3 計数値の解析

3.1 ２項分布

テキスト

芳賀敏郎（2016）医薬品開発のための統計解析
第３部　非線形モデル 改訂版、サイエンティスト社、p.288

第３部　非線形モデル

1. 非線形最小２乗法（基礎）

    1.1 線形と非線形、1.2 非線形最小２乗法の基本的な考え方、1.3 指数曲線のあてはめ、

    1.4 Emaxモデルとロジスティック曲線

2. 非線形最小２乗法（応用）

    2.1 誤差を考慮した解析、2.2 効力比、2.3 併用効果（相乗・拮抗効果）、

    2.4 モデルの探索（複数の曲線の同時あてはめ）、2.5 薬物動態の解析

3. 計数値の解析

    3.1 ２項分布、3.2 割合の推定・検定と区間推定、3.3 割合の差の推定・検定と区間推定、

    3.4 多項分布（名義尺度）、3.5 多項分布（順序尺度）、3.6 要因が複数の場合

4. ロジスティック回帰分析

    4.1 復習、4.2 ロジスティック回帰分析（基本）、4.3 ロジスティック回帰分析（応用）

第３章　計数値の解析

テキストの該当ページ

p.137

●データと変数

データの各列（因子、結果（特性））を変数として扱う

(1)

薬剤	反応
薬剤A	有効
薬剤A	有効
薬剤A	有効
薬剤A	有効
薬剤A	無効
薬剤B	有効
薬剤B	無効
薬剤B	無効
薬剤B	無効
薬剤B	有効

質的変数（名義尺度）　質的変数（名義尺度）

(2)

薬剤	反応	度数
薬剤A	有効	4
薬剤A	無効	1
薬剤B	有効	2
薬剤B	無効	3

質的変数（名義尺度）　質的変数（名義尺度）　量的変数 計数値（連続尺度）

(3)

薬剤	発疹程度
薬剤A	−
薬剤A	＋
薬剤A	−
薬剤A	＋＋
薬剤A	＋＋
薬剤B	＋
薬剤B	−
薬剤B	−
薬剤B	＋
薬剤B	−

質的変数（名義尺度）　質的変数（順序尺度）

(4)

薬剤	発疹数
薬剤A	43
薬剤A	39
薬剤A	28
薬剤A	39
薬剤A	26
薬剤B	53
薬剤B	42
薬剤B	61
薬剤B	30
薬剤B	46

質的変数（名義尺度）　量的変数 計数値（連続尺度）

(5)

投与量	血中濃度
0	100.0
1	90.5
2	81.9
3	74.1
5	60.7
7	49.7
10	36.8
15	22.3
20	13.5
25	8.2

量的変数 計量値（連続尺度）　量的変数 計量値（連続尺度）

第３章　計数値の解析

テキストの該当ページ p.137

●データと変数

データの各列（因子、結果（特性））を変数として扱う

(1)

薬剤	反応
薬剤A	有効
薬剤A	有効
薬剤A	有効
薬剤A	有効
薬剤A	無効
薬剤B	有効
薬剤B	無効
薬剤B	無効
薬剤B	無効
薬剤B	有効
質的変数 (名義尺度)	質的変数 (名義尺度)

(2)

薬剤	反応	度数
薬剤A	有効	4
薬剤A	無効	1
薬剤B	有効	2
薬剤B	無効	3
質的変数 (名義尺度)	質的変数 (名義尺度)	量的変数 計数値 (連続尺度)

(3)

薬剤	発疹程度
薬剤A	−
薬剤A	＋
薬剤A	−
薬剤A	＋＋
薬剤A	＋＋
薬剤B	＋
薬剤B	−
薬剤B	−
薬剤B	＋
薬剤B	−
質的変数 (名義尺度)	質的変数 (順序尺度)

(4)

薬剤	発疹数
薬剤A	43
薬剤A	39
薬剤A	28
薬剤A	39
薬剤A	26
薬剤B	53
薬剤B	42
薬剤B	61
薬剤B	30
薬剤B	46
質的変数 (名義尺度)	量的変数 計数値 (連続尺度)

(5)

投与量	血中濃度
0	100.0
1	90.5
2	81.9
3	74.1
5	60.7
7	49.7
10	36.8
15	22.3
20	13.5
25	8.2
量的変数 計量値 (連続尺度)	量的変数 計量値 (連続尺度)

第３章　計数値の解析

●量的変数と質的変数（カテゴリー変数、カテゴリカル変数）

JMPでの指定

質的変数：２つのカテゴリー　薬の効果が「有効・無効」（２値データ）　：名義尺度

３つ以上のカテゴリー（カテゴリーの順序に意味がある場合とない場合）

順序に意味がない場合（薬の副作用が「発熱・発疹・なし」）　：名義尺度

順序に意味がある場合（薬の効果が「無効・有効・著効」）　：順序尺度

場合によっては、数値化して連続尺度で解析する

（第１部§3.7、第２部§1.5）

量的変数：計量値（連続変数）　重さ、長さ、濃度などの連続的なデータ　：連続尺度

計数値（離散変数）　個数、人数、度数などの離散的なデータ　：連続尺度

連続変数とみなして解析する場合もある

第３章　計数値の解析

p.137

●量的変数と質的変数（カテゴリー変数、カテゴリカル変数）

JMPでの指定

質的変数：２つのカテゴリー　薬の効果が「有効・無効」（２値データ）　：名義尺度

３つ以上のカテゴリー（カテゴリーの順序に意味がある場合とない場合）

順序に意味がない場合（薬の副作用が「発熱・発疹・なし」）　：名義尺度

順序に意味がある場合（薬の効果が「無効・有効・著効」）　：順序尺度

場合によっては、数値化して連続尺度で解析する

（第１部§3.7、第２部§1.5）

量的変数：計量値（連続変数）　重さ、長さ、濃度などの連続的なデータ　：連続尺度

計数値（離散変数）　個数、人数、度数などの離散的なデータ　：連続尺度

連続変数とみなして解析する場合もある

p.137

●データと変数

データの各列（因子、観測値など）を変数として扱う

(1)

薬剤	反応
薬剤A	有効
薬剤A	有効
薬剤A	有効
薬剤A	有効
薬剤A	無効
薬剤B	有効
薬剤B	無効
薬剤B	無効
薬剤B	無効
薬剤B	有効
質的変数 (名義尺度)	質的変数 (名義尺度)

(2)

薬剤	反応	度数
薬剤A	有効	4
薬剤A	無効	1
薬剤B	有効	2
薬剤B	無効	3
質的変数 (名義尺度)	質的変数 (名義尺度)	量的変数 計数値 (連続尺度)

離散変数
計数値

(3)

薬剤	発疹程度
薬剤A	－
薬剤A	＋
薬剤A	－
薬剤A	＋＋
薬剤A	＋＋
薬剤B	＋
薬剤B	－
薬剤B	－
薬剤B	＋
薬剤B	－
質的変数 (名義尺度)	質的変数 (順序尺度)

(4)

薬剤	発疹数
薬剤A	43
薬剤A	39
薬剤A	28
薬剤A	39
薬剤A	26
薬剤B	53
薬剤B	42
薬剤B	61
薬剤B	30
薬剤B	46
質的変数 (名義尺度)	量的変数 計数値 (連続尺度)

(5)

投与量	血中濃度
0	100.0
1	90.5
2	81.9
3	74.1
5	60.7
7	49.7
10	36.8
15	22.3
20	13.5
25	8.2
量的変数 計量値 (連続尺度)	量的変数 計量値 (連続尺度)

第３章　計数値の解析

p.137

●本章の取り扱うデータ

n 匹の動物に薬剤を投与して、有効・無効で評価（＝質的変数）

有効であった匹数（f）は 0 ～ n の整数値を取る（＝量的変数、計数値）

f / n という割合（有効率）で表示できる（f ：Frequencyの頭文字が由来）

ここでは､｢率｣､｢割合｣を区別せずに使用

●本章の内容

§3.1　２項分布・・・・・・・・・・・・・・計数値の基本的な分布、尤度

§3.2　割合の推定・検定と区間推定・・・・１つの割合　（第１部§2 に対応）

§3.3　割合の差の推定・検定と区間推定・・２つの割合、オッズ　（第１部§3に対応）

§3.4　多項分布（名義尺度）・・・・・・・ $a \times b$ 分割表

§3.5　多項分布（順序尺度）　（第２部§1「質的因子の１因子実験」に対応）

§3.6　要因が複数の場合・・・・・・・・・・（第２部§5 「２因子実験」に対応）

第３章　計数値の解析

p.137

●本章の取り扱うデータ

n 匹の動物に薬剤を投与して、有効・無効で評価（＝質的変数）
有効であった匹数（f）は $0 \sim n$ の整数値を取る（＝量的変数、計数値）
f/n という割合（有効率）で表示できる（f ：Frequencyの頭文字が由来）

全体の内容を頭に留めて、
第１部、第２部との対応を確認

●本章の内容

§3.1　２項分布・・・・・・・・・・・・・・計数値の基本的な分布、尤度
§3.2　割合の推定・検定と区間推定・・・・１つの割合　（第１部§2 に対応）
§3.3　割合の差の推定・検定と区間推定・・２つの割合、オッズ　（第１部§3に対応）
§3.4　多項分布（名義尺度）・・・・・・・・$a \times b$ 分割表
§3.5　多項分布（順序尺度）　（第２部§1「質的因子の１因子実験」に対応）
§3.6　要因が複数の場合・・・・・・・・・・（第２部§5 「２因子実験」に対応）

3.1 ２項分布

p.137

（1）簡単な例
（2）シミュレーション
（3）２項分布
（4）Excel による確率の計算
（5）JMP による確率の計算
（6）n, π による２項分布の変化
（7）２項分布の期待値と分散
（8）２項分布の正規近似
（9）確率と尤度

使用するファイル　Excelファイル「改3計数値.xlsm」
JMP 10.02 を使用した結果を表示

★プレゼンテーションの
スピーカーノートを、
PDF の注釈に変換してあります

はじめに

p.138

●課題 3.1

ある病院でこれまでに 100 例の手術を実施し，失敗は 1 例もなく，成功率は100% であった。
このような事例が起こる確率 π （成功率）はいくらになるであろうか？
また，成功率の信頼区間はどのように推定されるで あろうか？
（カテゴリー：成功、失敗・・・二者択一とする）

この事実から、
この手術は完璧で、
真の成功率が 100% だとは
言えない

●本節の内容

ある事象の起こる確率が π であるとき，
n 回の試行で、その事象の起こった回数を f とする
本節では，n と f が得られたとき，それを使って π の仮説検定や区間推定をする方法を学ぶ

具体例：n 匹の動物にある薬を投与したところ、副作用の見られた匹数は f であった
その薬の副作用の起こる確率 π の仮説検定や区間推定を行う

はじめに

●課題 3.1

ある病院でこれまでに 100 例の手術を実施し，失敗は 1 例もなく，成功率は100% であった。
このような事例が起こる確率 π （成功率）はいくらになるであろうか？
また，成功率の信頼区間はどのように推定されるで あろうか？
（カテゴリー：成功、失敗）

π：真の値、推定する対象、観測値は $p = f / n$
正しいサイコロの目の出る確率は π=1/6
実際のサイコロの p から π を推定
（π は英字の p に対応するギリシャ文字）

●本節の内容

ある事象の起こる確率が π であるとき，
n 回の試行で、その事象の起こった回数を f とする
本節では，n と f が得られたとき，それを使って π の仮説検定や区間推定をする方法を学ぶ

具体例：n 匹の動物にある薬を投与したところ、副作用の見られた匹数は f であった
その薬の副作用の起こる確率 π の仮説検定や区間推定を行う

p.138

（1）簡単な例

手術の「成功」「失敗」
100例中、全て成功する場合の確率

簡単な例

p.138

●事例

ある成功率が高い手術を実施し、成功・失敗のカテゴリーで結果を評価・・・２値データ

（1）100人で全て手術は成功した。
この手術の成功率は100%といえるか？
（失敗する危険は全くないといえるか？）

（2）100人で全て手術は成功した。
失敗率は１%以下（成功率は 99%を超える）といえるか？

テキストでは、
「超える」を「以上」として
表現している部分がある

（3）仮に、本当の成功率を 99% とした場合
100人の患者に手術して、100人全てが成功する確率はいくらか？

簡単な例

p.138

●事例

ある成功率が高い手術を実施し、成功・失敗のカテゴリーで結果を評価・・・２値データ

（1）100人で全て手術は成功した。
この手術の成功率は100%といえるか？
（失敗する危険は全くないといえるか？）

（2）100人で全て手術は成功した。
失敗率は１%以下（成功率は 99%を超える）とい

（3）仮に、本当の成功率を 99% とした場合
100人の患者に手術して、100人全てが成功する確率はいくらか？

手術を無限に行えば
100人あたり
99人が成功し、
１人が失敗する確率

100人単位で手術を無限
に行った場合
100人全員が成功する確率

簡単な例

p.138

●事例

ある成功率が高い手術を実施し、成功・失敗のカテゴリーで結果を評価・・・２値データ

（1）100人で全て手術は成功した。
この手術の成功率は100%といえるか？
（失敗する危険は全くないといえるか？）

101例目で失敗するかもしれない
失敗する危険は全くないと
言えないことは自明のこと

（2）100人で全て手術は成功した。
失敗率は１%以下（成功率は 99%を超える）といえるか？

（3）仮に、本当の成功率を 99% とした場合
100人の患者に手術して、100人全てが成功する確率はいくらか？

（2）と（3）は関連した内容、実際に計算を試みる

簡単な例

p.138

●考え方

（3）仮に、本当の成功率を 99% とした場合

100人の患者に手術して、100人全てが成功する確率はいくらか？

本当の成功率が99%（$\pi = 0.99$）の場合

100人（$n = 100$）の患者に手術して、100人全てが成功（$f = 100$）する確率（p_{100}）は？

$n = 1$ で、全て成功（f=1）する確率 $p_1 = \pi^n = 0.99^1 = 0.99$

$n = 2$ で、全て成功（f=2）する確率 $p_2 = \pi^n = 0.99^2 = 0.980$

$n = 3$ で、全て成功（f=3）する確率 $p_3 = \pi^n = 0.99^3 = 0.970$

$n = 100$ で、全て成功（f=100）する確率 $p_{100} = \pi^n = 0.99^{100} = 0.366$ （3.1.1）

100 人で全て成功する確率は 0.366・・・しばしば起こる確率（約 1/3）、珍しくない

簡単な例

p.138

●考え方

（3）仮に、本当の成功率を 99% とした場合

100人の患者に手術して、100人全てが成功する確率はいくらか？

本当の成功率が99%（π = 0.99）の場合

100人（n = 100）の患者に手術して、100人全てが成功（f = 100）する確率（p_{100}）は？

n = 1 で、全て成功（f=1）する確率 $p_1 = \pi^n = 0.99^1 = 0.99$

n = 2 で、全て成功（f=2）する確率 $p_2 = \pi^n = 0.99^2 = 0.980$

n = 3 で、全て成功（f=3）する確率 $p_3 = \pi^n = 0.99^3 = 0.970$

n = 100 で、全て成功（f=100）する確率 $p_{100} = \pi^n = 0.99^{100} = 0.366$ （3.1.1）

100 人で全て成功する確率は 0.366・・・しばしば起こる確率（約 1/3）、珍しくない

（2）100人で全て手術は成功した。失敗率は 1 %以下（成功率は 99%を超える）といえるか？

$\pi > 0.99$ とはいえない（H_0：π=0.99、H_1：π > 0.99 、帰無仮説 H_0 を棄却できない）

簡単な例

●Excelファイルの読み込みと表示

Excel ファイル「改３計数値.xlsm」、名前ボックスから「表示3.1.1」（Fig31_01）を選択

●確率の計算

$f = n$ の確率：n 人が全員成功する確率（f は成功する人数）

表示3.1.1　$f = n$ の確率の計算表

	A	B	C	D
2	失敗率	成功率		
3	$1-\pi$	π	n	f=n の確率
4	0.01	0.99	2	0.9801
5	0.01	0.99	100	0.3660
6	0.03	0.97	100	0.0476
7				D4: =B4^C4

$p_n = \pi^n$

真の成功率が99%（$\pi = 0.99$）の場合
$n = 2$ で、全員が成功（f=2）する確率
$p_2 = \pi^n = 0.99^2 = 0.980$

真の成功率が99%（$\pi = 0.99$）の場合
n =100 で、全員が成功（f=100）する確率
$p_{100} = \pi^n = 0.99^{100} = 0.366$

簡単な例

p.139

●仮説検定

（2）100人で全て手術は成功した。成功率は99%を超えるといえるか（片側検定）。

帰無仮説 H_0：π=0.99、対立仮説 H_1：$\pi > 0.99$

p = 0.366・・・帰無仮説が正しいという前提の下、（第１部§1.4）

得られた値（f）とそれ以上に出現確率の小さい値が得られる確率

（この事例では、得られた値 f 以上に大きな値がないので、f=n の確率が p 値になる）

表示3.1.1　$f = n$ の確率の計算表

	A	B	C	D
2	失敗率	成功率		
3	$1-\pi$	π	n	f=n の確率
4	0.01	0.99	2	0.9801
5	0.01	0.99	100	0.3660
6	0.03	0.97	100	0.0476
7				D4: =B4^C4

$p_n = \pi^n$

有意水準 0.05 で、帰無仮説を棄却できない

（成功率99% を超えるとはいえない）

真の成功率が99%（π = 0.99）の場合

n =100 で、全員が成功（f=100）する確率

$p_{100} = \pi^n = 0.99^{100} = 0.366$

簡単な例

p.139

●逆推定（演習3.1.1）

成功率が0.99 を超える（$\pi > 0.99$）というためには，何例の手術で全例成功になればよいか？

試行錯誤による方法：セルC4 に適当な数値を入力し、セル D4 が 0.05 になるように探索

表示3.1.1　$f = n$ の確率の計算表

	A	B	C	D
2	失敗率	成功率		
3	$1-\pi$	π	n	$f=n$ の確率
4	0.01	0.99	2	0.9801
5	0.01	0.99	100	0.3660
6	0.03	0.97	100	0.0476
7			D4: =B4^C4	

セルC4 に適当な数値を入力
セルD4 を 0.05 以下にする最小値を求める

簡単な例

p.139

● 逆推定（演習3.1.1）

成功率が0.99 を超える（$\pi > 0.99$）というためには，何例の手術で全例成功になればよいか？

試行錯誤による方法：セルC4 に適当な数値を入力し、セル D4 が 0.05 になるように探索

299 で 0.05 以下になる → 299 例で全て成功すれば0.99よりも成功率は高いといえる

帰無仮説 H_0：π=0.99 を棄却し、対立仮説 H_1：$\pi > 0.99$ を採択

表示3.1.1　$f = n$ の確率の計算表

	A	B	C	D
2	失敗率	成功率		
3	1－π	π	n	f=n の確率
4	0.01	0.99	2	0.9801
5	0.01	0.99	100	0.3660
6	0.03	0.97	100	0.0476
7			D4: =B4^C4	

失敗率	成功率		
1－π	π	n	f=n の確率
0.01	0.99	100	0.36603
0.01	0.99	200	0.13398
0.01	0.99	250	0.08106
0.01	0.99	270	0.06630
0.01	0.99	290	0.05423
0.01	0.99	295	0.05157
0.01	0.99	298	0.05004
0.01	0.99	299	0.04954

簡単な例

p.226

●逆推定（演習3.1.1）

成功率は0.99 を超える（$\pi > 0.99$）というためには，何例の手術で全例成功になればよいか？

ゴールシークによる方法：π に0.99 を入力し，$f = n$ の確率が0.05 になる n を求める

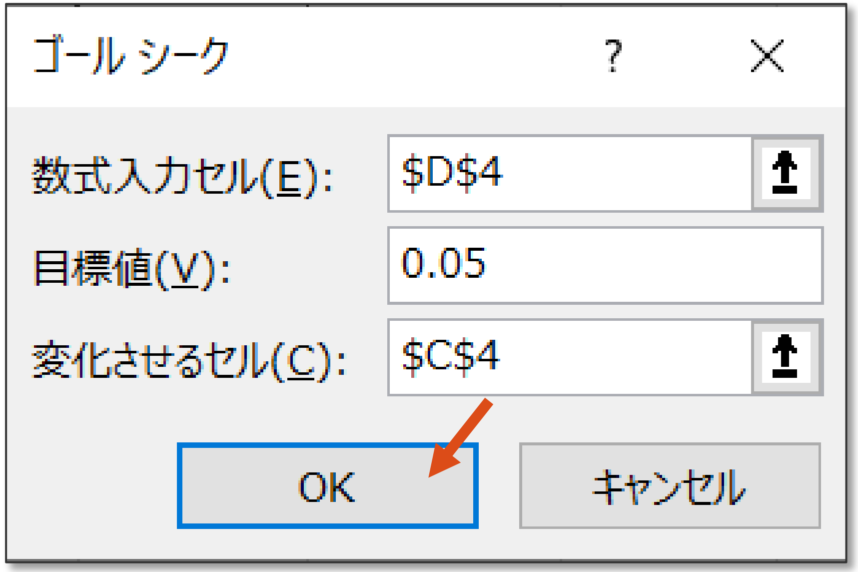
［データ］＞［What -If 分析］＞［ゴールシーク］、

［数式入力セル］にセル D4、［目標値］に0.05、［変化させるセル］にセルC4 を設定

セルC4に 297.95、セルD4 に 0.050061 が入力される

表示3.1.1　$f = n$ の確率の計算表

	A	B	C	D
2	失敗率	成功率		
3	$1-\pi$	π	n	$f=n$ の確率
4	0.01	0.99	297.952	0.050061
5	0.01	0.99	100	0.3660
6	0.03	0.97	100	0.0476
7			D4: =B4^C4	



