

体験統計学

～第10回～

[本稿のWebページ](#)

古橋 武

小テスト9.1 解答

マウス10匹に対する投与実験を行い、薬の効き具合 p と薬の効き具合の標準偏差の近似値

$$\hat{\sigma} = \sqrt{p(1-p)/(n-1)}$$

を求めよ。これを5000回繰り返したときの p の平均値と $\hat{\sigma}$ の平均値を求めよ。

p の平均値が真の有効率 p_0 に近い値となり、

$\hat{\sigma}$ の平均値が真の標準偏差 $\sigma = \sqrt{p_0(1-p_0)/n}$ に近い値となることを確認せよ。

有効率_5000試行のσの近似値の分布_(n=10).xlsx - Microsoft Excel

ホーム 挿入 ページレイアウト 数式 データ 校閲 表示 開発 アドイン Help Acrobat チーム

貼り付け MS Pゴシック 11 標準 挿入 削除 書式 並べ替えと検索とフィルタ

クリップボード フォント 配置 数値 セル 編集

123

1 薬の効き具合

2

3 薬の効く確率 p₀ = 0.5

4 マウスの数 n = 10 匹

5 真の標準偏差 σ = 0.1581139

6

7 薬を投与した結果 第1組 薬を投与した結果 第2組 薬を投与した結果 第3組 薬を投与した結果 第4組 薬を投与した結果 第5組 薬を投与した結果 第6組 薬を投与した結果 第7組 薬を投与した結果 第8組 薬を投与した結果 第9組 薬を投与した結果 第10組

8 0 1 0 1 0 1 1 0 1 0

9 0 0 0 1 1 0 1 1 1 0

10 1 0 0 0 0 1 0 1 1 1

11 0 0 1 0 1 0 0 0 0 0

12 0 1 1 1 0 1 1 0 0 0

13 1 1 1 0 0 0 0 0 0 0

14 0 0 0 0 0 0 0 0 0 0

15 0 1 1 1 1 1 1 1 1 1

16 0 1 1 0 0 0 0 0 0 0

17 1 1 0 1 0 0 0 0 0 0

18

19 p値 0.3 0.5 0.4 0.6 0.4 0.6 0.3 0.6 0.4

20 σの近似値 0.15 0.16 0.17 0.16 0.16 0.16 0.15 0.16 0.16

21

22 p値の平均値 0.49694

23 σの近似値の平均値 0.15621

Sheet1 Sheet2 Sheet3

コマンド

$$\sigma = \sqrt{p_0(1-p_0)/n}$$

$$=SQRT(C3*(1-C3)/C4)$$

$$\hat{\sigma} = \sqrt{p(1-p)/(n-1)}$$

$$=ROUND(SQRT(B19*(1-B19)/(C4-1)),2)$$

$$=AVERAGE(B19:GJ19)$$

$$=AVERAGE(B20:GJ20)$$

有効率_5000試行のσの近似値の分布_(n=10).xlsx - Microsoft Excel

ホーム 挿入 ページレイアウト 数式 データ 校閲 表示 開発 アドイン Help Acrobat チーム

MS Pゴシック 11

標準 挿入 Σ 削除 書式 並べ替えと検索と選択

① 真の有効率 p_0 を設定して

	A	B	C	D	E	F	G	H	I	J	K
1	薬の効き具合										
2											
3	薬の効く確率	$p_0 =$	0.5								
4	マウスの数	$n =$	10 匹								
5	真の標準偏差	$\sigma =$	0.1581139								
6		薬を投与した結果	薬を投与した結果	薬を投与した結果	薬を投与した結果	薬を投与した結果	薬を投与した結果	薬を投与した結果	薬を投与した結果	薬を投与した結果	薬を投与した結果
7		第1組	第2組	第3組	第4組	第5組	第6組	第7組	第8組	第9組	第10組
8		0	1	0	1	0	1	1	1	0	
9		0									
10		1									
11		0									
12		0									
13		1									
14		0									
15		0									
16		0	1	1	1	1	0	1	0	0	0
17		1	1	0	0	0	1	0	0	1	0
18											
19	p値	0.3	0.6	0.5	0.4	0.6	0.4	0.6	0.3	0.6	0.4
20	σの近似値	0.15	0.16	0.17	0.16	0.16	0.16	0.16	0.15	0.16	0.16
21											
22	p値の平均値	0.49694									
23	σの近似値の平均値	0.15621									

② =INT(RAND()+ p_0)により有効率 p_0 に応じて薬の有効(1), 無効(0)をシミュレーションする.

③ 有効率 p を求める

④ σの近似値を求める

我々は

しかない。

有効率_5000試行のσの近似値の分布_(n=10).xlsx - Microsoft Excel

ホーム 挿入 ページレイアウト 数式 データ 校閲 表示 開発 アドイン Help Acrobat チーム

MS Pゴシック 11

標準

挿入 削除 書式

並べ替えと検索とフィルタ 選択

クリップボード

貼り付け

フォント

配置

I23

薬の効き具合

薬の効く確率 $p_0 = 0.5$

マウスの数 $n = 10$ 匹

真の標準偏差 $\sigma = 0.1581139$

	薬を投与した結果 第1組	薬を投与した結果 第2組	薬を投与した結果 第3組	薬を投与した結果 第4組	薬を投与した結果 第5組	薬を投与した結果 第6組	薬を投与した結果 第7組	薬を投与した結果 第8組	薬を投与した結果 第9組	薬を投与した結果 第10組
7	0	1	0	1	0	1	1	1	0	
8	0									
9	0									
10	1									
11	0									
12	0									
13	1									
14	0									
15	0									
16	0	1	1	1	1	0	1	0	0	0
17	1	1	0	0	0	1	0	0	1	0
18										
19	p値	0.3	0.6	0.5	0.4	0.6	0.4	0.6	0.3	0.6
20	σの近似値	0.15	0.16	0.17	0.16	0.16	0.16	0.16	0.15	0.16
21										
22	p値の平均値	0.49694								
23	σの近似値の平均値	0.15621								

コマンド

神のみぞ知る

① 真の有効率 p_0 を設定して

我々は
しかない。

② $=INT(RAND()+ p_0)$ により有効率 p_0 に応じて薬の有効(1), 無効(0)をシミュレーションする.

③ 有効率 p を求める

④ σ の近似値を求める

神のみぞ知る

我々は得られた結果から p_0 を推定するしかない。

① 真の有効率 p_0 を設定して

1	薬の効き具合										
2											
3	薬の効く確率	$p_0 =$	0.5								
4	マウスの数	$n =$	10匹								
5	真の標準偏差	$\sigma =$	0.1581139								
6											
7		薬を投与した結果	薬を投与した結果	薬を投与した結果	薬を投与した結果	薬を投与した結果	薬を投与した結果	薬を投与した結果	薬を投与した結果	薬を投与した結果	薬を投与した結果
8		第1組	第2組	第3組	第4組	第5組	第6組	第7組	第8組	第9組	第10組

8	0
9	0
10	1
11	0
12	0
13	1
14	0
15	0
16	0
17	1

② $=INT(RAND()+ p_0)$ により有効率 p_0 に応じて薬の有効(1), 無効(0)をシミュレーションする。

③ 有効率 p を求める

④ σ の近似値を求める

19	p値	0.3	0.6	0.5	0.4	0.6	0.4	0.6	0.3	0.6	0.4
20	σ の近似値	0.15	0.16	0.17	0.16	0.16	0.16	0.16	0.15	0.16	0.16
21											
22	p値の平均値	0.49694									
23	σ の近似値の平均値	0.15621									

小テスト9.2 解答

真の有効率 p_0 の 95% 信頼区間

今、10匹のマウスにある薬を投与して、6匹に効き目があったとしよう

$$p - 1.96\sqrt{\frac{p(1-p)}{n-1}} \leq p_0 \leq p + 1.96\sqrt{\frac{p(1-p)}{n-1}}$$

$$\frac{6}{10} - 1.96\sqrt{\frac{\frac{6}{10}\left(1 - \frac{6}{10}\right)}{10-1}} \leq p_0 \leq \frac{6}{10} + 1.96\sqrt{\frac{\frac{6}{10}\left(1 - \frac{6}{10}\right)}{10-1}}$$

$$0.28 \leq p_0 \leq 0.92$$

小テスト9.2 解答

真の有効率 p_0 の 95% 信頼区間

10匹のマウスにある薬を投与して、6匹に効き目があったとしよう

真の有効率は95%の確率で

$$0.28 < p_0 < 0.92$$

の間にある.

100匹のマウスにある薬を投与して、60匹に効き目があったとしよう

真の有効率は95%の確率で

$$0.503 < p_0 < 0.697$$

の間にある.

実験のサンプルサイズが大きくなると、信頼区間は狭まる. →

小テスト9.2 解答

真の有効率 p_0 の 95% 信頼区間

10匹のマウスにある薬を投与して、6匹に効き目があったとしよう

真の有効率は95%の確率で

$$0.28 < p_0 < 0.92$$

の間にある.

100匹のマウスにある薬を投与して、60匹に効き目があったとしよう

真の有効率は95%の確

$$0.503 < p_0 < 0.697$$

の間にある.

