

2016年度秋学期 統計学 第11回

分布の「型」を考える — 確率分布モデルと正規分布

浅野 晃
関西大学総合情報学部



ちょっと前回の復習

「統計的推測」とは

調べたい集団の、データ全体を調べられるか？

日本男性全員の身長を調べられるか？

データの**一部**を調べて

度数分布を推測する

いや、せめて平均や分散を推測する

統計的推測

無作為抽出

データの集団から、いくつかの数値を
公平なくじびきで選ぶ

【無作為標本抽出】 という

調べたい（が全部を調べるの
は無理な）集団 **【母集団】**

調べられる程度のデータ
【標本（サンプル）】

度数分布で考えると

母集団の度数分布		無作為抽出	標本の [確率分布]	
階級値	相対度数		階級値	選ばれる確率
162.5	15%	→	162.5	15%
167.5	20%		167.5	20%
172.5	20%		172.5	20%
177.5	10%		177.5	10%

2016年度秋学期 統計学

確率分布と確率変数

つまり

母集団の度数分布 (母集団分布)		=	標本の確率分布	
階級値	選ばれる確率		階級値	選ばれる確率
162.5	15%		162.5	15%
167.5	20%		167.5	20%
172.5	20%		172.5	20%
177.5	10%		177.5	10%

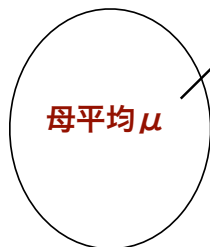
いくらかは決まってい
ないが、
確率分布が
決まっている

[確率変数]
という

2016年度秋学期 統計学

母平均の推定

母集団
(日本男性全体)



標本として数値を
いくつか取り出して、
それらの平均
[標本平均]

標本平均は母平均に
近い値になるか？

母平均が知りたい
が、日本男性全員は調べられない

2016年度秋学期 統計学

母平均の推定

母集団	サイズnの標本1セット	標本平均
母平均 μ 母分散 σ^2	X_1 X_2 ... X_n	\bar{X}
	X_1 X_2 ... X_n	\bar{X}
	X_1 X_2 ... X_n	\bar{X}
	\vdots	\vdots