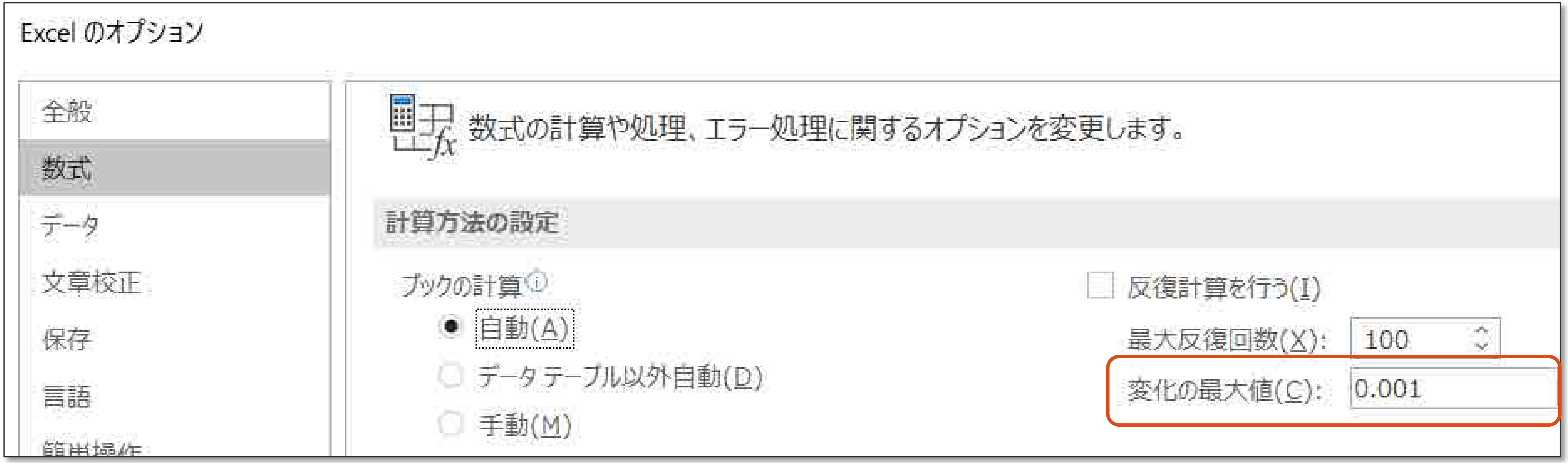
簡単な例

p.226

●ゴールシークの精度の設定

目的の精度が得られない場合、計算精度を高める

［ファイル］＞［オプション］＞［数式］＞［変化の最大値］　　0.001 → 0.00001



簡単な例

p.226

●逆推定

成功率は0.99を超える（$\pi > 0.99$）というためには，何例の手術で全例成功になればよいか？

ゴールシークによる方法：π に0.99を入力し，$f = n$ の確率が0.05になる n を求める
［データ］＞［What -If 分析］＞［ゴールシーク］

表示3.1.1　$f = n$ の確率の計算表

	A	B	C	D
2	失敗率	成功率		
3	$1-\pi$	π	n	f=nの確率
4	0.01	0.99	297.952	0.050061
5	0.01	0.99	100	0.3660
6	0.03	0.97	100	0.0476
7			D4: =B4^C4	

	A	B	C	D
2	失敗率	成功率		
3	$1-\pi$	π	n	f=nの確率
4	0.01	0.99	298.074	0.049999
5	0.01	0.99	100	0.3660
6	0.03	0.97	100	0.0476
7			D4: =B4^C4	

簡単な例

p.139

●区間推定

100 人を手術して、全員が成功した。成功率 π は何% を超えるといえるか。

試行錯誤：n に100 を入力し，$f = n$ の確率が0.05 になる π を求めると 0.970 が得られる
ゴールシークを利用：確率 0.05 になる解 π=0.970 が得られる
区間推定：$\pi > 0.97$ （片側信頼区間、信頼率 $1-\alpha = 0.95$）
仮説検定：帰無仮説 H_0：π=0.97 を棄却し、対立仮説 H_1：$\pi > 0.97$ を採択（片側検定、$\alpha = 0.05$）

表示3.1.1　$f = n$ の確率の計算表

	A	B	C	D
2	失敗率	成功率		
3	1－π	π	n	f=n の確率
4	0.01	0.99	298.074	0.049999
5	0.01	0.99	100	0.3660
6	0.02951	0.97049	100	0.0500

ゴールシークの解

D4: =B4^C4

試行錯誤で入力

失敗率	成功率		
1－π	π	n	f=n の確率
0.031	0.969	100	0.0429
0.030	0.970	100	0.0476
0.029	0.971	100	0.0527
0.025	0.975	100	0.0795
0.020	0.980	100	0.1326
0.010	0.990	100	0.3660

簡単な例

p.139

- fが n または 0 ではない場合

  p 値：帰無仮説が正しいという前提の下、
  得られた値（f）とそれ以上に出現確率の小さい値が得られる確率（第 1 部§1.4）

  $f = n$ 、$f = 0$ の場合
  100 人（n）に手術して、100人（f）で成功した。成功率は 99% を超えるといえるか
  $f = n$ の場合、f が n よりも大きい値は取り得ない
  $n = 100$ で、$f = 100$ の確率が p 値になる
  （$f = 0$ の場合も、逆に考えれば同様に計算できる）

  $0< f < n$ の場合
  100 人（n）に手術して、99人（f）で成功した。成功率は 90% を超えるといえるか
  p 値は、n =100で、$f = 99$ と $f = 100$ になる確率の和
  $f = 99$ になる確率を求めるには、fの確率分布（2 項分布）の知識が必要（後述）

Rule of 3

●Rule of 3 の考え方

表示3.1.1　$f = n$ の確率の計算表

	A	B	C	D
2	失敗率	成功率		
3	1-π	π	n	f=n の確率
4	0.01	0.99	10	0.90438
6	0.01	0.99	100	0.36603
7	0.001	0.999	1000	0.36770
8	0.0001	0.9999	10000	0.36786
9	0.00001	0.99999	100000	0.36788

$\frac{1}{n}$

$1-\frac{1}{n}$

設定

→ 表示 3.1.2 Rule of 3　左の部分

$$(\mathbf{1}-\boldsymbol{\pi}')\boldsymbol{n}=\mathbf{1}$$

n が大きくなると $f = n$ の確率は 0.36788 に接近

$0.99999^{100000} = 0.36788$

$\frac{1}{e} = \frac{1}{2.71828} = 0.36788$

e = 2.71828・・・自然対数の底

Rule of 3

●Rule of 3 の考え方

表示 3.1.2 Rule of 3

$1-\pi$	π	n	$f=n$ の確率
0.1	0.9	10	0.34868
0.01	0.99	100	0.36603
0.001	0.999	1000	0.36770
0.0001	0.9999	10000	0.36786
0.00001	0.99999	100000	0.36788
		1/e	0.36788

$\frac{1}{n}$　$1-\frac{1}{n}$　設定

$$(\mathbf{1}-\boldsymbol{\pi}')\boldsymbol{n}=\mathbf{1}$$

$(1-\pi)\times 3$

$1-\pi'$	π'	n	$f=n$ の確率
0.3	0.7	10	0.02825
0.03	0.97	100	0.04755
0.003	0.997	1000	0.04956
0.0003	0.9997	10000	0.04976
0.00003	0.99997	100000	0.04978
		1/(e^3)	0.04979

$$(\mathbf{1}-\boldsymbol{\pi}')\boldsymbol{n}=\mathbf{3}$$

0.05 に接近

$n\times 3$

$1-\pi$	π	n'	$f=n'$ の確率
0.1	0.9	30	0.04239
0.01	0.99	300	0.04904
0.001	0.999	3000	0.04971
0.0001	0.9999	30000	0.04978
0.00001	0.99999	300000	0.04979
		1/(e^3)	0.04979

$$(\mathbf{1}-\boldsymbol{\pi})\boldsymbol{n}'=\mathbf{3}$$

0.05 に接近

Rule of 3

p.140

● Rule of 3 の考え方

表示 3.1.2 Rule of 3（改変）

$1\text{-}\pi$	π	n	$f=n'$の確率
0.02190	0.9781	137	0.04815
0.01000	0.9900	300	0.04904
0.00335	0.9966	895	0.04954
0.00200	0.9980	1500	0.04964
0.00053	0.9995	5675	0.04975

$\frac{1}{n}$

設定
>100

ほぼ 0.05
< 0.05

$$(1-\pi)n = 3$$

仮説検定（α=0.05 の片側検定）と
片側区間推定（信頼率：1-α=0.95）を
近似的に行うことができる

一般化

$(1-\pi)\times 3$

この関係にあるとき、
n が大きければ、
f=n の確率は 0.05

$1\text{-}\pi'$	π'	n	$f=n$の確率
0.3	0.7	10	0.02825
0.03	0.97	100	0.04755
0.003	0.997	1000	0.04956
0.0003	0.9997	10000	0.04976
0.00003	0.99997	100000	0.04978
		1/(e^3)	0.04979

$$(1-\pi')n = 3$$

0.05 に接近

$n \times 3$

$1\text{-}\pi$	π	n'	$f=n'$の確率
0.1	0.9	30	0.04239
0.01	0.99	300	0.04904
0.001	0.999	3000	0.04971
0.0001	0.9999	30000	0.04978
0.00001	0.99999	300000	0.04979
		1/(e^3)	0.04979

$$(1-\pi)n' = 3$$

0.05 に接近

Rule of 3

p.140

●Rule of 3 の応用

表示 3.1.2 Rule of 3（改変）

$1\text{-}\pi$	π	n	$f=n'$の確率
0.02190	0.9781	137	0.04815
0.01000	0.9900	300	0.04904
0.00335	0.9966	895	0.04954
0.00200	0.9980	1500	0.04964
0.00053	0.9995	5675	0.04975

$\frac{1}{n}$

設定
>100

ほぼ 0.05
< 0.05

$$(\mathbf{1}-\boldsymbol{\pi})\boldsymbol{n} = \mathbf{3}$$

仮説検定（α=0.05 の片側検定）と
片側区間推定（信頼率：1-α=0.95）を
近似的に行うことができる

300 例で全例成功したとき（$f = n = 300$）、
失敗率は 0.01 未満である

$$(1-\pi)\times 300 = 3$$
$$(1-\pi) = 0.01$$

片側検定（$\alpha = 0.05$）
H_0：$1\text{-}\pi=0.01$ を棄却し、H_1：$1\text{-}\pi < 0.01$ を採択
$p = 0.04904$

片側信頼区間（信頼率：$1\text{-}\alpha = 0.95$）
$1\text{-}\pi < 0.01$

失敗率が 0.2% 未満であることを示すためには、
1500 例で全例成功しなければならない

$$0.002\times n = 3$$
$$n = 1500$$

Rule of 3

p.140

●Rule of 3 の応用

表示 3.1.2 Rule of 3（改変）

$1\text{-}\pi$	π	n	$f=n'$の確率
0.02190	0.9781	137	0.04815
0.01000	0.9900	300	0.04904
0.00335	0.9966	895	0.04954
0.00200	0.9980	1500	0.04964
0.00053	0.9995	5675	0.04975
0.00030	0.9997	10000	0.04976
0.00003	0.99997	100000	0.04978

$\frac{1}{n}$

設定
>100

ほぼ 0.05
< 0.05

$$(\mathbf{1}-\boldsymbol{\pi})\boldsymbol{n} = \mathbf{3}$$

300 例で全例成功したとき（$f = n = 300$）、
失敗率は 0.01 未満である

$$(1-\pi) \times 300 = 3$$
$$(1-\pi) = 0.01$$

片側検定（$\alpha = 0.05$）

H_0：$1\text{-}\pi=0.01$ を棄却し、H_1：$1\text{-}\pi < 0.01$ を採択

$$p = 0.04904$$

片側信頼区間（信頼率：$1\text{-}\alpha = 0.95$）

$1\text{-}\pi < 0.01$

失敗率→失敗する人数
300×0.01 = 3

10000 例で全例成功したとき（$f = n = 10000$）、
失敗率は 0.0003 未満である

$$(1-\pi) \times 10000 = 3$$
$$(1-\pi) = 0.0003$$

失敗率→失敗する人数
10000×0.0003 = 3

Rule of 3

●Rule of 3 の応用（補足）

n 人調べて事象が観測されなくても，
他の n 人中の 3 人に事象が観測される可能性がある
ここで重要な点は，n が 100 でも 1,000,000 でもこの法則が（近似的に）成り立つことである

（薬の副作用に関して）
すなわち，100 人調べて 1 度も観測されなくても
別の 100 人中 3 人に事象が生起する可能性があり，
100 万人調べて 0 でも、なお他の 100 万人中 3 人に生起する可能性は否定できない
この 法則は，実際家に対する戒めとして簡潔にして要領を得ており，
覚えやすいという点からもきわ めて優れたものである

（岩崎・吉田, 2005 ）

（2）シミュレーション

コインを投げて「表」「裏」が出る事象（$\pi = 0.5$ の事象）
10回中、「表」が５回出る確率、４回出る確率
（すべて「表」が出る確率ではない）

シミュレーション

p.140

●コイン実験

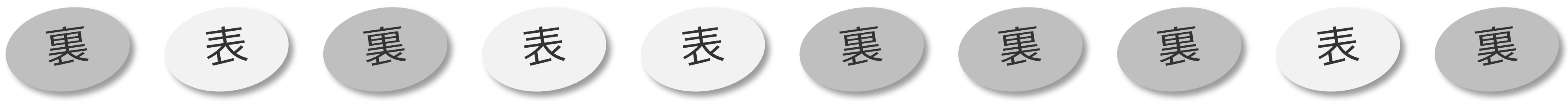
変形していない新しいコインを投げて、「裏」か「表」を得る
結果のカテゴリーは「表」「裏」の２種類（２値データ）
カテゴリーの生起確率はそれぞれ 1/2（0.5）
10回投げて、「表」なら 1、「裏」なら 0 として、「表の個数」を記入
これを100回繰り返す

表示 3.1.3 コイン実験（一部）

実験番号	10回投げた結果										表の個数
	1	2	3	4	5	6	7	8	9	10	
1	0	1	0	1	1	0	0	0	1	0	4

実験番号 1

0 1 0 1 1 0 0 0 1 0



シミュレーション

p.141

●コイン実験

10列×100行

100 個

表示3.1.3 コイン実験（一部）

実験番号	10回投げた結果 1	2	3	4	5	6	7	8	9	10	表の個数
1	0	1	0	1	1	0	0	0	1	0	4
2	0	0	0	0	0	0	1	0	1	0	2
3	0	0	1	0	1	1	1	0	0	0	4
4	0	1	1	0	1	1	0	1	0	0	5
5	1	0	1	0	1	0	1	1	0	1	6
6	0	1	1	1	1	1	0	1	0	0	6
7	1	1	1	0	1	0	0	1	0	1	6
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
99	0	0	1	0	1	1	1	0	0	0	4
100	0	1	1	0	1	1	0	1	0	0	5

1000 回に 1 回起こる確率

表示3.1.4 「表の個数」のまとめ

表の個数	度数	ヒストグラム	理論値(%)
0	0	|	0.10
1	1	|*	0.98
2	4	|****	4.39
3	14	|**************	11.72
4	13	|*************	20.51
5	26	|**************************	24.61
6	22	|**********************	20.51
7	11	|***********	11.72
8	8	|********	4.39
9	1	|*	0.98
10	0	|	0.10
合計	100		100.00

度数 0

最多

86 個

度数分布

シミュレーション

p.142

●コイン実験の疑似体験

表示3.15
Excel によるシミュレーション

一様乱数の
出力先：B5
（左上隅）

10 列×100 行
一様乱数
0 ～ 1

10列100行の一様乱数が、
コインの10回投げる実験を
100回繰り返すことに対応

	A	B	C	D	E	F	G	H	I	J	K	L
3											π =	0.5
4	実験番号	1	2	3	4	5	6	7	8	9	10	表の個数
5	1	0.44	0.18	0.63	0.99	0.12	0.55	0.90	0.14	0.50	0.81	5
	2	0.38	0.71	0.54	0.17	0.46	0.82	0.42	0.06	0.70	0.94	5
7	3	0.81	0.72	0.98	0.03	0.56	0.92	0.62	0.80	0.06	0.39	3
8	4	0.13	0.06	0.30	0.36	0.75	0.77	0.93	0.95	0.43	0.68	5
9	5	0.30	0.68	0.41	0.28	0.87	0.88	0.03	0.94	0.86	0.36	5
10	6	0.95	0.87	0.59	0.01	0.48	0.78	0.40	0.65	0.09	0.03	5
11	7	0.58	0.96	0.75	0.05	0.10	0.69	0.46	0.65	0.80	0.03	4
12	8	0.27	0.35	0.61	0.31	0.28	0.05	0.50	0.87	0.12	0.35	8
13	9	0.90	0.43	0.06	0.72	0.08	0.64	0.43	0.85	0.30	0.13	6
14	10	0.55	0.57	0.24	0.91	0.77	0.34	0.58	0.26	0.75	0.94	3
・	・	・	・	・	・	・	・	・	・	・	・	・
・	・	・	・	・	・	・	・	・	・	・	・	・
・	・	・	・	・	・	・	・	・	・	・	・	・
102	98	0.58	0.48	0.41	0.97	0.94	0.47	0.00	0.22	0.08	0.39	7
103	99	0.99	0.55	0.48	0.29	0.39	0.10	0.66	0.09	0.28	0.39	7
104	100	0.20	0.34	0.87	0.90	0.93	0.04	0.99	0.08	0.19	0.71	5

シミュレーション

p.142

● コイン実験の疑似体験

表示3.15
Excel によるシミュレーション

	A	B	C	D	E	F	G	H	I	J	K	L
3											$\pi =$	0.5
4	実験番号	1	2	3	4	5	6	7	8	9	10	表の個数
5	1	0.44	0.18	0.63	0.99	0.12	0.55	0.90	0.14	0.50	0.81	5
[illegible]	2	0.38	0.71	0.54	0.17	0.46	0.82	0.42	0.06	0.70	0.94	5
7	3	0.81	0.72	0.98	0.03	0.56	0.92	0.62	0.80	[illegible]	0.39	3
8	4	0.13	0.06	0.30	0.36	0.75	0.77	0.93	[illegible]	0.43	0.68	5
9	5	0.30	0.68	0.41	0.28	0.87	[illegible]	[illegible]	0.94	0.86	0.36	5
[illegible]	6	0.95	0.87	0.59	0.01	[illegible]	0.78	0.40	0.65	0.09	0.03	5
11	7	0.58	0.96	[illegible]	[illegible]	0.10	0.69	0.46	0.65	0.80	0.03	4
12	8	0.27	[illegible]	[illegible]	0.31	0.28	0.05	0.50	0.87	0.12	0.35	8
13	9	[illegible]	[illegible]	0.06	0.72	0.08	0.64	0.43	0.85	0.30	0.13	6
14	[illegible]	[illegible]	0.57	0.24	0.91	0.77	0.34	0.58	0.26	0.75	0.94	3
		・	・	・	・	・	・	・	・	・	・	・
	・	・	・	・	・	・	・	・	・	・	・	・
・	・	・	・	・	・	・	・	・	・	・	・	・
102	98	0.58	0.48	0.41	0.97	0.94	0.47	0.00	0.22	0.08	0.39	7
103	99	0.99	0.55	0.48	0.29	0.39	0.10	0.66	0.09	0.28	0.39	7
104	100	0.20	0.34	0.87	0.90	0.93	0.04	0.99	0.08	0.19	0.71	5

一様乱数の
出力先：B5
（左上隅）

10 列×100 行
一様乱数
0 ～ 1

B5:K5 （ブルー枠）からセル L3（黄）
以下の数値を「表（オモテ）」として計数
旧 =FREQUENCY(B5:K5, L3)
新 {=FREQUENCY(B5:K5, L3)}
配列数式として入力