実験のサンプルサイズが大きくなると、信頼区間は狭まる. → **有効率の推定値の信頼度が高まる.**

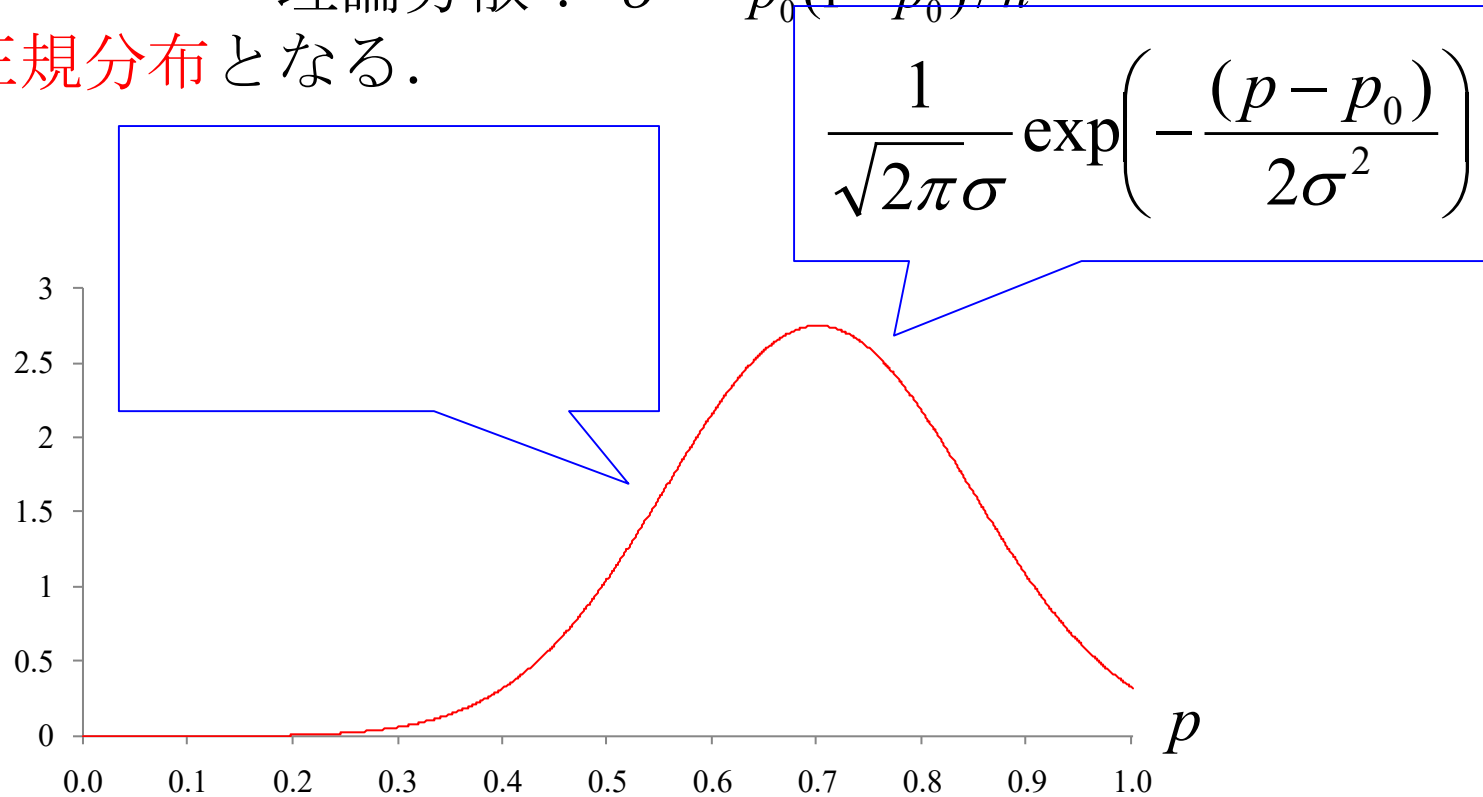
確率分布

n 匹のマウスへの薬の効き具合 \bar{p} の分布は

理論平均： p_0

理論分散： $\sigma^2 = p_0(1-p_0)/n$

の正規分布となる。



10匹のマウスに薬を投与して7匹に効けば、薬の効き具合 p は

確率分布

n匹のマウスへの薬の効き具合の分布は

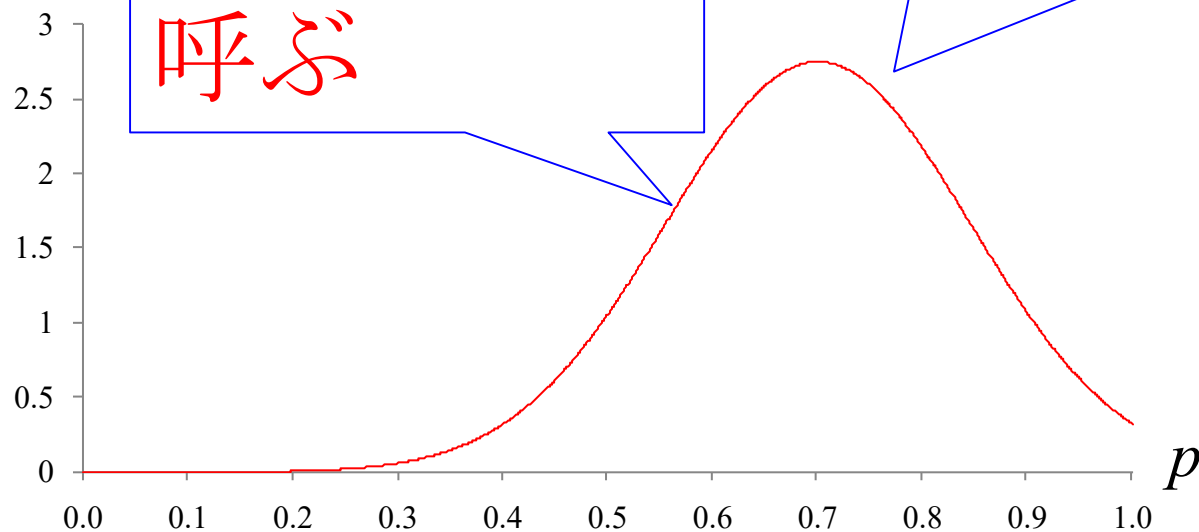
理論平均： p_0

理論分散： $\sigma^2 = p_0(1-p_0)/n$

の正規分布となる。

確率分布と
呼ぶ

$$\frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(p-p_0)^2}{2\sigma^2}\right)$$



10匹のマウスに薬を投与して7匹に効けば、薬の効き具合 p は

確率分布

n匹のマウスへの薬の効き具合の分布は

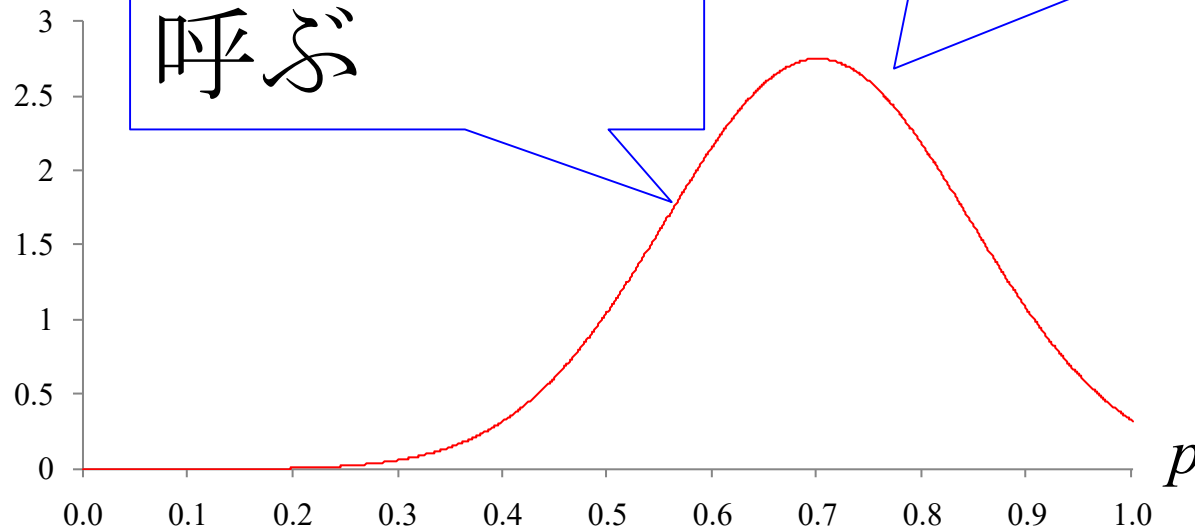
理論平均： p_0

理論分散： $\sigma^2 = p_0(1-p_0)/n$

の正規分布となる。

確率分布と
呼ぶ

$$\frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(p-p_0)^2}{2\sigma^2}\right)$$



