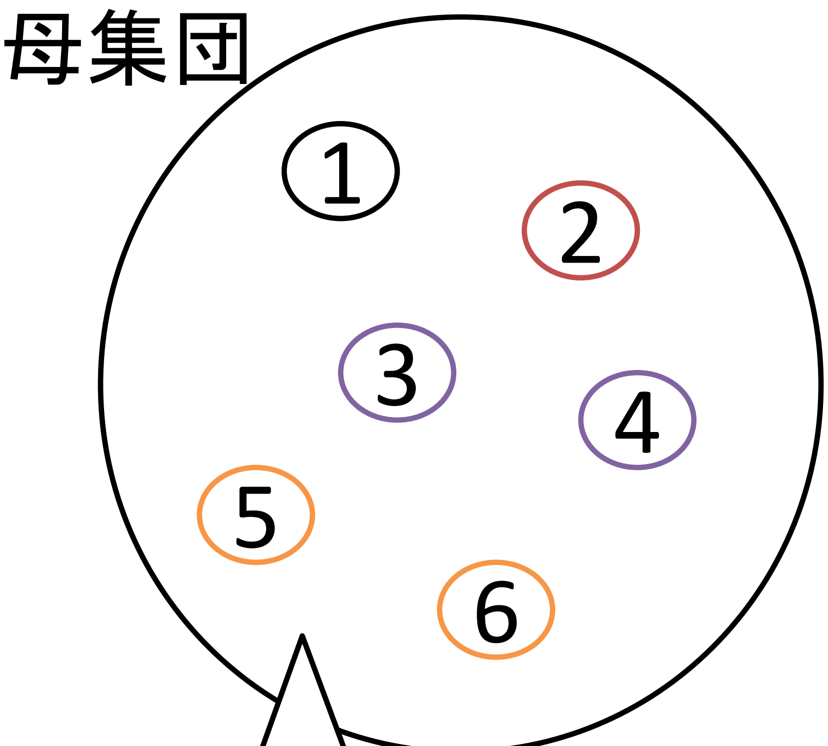
サイコロをn回投げたときの目の和の確率分布を求めよう

サイコロを１回投げるとき
サイコロ投げ有限母集団説を採用する



割合はそれぞれ「６分の１」である。

- 母集団から1個ランダムに取り出し数字を確認する作業を、X（確率変数）とする。

確率は、それぞれ６分の１である。

サイコロの目の期待値と分散・標準偏差

a_k	p_k	$a_k p_k$	$(a_k-\mu)^2 p_k$
1	$1/6$		
2	$1/6$		
3	$1/6$		
4	$1/6$		
5	$1/6$		
6	$1/6$		
合計	$(1/6)\times 6=1$		

サイコロの目の期待値と分散・標準偏差

a_k	p_k	$a_k p_k$	$(a_k - \mu)^2 p_k$
1	$1/6$	$1\times(1/6)$	$(1-3.5)^2\times(1/6)$
2	$1/6$	$2\times(1/6)$	$(2-3.5)^2\times(1/6)$
3	$1/6$	$3\times(1/6)$	$(3-3.5)^2\times(1/6)$
4	$1/6$	$4\times(1/6)$	$(4-3.5)^2\times(1/6)$
5	$1/6$	$5\times(1/6)$	$(5-3.5)^2\times(1/6)$
6	$1/6$	$6\times(1/6)$	$(6-3.5)^2\times(1/6)$
合計	$(1/6)\times 6=1$	$21/6=3.5$	$17.5\times(1/6)=2.92$

平均：$\mu = E[X] = 3.5$

分散：$\sigma^2 = V[X] = E[(X-\mu)^2] \cong 2.92$

標準偏差：$\sigma = \sqrt{V[X]} = \sqrt{E[(X-\mu)^2]} \cong 1.71$

サイコロを2回投げるとき、

X_1 :1 回 目 の 目

X_2 :2回 目 の 目

度数		X_1					
		1	2	3	4	5	6
X_2	1						
	2						
	3						
	4						
	5						
	6						

$X_1 + X_2$		X_1					
		1	2	3	4	5	6
X_2	1						
	2						
	3						
	4						
	5						
	6						

課題

サイコロを2回投げるとき、

度数		X_1					
		1	2	3	4	5	6
X_2	1	1	1	1	1	1	1
	2	1	1	1	1	1	1
	3	1	1	1	1	1	1
	4	1	1	1	1	1	1
	5	1	1	1	1	1	1
	6	1	1	1	1	1	1