母集団と同じ
期待値 μ
分散 σ^2

極端な値はあまりないので
分散が小さくなる
期待値 μ
分散 σ^2/n

2016年度秋学期 統計学

母平均の推定

母平均が μ のとき、
母分散が σ^2 のとき、
標本平均の期待値が μ
標本平均の分散が σ^2/n

仮に何度も標本を抽出して、
何度も標本平均を計算したとすると

分散が小さくなっているので、
たいてい、ほぼ母平均に近い値になる

いま1回だけ計算した標本平均も、
おそらく、ほぼ母平均に近い値だろう

母平均の推定

いま1回だけ計算した標本平均も、
おそらく、ほぼ母平均に近い値だろう

どのくらい近い？

どのくらいの確率で？
はずれる確率は？

このあたりを
今回から考える

分布の「型」を考える

母平均の推定

いま1回だけ計算した標本平均も、
おそらく、ほぼ母平均に近い値だろう

どのくらい近い？

どのくらいの確率で？
はずれる確率は？

計算するには、
式で表されてない
といけない

母集団分布は

つまり

母集団の度数分布
(母集団分布) = 標本の確率分布

階級値	選ばれる確率
162.5	15%
167.5	20%
172.5	20%
177.5	10%

これは式ではなく
数値の集まり、
計算できない

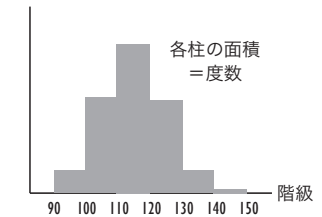
2016年度秋学期 統計学

式で表す

度数分布を

階級値	選ばれる確率
162.5	15%
167.5	20%
172.5	20%
177.5	10%

ヒストグラムが

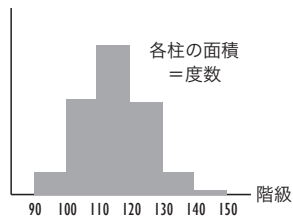


何かの式で書ける
ものと仮定する

何かの式で表される
関数のグラフであると
仮定する

2016年度秋学期 統計学

確率分布モデルとパラメータ

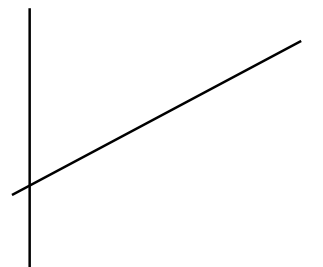


何かの式のグラフで
あると仮定する

式 = 確率分布モデル

パラメータを推定す
ればグラフが描ける

直線のモデル



$$y = ax + b$$

パラメータ

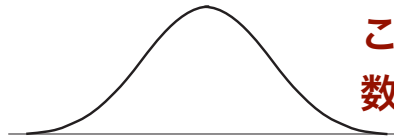
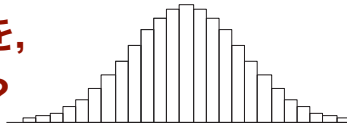
2016年度秋学期 統計学

連続型確率分布

2016年度秋学期 統計学

ヒストグラムを式で表す

こんなヒストグラムを、
式で書けるだろうか？



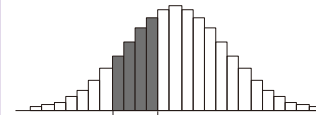
これを表す式のほうが
数学は簡単。

階級の区切り方がどん
どん細かくなって、見
えなくなったと考える

[連続型確率分布]

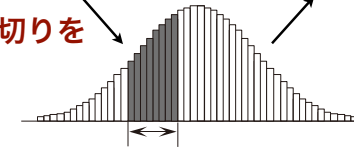
連続型確率分布

ヒストグラム



ある範囲に入る確率
= 柱の面積の合計

階級の区切りを
細かく



同じ範囲なら
柱の面積の合計は同じ

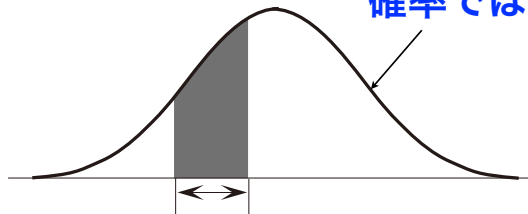
ヒストグラムの上の縁=
[確率密度関数]



同じ範囲なら
柱の面積の合計は同じ

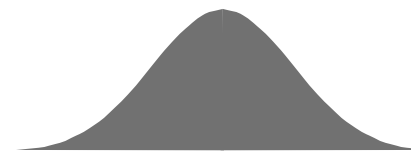
確率密度関数

ヒストグラムの上の縁=
[確率密度関数]
確率ではない

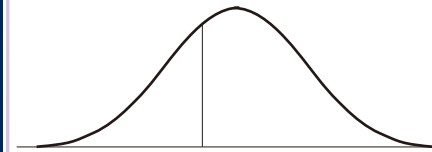


この範囲に入る確率 = この面積
= 確率密度関数の積分

連続型確率変数が
すべての実数のうちの
どれかになる確率
= 1 (100%)



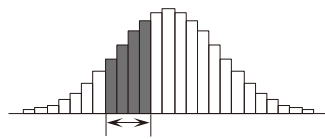
連続型確率変数が
ある特定の値 a になる確率
= 0



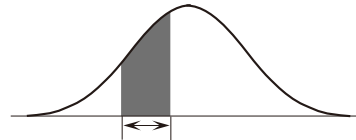
幅が0だから、面積も0

なんかヘン？
演習の解答例の付録で

連続型確率分布は、数学の都合



こんなのより



こんなのの方が
数式にしやすい

実際のデータは、有限の桁数の数字
で表されている限り、必ず離散的。

正規分布モデル

正規分布モデル

世の中には、[正規分布モデル]で表せる
ような母集団分布がたくさんある

長さの測定値の分布

センター試験の成績の分布 …

[中心極限定理]