10匹のマウスに薬
の効き具合 p は

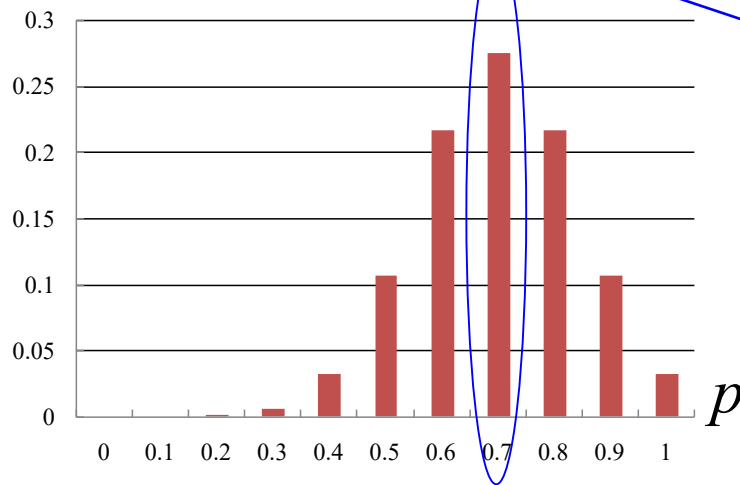
$$p = \frac{7}{10} = 0.7$$

けば、薬

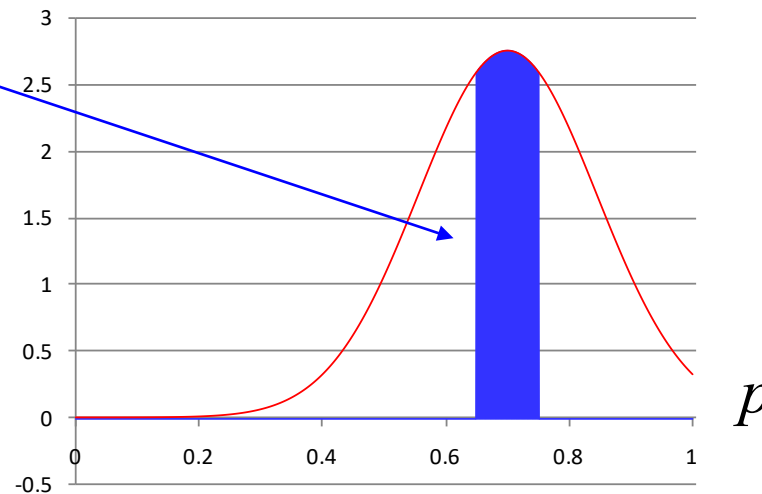
頻度グラフと確率分布の意味

p が の範囲の値となった
のは5000組中の であった。

p が の範囲の値をとる確率
は である。



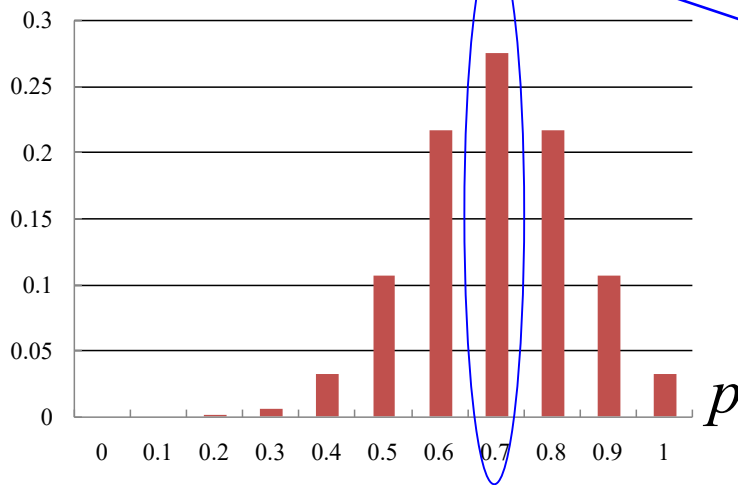
頻度分布



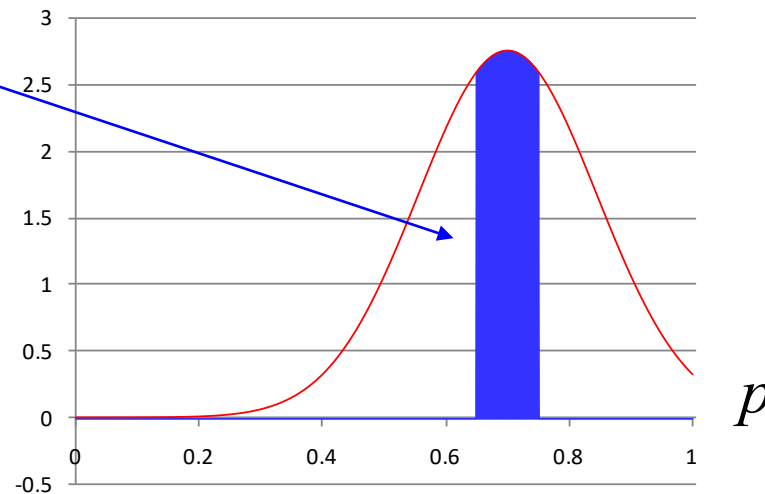
確率分布

ステップ
 p が**0.65~0.75**の範囲の値と
なったのは5000組中の**約27%**で
あった。

p が
は
の範囲の値をとる確率
である。



頻度分布

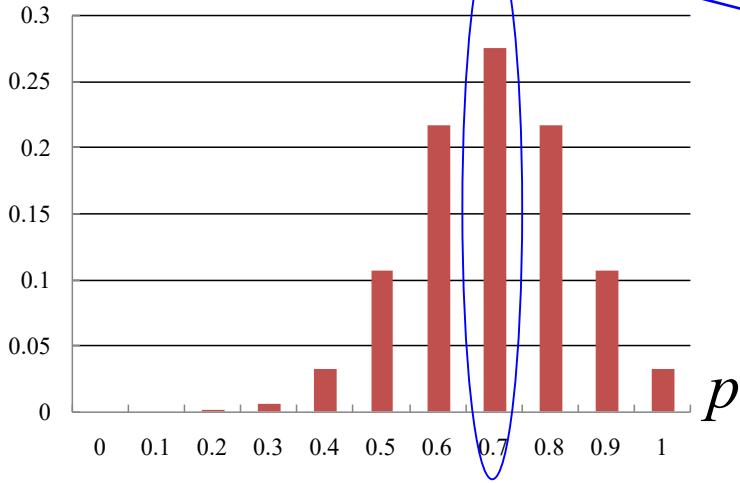


確率分布

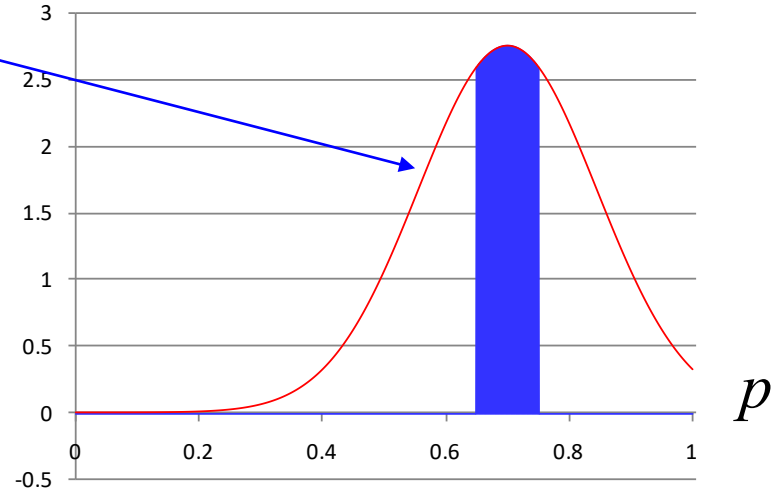
歩

p が **0.65~0.75** の範囲の値となったのは5000組中の**約27%**であった。

p が **0.65~0.75** の範囲の値をとる確率は**0.27**である。 **青い部分の面積が0.27**になる。



頻度分布

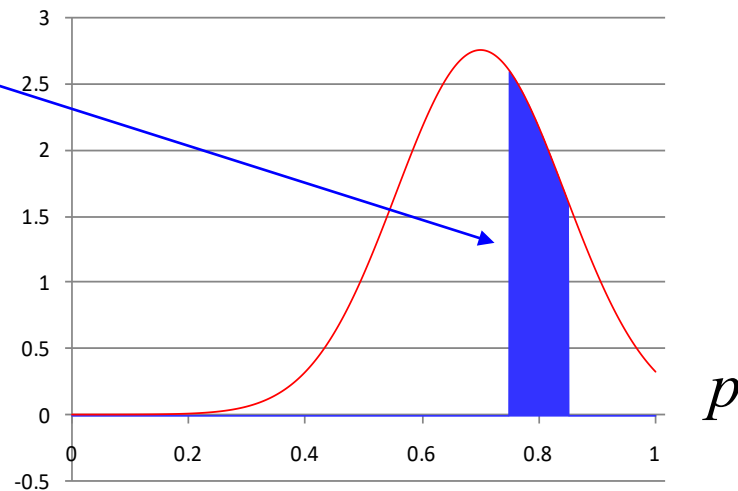
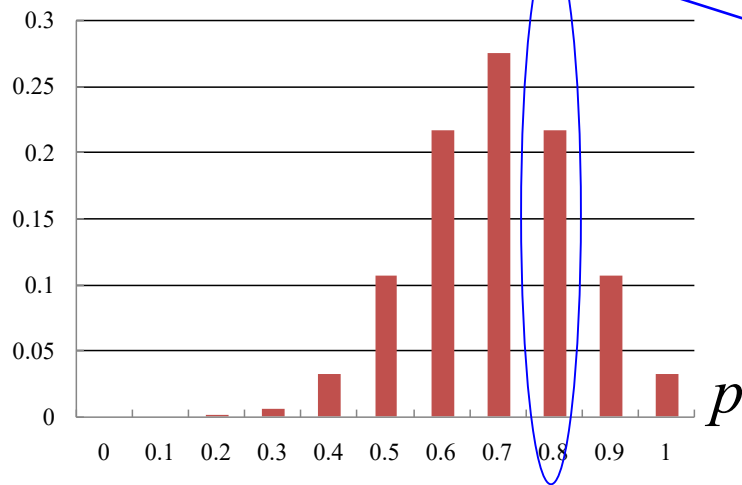


確率分布

頻度グラフと確率分布の意味

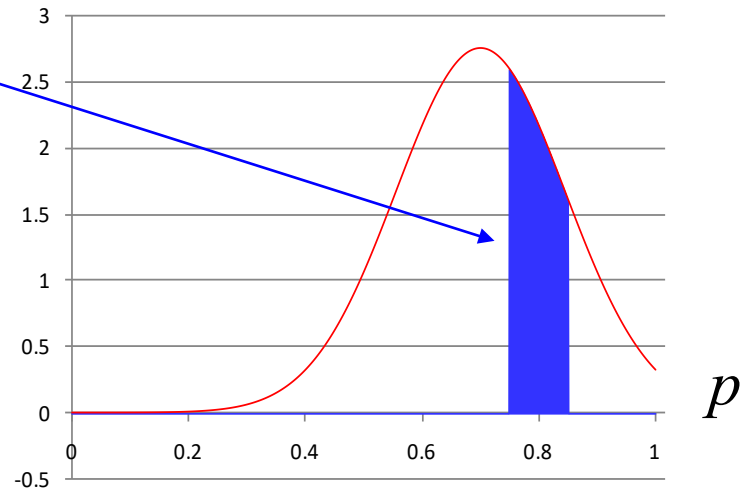
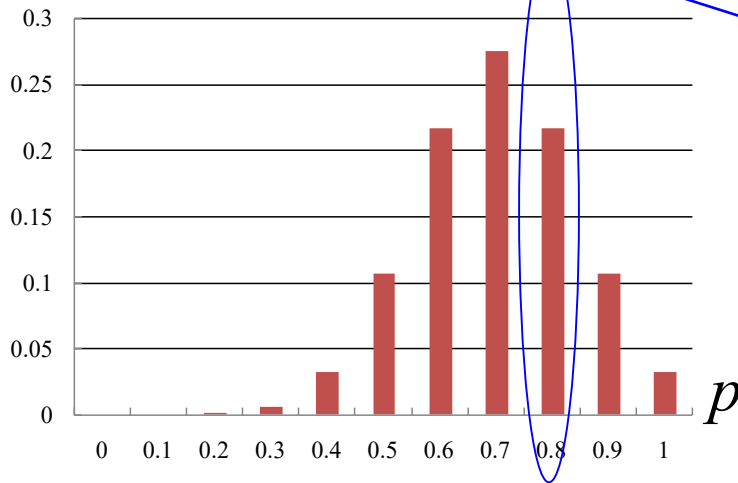
p が の範囲の値となった
のは6000組中の であつた。

p が の範囲の値をとる
確率は である。



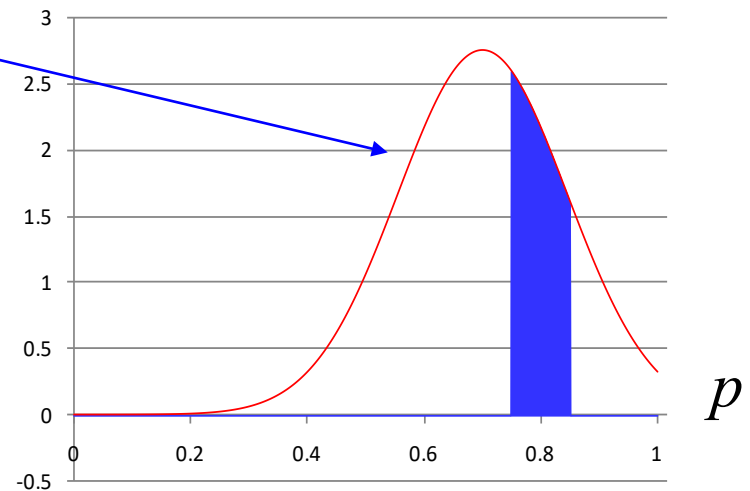
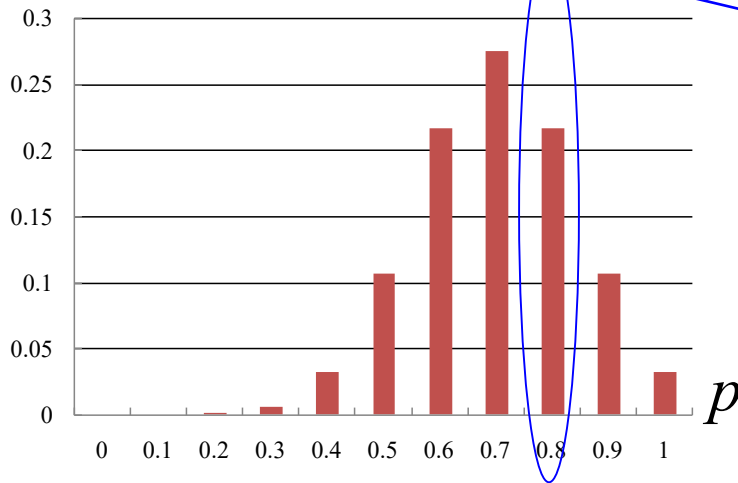
ステップ p が **0.75~0.85** の範囲の値と
なったのは6000組中の**約21%**で
あった。