X_1 :1 回目の目

X_2 :2回目の目

X_1+X_2		X_1					
		1	2	3	4	5	6
X_2	1	2	3	4	5	6	7
	2	3	4	5	6	7	8
	3	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

サイコロを2回投げるとき、

度数		X_1					
		1	2	3	4	5	6
X_2	1	1	1	1	1	1	1
	2	1	1	1	1	1	1
	3	1	1	1	1	1	1
	4	1	1	1	1	1	1
	5	1	1	1	1	1	1
	6	1	1	1	1	1	1

X_1 :1 回目の目

X_2 :2回目の目

$X_1 + X_2$		X_1					
		1	2	3	4	5	6
X_2	1	2	3	4	5	6	7
	2	3	4	5	6	7	8
	3	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

課題

$X_1 + X_2$ の確率分布

a_k	f_k	p_k	$a_k p_k$	$(a_k - \mu)^2 p_k$
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
Σ				

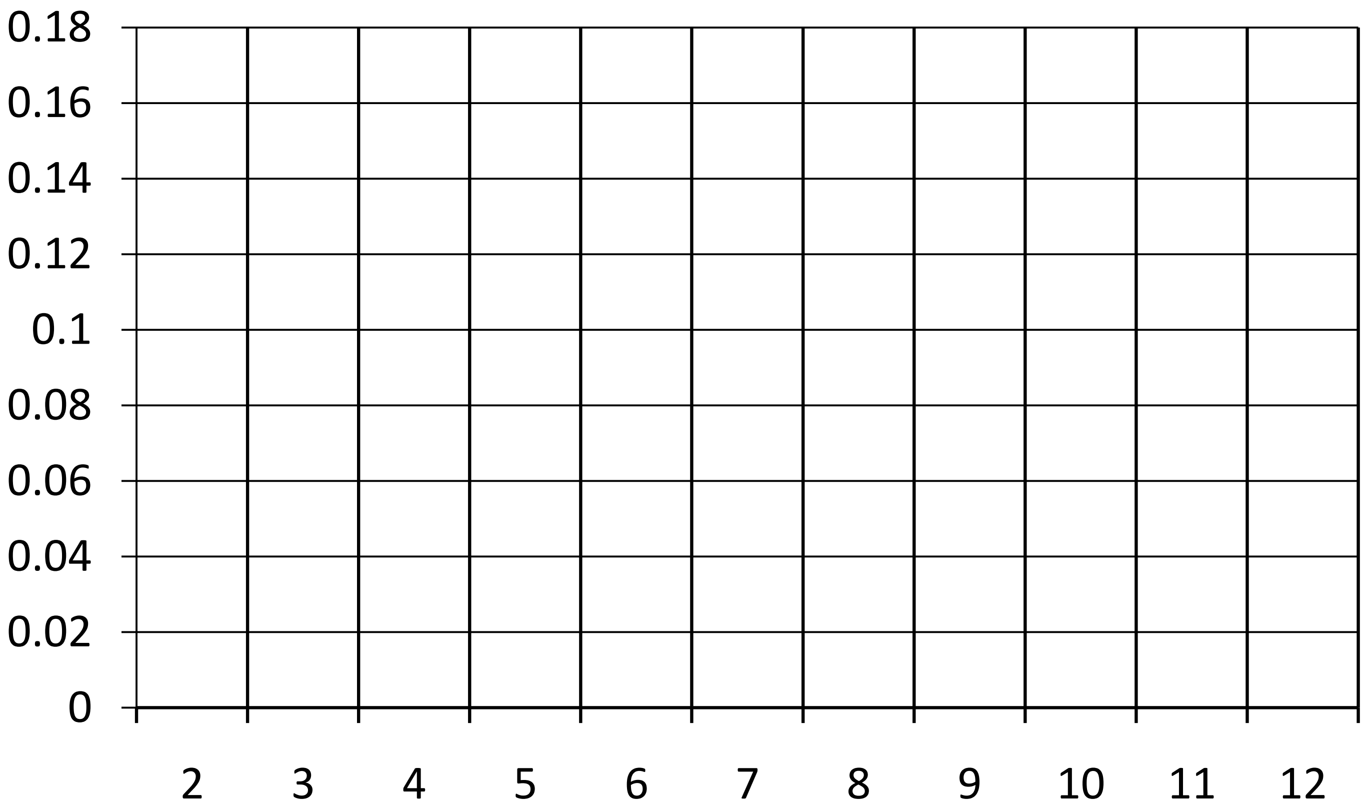
$X_1 + X_2$ の確率分布

a_k	f_k	p_k	$a_k p_k$	$(a_k - \mu)^2 p_k$
2	1	1/36		
3	2	2/36		
4	3	3/36		
5	4	4/36		
6	5	5/36		
7	6	6/36		
8	5	5/36		
9	4	4/36		
10	3	3/36		
11	2	2/36		
12	1	1/36		
Σ	36	1		