母集団のばらつきの原因が

無数の独立な原因の和のとき、

母集団分布は概ね正規分布になる

正規分布の特徴

パラメータが平均（期待値）と分散
 μ σ^2

(わかりやすいものを推定すればよい
ので都合がいい)

確率変数 X の確率分布が

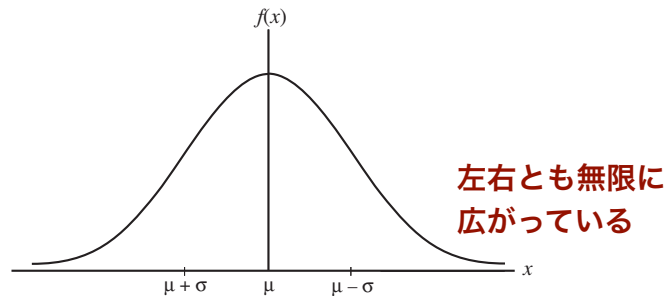
期待値 μ 、分散 σ^2 の正規分布であることを

確率変数 X が $N(\mu, \sigma^2)$ にしたがう という

正規分布の特徴

パラメータが平均（期待値）と分散
 μ σ^2

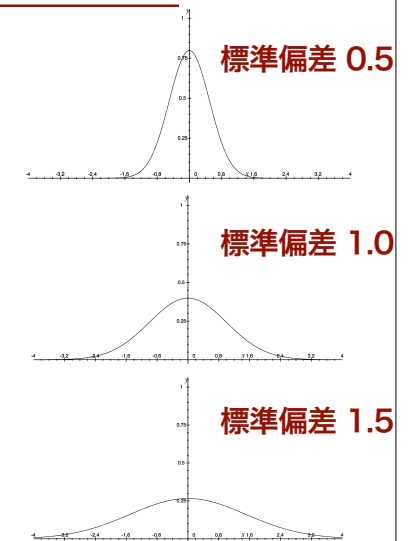
確率密度関数はこんな形



2016年度秋学期 統計学

正規分布の特徴

期待値0の正規分布の
確率密度関数

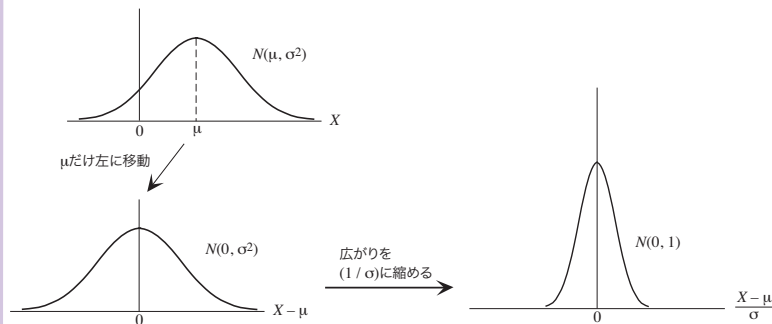


標準偏差が大きくなると
中央部の広がりが大きくなり
高さが低くなる

2016年度秋学期 統計学

正規分布の性質 1

確率変数 X が $N(\mu, \sigma^2)$ にしたがう とき



$(X - \mu) / \sigma$ は $N(0, 1)$ にしたがう

2016年度秋学期 統計学

正規分布の性質 1

確率変数 X が $N(\mu, \sigma^2)$ にしたがう とき
 $(X - \mu) / \sigma$ は $N(0, 1)$ にしたがう

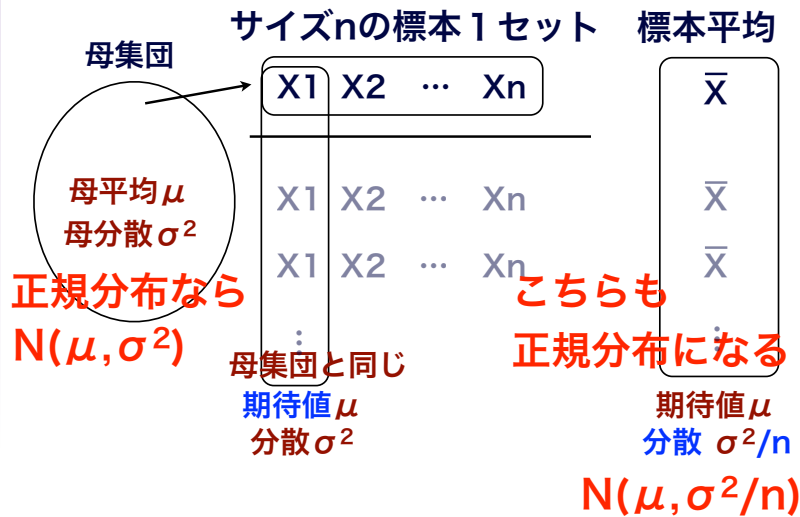
「標準得点」と同じ

変換しても、
やはり正規分布になる

$N(0, 1)$ を [標準正規分布] という

2016年度秋学期 統計学

正規分布の性質 2



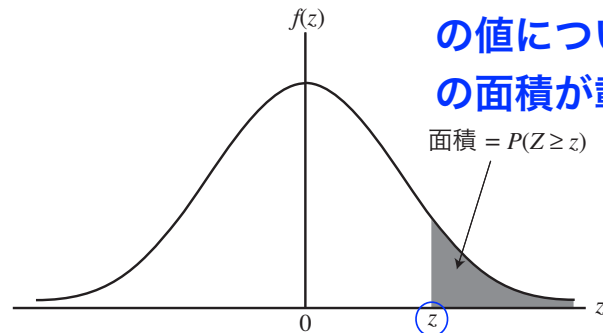
2016年度秋学期 統計学