p が
確率は
の範囲の値をとる
である。



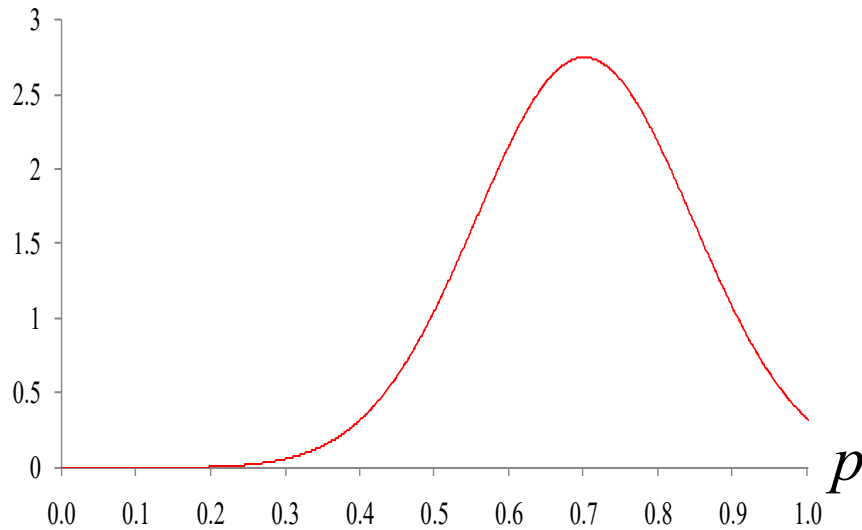
ステップ p が $0.75 \sim 0.85$ の範囲の値となったのは6000組中の約21%であった。

p が $0.75 \sim 0.85$ の範囲の値をとる確率は0.21である。青い部分の面積が0.21になる。

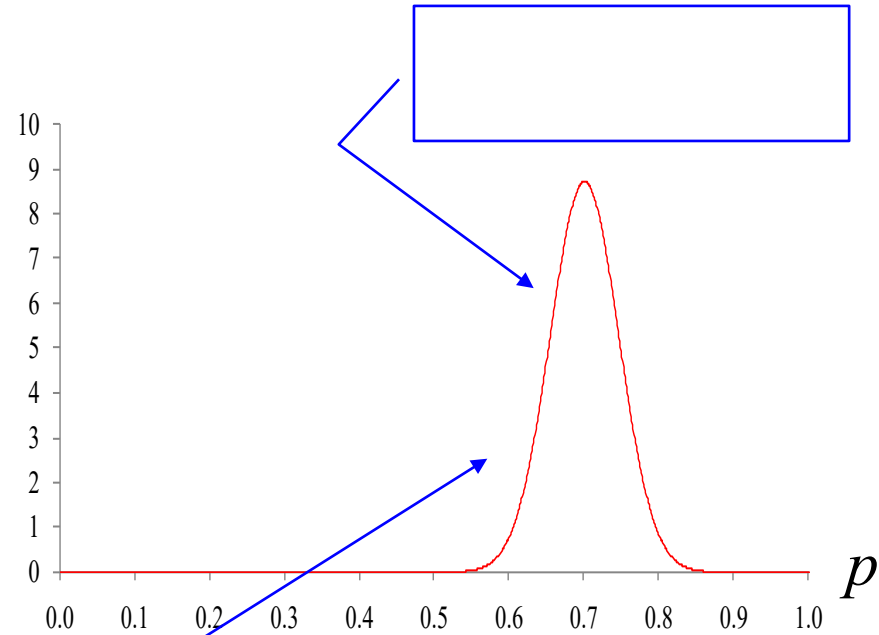


σ による分布形状の違い

$n = 10, p_0 = 0.7$ の場合



$n = 100, p_0 = 0.7$ の場合



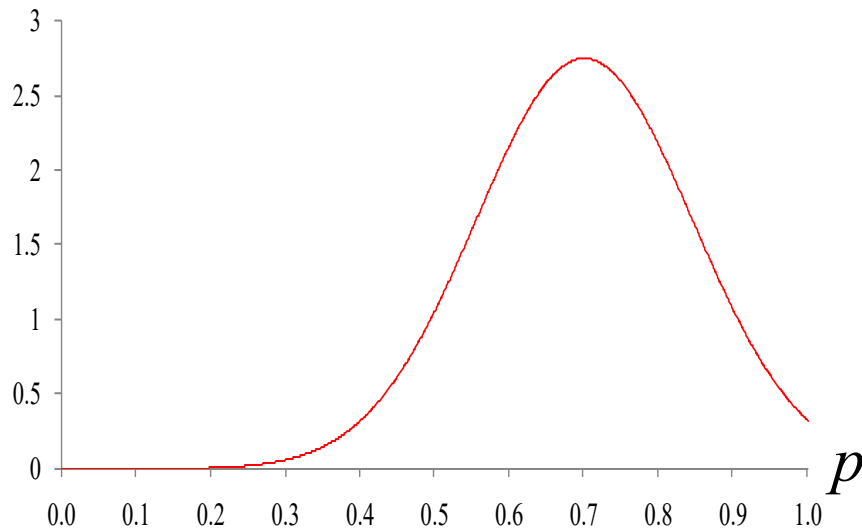
$$\sigma = \sqrt{p_0(1 - p_0) / n}$$

100匹で実験すれば、薬の効き具合 p

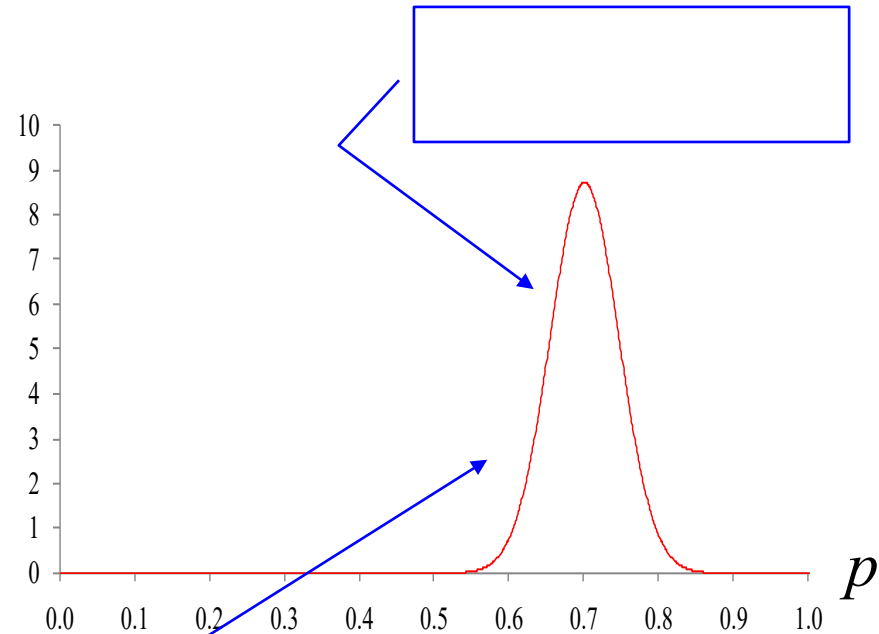
σ による分布形状の違い

$n = 10, p_0 = 0.7$ の場合

$$\sigma = 0.145$$



$n = 100, p_0 = 0.7$ の場合



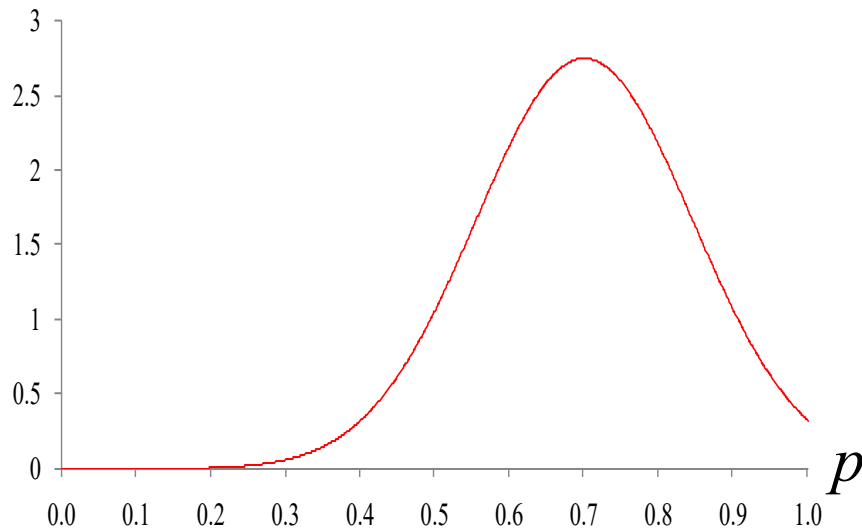
$$\begin{aligned}\sigma &= \sqrt{p_0(1-p_0)/n} \\ &= \sqrt{0.7 \times (1-0.7)/10} \\ &= 0.145\end{aligned}$$