サイコロを2回投げるときの目の和の分布

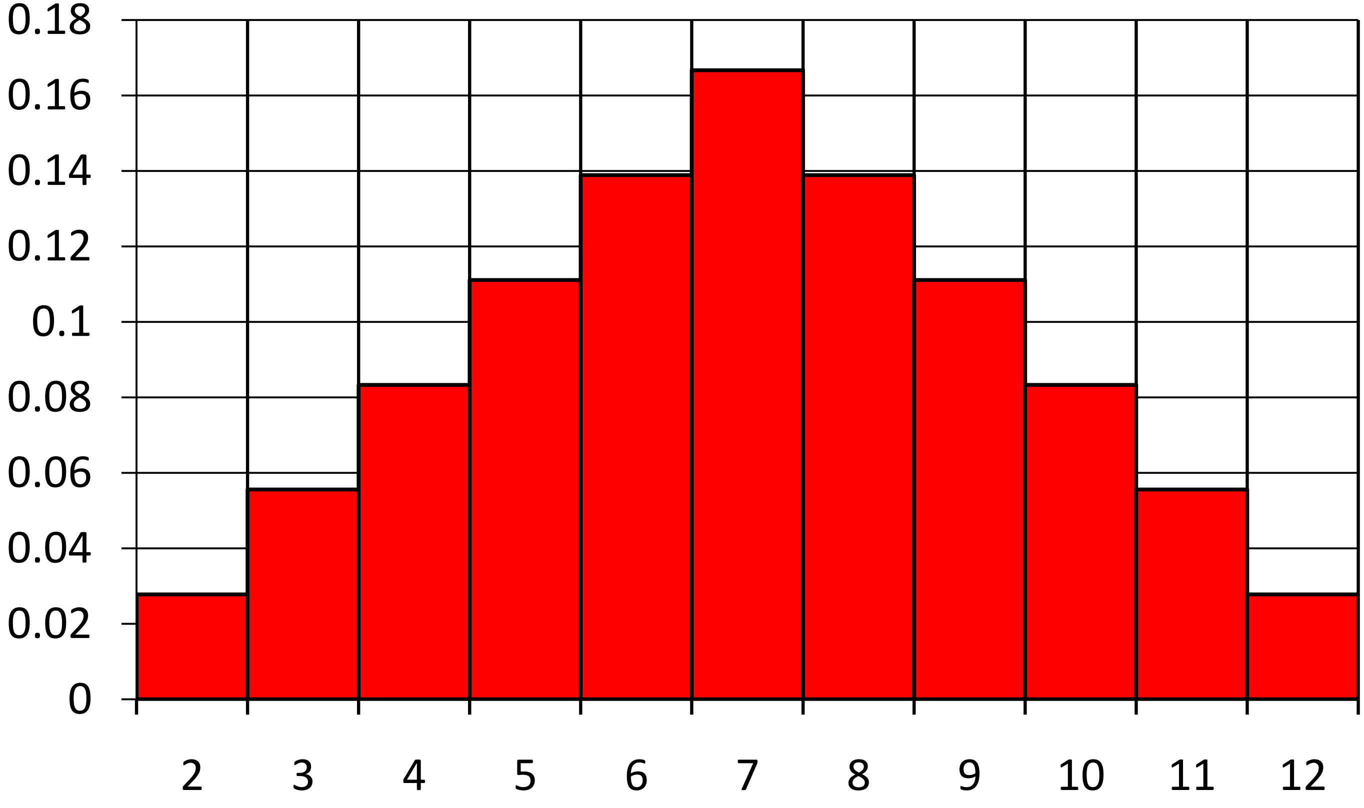
課題

$X_1 + X_2$ の確率分布



サイコロを2回投げるときの目の和の分布

$X_1 + X_2$ の確率分布



サイコロを3回投げるとき、

課題

度数		X_1+X_2										
		2	3	4	5	6	7	8	9	10	11	12
X_3	1											
	2											
	3											
	4											
	5											
	6											