正規分布表の見方

正規分布にもとづく計算

正規分布にしたがう確率変数が
ある範囲に入る確率
数表を使って求める

標準正規分布の場合に、いろいろなzの値について、この面積が載っている



2016年度秋学期 統計学

正規分布にもとづく計算

例) 確率変数Xが $N(50, 10^2)$ にしたがうとき、Xが60以上である確率を求めよ。

性質1により、 $Z = (X - 50) / 10$ と変換
Zは標準正規分布にしたがう

X=60のとき、 $Z = (60 - 50) / 10 = 1$

よって、求めるのは、Zが1以上である確率
 $P(Z \geq 1)$

2016年度秋学期 統計学

正規分布にもとづく計算

$P(Z \geq z)$ を求める

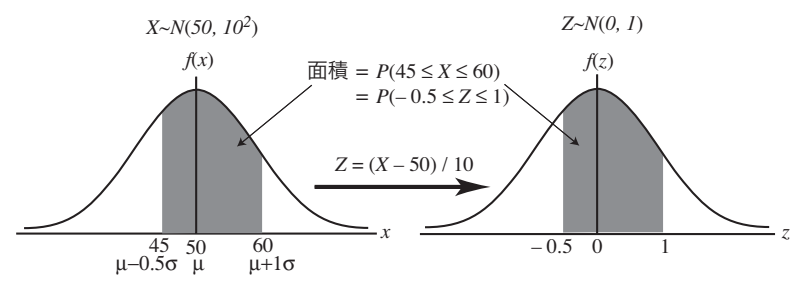
zの小数第2位まで
zの小数第1位まで

$P(Z \geq 1)$

	0.00	0.01	0.02	0.03	0.04	0.05	
0.0	0.50000	0.49601	0.49202	0.48803	0.48405	0.48006	
0.1	0.46017	0.45620	0.45224	0.44828	0.44433	0.44038	...
0.2	0.42074	0.41683	0.41294	0.40905	0.40517	0.40129	...
0.3	0.38209	0.37828	0.37448	0.37070	0.36693	0.36317	...
0.4	0.34458	0.34090	0.33724	0.33360	0.32997	0.32636	...
			⋮				
1.0	0.15866	0.15625	0.15386	0.15151	0.14917	0.14686	...

正規分布にもとづく計算

演習の2



正規分布にもとづく計算

演習の2