シミュレーション

p.142

- コイン実験の疑似体験

表示3.15
Excel によるシミュレーション

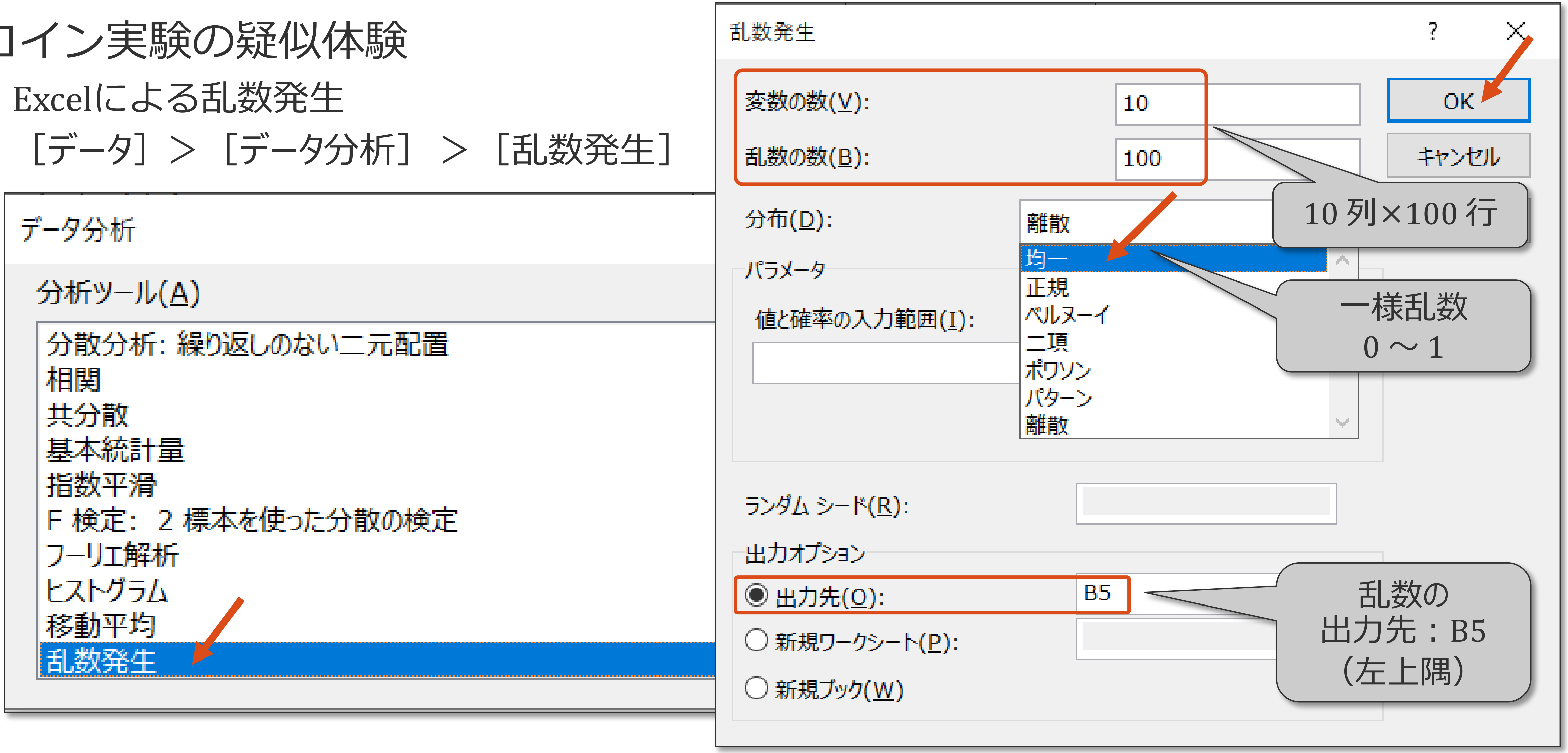
	A	B	C	D	E	F	G	H	I	J	K	L
3											π =	0.5
4	実験番号	1	2	3	4	5	6	7	8	9	10	表の個数
5	1	0.44	0.18	0.63	0.99	0.12	0.55	0.90	0.14	0.50	0.81	5
6	2	0.38	0.71	0.54	0.17	0.46	0.82	0.42	0.06	0.70	0.94	5
7	3	0.81	0.72	0.98	0.03	0.56	0.92	[illegible]	0.80	0.06	0.39	3
8	4	0.13	0.06	0.30	0.36	0.75	[illegible]	0.93	0.95	0.43	0.68	5
9	5	0.30	0.68	0.41	0.28	[illegible]	0.88	0.03	0.94	0.86	0.36	5
10	6	0.95	0.87	0.59	[illegible]	0.48	0.78	0.40	0.65	0.09	0.03	5
11	7	0.58	0.96	[illegible]	0.05	0.10	0.69	0.46	0.65	0.80	0.03	4
12	8	0.27	[illegible]	0.61	0.31	0.28	0.05	0.50	0.87	0.12	0.35	8
13	9	[illegible]	0.43	0.06	0.72	0.08	0.64	0.43	0.85	0.30	0.13	6
14	[illegible]	0.55	0.57	0.24	0.91	0.77	0.34	0.58	0.26	0.75	0.94	3
[illegible]	・	・	・	・	・	・	・	・	・	・	・	・
[illegible]	・	・	・	・	・	・	・	・	・	・	・	・
・	・	・	・	・	・	・	・	・	・	・	・	・
102	98	0.58	0.48	0.41	0.97	0.94	0.47	0.00	0.22	0.08	0.39	7
103	99	0.99	0.55	0.48	0.29	0.39	0.10	0.66	0.09	0.28	0.39	7
104	100	0.20	0.34	0.87	0.90	0.93	0.04	0.99	0.08	0.19	0.71	5

セル L3（黄）は 0.5
　表の出る確率が 0.5 の正しいコイン
セル L3（黄）を 0.4 に変更
　表の出る確率が0.4 、
　裏の出る確率が 0.6 の正しくないコイン

シミュレーション

p.142

- ●コイン実験の疑似体験
  - Excelによる乱数発生
  - ［データ］ ＞ ［データ分析］ ＞ ［乱数発生］



シミュレーション

●コイン実験の疑似体験（演習3.1.2）

B5:K104 の乱数を更新して、新しいシミュレーションを行う

π の値（$1 < \pi < 1$）を変えて「表」「裏」の出る確率が異なるコイン実験を行える（演習3.1.2）

表示3.1.5 Excelによるシミュレーション

0.5 → 0.2

$\pi =$ 0.5

実験番号	1	2	3	4	5	6	7	8	9	10	表の個数
1	0.44	0.18	0.63	0.99	0.12	0.55	0.90	0.14	0.50	0.81	5
2	0.38	0.71	0.54	0.17	0.46	0.82	0.42	0.06	0.70	0.94	5
3	0.81	0.72	0.98	0.03	0.56	0.92	0.62	0.80	0.06	0.39	3
4	0.13	0.06	0.30	0.36	0.75	0.77	0.93	0.95	0.43	0.68	5
5	0.30	0.68	0.41	0.28	0.87	0.88	0.03	0.94	0.86	0.36	5
6	0.95	0.87	0.59	0.01	0.48	0.78	0.40	0.65	0.09	0.03	5
7	0.58	0.96	0.75	0.05	0.10	0.69	0.46	0.65	0.80	0.03	4
8	0.27	0.35	0.61	0.31	0.28	0.05	0.50	0.87	0.12	0.35	8
9	0.90	0.43	0.06	0.72	0.08	0.64	0.43	0.85	0.30	0.13	6
10	0.55	0.57	0.24	0.91	0.77	0.34	0.58	0.26	0.75	0.94	3
11	0.07	0.08	0.76	0.32	0.16	0.76	0.02	0.14	0.38	0.55	7
•	•	•	•	•	•	•	•	•	•	•	•

表示3.1.4 「表の個数」のまとめ

表の個数	度数	ヒストグラム	理論値(%)
0	0	|	0.10
1	1	|*	0.98
2	4	|****	4.39
3	14	|**************	11.72
4	13	|*************	20.51
5	26	|**************************	24.61
6	22	|**********************	20.51
7	11	|***********	11.72
8	8	|********	4.39
9	1	|*	0.98
10	0	|	0.10
合計	100		100.00

シミュレーション

p.227

●コイン実験の疑似体験（演習3.1.2）

π = 0.5 の場合

「表の個数」 5で度数が最大

π = 0.2 の場合

「表の個数」 2で度数が最大

π=0.5 より、ピークが高く、変化の幅が狭い

表示3.1.4 「表の個数」のまとめ（π = 0.5）

表の個数	度数	ヒストグラム	理論値(%)
0	0		0.10
1	1	*	0.98
2	4	****	4.39
3	14	**************	11.72
4	13	*************	20.51
5	26	**************************	24.61
6	22	**********************	20.51
7	11	***********	11.72
8	8	********	4.39
9	1	*	0.98
10	0		0.10
合計	100		100.00

表示3.8.2 「表の個数」のまとめ（π = 0.2）

表の個数	度数	ヒストグラム	理論値(%)
0	5	*****	10.74
1	20	********************	26.84
2	40	**	30.20
3	21	*********************	20.13
4	10	**********	8.81
5	3	***	2.64
6	1	*	0.55
7	0		0.08
8	0		0.01
9	0		0.00
10	0		0.00
合計	100		100.00

ピークが高く
変化の幅が狭い

p.142

（3） 2 項分布

π = 0.5 ではない事例から 2 項分布を説明

２項分布

p.142

●事例（$\pi \neq 0.5$）

コインを投げて、「表」が出れば勝ち

１回について勝つ確率：$\pi = 1/2 = 0.5$

サイコロを投げて、１か６が出れば勝ち

１回について勝つ確率：$\pi =$ 1/3 （2/6＝1/3）

これを３回繰り返した場合の勝ち方は４パターン

勝数 f=0：３回とも負け

勝数 f=1：３回のうち、１回勝ち

勝数 f=2：３回のうち、２回勝ち

勝数 f=3：３回とも勝ち

↓

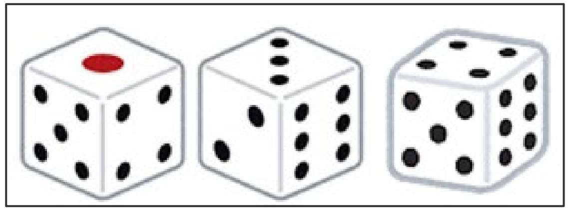
サイコロを 3回（n）投げて、

１か６が出る回数（f）とその確率（p_f）を考える（1か６が勝ち、それ以外は負けの２値）

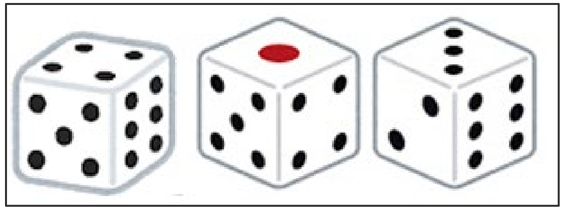
１投目　２投目　３投目

勝数 f=1 の場合

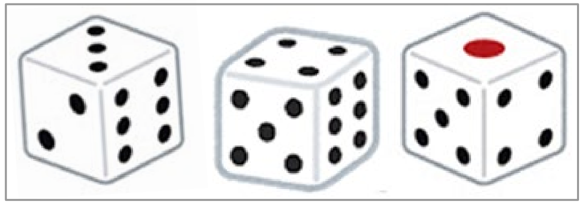
①勝、負、負



②負、勝、負



③負、負、勝



勝数 f=1でも、勝ち方は３種類ある

（１の目と６の目は区別しない）

2 項分布

p.142

●事例（$\pi \neq 0.5$）

コインを投げて、「表」が出れば勝ち

1 回について勝つ確率：$\pi = 1/2 = 0.5$

サイコロを投げて、１か６が出れば勝ち

1 回について勝つ確率：$\pi =$ 1/3 （2/6＝1/3）

これを３回繰り返した場合の勝ち方は４パターン

勝数 f=0：３回とも負け

勝数 f=1：３回のうち、１回勝ち

勝数 f=2：３回のうち、２回勝ち

勝数 f=3：３回とも勝ち

↓

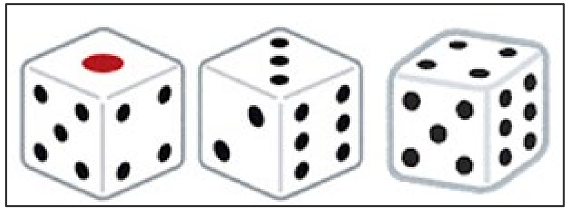
サイコロを 3回（n）投げて、

１か６が出る回数（f）とその確率（p_f）を考える（1か６が勝ち、それ以外は負けの２値）

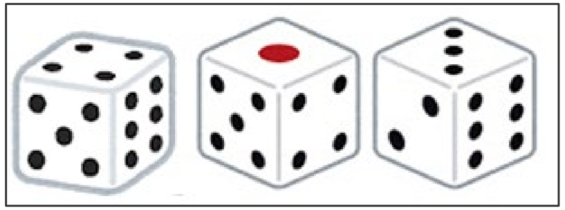
１投目　２投目　３投目

勝数 f=1 の場合

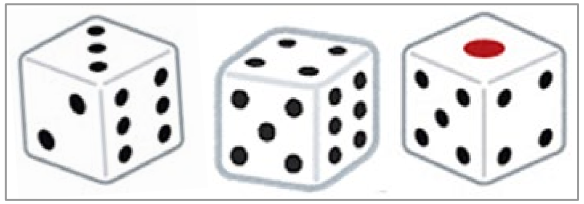
①勝、負、負



②負、勝、負



③負、負、勝



勝数 f=1でも、勝ち方は３種類ある

（１の目と６の目を区別しない）

2 項分布

p.143

●組合せの生起確率

勝ち方のパターン別に
生起確率を計算

(負：0)　(勝：1)　(負：0)

3 回のうちの
勝数

表示3.16

	組合せ番号	勝ち（1）負け（0） 1回目	2回目	3回目	1,6の個数 f	組合せの数	確率（p_f）
勝数０：３回とも負け	(1)	0	0	0	0	1	$(2/3)^3 = 8/27 = p_0$
勝数１：３回のうち、１回勝ち	(2)	1	0	0	1	3	$3 \times (1/3)(2/3)^2 = 12/27 = p_1$
	(3)	0	1	0	1		
	(4)	0	0	1	1		
勝数２：３回のうち２回勝ち	(5)	1	1	0	2	3	$3 \times (1/3)^2(2/3) = 6/27 = p_2$
	(6)	1	0	1	2		
	(7)	0	1	1	2		
勝数３：３回とも勝ち	(8)	1	1	1	3	1	$(1/3)^3 = 1/27 = p_3$
	計					8	$27/27 = 1$

2項分布

p.143

●組合せの生起確率

3回の
勝数

勝つ確率　1/3
負ける確率 1－1/3=2/3

表示3.16

	組合せ番号	勝ち（1）負け（0）1回目	2回目	3回目	1,6の個数 f	組合せの数	確率（p_f）
勝数0：3回とも負け	(1)	0	0	0	0	1	$(2/3)^3 = 8/27 = p_0$
勝数1：3回のうち、1回勝ち	(2)	1	0	0	1	3	$3\times(1/3)(2/3)^2 = 12/27 = p_1$
	(3)	0	1	0	1		
	(4)	0	0	1	1		
勝数2：3回のうち2回勝ち	(5)	1	1	0	2	3	$3\times(1/3)^2(2/3) = 6/27 = p_2$
	(6)	1	0	1	2		
	(7)	0	1	1	2		
勝数3：3回とも勝ち	(8)	1	1	1	3	1	$(1/3)^3 = 1/27 = p_3$
	計					8	$27/27 = 1$

2項分布

p.143

●組合せの生起確率

$$p_f = {}_nC_f \pi^f (1-\pi)^{n-f} \qquad (3.1.2)$$

生起確率 π の事象が、n 回の試行の中で f 回起こる確率 P_f

表示3.16

組合せ番号	勝ち（1）負け（0） 1回目	2回目	3回目	1,6の個数 f	組合せの数	確率（p_f）
(1)	0	0	0	0	1	$1\times(1/3)^0(2/3)^3 = 8/27 = p_0$
(2)	1	0	0	1	3	$3\times(1/3)^1(2/3)^2 = 12/27 = p_1$
(3)	0	1	0	1		
(4)	0	0	1	1		
(5)	1	1	0	2	3	$3\times(1/3)^2(2/3)^1 = 6/27 = p_2$
(6)	1	0	1	2		
(7)	0	1	1	2		
(8)	1	1	1	3	1	$1\times(1/3)^3(1/3)^0 = 1/27 = p_3$
計					8	$27/27 = 1$

2 項分布

● 2 項分布の計算

生起確率 π の事象が、n 回の試行の中で f 回起こる確率 P_f

生起確率：ある事象が起こる確率

$$p_f = {}_nC_f \pi^f (1-\pi)^{n-f} \qquad (3.1.2)$$

${}_nC_f$ は、n 個から f 個を選ぶ組合せ（Combination）の数・・・2 項係数

$$ {}_nC_f = \frac{n!}{f!\,(n-f)!} \qquad (3.1.3)$$

$$= \frac{n(n-1)(n-2)\cdots(f+1)f(f-1)\cdots 2\cdot 1}{\{f(f-1)(f-2)\cdots 2\cdot 1\}\{(n-f)(n-f-1)(n-f-2)\cdots 2\cdot 1\}}$$

$$= \frac{n(n-1)(n-2)\cdots(f+1)}{(n-f)(n-f-1)(n-f-2)\cdots 2\cdot 1}$$

$$n! = n(n-1)(n-2)\cdots 3\cdot 2\cdot 1 \quad 0! = 1$$

n の階乗（Factorial）

2 項分布

p.143

● 2 項分布の計算

生起確率 π の事象が、n 回の試行の中で f 回起こる確率 P_f

生起確率：ある事象が起こる確率

$$p_f = {}_nC_f \pi^f (1-\pi)^{n-f} \qquad (3.1.2)$$

2 種類の確率
生起確率　π
n 回の試行の中で
f 回起こる確率　P_f

${}_nC_f$ は、n 個から f 個を選ぶ組合せ（Combination）の数・・・2 項係

$${}_nC_f = \frac{n!}{f!\,(n-f)!} \qquad (3.1.3)$$

$$= \frac{n(n-1)(n-2)\cdots(f+1)f(f-1)\cdots 2\cdot 1}{\{f(f-1)(f-2)\cdots 2\cdot 1\}\{(n-f)(n-f-1)(n-f-2)\cdots 2\cdot 1\}}$$

$$= \frac{n(n-1)(n-2)\cdots(f+1)}{(n-f)(n-f-1)(n-f-2)\cdots 2\cdot 1}$$

$n! = n(n-1)(n-2)\cdots 3\cdot 2\cdot 1 \quad 0! = 1$

n の階乗（Factorial）

2 項分布

サイコロの事例

p.143

● 2 項係数

n 個から f 個を取り出す組合せの総数

サイコロを３回投げる事例（１か６が出れば勝ち）

３回の試行の中で０個を取り出す（f=0）　$_3C_0 = 1$

３回の試行の中で１個を取り出す（f=1）　$_3C_1 = 3$

３回の試行の中で２個を取り出す（f=2）　$_3C_2 = 3$

３回の試行の中で３個を取り出す（f=3）　$_3C_3 = 1$

$$_nC_f = \frac{n!}{f!\,(n-f)!} \qquad (3.1.3)$$

$$= \frac{n(n-1)(n-2)\cdots(f+1)}{(n-f)(n-f-1)(n-f-2)\cdots 2\cdot 1}$$

試行の回数 n	1	2	3	4	5
事象が起こる回数（取り出す回数） f					0
				0	
			0		1
		0		1	
	0		1		2
		1		2	
	1		2		3
		2		3	
			3		4
				4	
					5
2 項係数 $_nC_f$					1
				1	
			1		5
		1		4	
	1		3		10
		2		6	
	1		3		10
		1		4	
			1		5
				1	
					1

パスカルの三角形

2項分布

p.144

● 2項分布と多項分布

2項分布：事象が起こる確率 π と、試行回数 n が与えられたとき、

試行回数 n の中で事象が起こる回数 f とその起こる確率 p_f との関係： 2項分布

$$p_f = {}_nC_f \pi^f (1-\pi)^{n-f} \quad (3.1.2)$$

条件：結果が2通り、確率 π が一定、すべての試行は互いに独立している

2項分布：結果が2通りの事象

コインの「裏、表」、サイコロの目の「奇数、偶数」、薬効の有無、副作用の有無

多項分布：結果が3通り以上

薬効が無効・有効・著効　　→ §3.4 「多項分布（名義尺度）」

2 項分布

p.144

● 2 項分布の計算例

成功率が99%の手術を100人に実施して全て成功する確率

$$p_{100} = {}_nC_f \pi^f (1-\pi)^{n-f} \qquad (3.1.2)$$

$$= {}_{100}C_{100} \times 0.99^{100} \times (1-0.99)^{100-100} = 1 \times 0.99^{100} = 0.366$$

2 種類の確率
生起確率　$\pi = 0.99$
100 回の試行の中で
100 回起こる確率
$P_{100} = 0.366$

$${}_{100}C_{100} = \frac{n!}{f!\,(n-f)!} = \frac{100!}{100!\,(100-100)!} = \frac{100!}{100!} = 1$$

コインを10回投げて、「表」が 3 回出る確率

$$p_3 = {}_nC_f \pi^f (1-\pi)^{n-f} \qquad (3.1.2)$$

$$= {}_{10}C_3 \times 0.5^3 \times (1-0.5)^{10-3} = 120 \times 0.1250 \times 0.0078 = 0.117$$

2 種類の確率
生起確率　$\pi = 0.5$
10 回の試行の中で
3 回起こる確率
$P_3 = 0.117$

$${}_{10}C_3 = \frac{n!}{f!\,(n-f)!} = \frac{10!}{3!\,(10-3)!} = \frac{10\times9\times8\times7\times6\times5\times4}{7\times6\times5\times4\times3\times2\times1} = 120$$

p.144

（4）Excel による確率の計算

Excel の２項分布に関する関数

Excel による確率の計算

p.144

- ２項分布の Excel 関数

コインを10回投げて、「表」が３回出る確率

$$p_f = {}_nC_f \pi^f (1-\pi)^{n-f} \qquad (3.1.2)$$

$$= {}_{10}C_3 \times 0.5^3 \times (1-0.5)^{10-3} = 120 \times 0.1250 \times 0.0078 = 0.117$$

=BINOMDIST(3, 10, 0.5, FALSE)
=0.117

$${}_{10}C_3 = \frac{n!}{f!\,(n-f)!} = \frac{10!}{3!\,(10-3)!} = \frac{10\times9\times8\times7\times6\times5\times4}{7\times6\times5\times4\times3\times2\times1} = 120$$

=COMBIN(10, 3)
=120

=COMBIN(n, f)・・・・・・・・・・・・２項係数（組み合わせ数）

=BINOM.DIST(f, n, π, オプション) ・・・２項分布の確率

FALSE：f となる確率を計算するように指定

TRUE：f 以下の下側累積確率を計算するように指定

新しい関数は
ピリオドが入る

Excel による確率の計算

- 2 項分布の Excel 関数

表示3.1.6

組合せ番号	勝ち（1）負け（0） 1 回目	2 回目	3 回目	1,6 の個数 f	組合せの数	確率（p_f）
(1)	0	0	0	0	1	$1 \times (1/3)^0 (2/3)^3 = 8/27 = p_0$
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	
[illegible]	0	1	[illegible]	1	3	$3 \times (1/3)^1 (2/3)^2 = 12/27 = p_1$
(4)	0	0	1	1		

累積確率は通常下側累積確率を表す
「下側」を省略することが多い

上側累積確率の略記

表示3.1.7

	A	B	C	D	E	F
3	f	n	π	確率	累積確率	上側確率
4	0	3	0.3333	0.2963	0.2963	1.0000
5	1	3	0.3333	0.4444	0.7407	0.7037
6	1	10	0.5	0.0098	0.0107	0.9990
7	8	10	0.5	0.0439	0.9893	0.0547
8						
9	D4: =BINOMDIST(A4,B4,C4,FALSE)					
10	E4: =BINOMDIST(A4,B4,C4,TRUE)					
11	F4: =IF(A4=0,1,1-BINOMDIST(A4-1,B4,C4,TRUE))					