$X_1+X_2+X_3$		X_1+X_2										
		2	3	4	5	6	7	8	9	10	11	12
X_3	1											
	2											
	3											
	4											
	5											
	6											

サイコロを3回投げるとき、

度数		X_1+X_2										
		2	3	4	5	6	7	8	9	10	11	12
X_3	1	1	2	3	4	5	6	5	4	3	2	1
	2	1	2	3	4	5	6	5	4	3	2	1
	3	1	2	3	4	5	6	5	4	3	2	1
	4	1	2	3	4	5	6	5	4	3	2	1
	5	1	2	3	4	5	6	5	4	3	2	1
	6	1	2	3	4	5	6	5	4	3	2	1

$X_1+X_2+X_3$		X_1+X_2										
		2	3	4	5	6	7	8	9	10	11	12
X_3	1	3	4	5	6	7	8	9	10	11	12	13
	2	4	5	6	7	8	9	10	11	12	13	14
	3	5	6	7	8	9	10	11	12	13	14	15
	4	6	7	8	9	10	11	12	13	14	15	16
	5	7	8	9	10	11	12	13	14	15	16	17
	6	8	9	10	11	12	13	14	15	16	17	18

サイコロを3回投げるとき、

度数		X_1+X_2										
		2	3	4	5	6	7	8	9	10	11	12
X_3	1	1	2	3	4	5	6	5	4	3	2	1
	2	1	2	3	4	5	6	5	4	3	2	1
	3	1	2	3	4	5	6	5	4	3	2	1
	4	1	2	3	4	5	6	5	4	3	2	1
	5	1	2	3	4	5	6	5	4	3	2	1
	6	1	2	3	4	5	6	5	4	3	2	1

$X_1+X_2+X_3$		X_1+X_2										
		2	3	4	5	6	7	8	9	10	11	12
X_3	1	3	4	5	6	7	8	9	10	11	12	13
	2	4	5	6	7	8	9	10	11	12	13	14
	3	5	6	7	8	9	10	11	12	13	14	15
	4	6	7	8	9	10	11	12	13	14	15	16
	5	7	8	9	10	11	12	13	14	15	16	17
	6	8	9	10	11	12	13	14	15	16	17	18

$X_1 + X_2 + X_3$ の確率分布

課題

a_k	f_k	p_k	$a_k p_k$	$(a_k - \mu)^2 p_k$
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
$\sum$				