100匹で実験すれば、薬の効き具合 p

σによる分布形状の違い

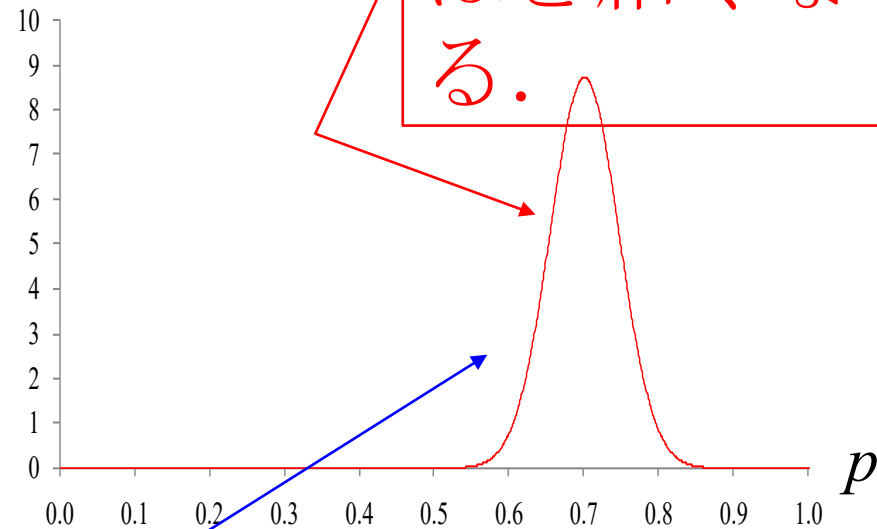
$n = 10, p_0 = 0.7$ の場合

$$\sigma = 0.145$$



$n = 100, p_0 = 0.7$ の場合

$$\sigma = 0.046$$



$$\sigma = \sqrt{p_0(1-p_0)/n}$$

$$= \sqrt{0.7 \times (1-0.7)/10}$$

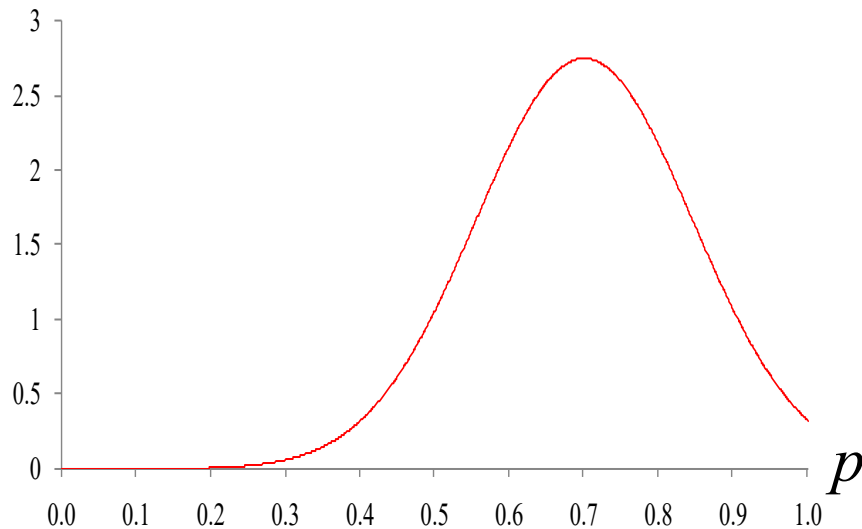
$$= 0.145$$

100匹で実験すれば、薬の効き具合 p

σ による分布形状の違い

$n = 10, p_0 = 0.7$ の場合

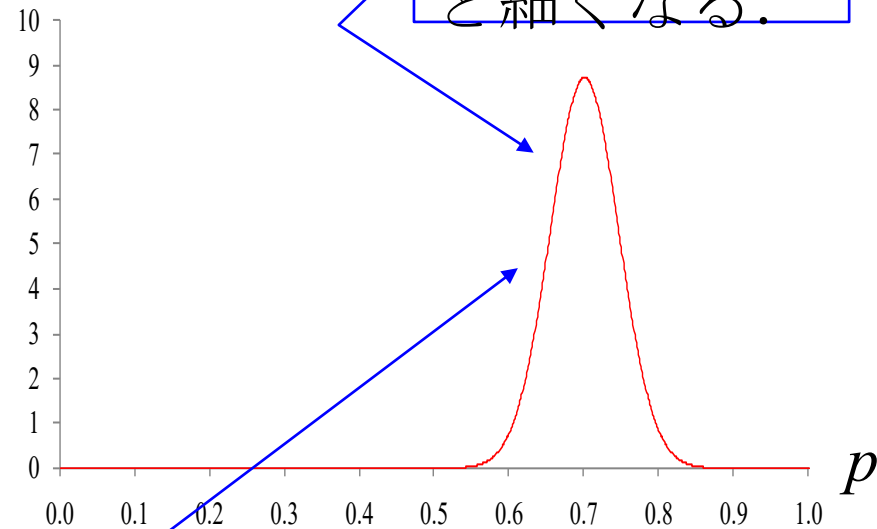
$$\sigma = 0.145$$



$n = 100, p_0 = 0.7$ の場合

$$\sigma = 0.046$$

σが小さいほど細くなる.



$$\begin{aligned}\sigma &= \sqrt{p_0(1-p_0)/n} \\ &= \sqrt{0.7 \times (1-0.7)/10} \\ &= 0.145\end{aligned}$$

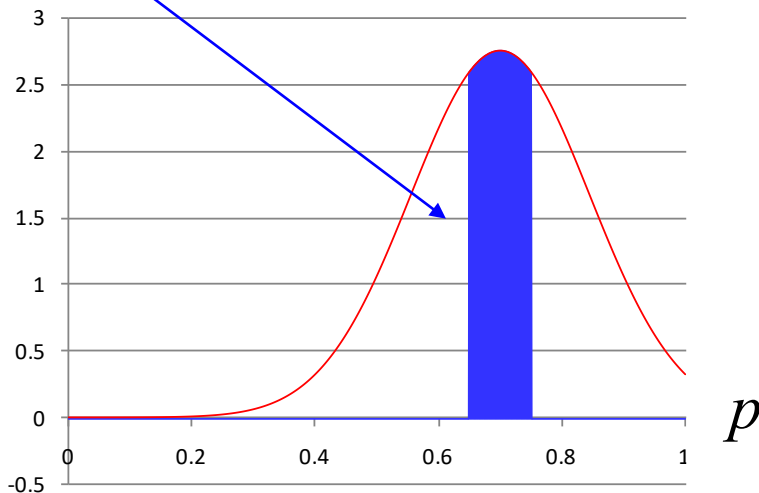
100匹で実験すれば、薬の効き具合 p のばらつき範囲は狭まる.

σ による分布形状の違い

$n = 10, p_0 = 0.7$ の場合

$\sigma = 0.145$

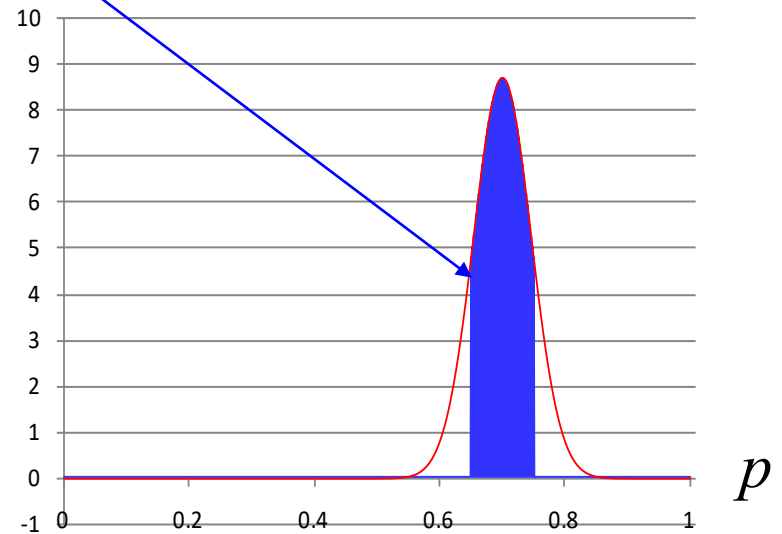
p が の範囲の値
をとる確率は である。



$n = 100, p_0 = 0.7$ の場合

$\sigma = 0.046$

p が の範囲の値を
とる確率は である。

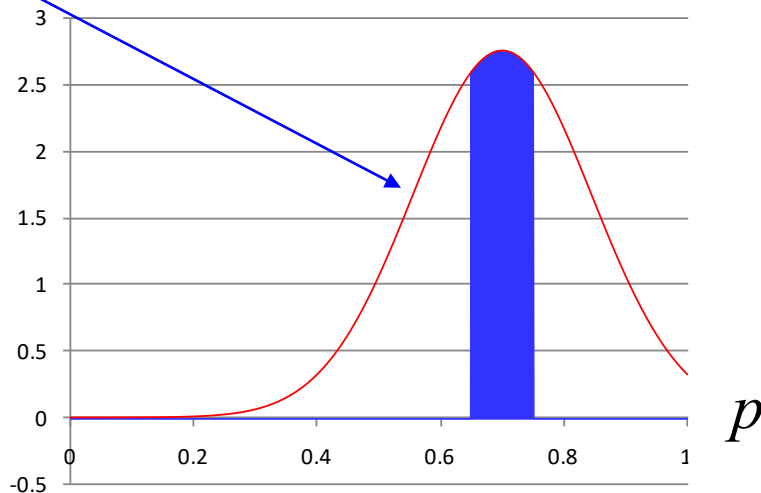


σ による分布形状の違い

$n = 10, p_0 = 0.7$ の場合

$\sigma = 0.145$

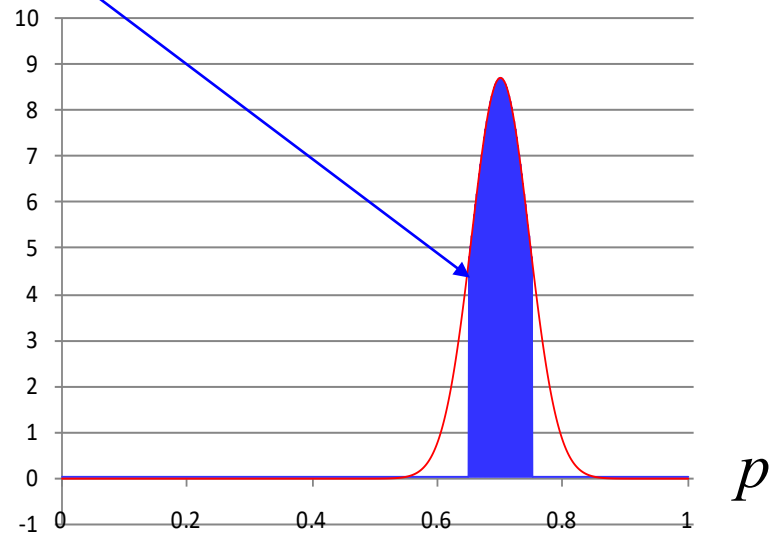
p が **0.65~0.75** の範囲
の値をとる確率は **0.27**
である。



$n = 100, p_0 = 0.7$ の場合

$\sigma = 0.046$

p が の範囲の値
をとる確率は である。

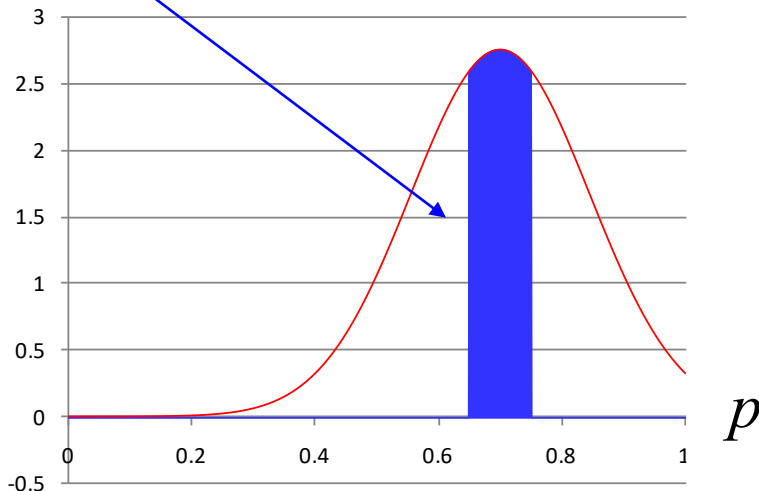


σ による分布形状の違い

$n = 10, p_0 = 0.7$ の場合

$\sigma = 0.145$

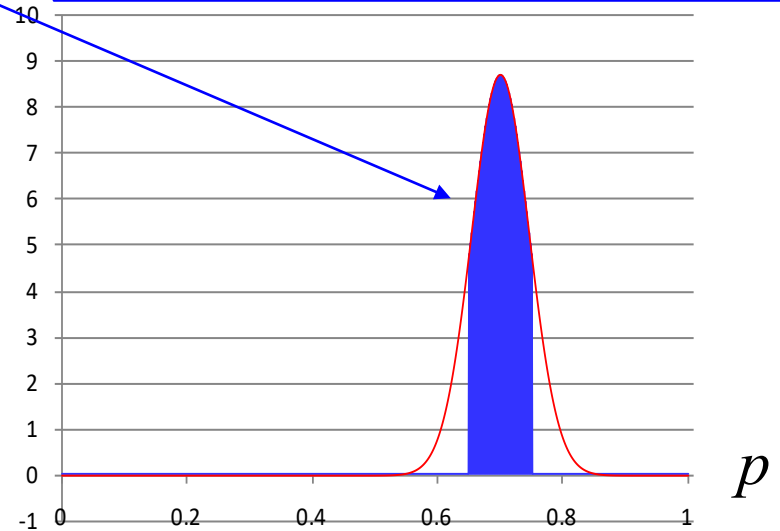
p が $0.65 \sim 0.75$ の範囲の値をとる確率は 0.27 である。