Excel による確率の計算

p.145

- 2 項分布の Excel 関数

表示3.1.6

8/27=0.2963

組合せ番号	勝ち（1）負け（0） 1 回目	2 回目	3 回目	1,6 の個数 f	組合せの数	確率（p_f）
(1)	0	0	0	0	1	$1\times(1/3)^0(2/3)^3 = 8/27 = p_0$
(2)	1	0	0	1	3	$3\times(1/3)^1(2/3)^2 = 12/27 = p_1$
(3)	0	1	0	1		
(4)	0	0	1	1		

表示3.1.7

	A	B	C	D	E	F
3	f	n	π	確率	累積確率	上側確率
4	0	3	0.3333	0.2963	0.2963	1.0000
5	1	3	0.3333	0.4444	0.7407	0.7037
6	1	10	0.5	0.0098	0.0107	0.9990
7	8	10	0.5	0.0439	0.9893	0.0547
8						
9	D4: =BINOMDIST(A4,B4,C4,FALSE)					
10	E4: =BINOMDIST(A4,B4,C4,TRUE)					
11	F4: =IF(A4=0,1,1-BINOMDIST(A4-1,B4,C4,TRUE))					

FALSE：f になる確率
TRUE：f 以下の下側累積確率

f 以上の上側累積確率は
1 から ($f-1$) の下側累積確率を引く
（IF 文で $f=0$ のエラーを回避）

Excel による確率の計算

- 2 項分布の Excel 関数

表示3.1.6

組合せ番号	勝ち（1）負け（0） 1 回目	2 回目	3 回目	1,6 の個数 f	組合せの数	確率（p_f）
(1)	0	0	0	0	1	$1\times(1/3)^0(2/3)^3 = 8/27 = p_0$
(2)	1	0	0	1	3	$3\times(1/3)^1(2/3)^2 = 12/27 = p_1$
(3)	0	1	0	1		
(4)	0	0	1	1		

12/27=0.4444

表示3.1.7

	A	B	C	D	E	F
3	f	n	π	確率	累積確率	上側確率
4	0	3	0.3333	0.2963	0.2963	1.0000
5	1	3	0.3333	0.4444	0.7407	0.7037
6	1	10	0.5	0.0098	0.0107	0.9990
7	8	10	0.5	0.0439	0.9893	0.0547
8						
9	D4: =BINOMDIST(A4,B4,C4,FALSE)					
10	E4: =BINOMDIST(A4,B4,C4,TRUE)					
11	F4: =IF(A4=0,1,1-BINOMDIST(A4-1,B4,C4,TRUE))					

0.2963 + 0.4444 == 0.7407

1 − 0.2963 = 0.7037

FALSE：f になる確率
TRUE：f 以下の下側累積確率

f 以上の上側累積確率は
1 から $(f-1)$ の下側累積確率を引く
（IF 文で $f = 0$ のエラーを回避）

Excel による確率の計算

● 2 項分布の確率

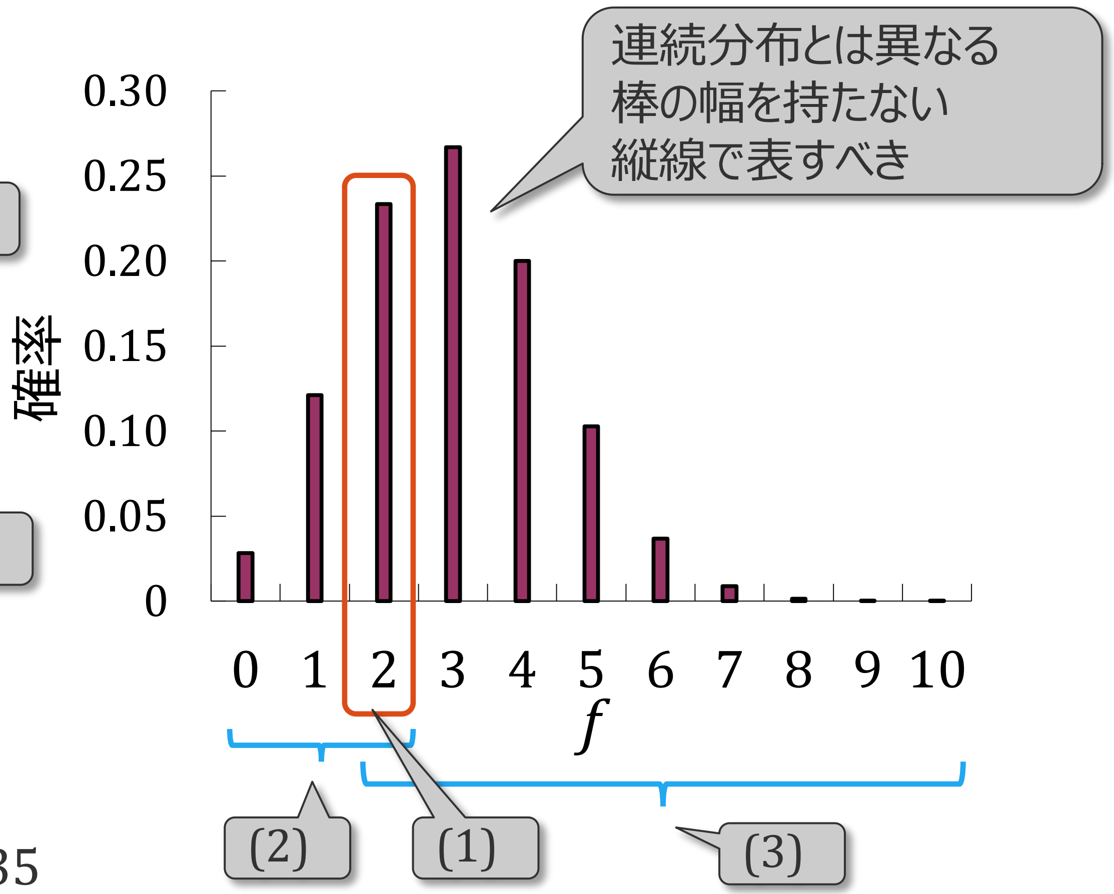
表示 3.1.8

f	n	π	確率
0	10	0.3	0.0282
1	10	0.3	0.1211
2	10	0.3	0.2335
3	10	0.3	0.2668
4	10	0.3	0.2001
5	10	0.3	0.1029
6	10	0.3	0.0368
7	10	0.3	0.0090
8	10	0.3	0.0014
9	10	0.3	0.0001
10	10	0.3	0.0000

(2)累積確率

(1)

(3)上側確率



(1) =BINOM.DIST(2, 10, 0.3, FALSE) = 0.2335

(2) =BINOM.DIST(2, 10, 0.3, TRUE) = 0.0282 + 0.1211 + 0.2335 =0.3828

(3) =1－BINOM.DIST(2-1, 10, 0.3, TRUE) =1－0.0282－0.1211 = 0.8507

Excel による確率の計算

p.145

● 2 項分布の確率

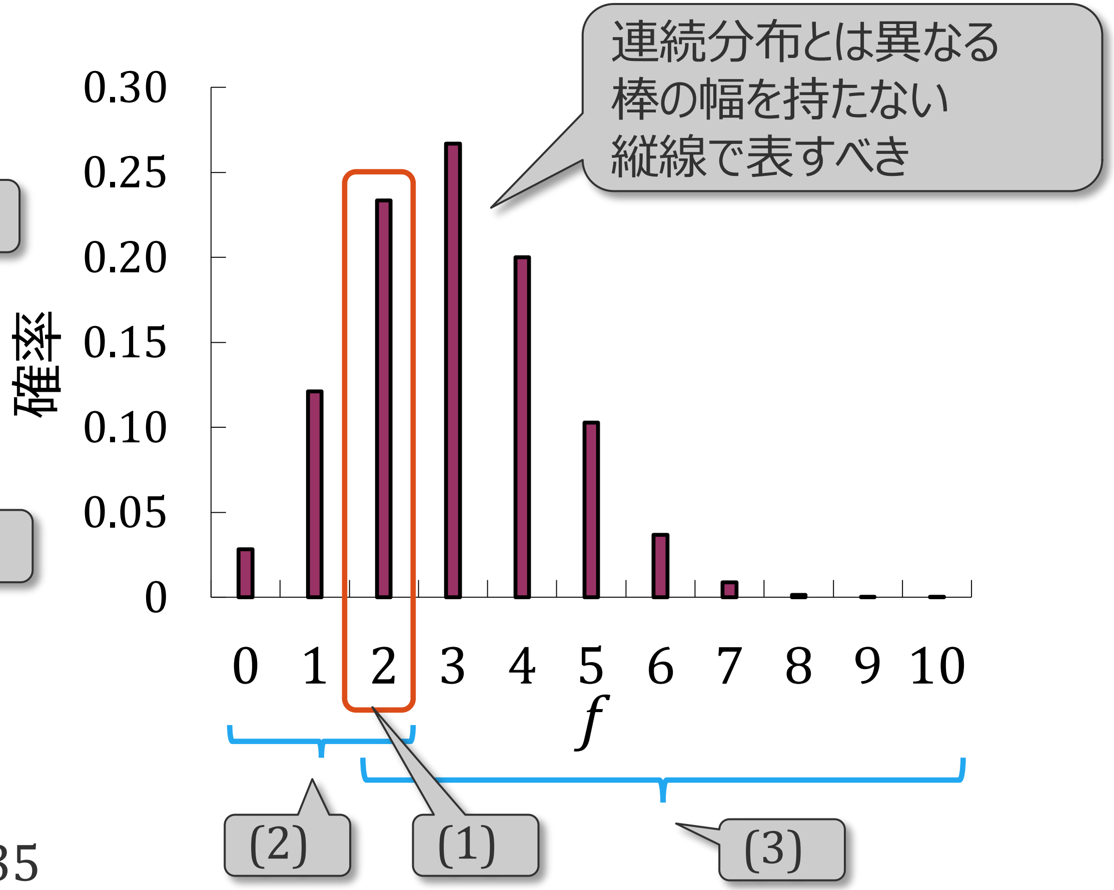
表示 3.1.8

f	n	π	確率
0	10	0.3	0.0282
1	10	0.3	0.1211
2	10	0.3	0.2335
3	10	0.3	0.2668
4	10	0.3	0.2001
5	10	0.3	0.1029
6	10	0.3	0.0368
7	10	0.3	0.0090
8	10	0.3	0.0014
9	10	0.3	0.0001
10	10	0.3	0.0000

(2)累積確率

(1)

(3)上側確率



(1) =BINOM.DIST(2, 10, 0.3, FALSE) = 0.2335

(2) =BINOM.DIST(2, 10, 0.3, TRUE) = 0.0282 + 0.1211 + 0.2335 =0.3828

(3) =1－BINOM.DIST(2-1, 10, 0.3, TRUE) =1－0.0282－0.1211 = 0.8507

Excel による確率の計算

● 2 項分布の確率

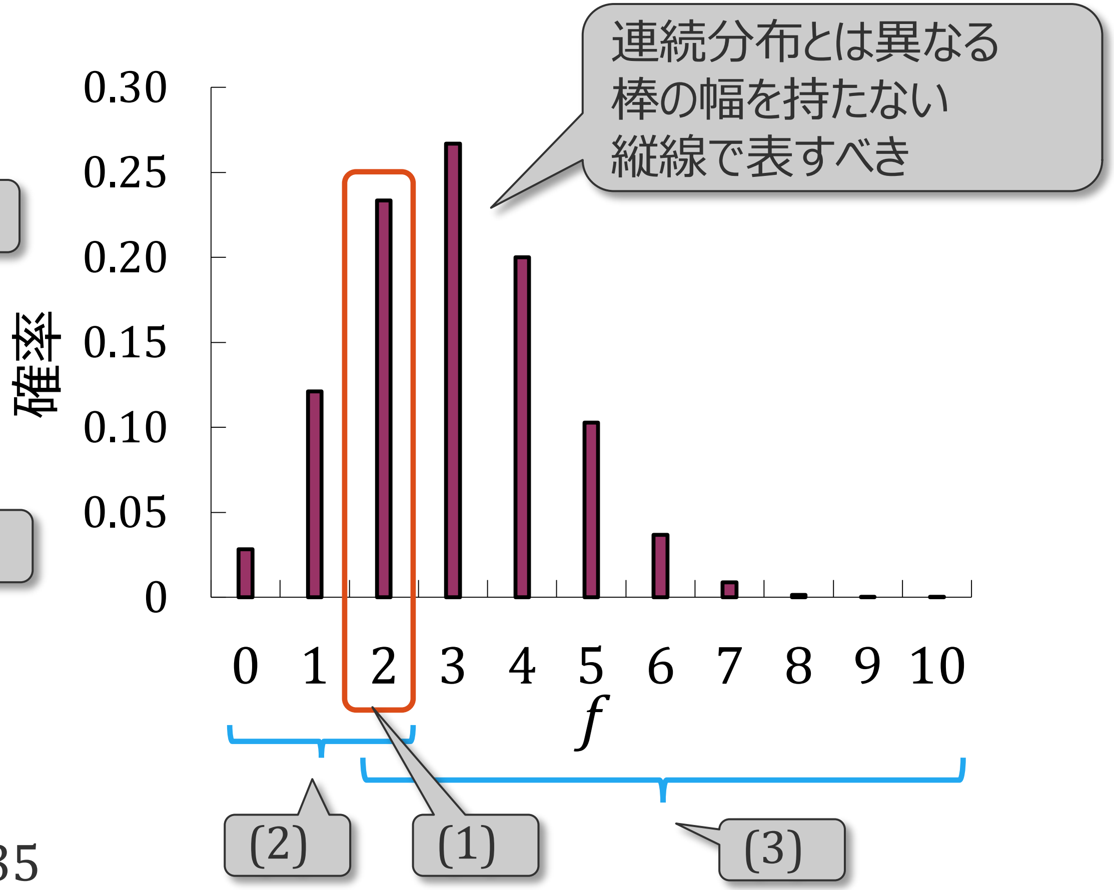
表示 3.1.8

f	n	π	確率
0	10	0.3	0.0282
1	10	0.3	0.1211
2	10	0.3	0.2335
3	10	0.3	0.2668
4	10	0.3	0.2001
5	10	0.3	0.1029
6	10	0.3	0.0368
7	10	0.3	0.0090
8	10	0.3	0.0014
9	10	0.3	0.0001
10	10	0.3	0.0000

(2)累積確率

(1)

(3)上側確率



(1) =BINOM.DIST(2, 10, 0.3, FALSE) = 0.2335

(2) =BINOM.DIST(2, 10, 0.3, TRUE) = 0.0282 + 0.1211 + 0.2335 =0.3828

(3) =1－BINOM.DIST(2-1, 10, 0.3, TRUE) =1－0.0282－0.1211 = 0.8507

合計は
1 より大きい

Excel による確率の計算

p.145

● 2 項分布の確率

$$p_f = {}_nC_f \pi^f (1-\pi)^{n-f} \qquad (3.1.2)$$
$$= {}_{10}C_f 0.5^f (1-0.5)^{10-f}$$

=BINOM.DIST(f, 10, 0.5, FALSE)*100

表示3.1.4 「表の個数」のまとめ

π = 0.5

実験番号	1	2	3	4	5	6	7	8	9	10	表の個数
1	0.44	0.18	0.63	0.99	0.12	0.55	0.90	0.14	0.50	0.81	5
2	0.38	0.71	0.54	0.17	0.46	0.82	0.42	0.06	0.70	0.94	5
3	0.81	0.72	0.98	0.03	0.56	0.92	0.62	0.80	0.06	0.39	3
4	0.13	0.06	0.30	0.36	0.75	0.77	0.93	0.95	0.43	0.68	5
5	0.30	0.68	0.41	0.28	0.87	0.88	0.03	0.94	0.86	0.36	5
6	0.95	0.87	0.59	0.01	0.48	0.78	0.40	0.65	0.09	0.03	5
7	0.58	0.96	0.75	0.05	0.10	0.69	0.46	0.65	0.80	0.03	4
8	0.27	0.35	0.61	0.31	0.28	0.05	0.50	0.87	0.12	0.35	8
9	0.90	0.43	0.06	0.72	0.08	0.64	0.43	0.85	0.30	0.13	6
10	0.55	0.57	0.24	0.91	0.77	0.34	0.58	0.26	0.75	0.94	3
11	0.07	0.08	0.76	0.32	0.16	0.76	0.02	0.14	0.38	0.55	7
•	•	•	•	•	•	•	•	•	•	•	•

π

f

n

表示3.1.4 「表の個数」のまとめ

表の個数	度数	ヒストグラム	理論値(%)
0	0	|	0.10
1	1	|*	0.98
2	4	|****	4.39
3	14	|**************	11.72
4	13	|*************	20.51
5	26	|**************************	24.61
6	22	|**********************	20.51
7	11	|***********	11.72
8	8	|********	4.39
9	1	|*	0.98
10	0	|	0.10
合計	100		100.00

p.146

（5）JMP による確率の計算

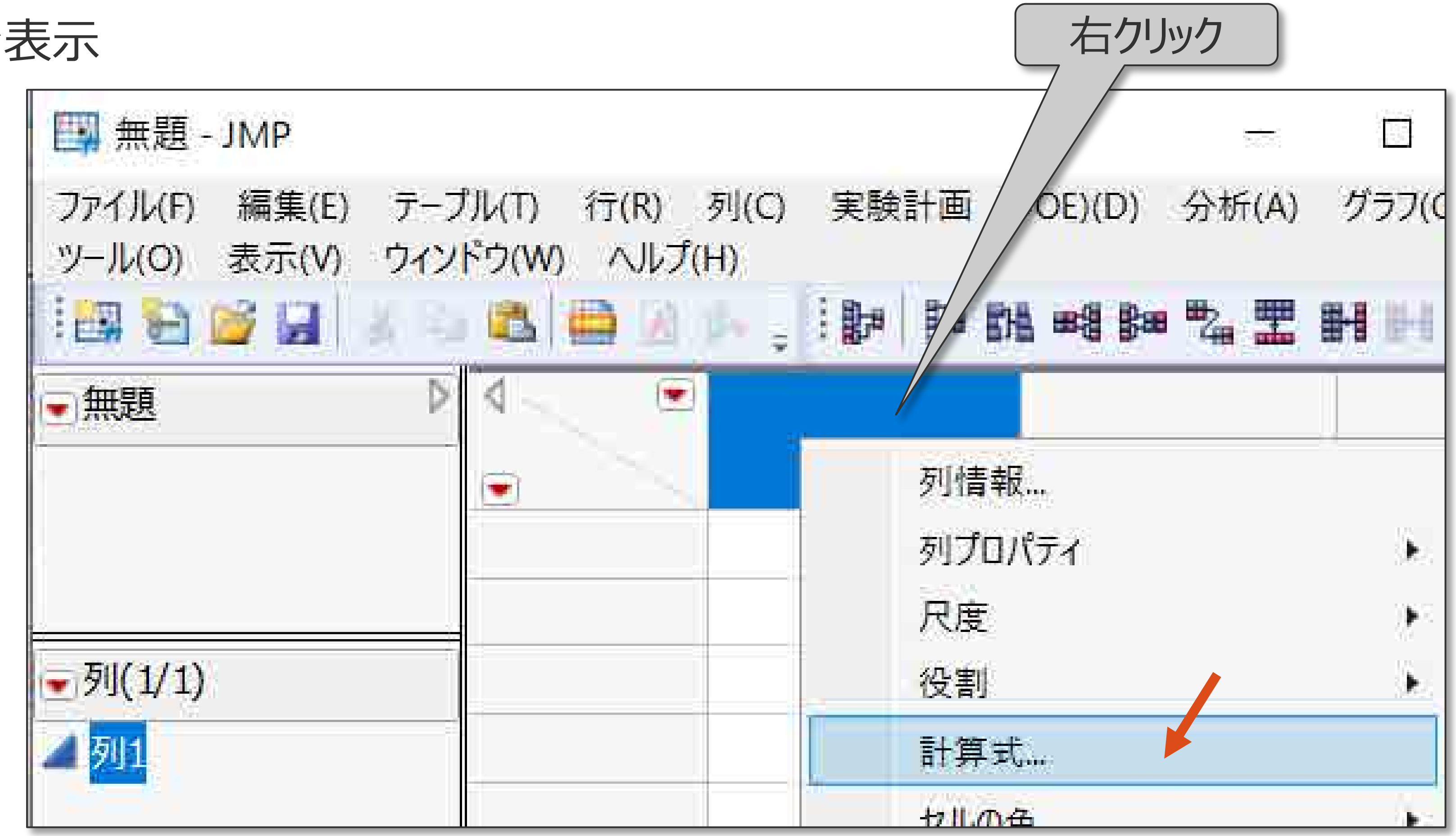
JMP の２項分布に関する関数

JMP による確率の計算

●JMP の計算式エディタの関数

JMP を立ち上げて、新規データテーブルを作成

適当な列で計算式エディタを表示

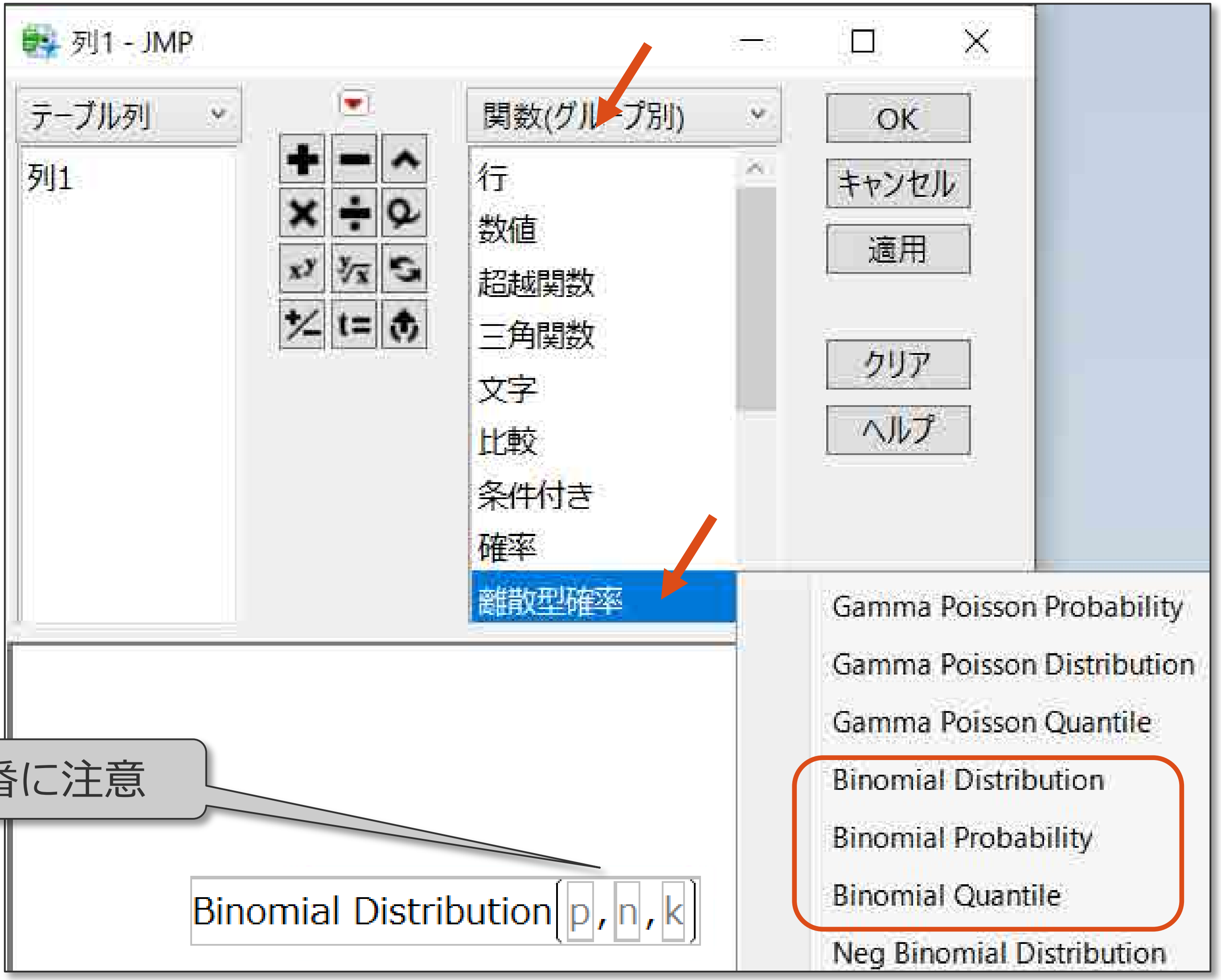


JMP による確率の計算

p.146

●JMP の
　２項分布に関する関数

［関数］＞［離散型確率］
Binomial Distribution
Binomial Probability
Binomial Quantile



引数の記号と順番に注意

JMP による確率の計算

p.146

●JMP の２項分布に関する関数

p：生起確率 π
n：試行回数 n
k：回数 f

表示3.1.9 JMP による２項分布の確率の計算

機能	関数名（引数）	引数の例	結果
確率	Binomial Probability (p, n, k)	(0.5, 10, 1)	0.0098
下側累積確率	Binomial Distribution (p, n, k)	(0.5, 10, 1)	0.0107
		(0.5, 10, 2)	0.0547
		(0.5, 10, 8)	0.9893
下側 % 点	Binomial Quantile (p, n, 累積確率)	(0.5, 10, 0.025)	2
		(0.5, 10, 0.975)	8