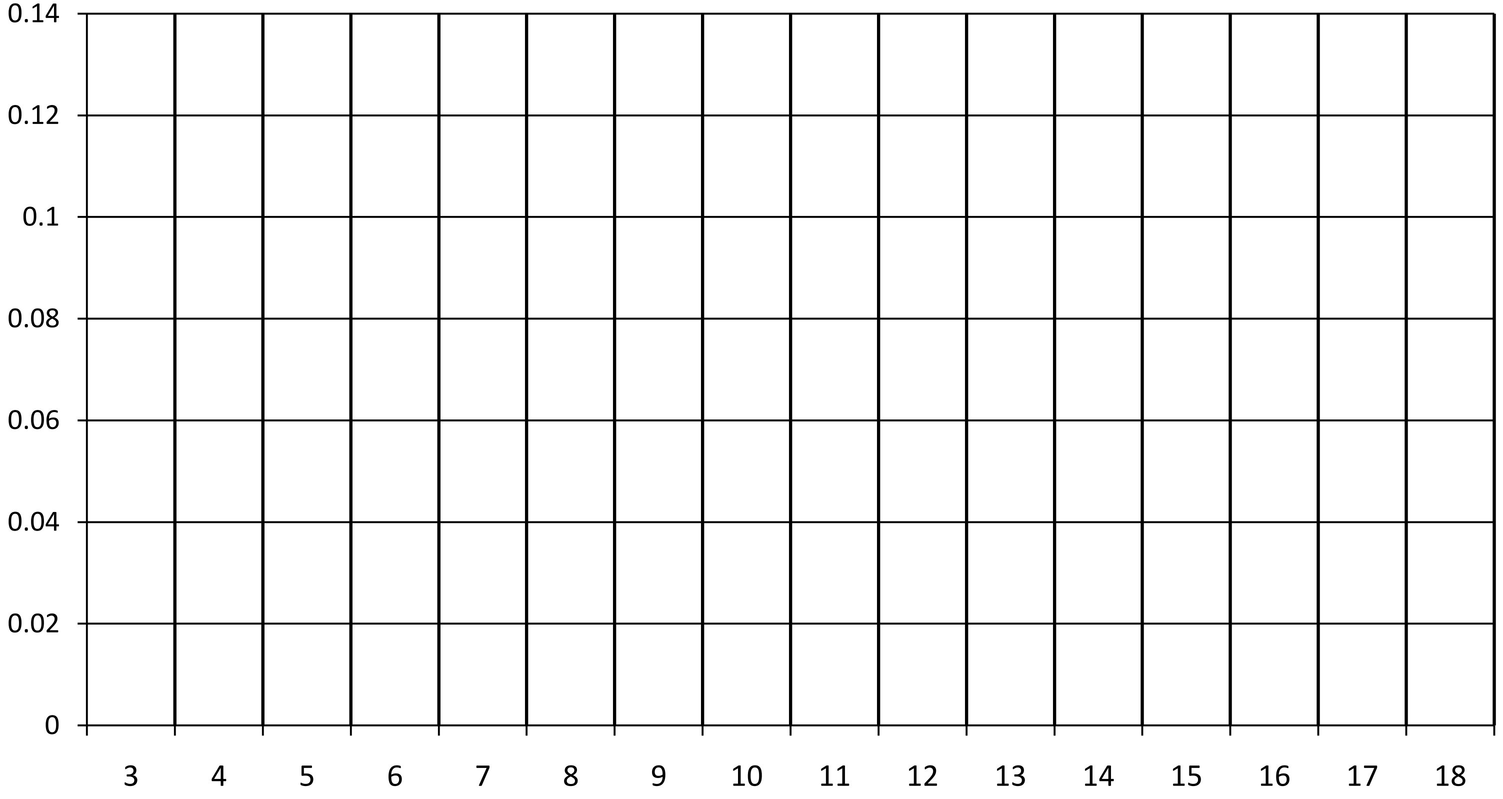
$X_1 + X_2 + X_3$ の確率分布

a_k	f_k	p_k	$a_k p_k$	$(a_k - \mu)^2 p_k$
3	1			
4	3			
5	6			
6	10			
7	15			
8	21			
9	25			
10	27			
11	27			
12	25			
13	21			
14	15			
15	10			
16	6			
17	3			
18	1			
$\sum$	216			

