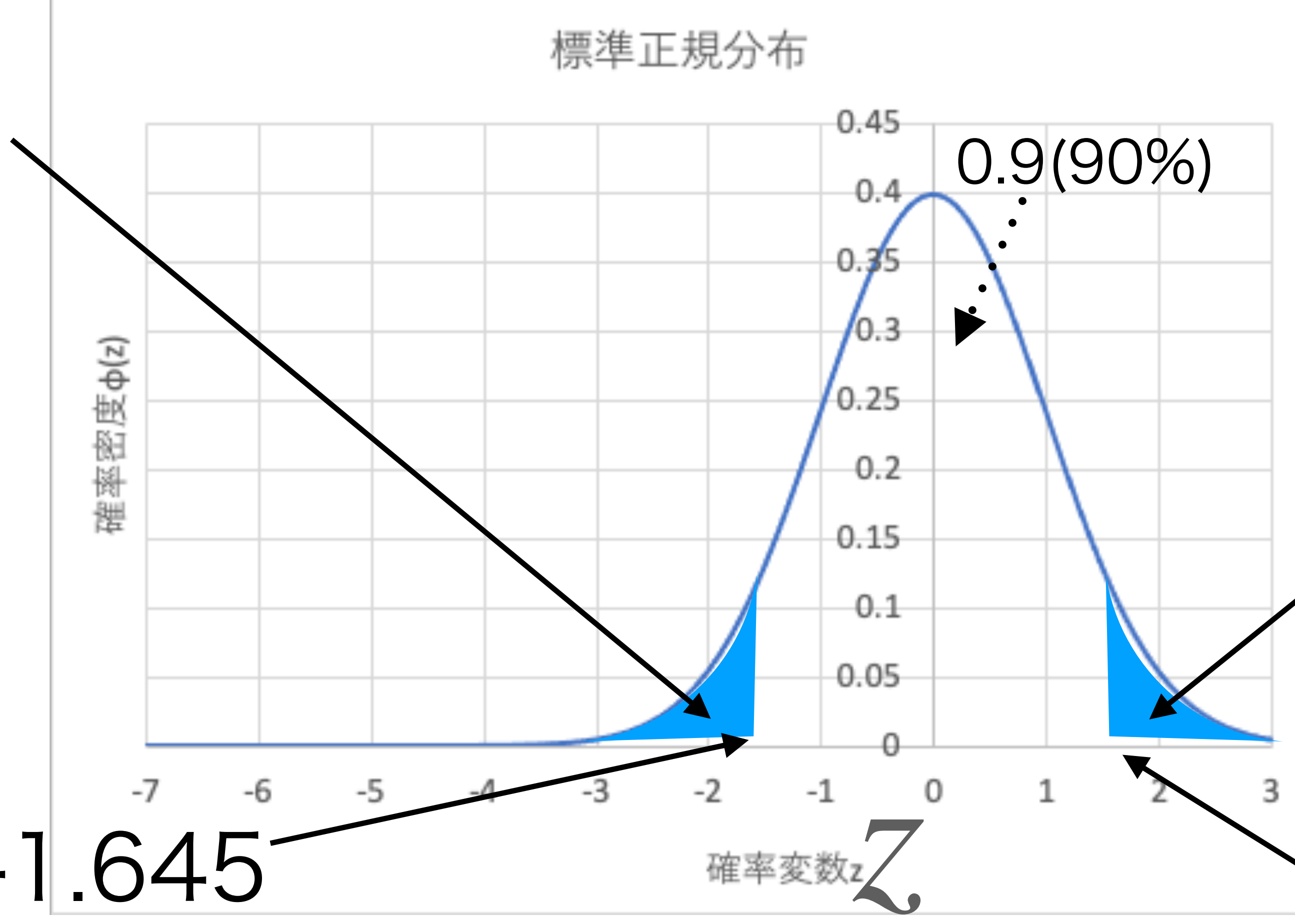
標本平均を標準化変換することで、分布の中のどこにあるかを検定する

μ :母平均
σ :母分散
n :標本数
$\bar{X}$:標本平均

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

今回はz=-1.382

0.05



0.05

z=-1.645

z=1.645

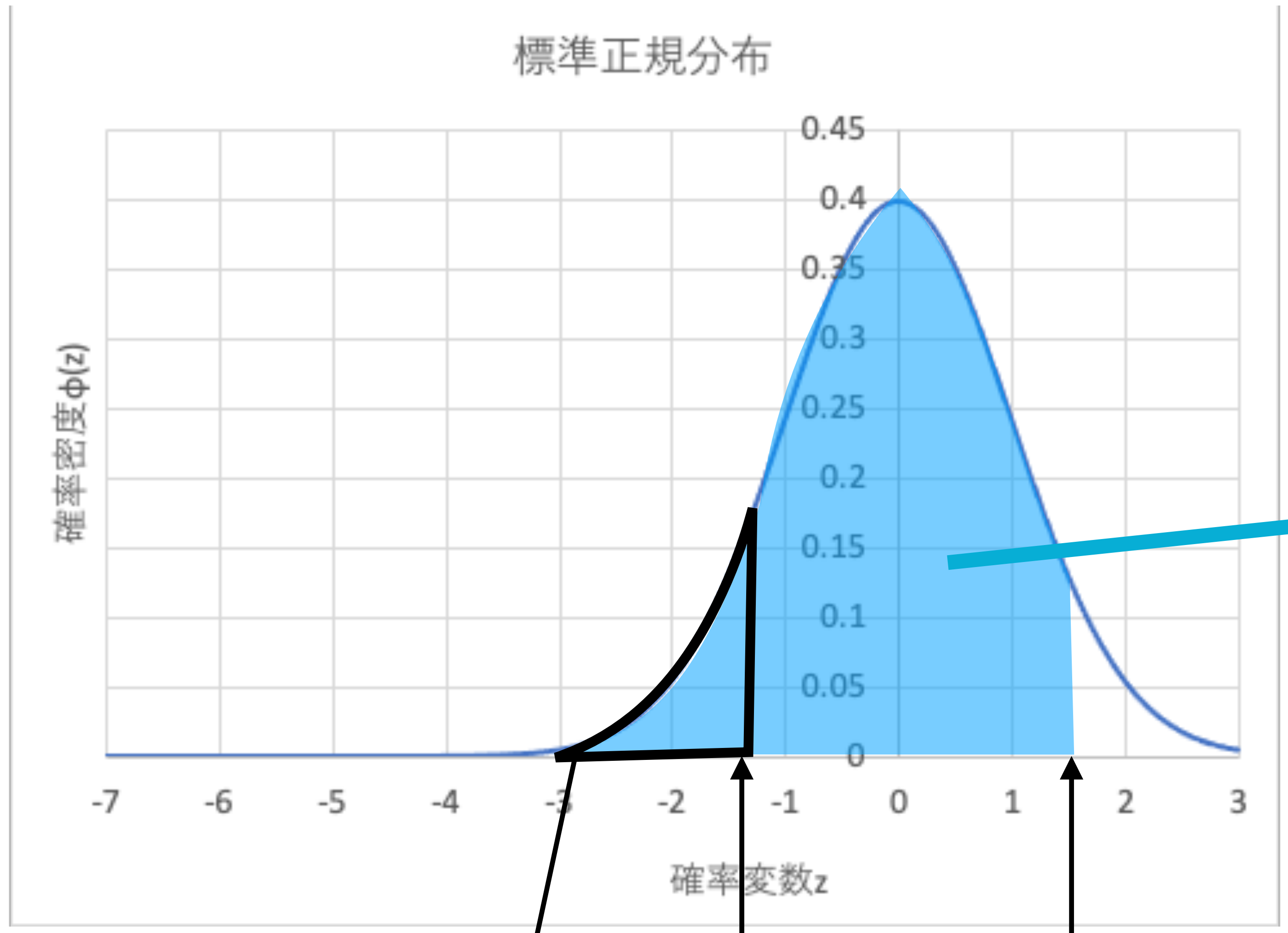
続き

$$\bar{X} = 34.77, \mu = 35.03, \sigma = 0.925, n = 25$$

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = -1.382$$

→NORM.S.DIST(z, TRUE)=p値

あるいは正規分布表の累積確率値を読み取っても良い



z=-1.38

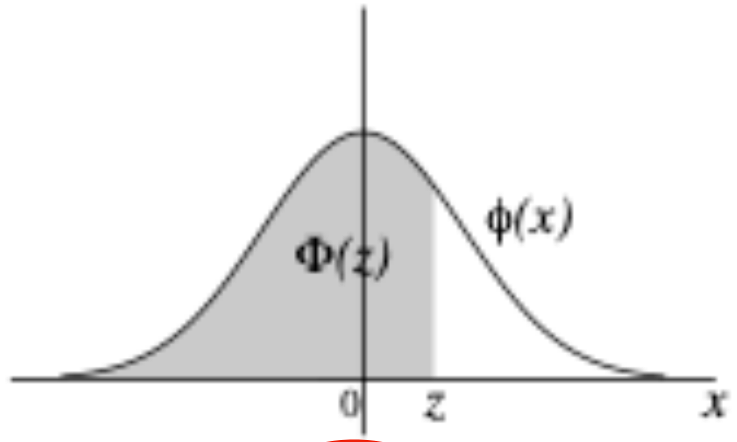
z=1.38

P値(面積)=1-0.91620=0.0838 > 0.05 →帰無仮説を棄却できない

B.2 正規分布表

正規分布表には積分範囲が異なるものがあるので注意すること．

$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$



	0	1	2	3	4	5	6	7	8	9
0.0	.50000	.50398	.50797	.51196	.51595	.51993	.52392	.52790	.53188	.53585
0.1	.53982	.54379	.54775	.55171	.55567	.55961	.56355	.56749	.57142	.57534
0.2	.57925	.58316	.58706	.59095	.59483	.59870	.60256	.60641	.61026	.61409
0.3	.61791	.62171	.62551	.62930	.63307	.63683	.64057	.64430	.64802	.65173