$n = 100, p_0 = 0.7$ の場合

$\sigma = 0.046$

p が $0.65 \sim 0.75$ の範囲の値をとる確率は 0.72 である。



では p が $p_0 - \sigma \sim p_0 + \sigma$ の範囲の値をとる確率は

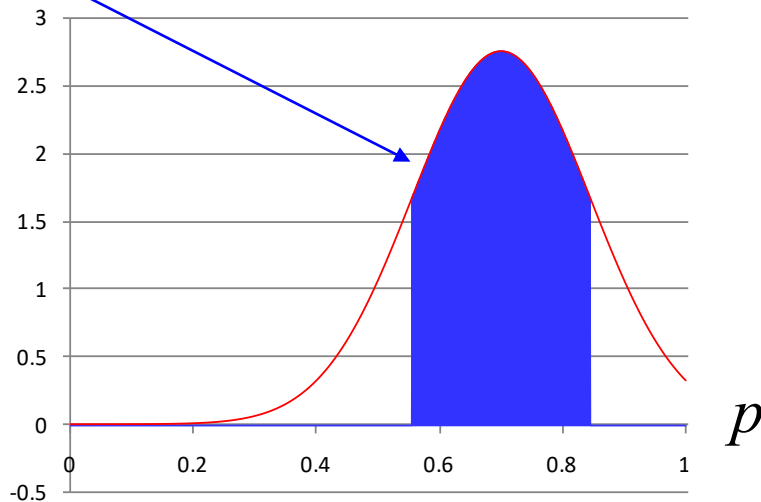
$n = 10, p_0 = 0.7$ の場合

$$\sigma = 0.145$$

$$p_0 - \sigma$$

$$p_0 + \sigma$$

p が \sim の範囲の値をとる確率は である。



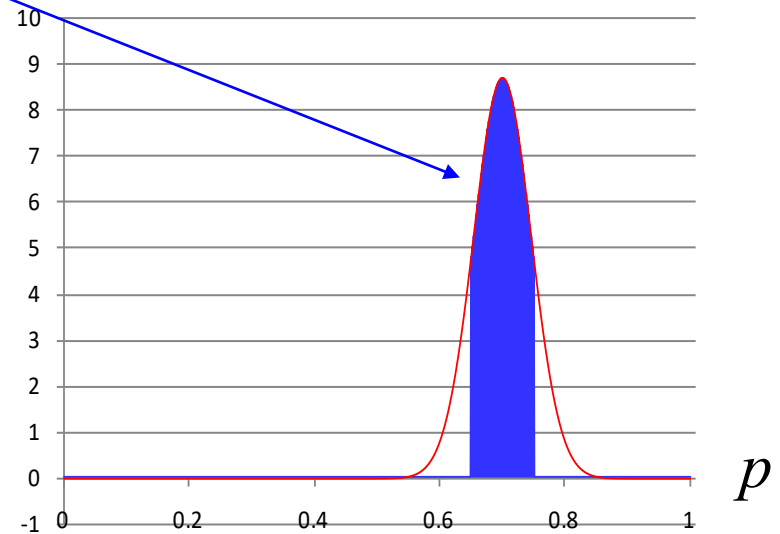
$n = 100, p_0 = 0.7$ の場合

$$\sigma = 0.046$$

$$p_0 - \sigma$$

$$p_0 + \sigma$$

p が \sim の範囲の値をとる確率は である。



では p が $p_0 - \sigma \sim p_0 + \sigma$ の範囲の値をとる確率は

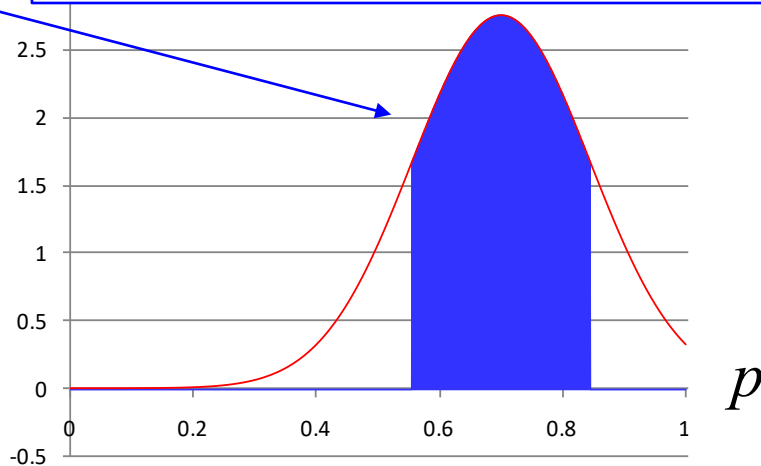
$n = 10, p_0 = 0.7$ の場合

$\sigma = 0.145$

$p_0 - \sigma$

$p_0 + \sigma$

p が **0.555** \sim **0.845** の範囲の値をとる確率は **0.68** である。



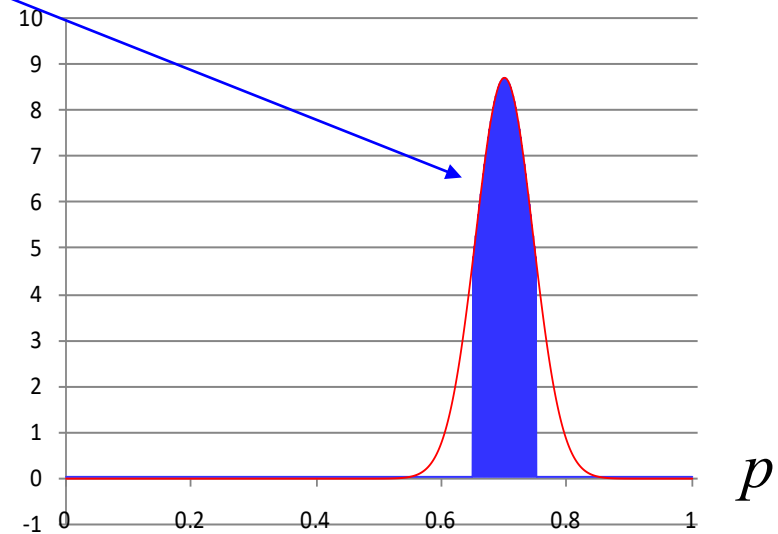
$n = 100, p_0 = 0.7$ の場合

$\sigma = 0.046$

$p_0 - \sigma$

$p_0 + \sigma$

p が \sim の範囲の値をとる確率は である。



では p が $p_0 - \sigma \sim p_0 + \sigma$ の範囲の値をとる確率は

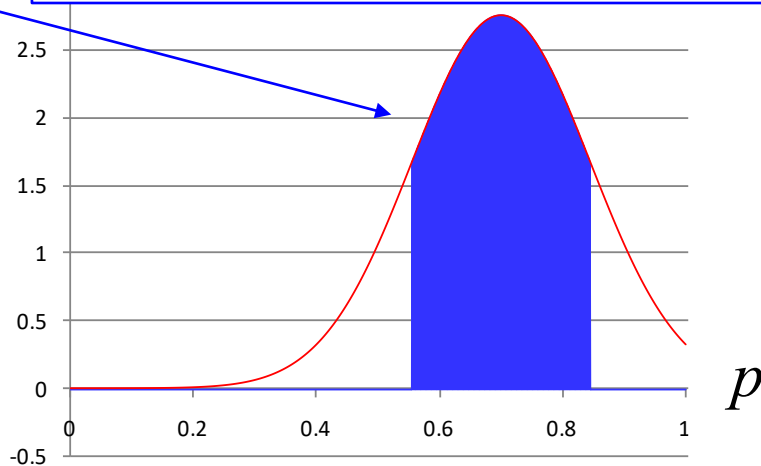
$n = 10, p_0 = 0.7$ の場合

$$\sigma = 0.145$$

$$p_0 - \sigma$$

$$p_0 + \sigma$$

p が $0.555 \sim 0.845$ の範囲の値をとる確率は 0.68 である。



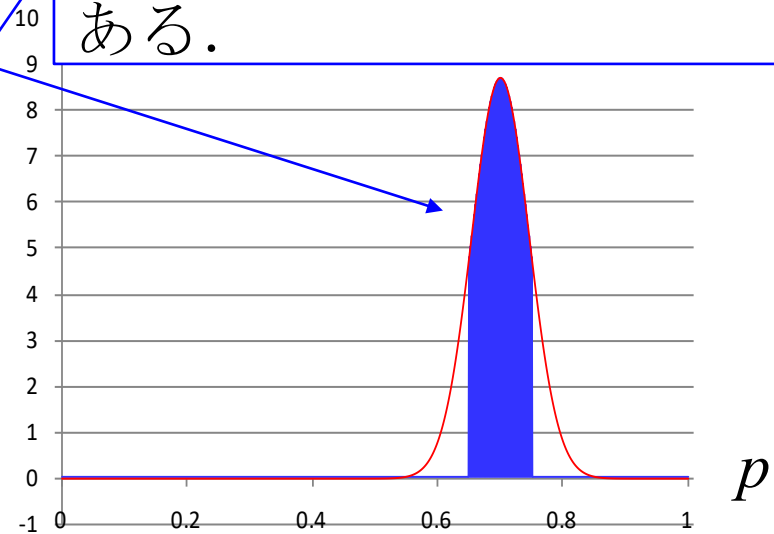
$n = 100, p_0 = 0.7$ の場合

$$\sigma = 0.046$$