サイコロを3回投げるときの目の和の分布

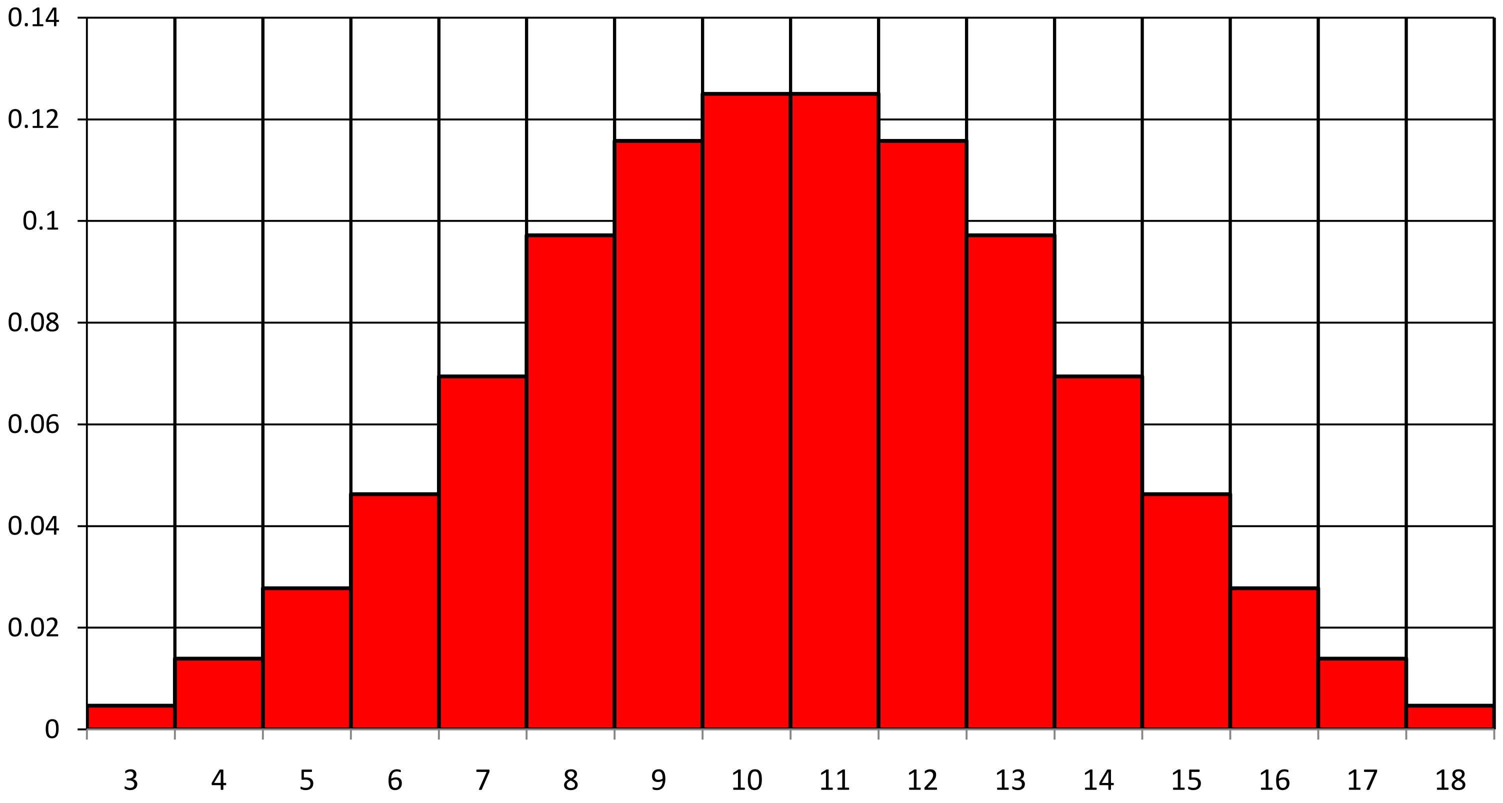
課題

$X_1 + X_2 + X_3$ の確率分布



サイコロを3回投げるときの目の和の分布

$X_1 + X_2 + X_3$ の確率分布



確率分布を求めることは次第に困難になってゆくが、

$\mu = E[X] \qquad \sigma^2 = V[X] = E\left[(X-\mu)^2\right]$ とおくとき、

これから講義で説明するが、以下の関係が成り立つ。

$$E[X_1 + X_2] = 2\mu, \quad V[X_1 + X_2] = 2\sigma^2$$

$$E[X_1 + X_2 + X_3] = 3\mu, \quad V[X_1 + X_2 + X_3] = 3\sigma^2$$

$$E[X_1 + X_2 + \cdots + X_n] = n\mu, \quad V[X_1 + X_2 + \cdots + X_n] = n\sigma^2$$