0.4	.65542	.65909	.66275	.66640	.67003	.67364	.67724	.68082	.68438	.68793
0.5	.69146	.69497	.69846	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72574	.72906	.73237	.73565	.73891	.74215	.74537	.74857	.75174	.75490
0.7	.75803	.76114	.76423	.76730	.77035	.77337	.77637	.77935	.78230	.78523
0.8	.78814	.79102	.79389	.79673	.79954	.80233	.80510	.80784	.81057	.81326
0.9	.81593	.81858	.82121	.82381	.82639	.82894	.83147	.83397	.83645	.83891
1.0	.84134	.84375	.84613	.84849	.85083	.85314	.85542	.85769	.85992	.86214
1.1	.86433	.86650	.86864	.87076	.87285	.87492	.87697	.87899	.88099	.88297
1.2	.88493	.88686	.88876	.89065	.89251	.89435	.89616	.89795	.89972	.90147
1.3	.90319	.90490	.90658	.90824	.90987	.91149	.91308	[illegible]	.91620	.91773
1.4	.91924	.92073	.92219	.92364	[illegible]	.92647	.92785	.92921	.93056	.93188
1.5	[illegible]	.93447	.93574	.93699	.93821	.93942	.94062	.94179	.94294	.94408
1.6	.94520	.94630	.94738	.94844	.94949	.95052	.95154	.95254	.95352	.95448