$$p_0 - \sigma$$

$$p_0 + \sigma$$

p が $0.654 \sim 0.746$ の範囲の値をとる確率は 0.68 である。



では p が $p_0 - 2\sigma \sim p_0 + 2\sigma$ の範囲の値をとる確率は

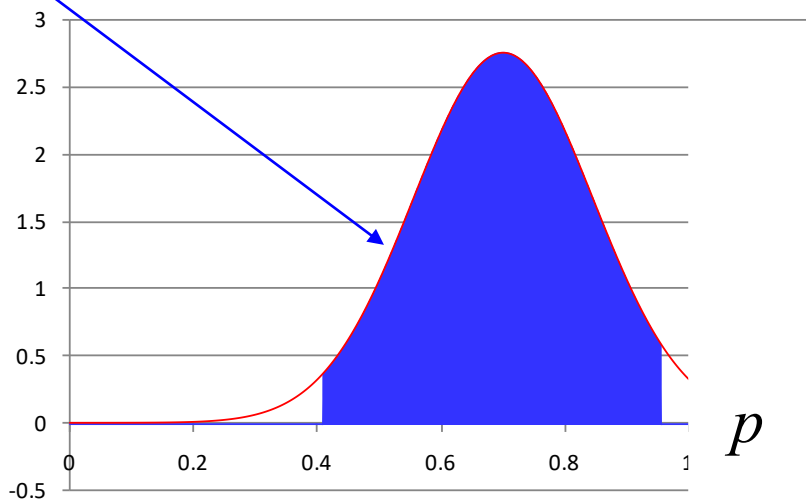
$n = 10, p_0 = 0.7$ の場合

$$\sigma = 0.145$$

$$p_0 - 2\sigma$$

$$p_0 + 2\sigma$$

p が \sim の範囲の値をとる確率は である。



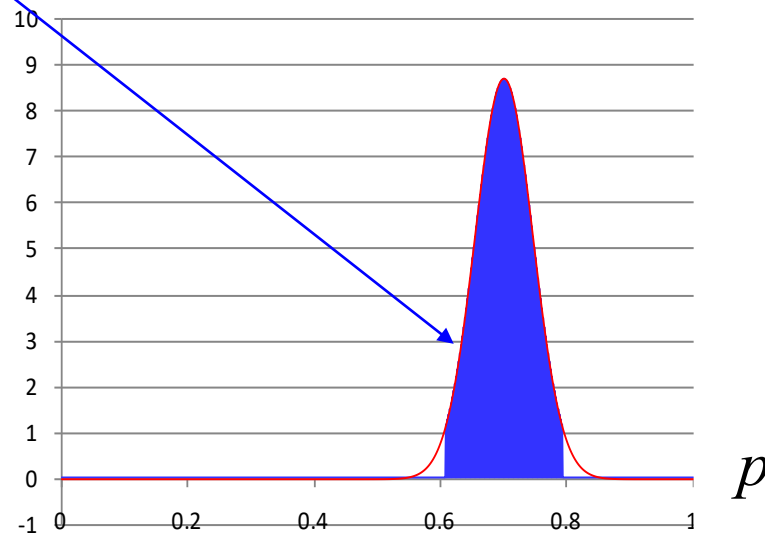
$n = 100, p_0 = 0.7$ の場合

$$\sigma = 0.046$$

$$p_0 - 2\sigma$$

$$p_0 + 2\sigma$$

p が \sim の範囲の値をとる確率は である。



では p が $p_0 - 2\sigma \sim p_0 + 2\sigma$ の範囲の値をとる確率は

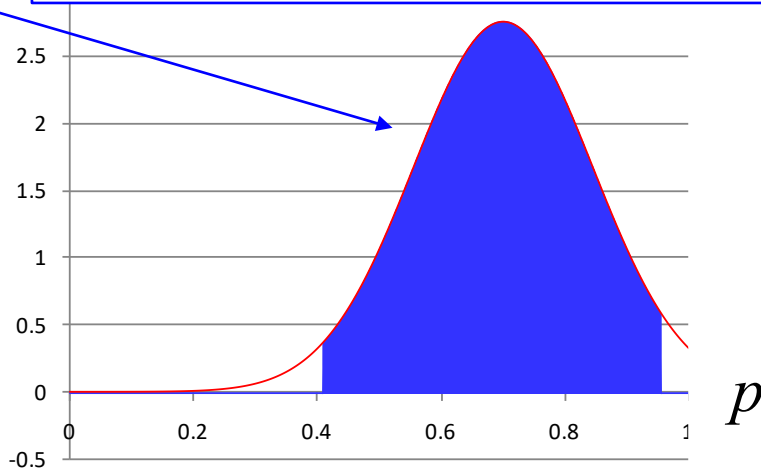
$n = 10, p_0 = 0.7$ の場合

$$\sigma = 0.145$$

$$p_0 - 2\sigma$$

$$p_0 + 2\sigma$$

p が $0.41 \sim 0.99$ の範囲の値をとる確率は 0.954 である。



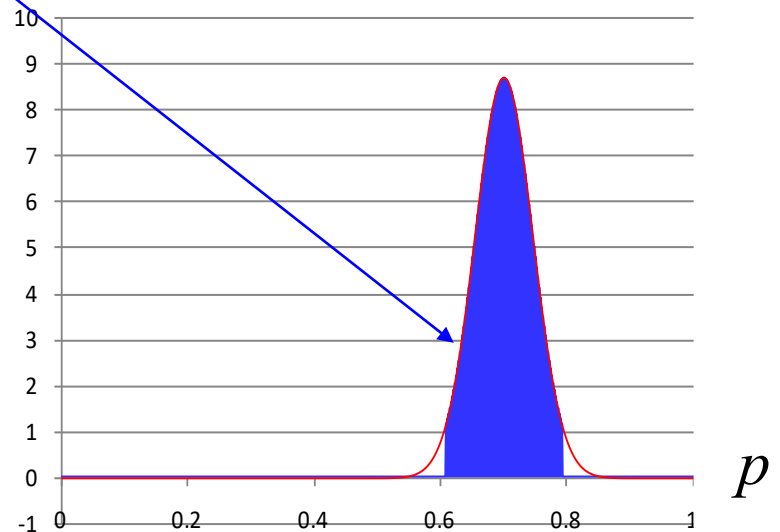
$n = 100, p_0 = 0.7$ の場合

$$\sigma = 0.046$$

$$p_0 - 2\sigma$$

$$p_0 + 2\sigma$$

p が \sim の範囲の値をとる確率は \sim である。



では p が $p_0 - 2\sigma \sim p_0 + 2\sigma$ の範囲の値をとる確率は

$n = 10, p_0 = 0.7$ の場合

$$\sigma = 0.145$$

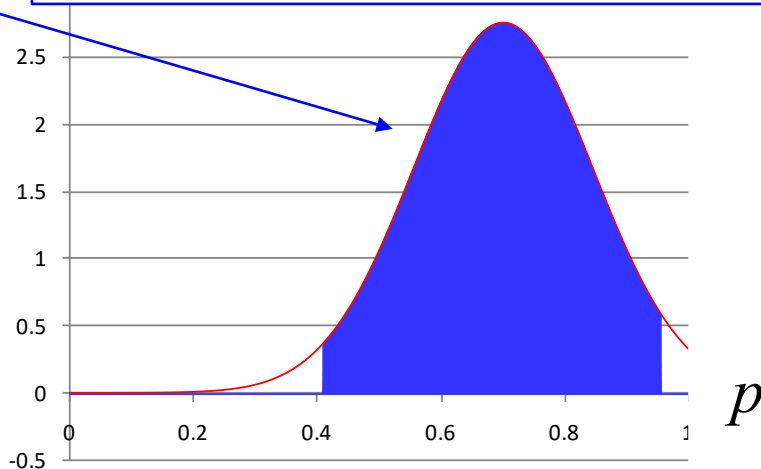
$n = 100, p_0 = 0.7$ の場合

$$\sigma = 0.046$$

$$p_0 - 2\sigma$$

$$p_0 + 2\sigma$$

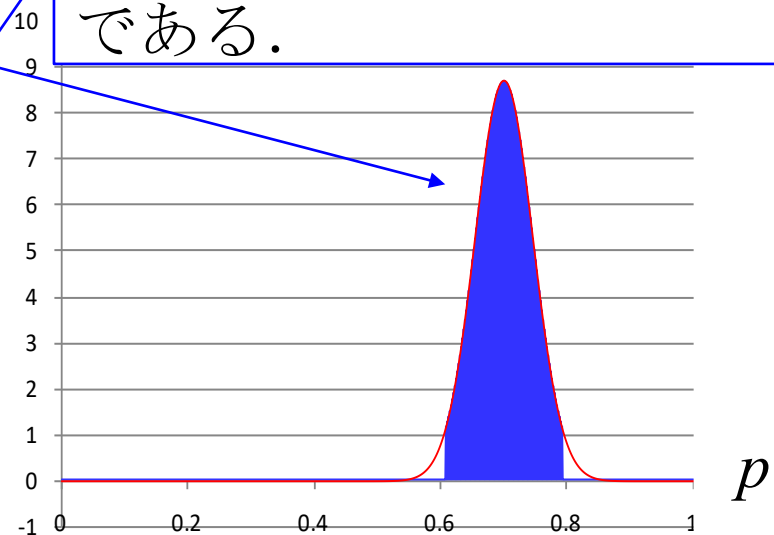
p が $0.41 \sim 0.99$ の範囲の値をとる確率は 0.954 である。