Excel 関数と引数の順番が逆
=BINOM.DIST(f, n, π, FALSE)

=BINOM.DIST(f, n, π, TRUE)

=BINOM.INV(n, π, 累積確率)
p 、n、下側累積確率を指定
得られた f が有意かを判断する
限界値を返す

上側累積確率　1－Binominal Distribution (p, n, k-1)
1－Binominal Distribution (0.5, 10, 0-1) = 1
k = 0 の場合 1 を返す（Excelではエラー）

JMP による確率の計算

p.146

●JMP の２項分布に関する関数

表示3.1.9

機能	関数名（引数）	引数の例	結果
下側 % 点	Binomial Quantile (p, n, 累積確率)	(0.5, 10, 0.025)	2
		(0.5, 10, 0.975)	8

指定した累積確率以上になる分位点のうち最小の整数
両側 α=0.05 →
累積確率 0.025 と 0.975（1－0.025） を指定

JMP による確率の計算

p.146

●JMP の２項分布に関する関数

表示3.1.9

機能	関数名（引数）	引数の例	結果
下側 ％ 点	Binomial Quantile (p, n, 累積確率)	(0.5, 10, 0.025)	2
		(0.5, 10, 0.975)	8

指定した累積確率以上になる分位点のうち最小の整数
両側 α=0.05　→
累積確率 0.025 と 0.975（1－0.025） を指定

コインを10回投げて「表」が２回出た。怪しい？
正しいコインか、不正なコインか
（p =0.2 は π=0.5 と有意に異なるか）
帰無仮説 H_0 ：$\pi = 0.5$、対立仮説 H_1 ：$\pi \neq 0.5$
n = 10、k＜2 または k＞８のとき H_0 を棄却（両側検定）
関数が返す値は、棄却できない範囲の両端を示す

JMP による確率の計算

p.146

●JMP の２項分布に関する関数

表示3.1.9

機能	関数名（引数）	引数の例	結果
下側 % 点	Binomial Quantile (p, n, 累積確率)	(0.5, 10, 0.025)	2
		(0.5, 10, 0.975)	8

k	n	p	確率	下側累積	上側累積
0	10	0.5	0.0010	0.0010	
1	10	0.5	0.0098	0.0107	
2	10	0.5	0.0439	0.0547	
3	10	0.5	0.1172	0.1719	
4	10	0.5	0.2051	0.3770	
5	10	0.5	0.2461	0.6230	
6	10	0.5	0.2051	0.8281	
7	10	0.5	0.1172	0.9453	0.1719
8	10	0.5	0.0439	0.9893	0.0547
9	10	0.5	0.0098	0.9990	0.0107
10	10	0.5	0.0010	1.0000	0.0010

この小さい方に
0.025 が位置する

この大きい方に
0.025 が位置する

指定した累積確率以上になる分位点のうち最小の整数
両側 α=0.05 →
累積確率 0.025 と 0.975（1－0.025） を指定

コインを10回投げて「表」が 2 回出た。怪しい？
正しいコインか、不正なコインか
（p =0.2 は π=0.5 と有意に異なるか）
帰無仮説 H_0 ：$\pi = 0.5$、対立仮説 H_1 ：$\pi \neq 0.5$
n = 10、k＜2 または k＞8 のとき H_0 を棄却（両側検定）
関数が返す値は、棄却できない範囲の両端を示す