1.7	.95543	.95636	.95728	.95818	.95907	.95994	.96079	.96163	.96246	.96327
1.8	.96406	.96485	.96562	.96637	.96711	.96784	.96855	.96925	.96994	.97062
1.9	.97128	.97193	.97257	.97319	.97381	.97441	.97500	.97558	.97614	.97670
2.0	.97724	.97778	.97830	.97882	.97932	.97981	.98030	.98077	.98123	.98169
2.1	.98213	.98257	.98299	.98341	.98382	.98422	.98461	.98499	.98537	.98573
2.2	.98609	.98644	.98679	.98712	.98745	.98777	.98808	.98839	.98869	.98898
2.3	.98927	.98955	.98982	.99009	.99035	.99061	.99086	.99110	.99134	.99157
2.4	.99180	.99202	.99223	.99245	.99265	.99285	.99305	.99324	.99343	.99361
2.5	.99379	.99396	.99413	.99429	.99445	.99461	.99476	.99491	.99505	.99520
2.6	.99533	.99547	.99560	.99573	.99585	.99597	.99609	.99620	.99631	.99642
2.7	.99653	.99663	.99673	.99683	.99692	.99702	.99710	.99719	.99728	.99736
2.8	.99744	.99752	.99759	.99767	.99774	.99781	.99788	.99794	.99801	.99807
2.9	.99813	.99819	.99824	.99830	.99835	.99841	.99846	.99851	.99855	.99860