$$p_0 - 2\sigma$$

$$p_0 + 2\sigma$$

p が $0.608 \sim 0.792$ の範囲の値をとる確率は 0.954 である。



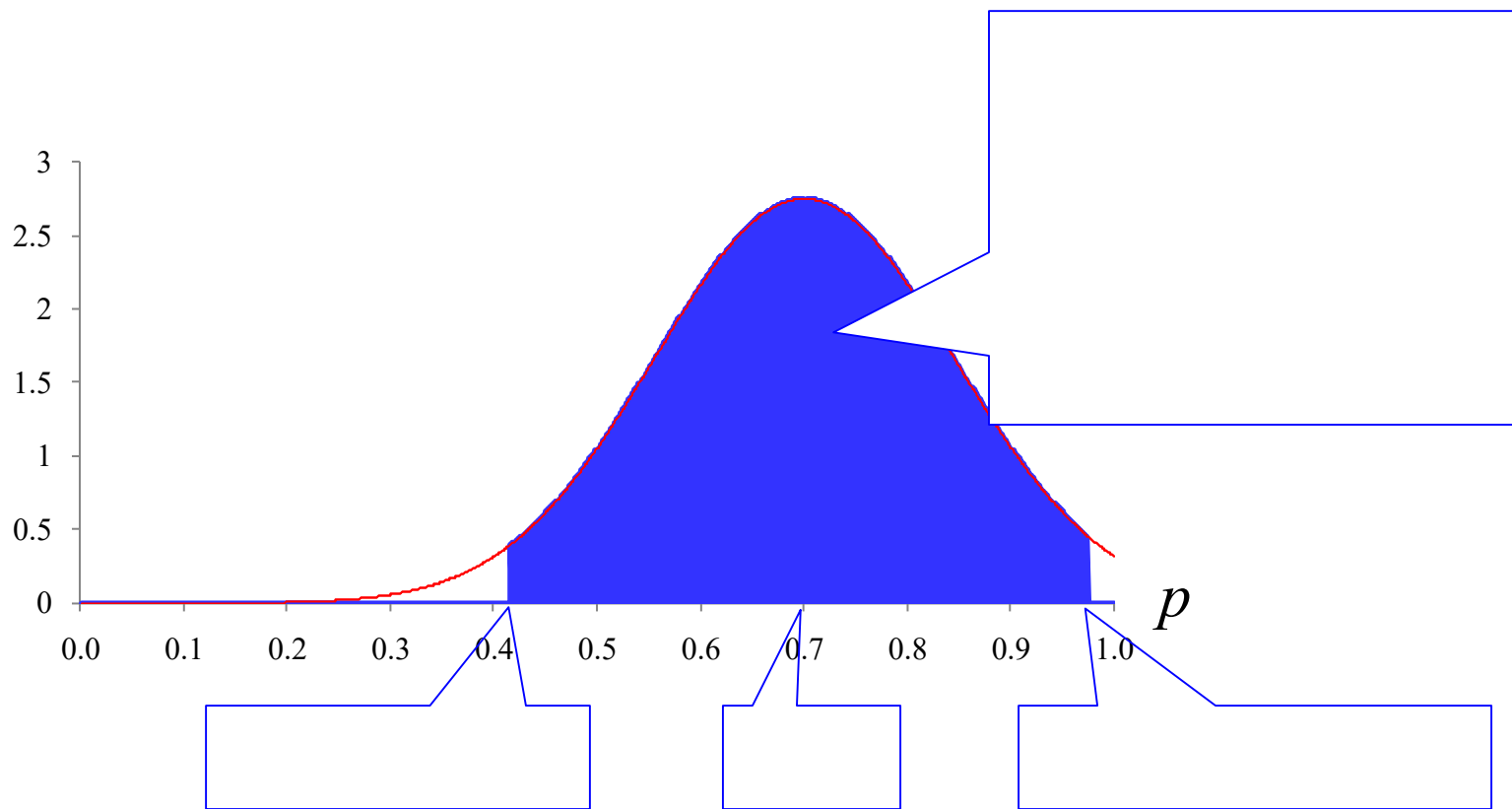
有効率の区間推定

n 匹のマウスへの薬の効き具合 p の分布は

理論平均： p_0

理論分散： $\sigma^2 = p_0(1-p_0)/n$

の正規分布となった。



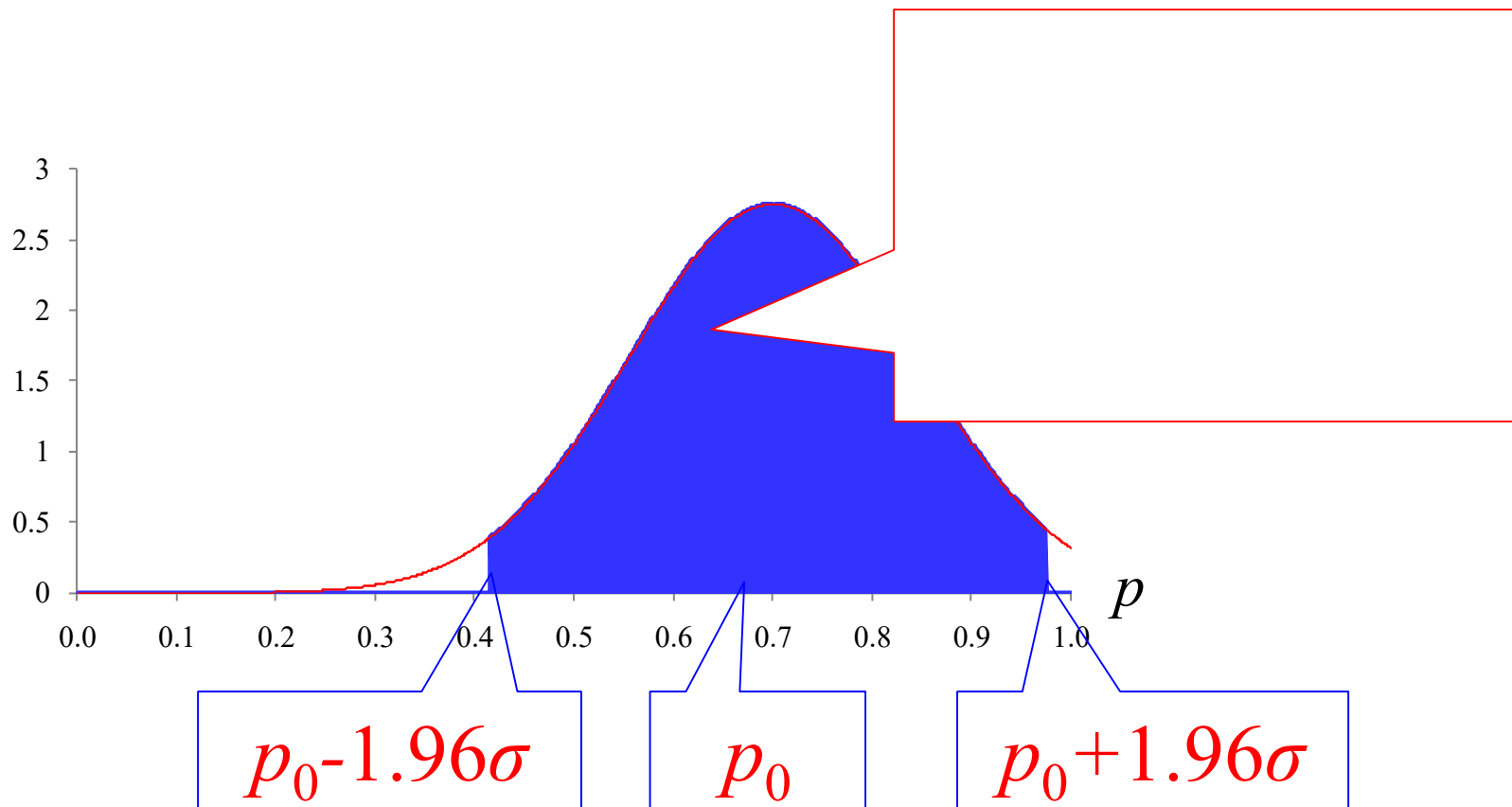
有効率の区間推定

n 匹のマウスへの薬の効き具合 p の分布は

理論平均： p_0

理論分散： $\sigma^2 = p_0(1-p_0)/n$

の正規分布となった。



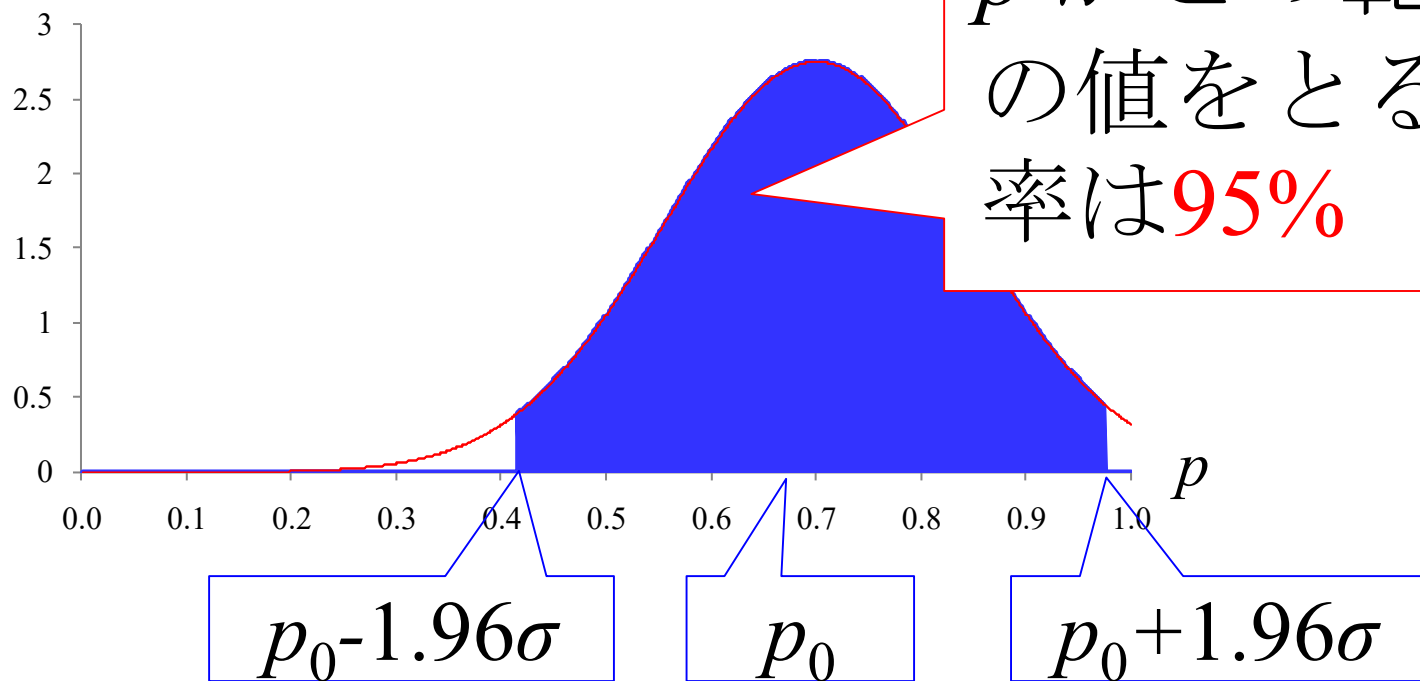
有効率の区間推定

n 匹のマウスへの薬の効き具合 p の分布は

理論平均： p_0

理論分散： $\sigma^2 = p_0(1-p_0)/n$

の正規分布となった。



■ 小テスト10.1 (選挙での出口調査の信頼度) ■

先のA市長選においてB氏の当選確実は、開票がまだ3%程度しか進んでいないにもかかわらず、ニュースで大々的に報道された。

これは投票所を出てきた人(約1000人)に対して、誰に投票したかを調べた(出口調査と呼ばれる)結果によっていた。

1000人程度の人に聞いただけで、A市の投票者(約100万人)全体の投票結果が果たして精確に分かるのであろうか？

B氏への真の支持率 $p_0 = 0.6$ 、それ以外の人への真の支持率 $1 - p_0 = 0.4$ として、 $n = 1000$ 人による投票のシミュレーションを行い、真の支持率 p_0 の95%信頼区間を求めよ。

■ 小テスト10.2 (テレビ番組の視聴率)

テレビ番組の制作者にとって視聴率を上げることは最優先課題である。

視聴率は例えば東海地区では600世帯がどのチャンネルを選んでいたかの調査結果で決められる。たったの600世帯で、東海地区数百万世帯がどのチャンネルを選んでいたかを精確に知ることができるのであろうか？

チャンネルAの真の視聴率 $p_0 = 0.2$ ，それ以外のチャンネルの視聴率 $1 - p_0 = 0.8$ として， $n = 600$ 世帯のチャンネル選択のシミュレーションを行い，真の視聴率 p_0 の95%信頼区間を求めよ。

2013年3月

著者： 古橋武
名古屋大学工学研究科計算理工学専攻
furuhashi@cse.nagoya-u.ac.jp