（6）n, π による２項分布の変化

２項分布の形

n, π による２項分布の変化

p.147

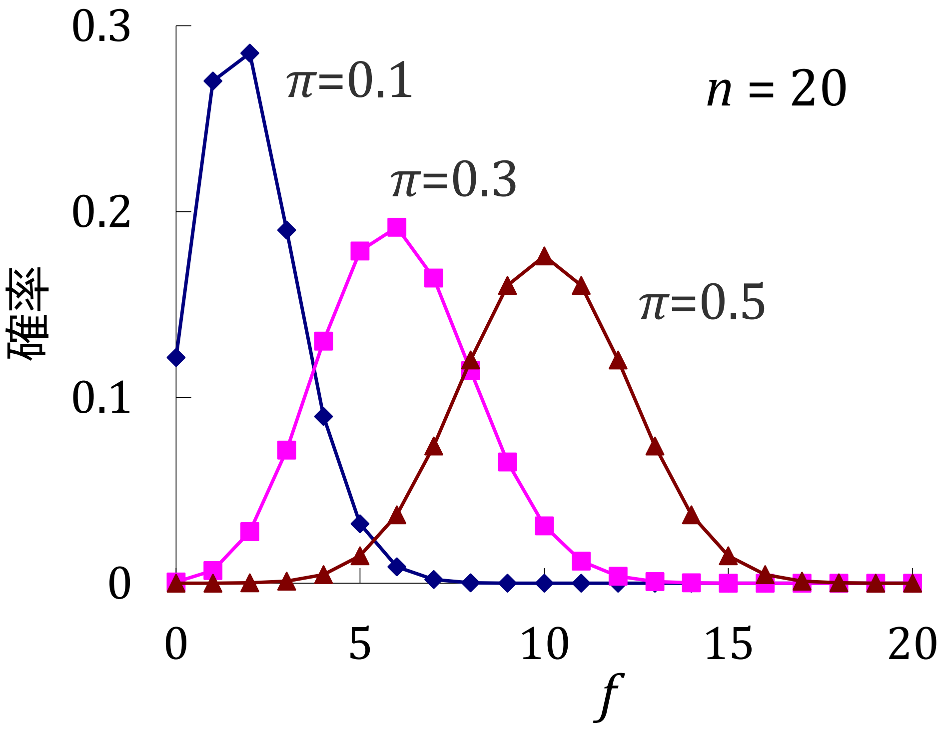
● ２項分布：n を固定

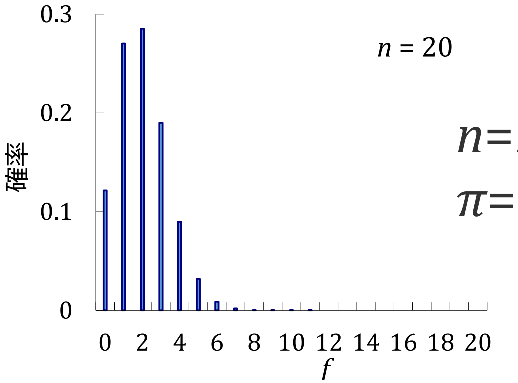
π が 0.5 のときに左右対称

π が 0.5 よりも小さくなると右に裾を引く非対称分布

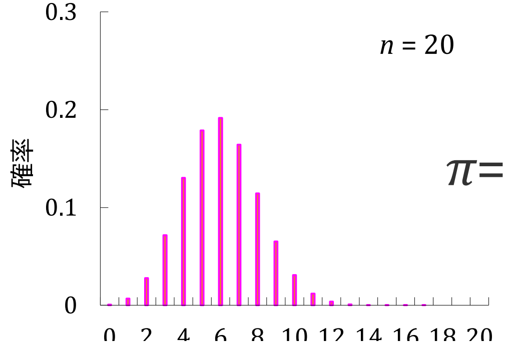
π が 0.5 よりも大きくなると左に裾を引く非対称分布

表示 3.1.10
n, π による
２項分布の
変化

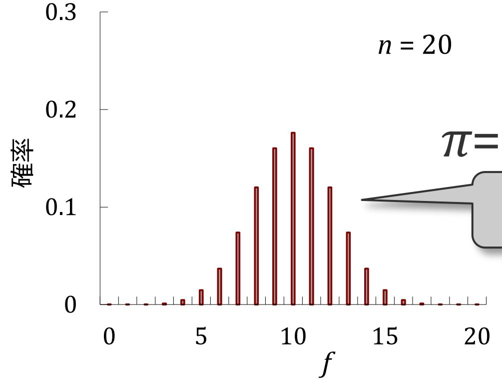




n=20
π=0.1



π=0.3



π=0.5

左右対称

n, π による２項分布の変化

p.147

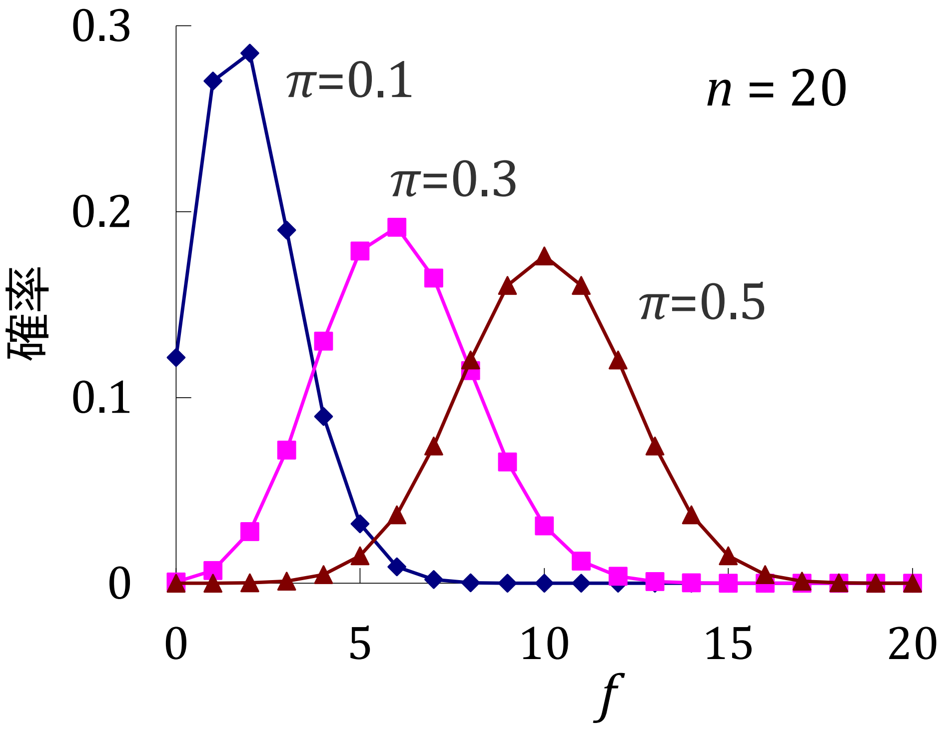
● ２項分布：n を固定

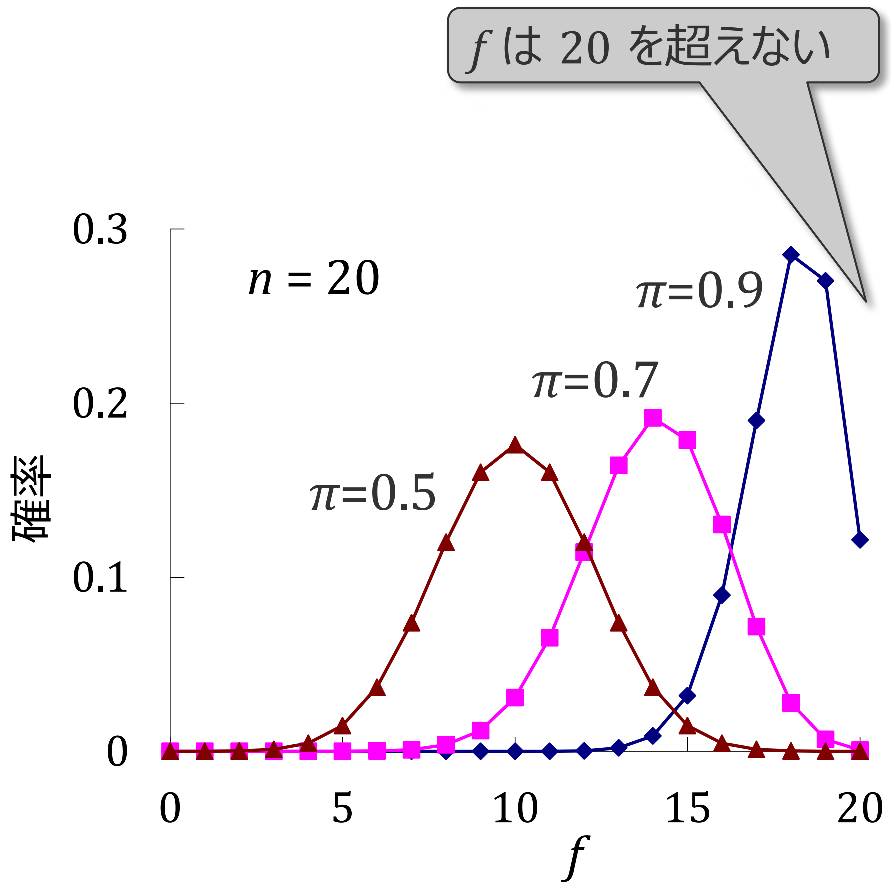
π が 0.5 のときに左右対称

π が 0.5 よりも小さくなると右に裾を引く非対称分布

π が 0.5 よりも大きくなると左に裾を引く非対称分布

表示 3.1.10
n, π による
２項分布の
変化





n, π による 2 項分布の変化

p.147

● 2 項分布：$n\pi$ が一定

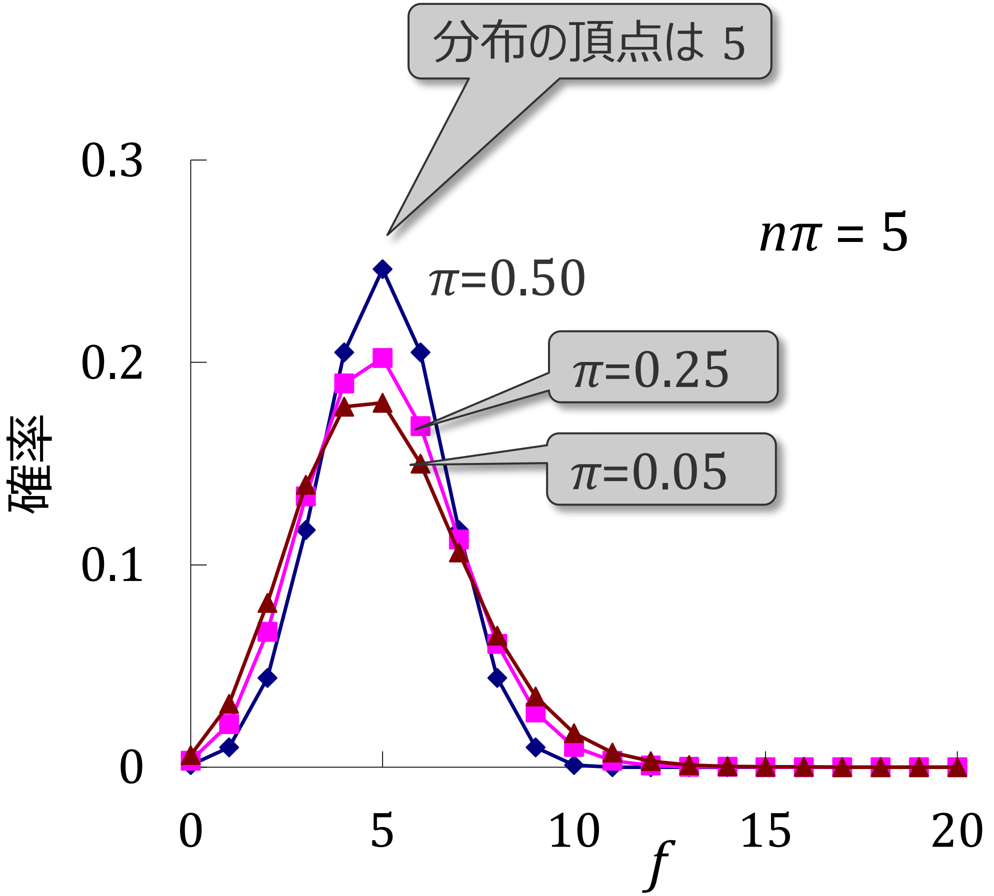
π が 0.5 のときに左右対称、頂点が最も高い

π が 0.5 から外れると非対称分布
その分、n を大きくしているので、
それほど対称からはずれない

いずれも、$f = n\pi = 5$ が頂点
（ $n\pi$ が整数の時のみ成立 ）

表示 3.1.10
n, π による
2 項分布の
変化

n	π	nπ
10	0.50	5
20	0.25	5
100	0.05	5



p.147

（7）２項分布の期待値と分散

２項分布の中心位置と
分布の広がり

2項分布の期待値と分散

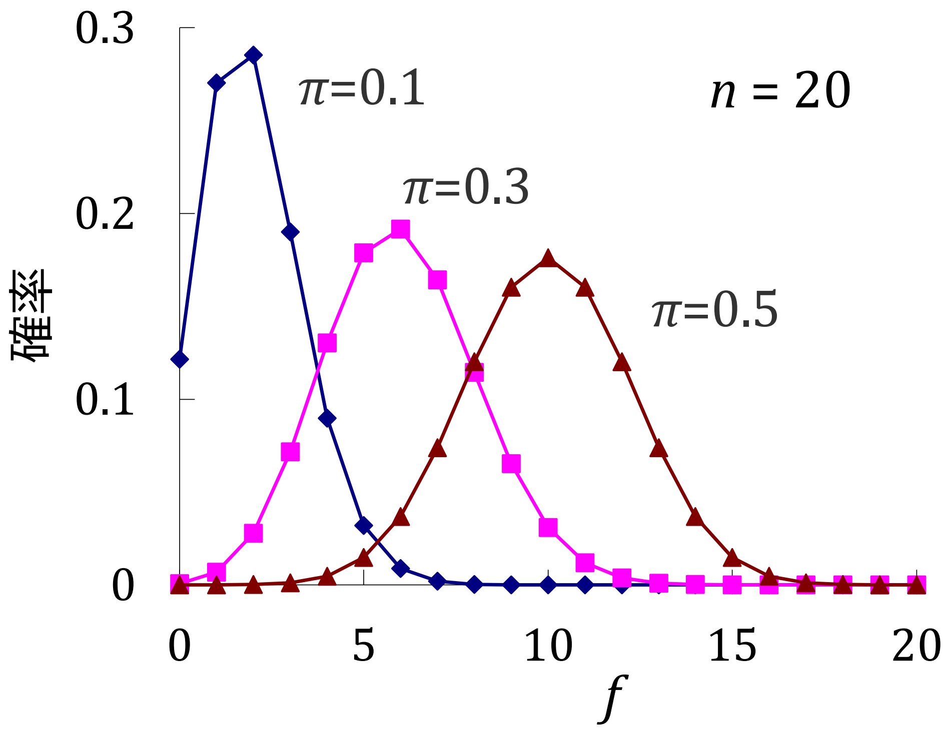
p.147

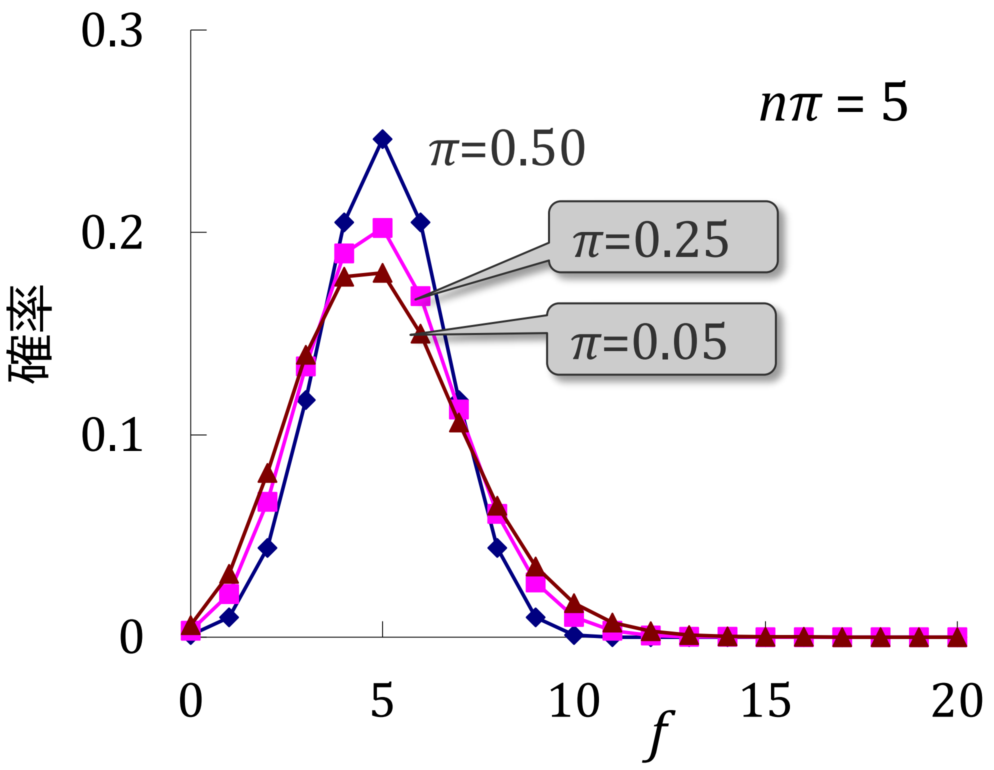
● 2項分布

2項分布の中心位置（期待値）と広がり（分散）は n と π で変化

f が、n と π の2項分布に従うとき、f の期待値 E[f] と分散 V[f] を求める

表示 3.1.10
n, π による2項分布の変化





2 項分布の期待値と分散

p.148

● f の期待値と分散

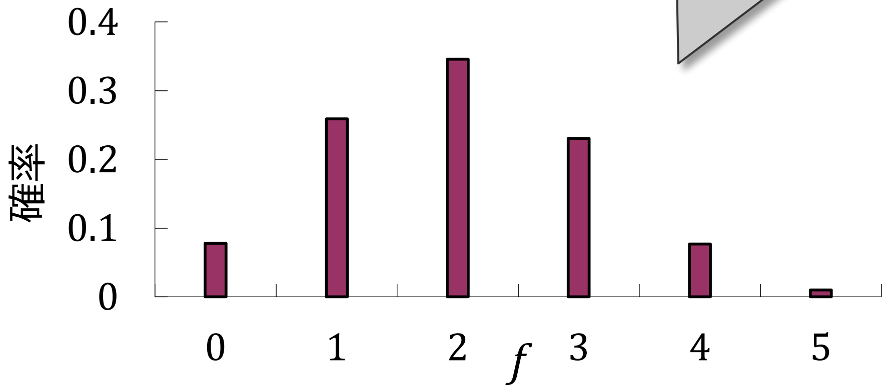
事例：fが n = 5 と π = 0.4 の 2 項分布に従う

表示 3.1.11 π = 0.4 ,n = 5 の期待値と分散の計算

f	n	π	確率 p	pf	e =f - E[f]	pe^2
0	5	0.40	0.0778	0.000	-2.000	0.311
1	5	0.40	0.2592	0.259	-1.000	0.259
2	5	0.40	0.3456	0.691	0.000	0.000
3	5	0.40	0.2304	0.691	1.000	0.230
4	5	0.40	0.0768	0.307	2.000	0.307
5	5	0.40	0.0102	0.051	3.000	0.092
		計	1.0000	2.000		1.200
			必ず 1	E[f]		V[f]

BINOM.DIST
(f, n, π, FALSE)

fの中心位置：期待値 E[f]
fの広がり　　：分散 V[f]



2 項分布の期待値と分散

● f の期待値と分散

事例：f が n = 5 と π = 0.4 の 2 項分布に従う

期待値 E[f] は、出現する値（f）と、出現確率（p）との積の和

分散 V[f] は、出現する値（f）と期待値の差（e）の 2 乗と、出現確率（p）との積の和

（第 1 部 §1.2）

表示 3.1.11 π = 0.4 ,n = 5 の期待値と分散の計算

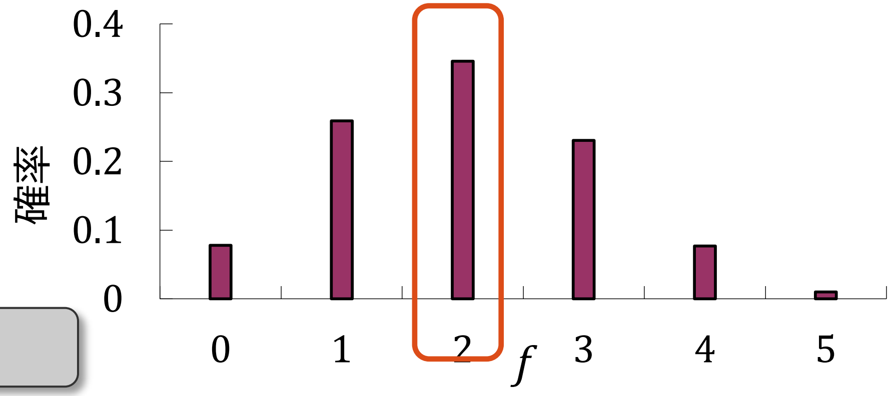
f	n	π	確率 p	pf	e =f - E[f]	pe^2
0	5	0.40	0.0778	0.000	-2.000	0.311
1	5	0.40	0.2592	0.259	-1.000	0.259
2	5	0.40	0.3456	0.691	0.000	0.000
3	5	0.40	0.2304	0.691	1.000	0.230
4	5	0.40	0.0768	0.307	2.000	0.307
5	5	0.40	0.0102	0.051	3.000	0.092
		計	1.0000	2.000		1.200
			必ず 1	E[f]		V[f]

BINOM.DIST
(f, n, π, FALSE)

期待度数 5×0.4 に一致

$$\mathrm{E}[f] = \sum pf = 2.0$$

$$\mathrm{V}[f] = \sum p(f - \mathrm{E}[f])^2 = \sum pe^2 = 1.2$$



2 項分布の期待値と分散

p.148

● f の期待値と分散

事例：f が n = 5 と π = 0.4 の 2 項分布に従う

期待値 E[f] は、出現する値（f）と、出現確率（p）との積の和

分散 V[f] は、出現する値（f）と期待値の差（e）の 2 乗と、出現確率（p）との積の和

（第 1 部 §1.2 ）

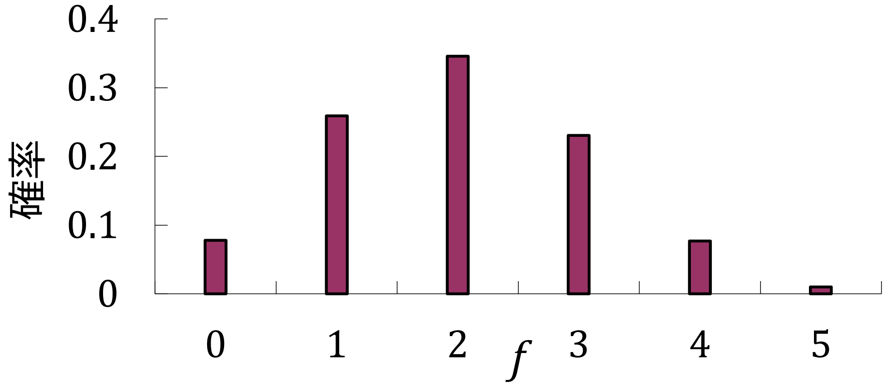
表示 3.1.11 $\pi = 0.4$,$n = 5$ の期待値と分散の計算

f	n	π	確率 p	pf	e = f - E[f]	pe^2
0	5	0.40	0.0778	0.000	-2.000	0.311
1	5	0.40	0.2592	0.259	-1.000	0.259
2	5	0.40	0.3456	0.691	0.000	0.000
3	5	0.40	0.2304	0.691	1.000	0.230
4	5	0.40	0.0768	0.307	2.000	0.307
5	5	0.40	0.0102	0.051	3.000	0.092
		計	1.0000	2.000		1.200
			必ず 1	E[f]		V[f]

BINOM.DIST
(f, n, π, FALSE)

$$\mathrm{E}[f] = \sum pf = 2.0$$

$$\mathrm{V}[f] = \sum p(f - \mathrm{E}[f])^2 = \sum pe^2 = 1.2$$



２項分布の期待値と分散

p.149

● f と p の期待値と分散（理論値）

２項分布に従う確率変数 f の期待値、分散、標準偏差 （$n = 5, \pi = 0.4$）

$$\mathrm{E}[f] = n\pi = 5 \times 0.4 = 2.0$$
$$\mathrm{V}[f] = n\pi(1-\pi) = 5 \times 0.4 \times (1-0.4) = 1.2$$
$$\mathrm{D}[f] = \sqrt{n\pi(1-\pi)} = \sqrt{1.2} = 1.10$$

割合（$p = f / n$）の期待値と分散（n 回の試行の中で f 回起こるときの割合）

p の期待値は f の期待値の $1/n$　$\mathrm{E}[p] = 2.0/5 = 0.4$

分散は f の分散の $1/n^2$　$\mathrm{V}[p] = 1.2/5^2 = 0.048$

標準偏差は分散の平方根　$D[p] = \sqrt{0.048} = 0.219$

$$\mathrm{E}[f] = n\pi \qquad \mathrm{V}[f] = n\pi(1-\pi) \qquad \mathrm{D}[f] = \sqrt{n\pi(1-\pi)} \tag{3.1.5}$$

$$\mathrm{E}[p] = \pi \qquad V[p] = \frac{\pi(1-\pi)}{n} \qquad \mathrm{D}[p] = \sqrt{\pi(1-\pi)/n} \tag{3.1.6}$$

割合の不偏推定

p.227

●割合 p の期待値は π（課題3.1.3）

n = 1 のとき、f = 1、f = 0 の確率は、式(3.1.2)より、

$$p_0 = {}_1C_0\pi^0(1-\pi)^{1-0} = 1\times 1\times(1-\pi) = 1-\pi$$

$$p_1 = {}_1C_1\pi^1(1-\pi)^{1-1} = 1\times\pi\times 1 = \pi$$

$$p_f = {}_nC_f\pi^f(1-\pi)^{n-f} \quad (3.1.2)$$

このときの f の期待値は、定義により、

$$\mathrm{E}[f] = \sum_{j=0}^{1} p_j f_j = p_0 f_0 + p_1 f_1 = (1-\pi)\times 0 + \pi\times 1 = \pi$$

n = 5 のとき、 n=1 の実験を 5 回繰り返すことと同じなので、各実験における度数を f_i とすると

$$f = f_1 + f_2 + f_3 + f_4 + f_5$$

$$\mathrm{E}[f] = \mathrm{E}[f_1] + \mathrm{E}[f_2] + \mathrm{E}[f_3] + \mathrm{E}[f_4] + \mathrm{E}[f_5] = \pi+\pi+\pi+\pi+\pi = 5\pi$$

実験回数を n 回とすると　$\mathrm{E}[f] = n\pi$

$p = f/n$ の期待値は f の期待値の $1/n$　$\mathrm{E}[p] = \mathrm{E}[f/n] = \mathrm{E}[f]/n = n\pi/n = \pi$　（3.1.6）

2 項分布の期待値と分散

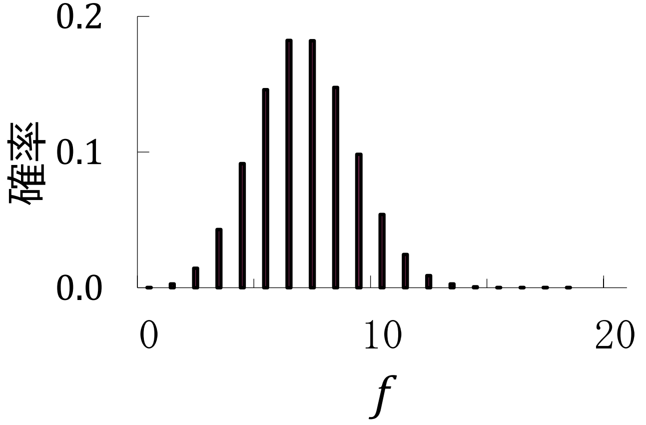
p.149

● f と p の期待値と分散（事例）

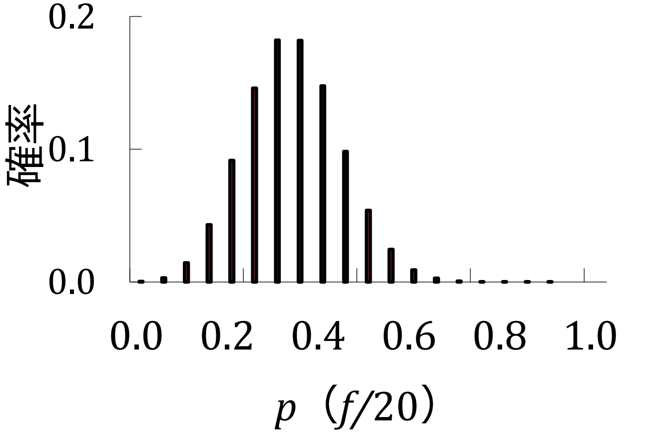
サイコロを20回（n）投げて、目の数 1 と 6 が出る回数 f とその割合 p（$f/20$）

（π = 1/3 = 0.333）

$$\mathrm{E}[f] = n\pi = 20 \times 0.333 = 6.667 \quad (3.1.5)$$
$$\mathrm{V}[f] = n\pi(1-\pi) = 20 \times 0.333 \times (1-0.333) = 4.444$$
$$\mathrm{D}[f] = \sqrt{n\pi(1-\pi)} = \sqrt{4.42} = 2.108$$



$$\mathrm{E}[p] = \pi = 0.333 \quad (3.1.6)$$
$$\mathrm{V}[p] = \frac{\pi(1-\pi)}{n} = \frac{0.333 \times (1-0.333)}{20} = 0.0111$$
$$\mathrm{D}[p] = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{0.0111} = 0.105$$



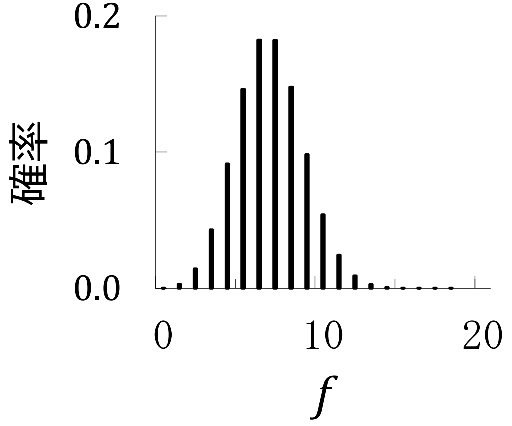
2 項分布の期待値と分散

p.149

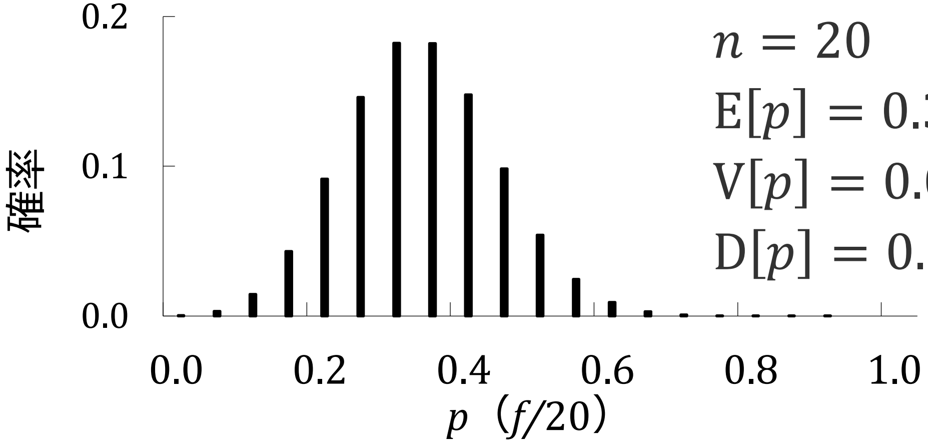
● f と p の期待値と分散（事例）

サイコロを20回または50回（n）投げて、目の数 1 と 6 が出る回数 fと割合 p（$f/20$, $f/50$）

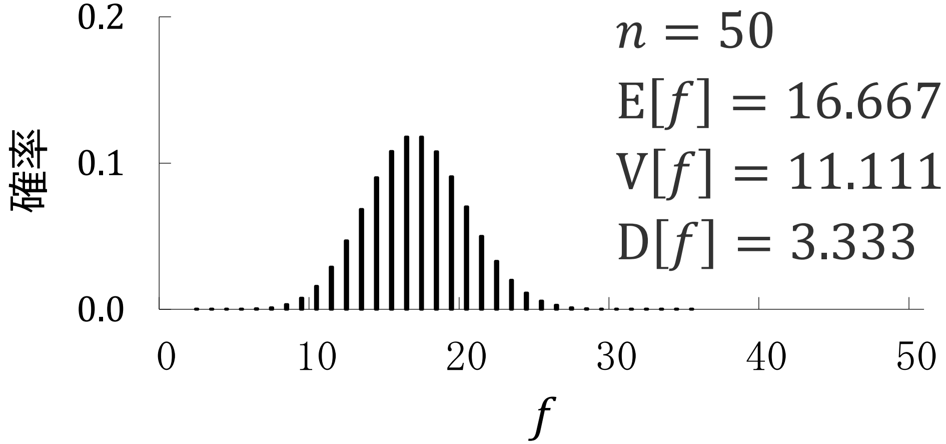
（π = 0.33



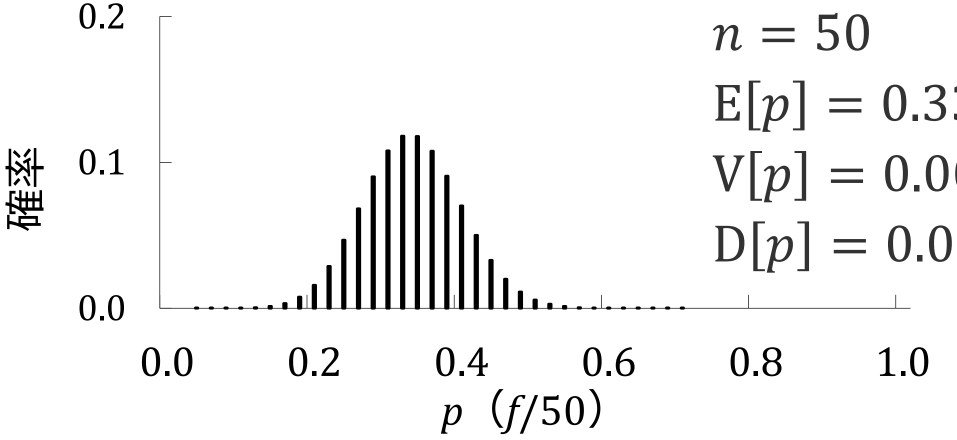
$n = 20$
$\mathrm{E}[f] = 6.667$
$\mathrm{V}[f] = 4.444$
$\mathrm{D}[f] = 2.108$



$n = 20$
$\mathrm{E}[p] = 0.333$
$\mathrm{V}[p] = 0.011$
$\mathrm{D}[p] = 0.105$



$n = 50$
$\mathrm{E}[f] = 16.667$
$\mathrm{V}[f] = 11.111$
$\mathrm{D}[f] = 3.333$



$n = 50$
$\mathrm{E}[p] = 0.333$
$\mathrm{V}[p] = 0.004$
$\mathrm{D}[p] = 0.067$

2 項分布の期待値と分散

p.149

● p の分散・標準偏差と π の関係

式(3.1.6)の曲線の形は
n によって変化しないので、
n=1 の条件で比較
（$n > 1$ の場合、左の縦軸をn で割る
右の縦軸を$\sqrt{n}$ で割る）