(以下次ページ)

補足

標本平均、標本分散、不偏分散

母分散：

$$\sigma^2 = \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu)^2$$

母平均：μ

n個の標本分散(m>>n)：

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

標本平均：

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n}X_i$$

不偏分散(標本分散が母分散に等しくなるよう補正)：

$$\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

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

母集団が平均μ、分散σ^2の
正規分布に従うとき

$$N[\mu, \sigma^2]$$

→

標本平均$\bar{X}$の従う分布

$$N[\mu, \frac{\sigma^2}{n}]$$

正規分布 → 標準正規分布

母集団は正規分布に従う: $N[\mu, \sigma^2]$

確率変数xが従う確率分布関数:

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

$f(x)$

$\boxed{\text{標準化変換}}$

①x軸方向に-μ平行移動
②σで割ってx軸の単位変換

$$\boxed{z = \frac{x-\mu}{\sigma}}$$

$N[0, 1^2]$

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

標本平均$\bar{X}$が従う分布

$$N\left[\mu, \frac{\sigma^2}{n}\right]$$

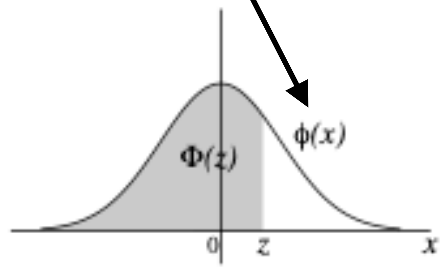
$\boxed{\text{標準化変換}}$

$$\boxed{z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}$$

B.2　正規分布表

正規分布表には積分範囲が異なるものがあるので注意すること.

$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$



	0	1	2	3	4	5	6	7	8	9
0.0	.50000	.50398	.50797	.51196	.51595	.51993	.52392	.52790	.53188	.53585
0.1	.53982	.54379	.54775	.55171	.55567	.55961	.56355	.56749	.57142	.57534
0.2	.57925	.58316	.58706	.59095	.59483	.59870	.60256	.60641	.61026	.61409
0.3	.61791	.62171	.62551	.62930	.63307	.63683	.64057	.64430	.64802	.65173
0.4	.65542	.65909	.66275	.66640	.67003	.67364	.67724	.68082	.68438	.68793
0.5	.69146	.69497	.69846	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72574	.72906	.73237	.73565	.73891	.74215	.74537	.74857	.75174	.75490
0.7	.75803	.76114	.76423	.76730	.77035	.77337	.77637	.77935	.78230	.78523
0.8	.78814	.79102	.79389	.79673	.79954	.80233	.80510	.80784	.81057	.81326
0.9	.81593	.81858	.82121	.82381	.82639	.82894	.83147	.83397	.83645	.83891
1.0	.84134	.84375	.84613	.84849	.85083	.85314	.85542	.85769	.85992	.86214
1.1	.86433	.86650	.86864	.87076	.87285	.87492	.87697	.87899	.88099	.88297
1.2	.88493	.88686	.88876	.89065	.89251	.89435	.89616	.89795	.89972	.90147
1.3	.90319	.90490	.90658	.90824	.90987	.91149	.91308	.91465	.91620	.91773
1.4	.91924	.92073	.92219	.92364	.92506	.92647	.92785	.92921	.93056	.93188
1.5	.93319	.93447	.93574	.93699	.93821	.93942	.94062	.94179	.94294	.94408
1.6	.94520	.94630	.94738	.94844	.94949	.95052	.95154	.95254	.95352	.95448
1.7	.95543	.95636	.95728	.95818	.95907	.95994	.96079	.96163	.96246	.96327
1.8	.96406	.96485	.96562	.96637	.96711	.96784	.96855	.96925	.96994	.97062
1.9	.97128	.97193	.97257	.97319	.97381	.97441	.97500	.97558	.97614	.97670
2.0	.97724	.97778	.97830	.97882	.97932	.97981	.98030	.98077	.98123	.98169
2.1	.98213	.98257	.98299	.98341	.98382	.98422	.98461	.98499	.98537	.98573
2.2	.98609	.98644	.98679	.98712	.98745	.98777	.98808	.98839	.98869	.98898
2.3	.98927	.98955	.98982	.99009	.99035	.99061	.99086	.99110	.99134	.99157
2.4	.99180	.99202	.99223	.99245	.99265	.99285	.99305	.99324	.99343	.99361
2.5	.99379	.99396	.99413	.99429	.99445	.99461	.99476	.99491	.99505	.99520
2.6	.99533	.99547	.99560	.99573	.99585	.99597	.99609	.99620	.99631	.99642
2.7	.99653	.99663	.99673	.99683	.99692	.99702	.99710	.99719	.99728	.99736
2.8	.99744	.99752	.99759	.99767	.99774	.99781	.99788	.99794	.99801	.99807
2.9	.99813	.99819	.99824	.99830	.99835	.99841	.99846	.99851	.99855	.99860

(以下次ページ)

母集団の平均、偏差、分散、標準偏差(1章)

平均： $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

平均は外れ値に弱い
(中央値：データの値を大きさ順に並べた時
中央に位置するデータの値→外れ値に強い)

偏差： $\delta x_i = x_i - \bar{x}$

データが平均値からどれだけずれているか
全てのデータの偏差の和はゼロになる

分散： $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n}\delta x_i^2$

$\sigma^2 = \overline{x^2} - \bar{x}^2$

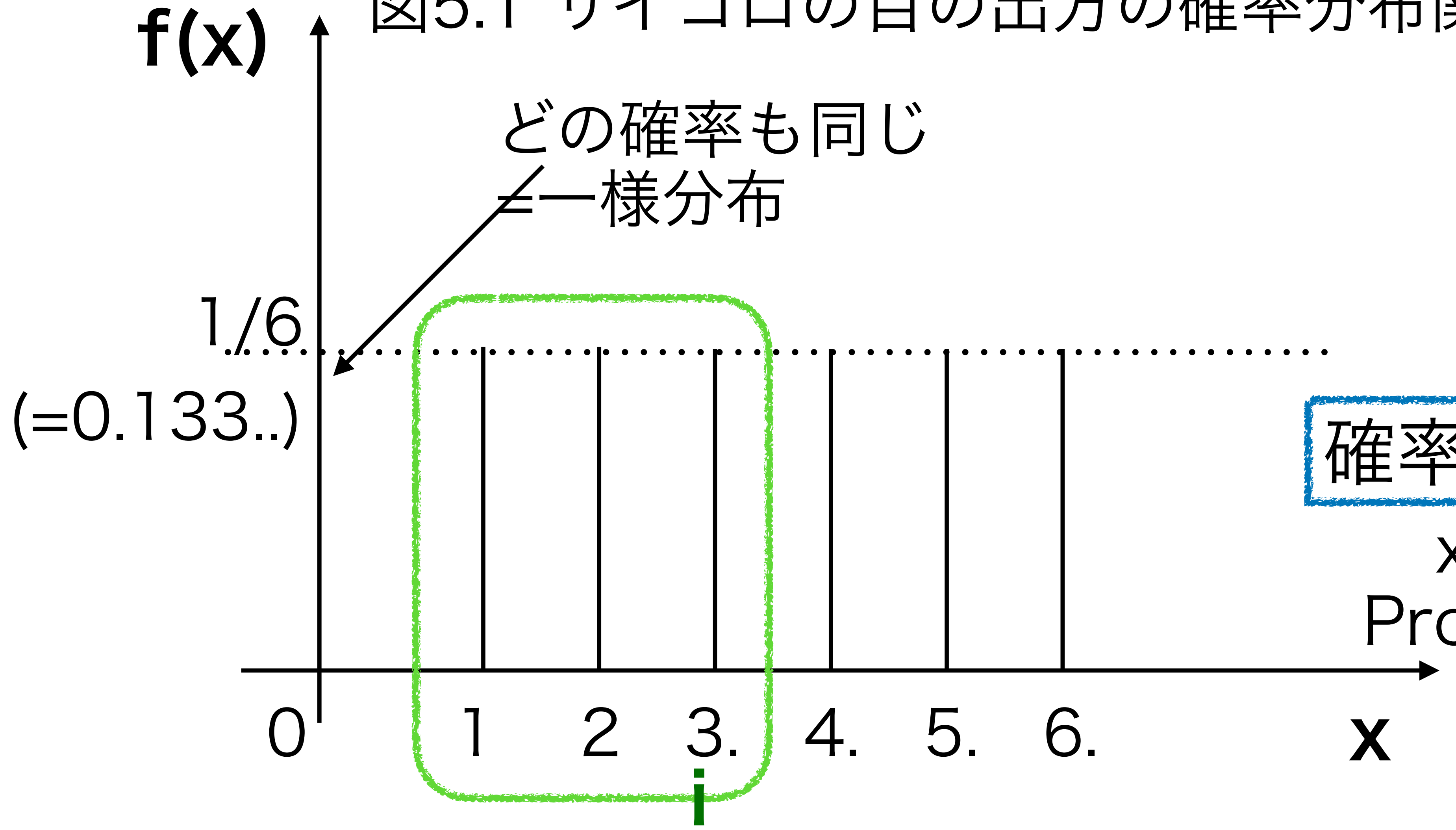
データのばらつきを示す量
分散は常にプラス（最小はゼロ）
データの２乗の単位

標準偏差： $\sigma = \sqrt{\sigma^2}$

データのばらつきを示す量
標準偏差も常にプラス（最小はゼロ）
データと同じの単位

離散的確率分布 → 連続的確率分布 (5章)

図5.1 サイコロの目の出方の確率分布関数



$$f(x)=\begin{cases}1/6 & (x=1,2,3,4,5,6)\\ 0 & (\text{それ以外})\end{cases}$$

確率変数xの多くは整数(離散的)