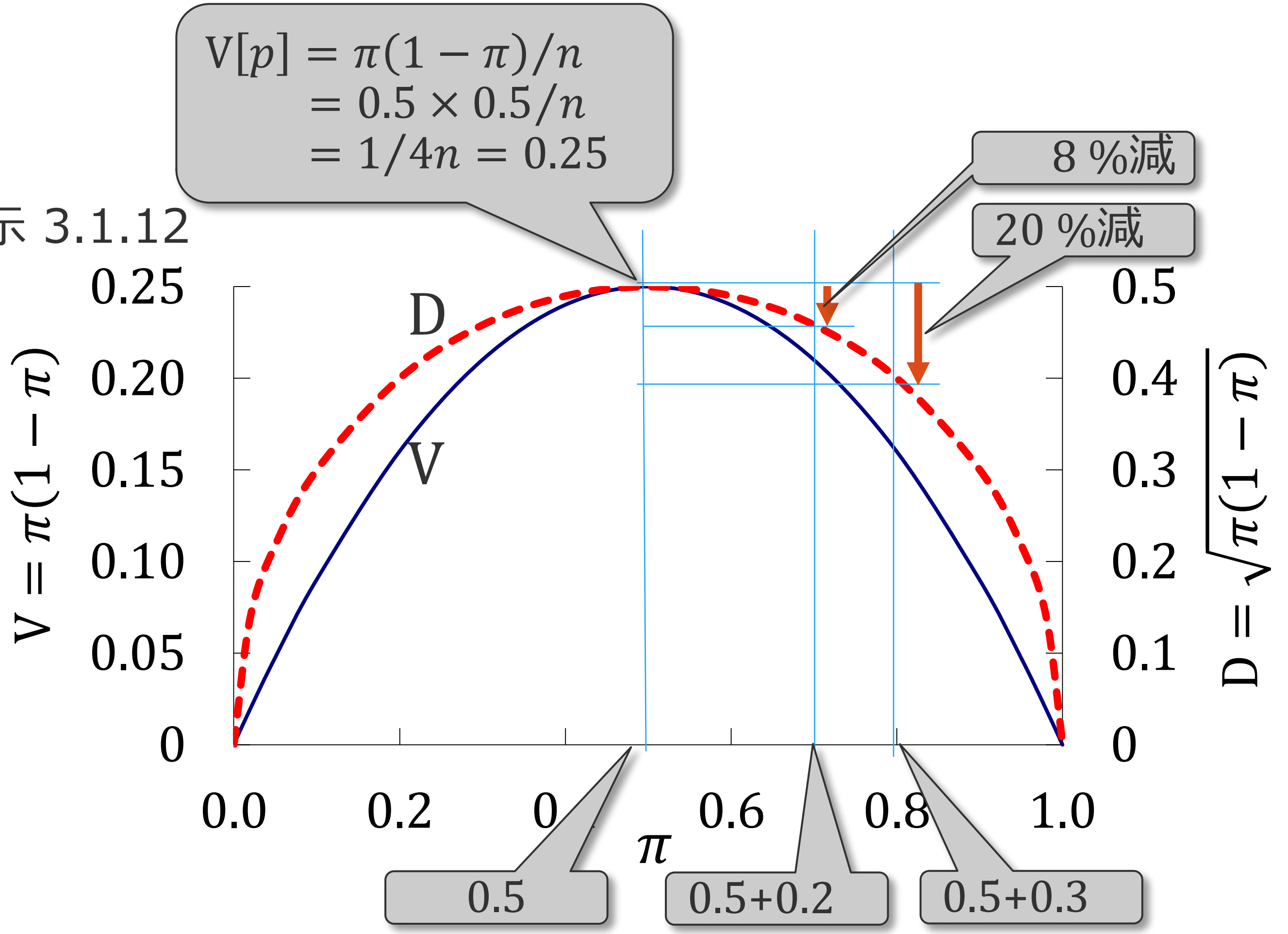
p の標準偏差は$0.1 < \pi < 0.9$ の範囲内で
あまり変化しない

$\pi = 0.5 \pm 0.2$ で約8% 減

$\pi = 0.5 \pm 0.3$ で約20% 減

この範囲の外では急激に減少

表示 3.1.12



$$\mathrm{E}[p] = \pi \qquad V[p] = \frac{\pi(1-\pi)}{n} \qquad \mathrm{D}[p] = \sqrt{\pi(1-\pi)/n} \qquad (3.1.6)$$

（8）２項分布の正規近似

２項分布を正規分布に近似させて検定・推定に利用

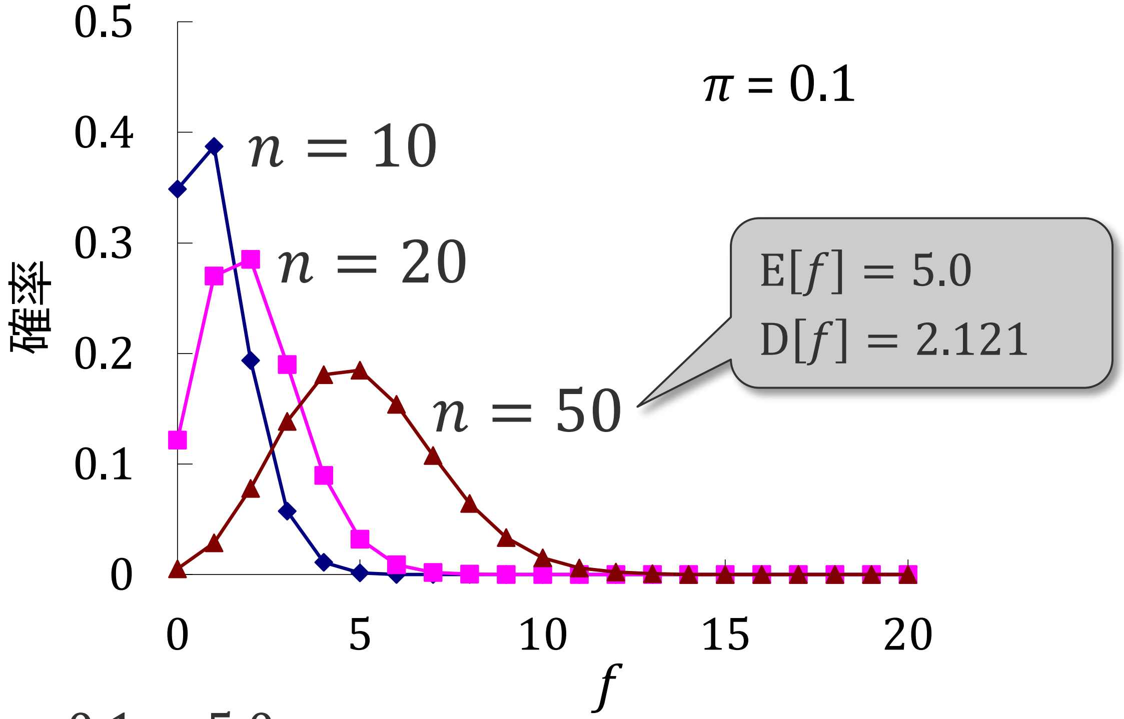
2 項分布の正規近似

p.149

● 2 項分布の特性

2 項分布は n が大きくなると正規分布に近づく（中心極限定理、第 1 部 §1.3 ）

正規分布に近似させて検定・区間推定で利用



事例

$\pi = 0.1$、$n = 50$ の 2 項分布

（$\mathrm{E}[f] = 5.0$、$\mathrm{D}[f] = 2.121$）

↓

正規分布に近似

（μ=5.0、σ=2.121）

$\pi = 0.1$

$n = 50$

$$\mathrm{E}[f] = n\pi = 50 \times 0.1 = 5.0$$

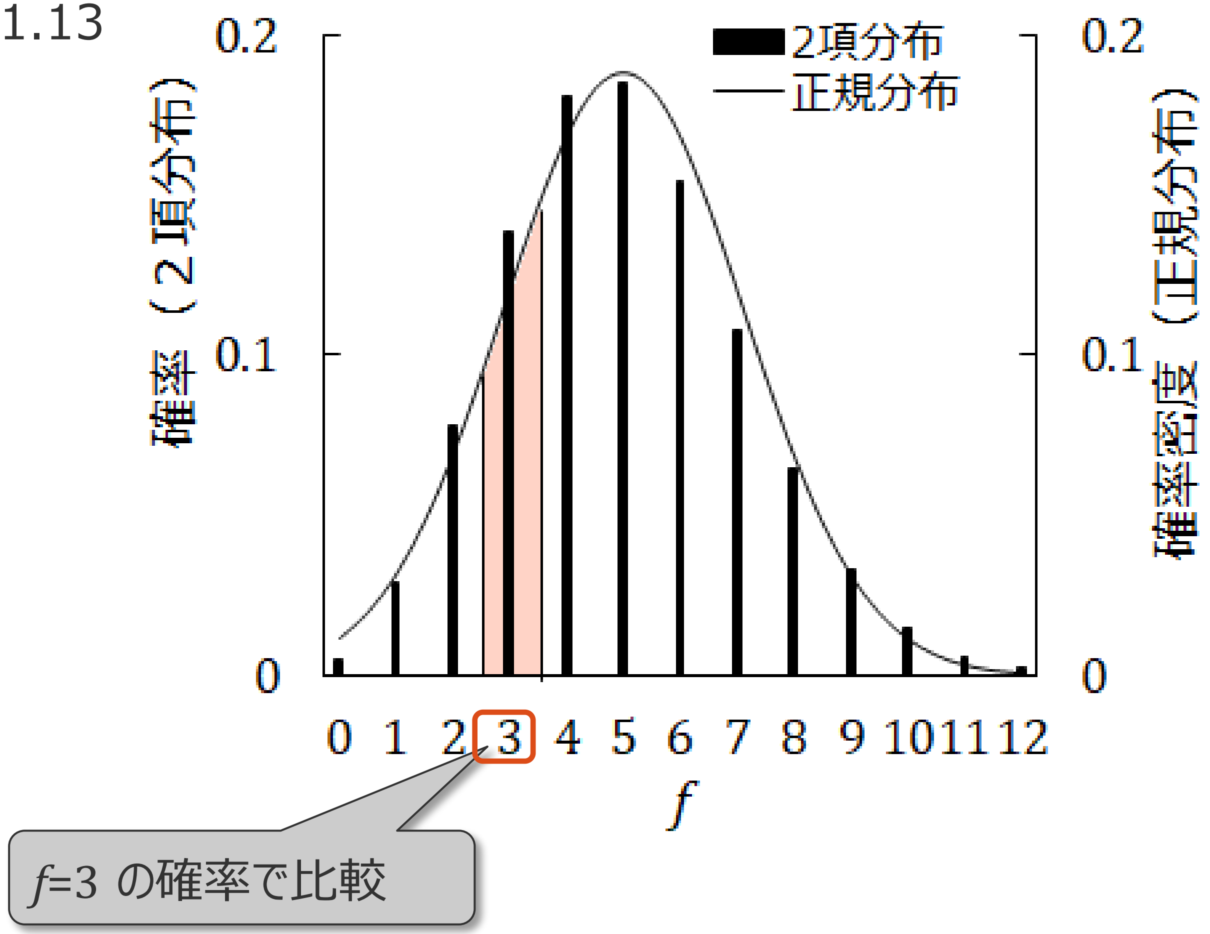
$$\mathrm{D}[f] = \sqrt{n\pi(1-\pi)} = \sqrt{50 \times 0.1 \times (1-0.1)} = \sqrt{4.5} = 2.121$$

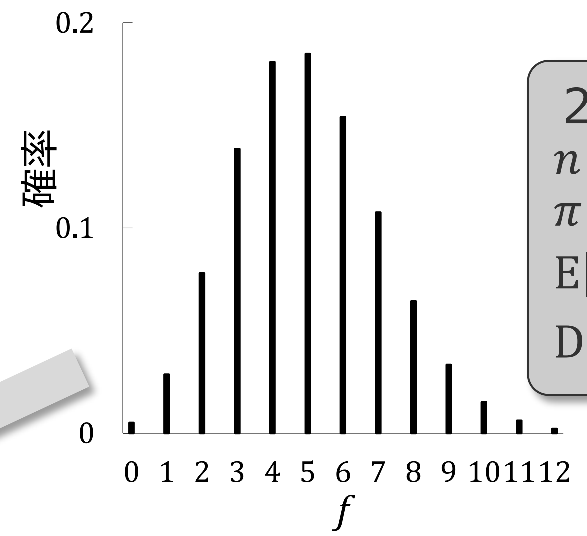
2 項分布の正規近似

p.150

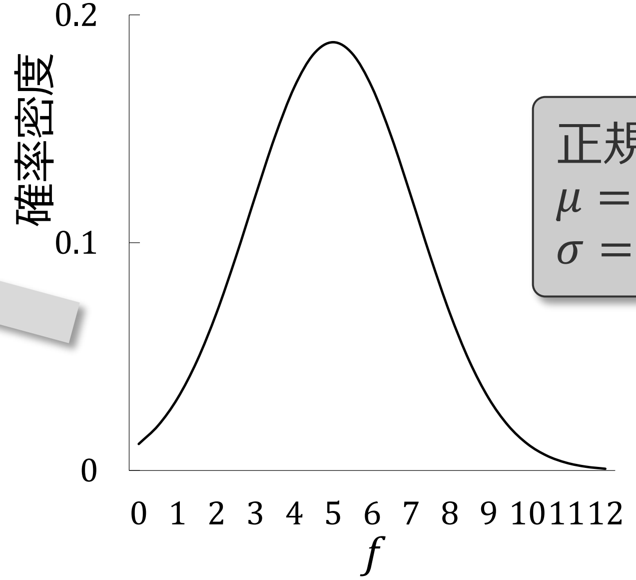
● 2 項分布と正規分布

表示 3.1.13





2 項分布
$n = 50$
$\pi = 0.1$
$\mathrm{E}[f] = 5.0$
$\mathrm{D}[f] = 2.121$

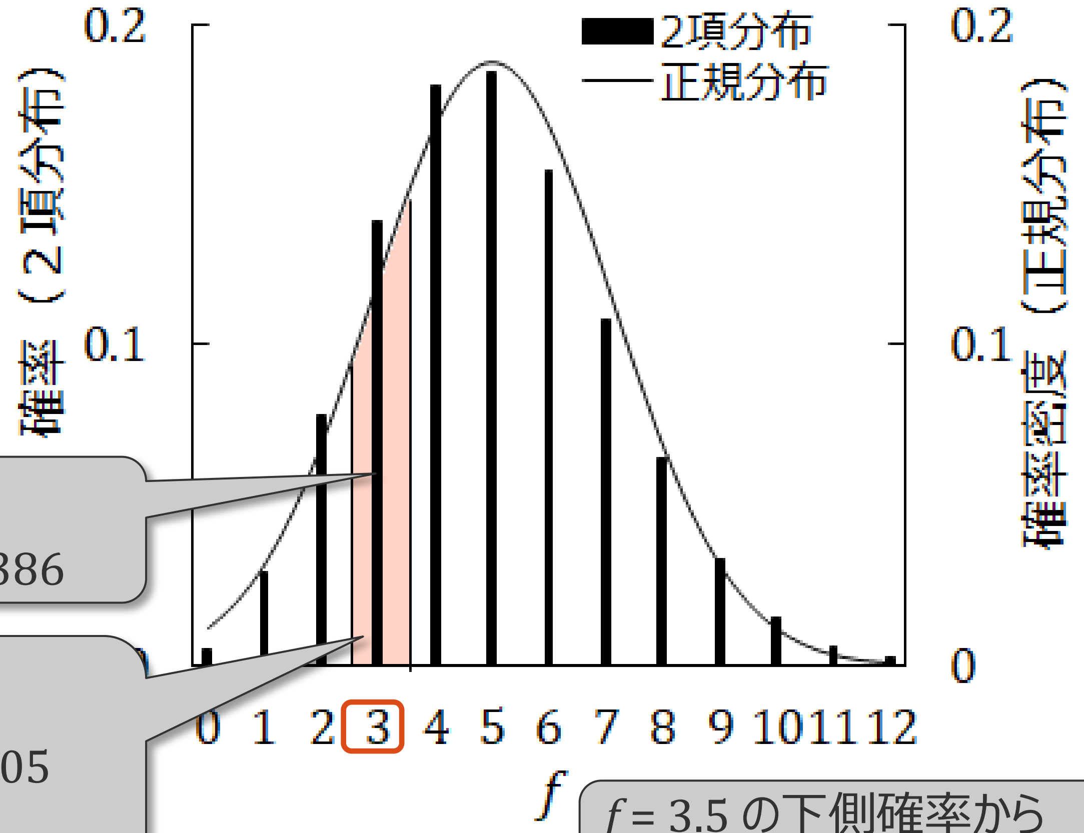


正規分布
$\mu = 5.0$
$\sigma = 2.121$

2 項分布の正規近似

●確率の正規近似

表示 3.1.13



2 項分布
棒の高さ=0.1386

正規分布：
面積=0.1205
f = [2.5, 3.5]

f = 3.5 の下側確率から
f = 2.5 の下側確率を引く

2 項分布

$\pi = 0.1$　　$\mathrm{E}[f] = 5.0$

$n = 50$　　$\mathrm{V}[f] = 2.121$

確率・・・縦軸

=BINOMDIST(f, n, π, FALSE)

=BINOMDIST(3, 50, 0.1, FALSE)=0.1386

正規分布

μ =5.0　σ = 2.121　　（第 1 部§1.3）

確率密度・・・縦軸

=NORMDIST(f, 5.0, 2.121, FALSE)

確率・・・面積

=NORMDIST(f + 0.5, μ, σ, TRUE)

－NORMDIST(f - 0.5, μ, σ, TRUE)

=NORMDIST(3+0.5, 5.0, 2.121, TRUE)

－NORMDIST(3-0.5, 5.0, 2.121, TRUE)

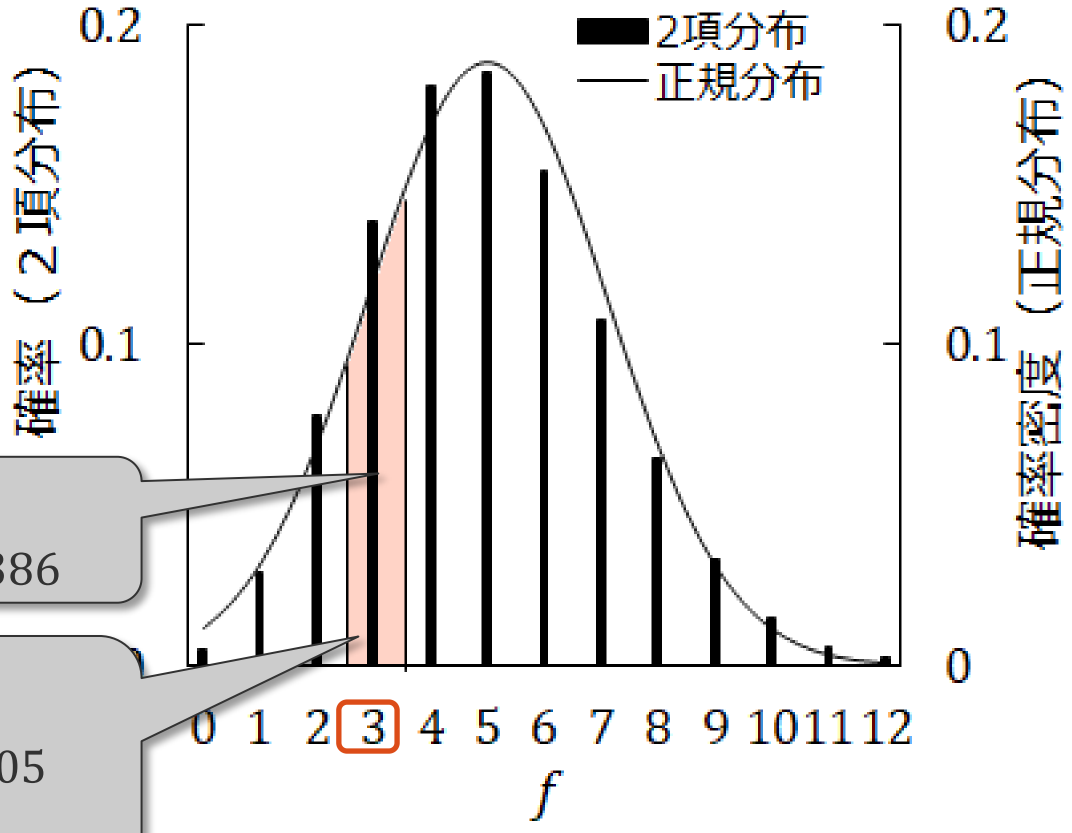
= 0.2397 - 0.1193 = 0.1205

2 項分布の正規近似

p.150

●確率の正規近似

表示 3.1.13



2 項分布
棒の高さ=0.1386

正規分布：
面積=0.1205
$f = [2.5, 3.5]$

n = 50
π = 0.10

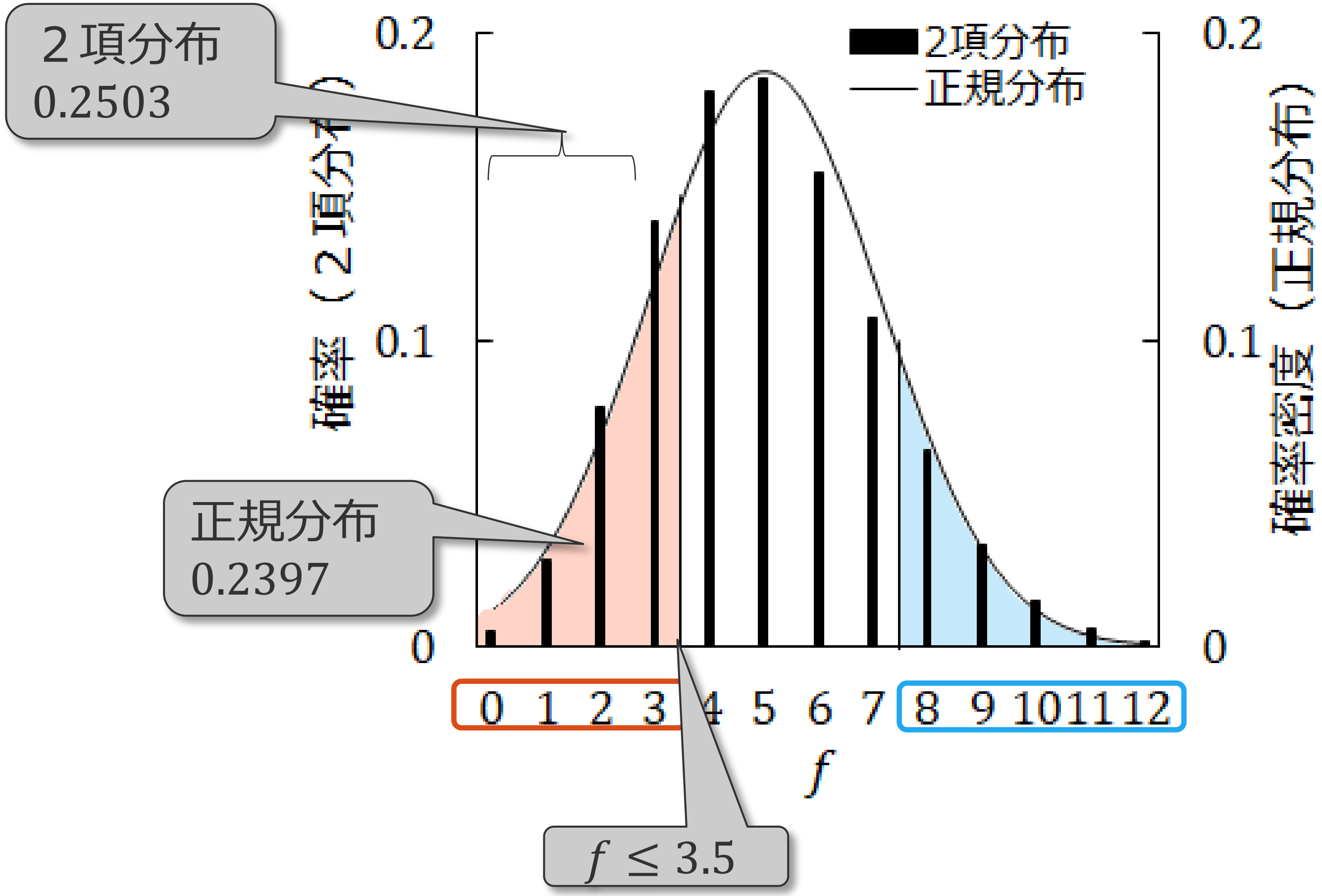
f	2項分布	正規分布	差	累積差
0	0.0052	0.0169	0.0118	0.0118
1	0.0286	0.0325	0.0039	0.0157
2	0.0779	0.0698	-0.0081	0.0076
3	**0.1386**	**0.1205**	**-0.0181**	-0.0105
4	0.1809	0.1671	-0.0138	-0.0244
5	0.1849	0.1863	0.0014	
6	0.1541	0.1671	0.0130	-0.0230
7	0.1076	0.1205	0.0128	-0.0100
8	0.0643	0.0698	0.0055	0.0028
9	0.0333	0.0325	-0.0008	0.0084
10	0.0152	0.0122	-0.0030	0.0076
11	0.0061	0.0037	-0.0025	0.0046
12	0.0022	0.0009	-0.0013	0.0021
期待値	5.0000			
標準偏差	2.1213			