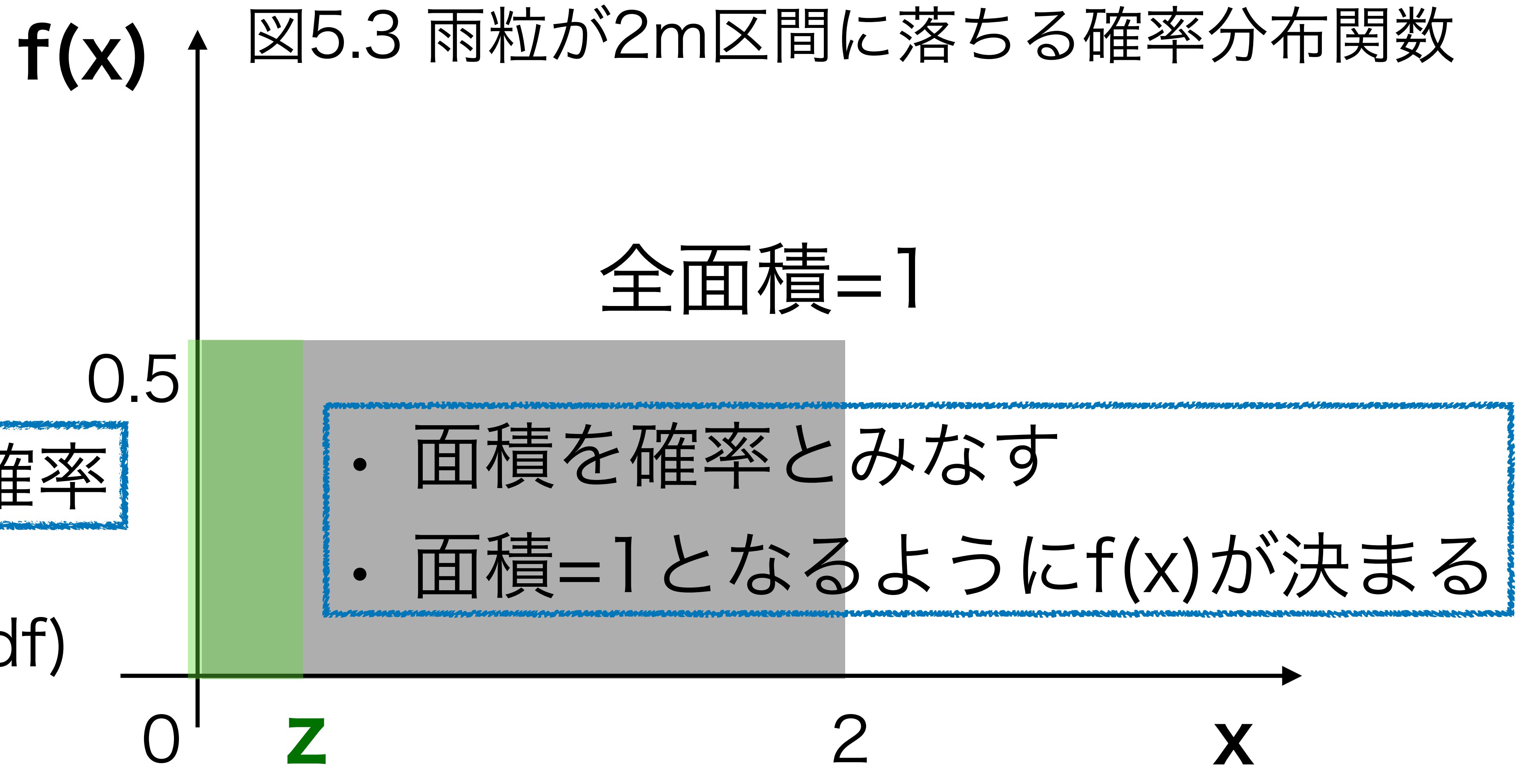
$$P(X=x_i)=f(x_i)$$

確率変数X: 取り得る値

確率関数・確率分布f(x): その確率

xが連続的な時は確率密度関数:
Probability density function(pdf)

図5.3 雨粒が2m区間に落ちる確率分布関数



$$f(x)=\begin{cases}0.5 & (0\le x\le 2)\\ 0 & (\text{それ以外})\end{cases}$$

確率変数xは実数(連続的)

全事象に対する確率の和が 1

$$\sum_{i=1}^{n} f(x_i)=1 \longrightarrow \int_{-\infty}^{\infty} f(x)dx=1$$

xが確率分布f(x)に従い、区間[a,b]に収まる確率

$$P(a\le X\le b)=\int_a^b f(x)dx$$

分布関数

$$F(x_i)=\sum_{k=1}^{i} f(x_k)$$

累積分布関数

$$\Phi(z)=\int_{-\infty}^{z} f(x)dx$$

確率関数の平均、分散 (5章)

離散的確率関数 → 連続的確率関数

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

全事象に対する確率の和が１

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

$$E[X] = \mu = \sum_{i}^{n} x_i f(x_i)$$

確率変数Xの期待値

$$E[X] = \mu = \int_{-\infty}^{\infty} xf(x)dx$$

$$V[X] = \sigma^2 = \sum_{i}^{n} (x_i - \mu)^2 f(x_i)$$

確率変数Xの分散

$$V[X] = \sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx$$

$$= E[X^2] - (E[X])^2$$

$$= E[X^2] - (E[X])^2$$

$$E[X^2] = \sum_{i}^{n} x_i^2 f(x_i)$$

$$\sum \longrightarrow \int dx$$

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f(x)dx$$

正規分布 → 標準正規分布

母集団が既知の場合

$N[\mu, \sigma^2]$

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

標準化変換

①x軸方向に-μ平行移動

②σで割ってx軸の単位変換

$$z = \frac{x-\mu}{\sigma}$$

$N[0, 1^2]$

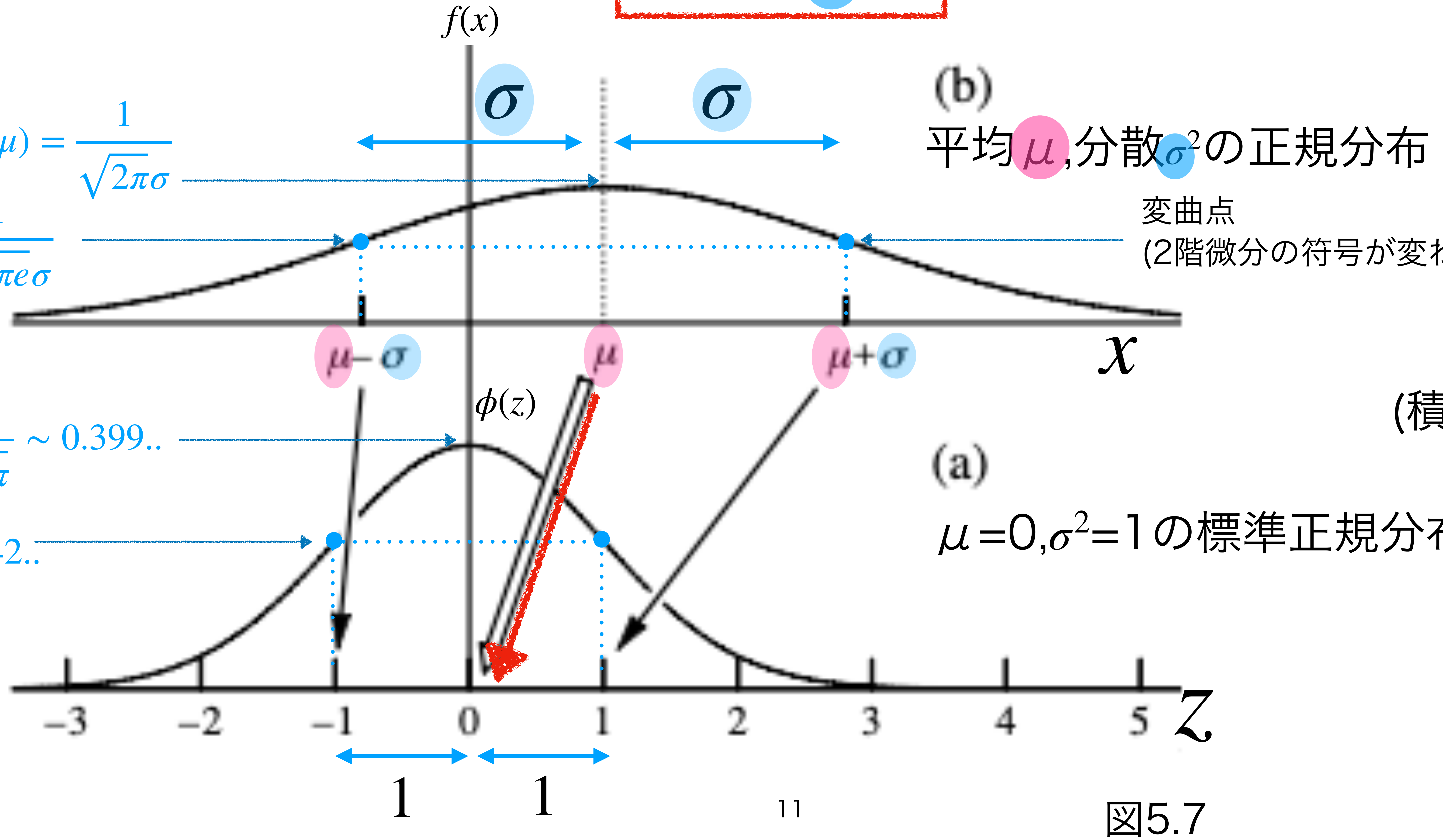
$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)$$

頂点 $f(\mu) = \dfrac{1}{\sqrt{2\pi}\sigma}$

変曲点 $f(\mu \pm \sigma) = \dfrac{1}{\sqrt{2\pi e}\sigma}$

頂点 $\phi(0) = \dfrac{1}{\sqrt{2\pi}} \sim 0.399..$

変曲点 $\phi(1) = \dfrac{1}{\sqrt{2\pi e}} \sim 0.242..$



変曲点
(2階微分の符号が変わるところ:数III)

μ ± σの範囲に
68%の確率で入る
(積分した面積が0.68)

図5.7

(補足)検定の種類の例

- 母平均の検定：z検定 (今回扱った例)
- 母分散が未知の場合の母平均の検定：t検定
- 母分散の推定：カイ２乗検定