２項分布の正規近似

p.151

● 累積確率の正規近似

$f \leq 3$ と $f \geq 8$ の累積確率の正規近似



２項分布
0.2503

正規分布
0.2397

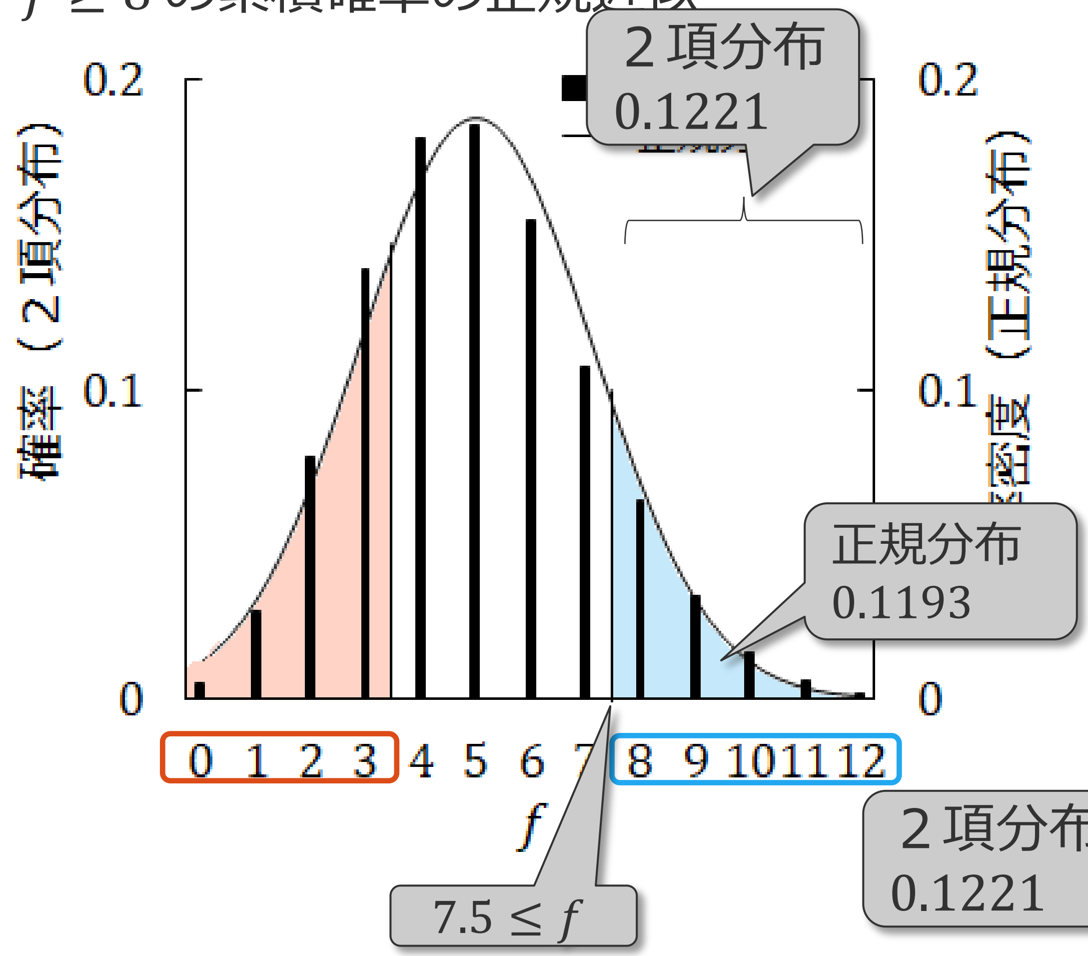
$n = 50$
$\pi = 0.10$

f	2項分布	正規分布	差	累積差
0	0.0052	0.0169	0.0118	0.0118
1	0.0286	0.0325	0.0039	0.0157
2	0.0779	0.0698	-0.0081	0.0076
3	0.1386	0.1205	-0.0181	-0.0105
4	0.1809	0.1671	-0.0138	-0.0244
5	0.1849	0.1863	0.0014	
6	0.1541	0.1671	0.0130	-0.0230
7	0.1076	0.1205	0.0128	-0.0100
8	0.0643	0.0698	0.0055	0.0028
9	0.0333	0.0325	-0.0008	0.0084
10	0.0152	0.0122	-0.0030	0.0076
11	0.0061	0.0037	-0.0025	0.0046
12	0.0022	0.0009	-0.0013	0.0021
期待値	5.0000			
標準偏差	2.1213			

2 項分布の正規近似

●累積確率の正規近似

$f \leq 3$ と $f \geq 8$ の累積確率の正規近似



$n = 50$

$\pi = 0.10$

f	2項分布	正規分布	差	累積差
0	0.0052	0.0169	0.0118	0.0118
1	0.0286	0.0325	0.0039	0.0157
2	0.0779	0.0698	-0.0081	0.0076
3	0.1386	0.1205	-0.0181	-0.0105
4	0.1809	0.1671	-0.0138	-0.0244
5	0.1849	0.1863	0.0014	
6	0.1541	0.1671	0.0130	-0.0230
7	0.1076	0.1205	0.0128	-0.0100
8	0.0643	0.0698	0.0055	0.0028
9	0.0333	0.0325	-0.0008	0.0084
10	0.0152	0.0122	-0.0030	0.0076
11	0.0061	0.0037	-0.0025	0.0046
12	0.0022	0.0009	-0.0013	0.0021
期待値	5.0000			
標準偏差	2.1213			

2 項分布
0.1221

正規分布
0.1193

2項分布の正規近似

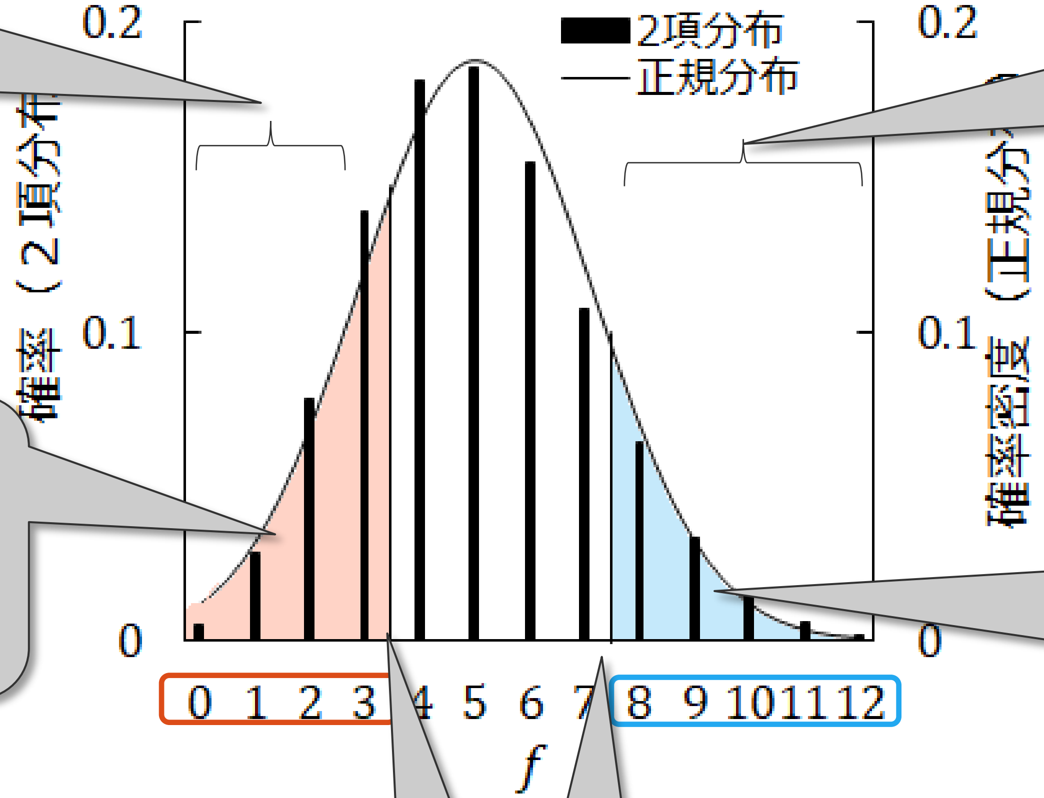
p.151

●累積確率の正規近似

$f \leq 3$ と $f \geq 8$ の累積確率の正規近似

2項分布
=BINOM.DIST
(3, 50, 0.1, TRUE)
=0.2503

正規分布
=NORM.DIST
(3.5, 5, 2.121, TRUE)
=0.2397



$f \leq 3.5$

$7.5 \leq f$

2項分布
= 1 − BINOM.DIST
(8-1, 50, 0.1, TRUE)
=0.1221

正規分布
=1 − NORM.DIST
(7.5, 5, 2.121, TRUE)
= 0.1193

2 項分布の正規近似

p.151

●確率と累積確率の正規近似

確率

2 項分布の f の確率に対して、正規分布の $f \pm 0.5$ の範囲の確率で近似

累積確率

2 項分布の f の上側確率に対して、正規分布の $f - 0.5$ の上側確率で近似

2 項分布の f の下側確率に対して、正規分布の $f + 0.5$ の下側確率で近似

連続修正：f が離散的に変化するとき、確率を求める範囲を0.5だけ広くとること

（ただし、連続修正を「する」あるいは「しない」の議論がある）

近似精度の精度

$n\pi$ が小さい（または $n(1–\pi)$ が小さい）場合，2 項分布が対称分布から大きく外れるので正規近似の精度は悪い（$n\pi \geq 5,\ n(1-\pi) \geq 5$ の場合に正規近似が使えるとされている）

（標準正規分布を使った計算、テキスト）

$f \pm 0.5$ を、偏差値 $u = -1.1785$、-0.7071 に変換して、標準正規分布に近似させている

$$f = 2.5, u = (2.5 - 5.0)/2.121 = -1.1785$$
$$f = 3.5, u = (3.5 - 5.0)/2.121 = -0.7071$$

p.151

（9）確率と尤度

同じものを別の角度から眺めたもの

確率と尤度

p.151

●確率、尤度

表示 3.1.14
2 項分布の確率と尤度

n = 10

(1) 縦方向

(2) 横方向

f	π .05	.10	.15	.20	.25	.30	.35	.40
0	.60	.35	.20	.11	.06	.03	.01	.01
1	.32	.39	.35	.27	.19	.12	.07	.04
2	.07	.19	.28	.30	.28	.23	.18	.12
3	.01	.06	.13	.20	.25	.27	.25	.21
4	.00	.01	.04	.09	.15	.20	.24	.25

=BINOM.DIST（f, 10, π, FALSE)

n = 10 のとき

(1) π が 0.20 である場合（縦方向）→ **確率**

f が 2 のとき確率は 0.30

f が 3 のとき確率は 0.20

(2) f が 3 である場合（横方向）

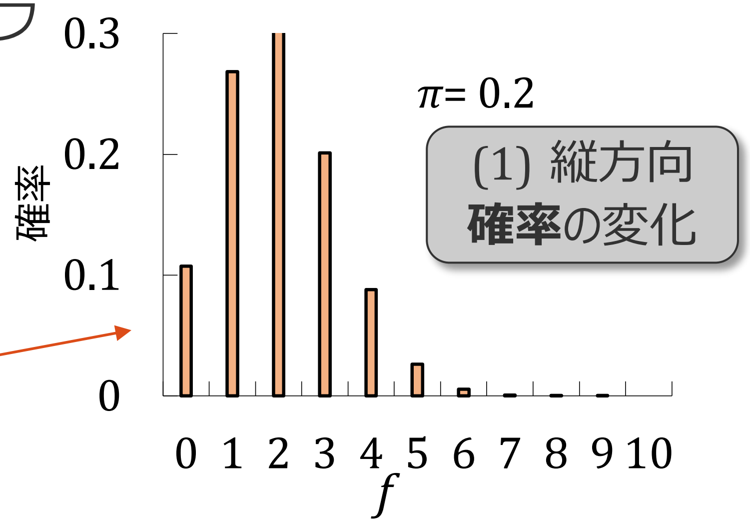
π が 0.2 のとき尤度は 0.20 → **尤度**・・・尤もらしさ（もっともらしさ）

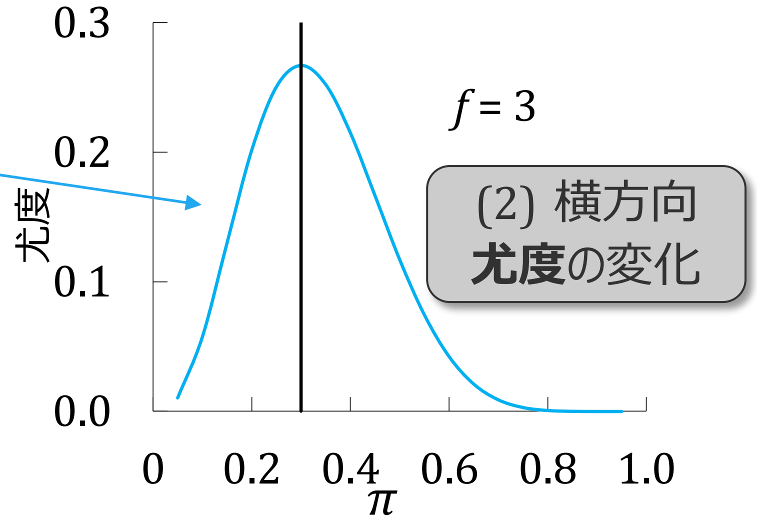
π が 0.3 のとき尤度は 0.27

確率と尤度

●確率、尤度

表示 3.1.15 確率と尤度のグラフ





表示 3.1.14 2 項分布の確率と尤度

(1) 縦方向

n = 10

f \ π	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50	.55	.60	.65	.70	.75	.80	.85	.90	.95
0	.60	.35	.20	.11	.06	.03	.01	.01	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
1	.32	.39	.35	.27	.19	.12	.07	.04	.02	.01	.00	.00	.00	.00	.00	.00	.00	.00	.00
2	.07	.19	.28	.30	.28	.23	.18	.12	.08	.04	.02	.01	.00	.00	.00	.00	.00	.00	.00
3	.01	.06	.13	.20	.25	.27	.25	.21	.17	.12	.07	.04	.02	.01	.00	.00	.00	.00	.00
4	.00	.01	.04	.09	.15	.20	.24	.25	.24	.21	.16	.11	.07	.04	.02	.01	.00	.00	.00
[illegible]	.00	.00	.01	.03	.06	.10	.15	.20	.23	.25	.23	.20	.15	.10	.06	.03	.01	.00	.00
[illegible]	[illegible]	[illegible]	.00	.01	.02	.04	.07	.11	.16	.21	.24	.25	.24	.20	.15	.09	.04	.01	.00
7	.00	.00	.00	.00	.00	.01	.02	.04	.07	.12	.17	.21	.25	.27	.25	.20	.13	.06	.01
8	.00	.00	.00	.00	.00	.00	.00	.01	.02	.04	.08	.12	.18	.23	.28	.30	.28	.19	.07
9	.00	.00	.00	.00	.00	.00	.00	.00	.00	.01	.02	.04	.07	.12	.19	.27	.35	.39	.32
10	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.01	.01	.03	.06	.11	.20	.35	.60

(2) 横方向

確率と尤度

p.152

●尤度と対数尤度

尤度（Likelihood）

πである尤もらしさを示す

対数尤度

尤度の自然対数を取った値

「－対数尤度」（テキストでは「対数尤度」）

事例

n =10、f = 3、π は不明、f が 2 項分布に従うとき

π を変えると尤度が変わる

π = 0.2 のとき　尤度：0.20、対数尤度：-ln(0.20) = 1.61

π = 0.3 のとき　尤度：0.27、対数尤度：-ln(0.27) = 1.31

尤度は 2 項分布に限って使われるものではない

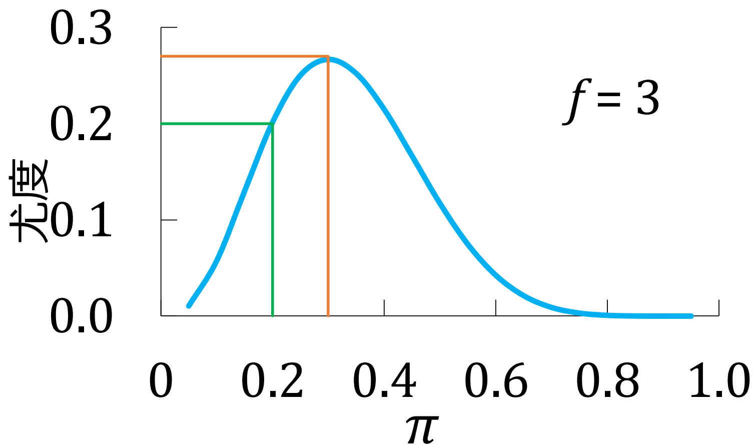
最尤推定や検定に利用、使い方は次節以降で順次説明

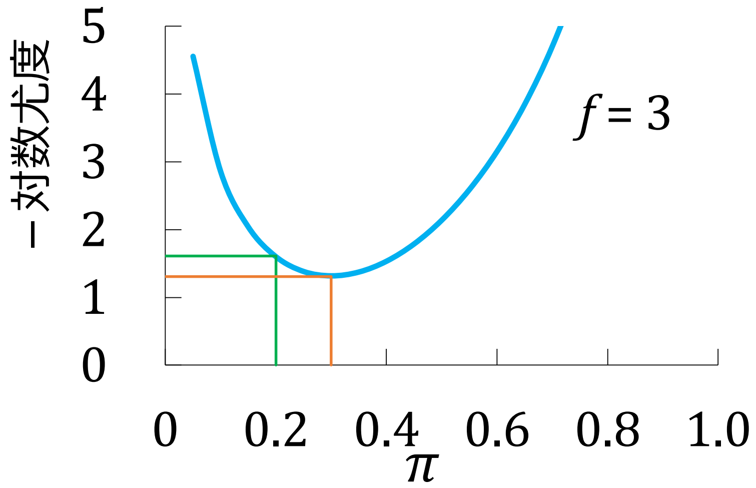
表示 3.1.14　2 項分布の確率と尤度

n = 10

	π						
f	.05	.10	.15	.20	.25	.30	.35
2	.07	.19	.28	.30	.28	.23	.18
3	.01	.06	.13	.20	.25	.27	.25
4	.00	.01	.04	.09	.15	.20	.24

表示 3.1.15 確率と尤度グラフ





ポアソン分布（補遺）

p.221

●概要

２項分布で、πn を一定にして、n を大きくした極限として導かれる

試行回数 n において出現数 f の平均（期待値）は $\mu = \pi n$、標準偏差は$\sqrt{\mu}$

一定の長さの時間、一定の大きさの空間において、ごくまれに起こる事象に適用

　　ある地域で起こる交通事故件数など

$\mu > 5$ のポアソン分布は、正規分布に近似できる

● 事例

ある疾病の発生率は、全国平均で1/8,000であるとする。

ある地域で20,000人を調査したところ、この疾病患者は見出されなかった。

この地域は全国平均よりも発生率が低いといえるか。

$\mu = \pi n = 20000 \times 1/8000 = 2.50$,　$f = 0$ の確率は

= POISSON.DIST（f，μ，オプション）= POISSON.DIST（0，2.50，TRUE）=0.082

この地域の発病率は全国の発病率と有意に異なるとはいえない（α=0.05 片側検定）

まとめ

p.152

- 質的変数と計数値（離散変数）

- ２項分布

  計数値の代表的な確率分布

  ２値データ

  生起確率 π の事象が、n 回の試行で f 回起こる、割合は f/n

- 尤度

  「確率」と「尤度」は、同じものを別の角度から眺めたもの

  「最小２乗法」と共に、統計解析の柱である「最尤法」の基本となるもの

● 引用文献

岩崎学・吉田清隆（2005）計量生物学 26：53–63.

● 作成　片瀬雅彦

● 監修　松本一彦、長谷文雄

● 作成時期　2020年7月2日

● 改訂　2021年3月3日、2022年8月1日