$$\boxed{\mu = E[X]} \qquad \boxed{\sigma^2 = V[X] = E\left[(X-\mu)^2\right]}$$

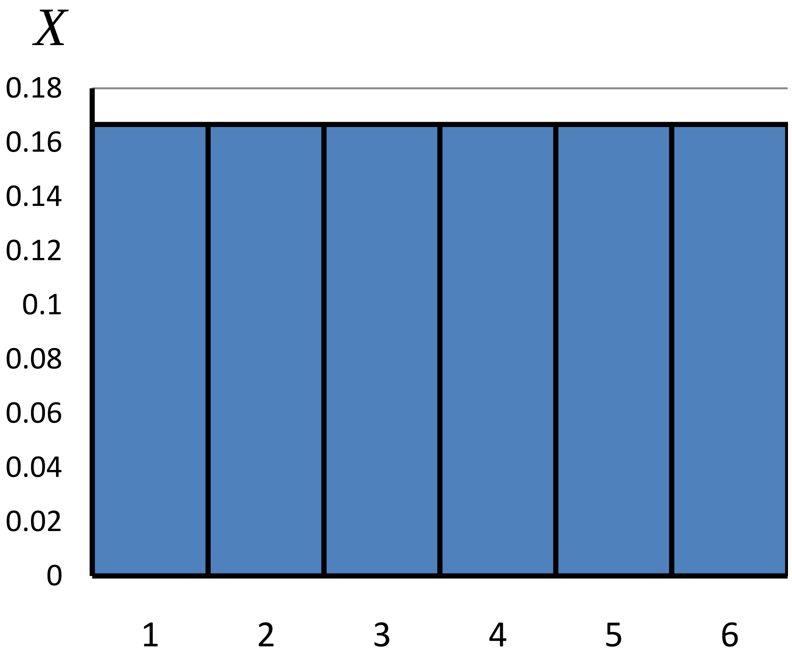
目の和ではなく、平均を考えると以下が成り立つ

$$E\left[\frac{X_1 + X_2}{2}\right] = \mu, \quad V\left[\frac{X_1 + X_2}{2}\right] = \frac{\sigma^2}{2}$$

$$E\left[\frac{X_1 + X_2 + X_3}{3}\right] = \mu, \quad V\left[\frac{X_1 + X_2 + X_3}{3}\right] = \frac{\sigma^2}{3}$$

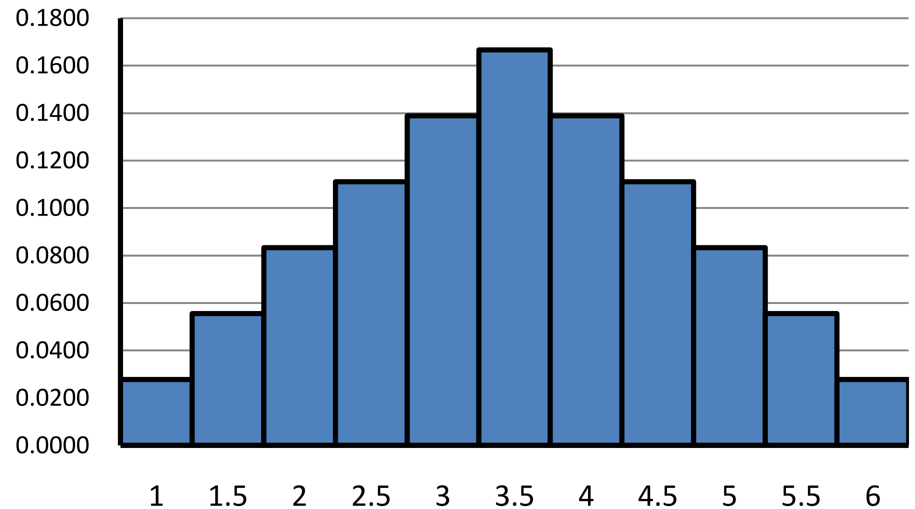
$$E\left[\frac{X_1 + X_2 + \cdots + X_n}{n}\right] = \mu, \quad V\left[\frac{X_1 + X_2 + \cdots + X_n}{n}\right] = \frac{\sigma^2}{n}$$

サイコロの目の平均の確率分布



$$E\left[X\right]=3.5, V\left[X\right]=2.92$$

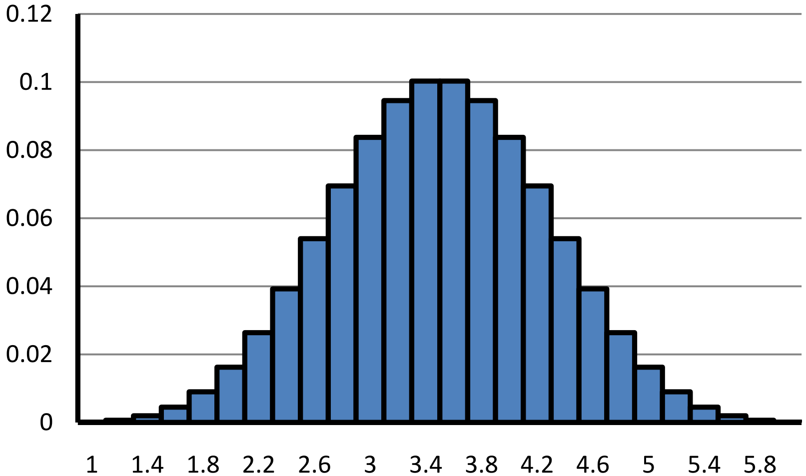
$$\overline{X}=\frac{X_1+X_2}{2}$$



$$E\left[\overline{X}\right]=3.5, V\left[\overline{X}\right]=\frac{2.92}{2}$$

サイコロの目の平均の確率分布

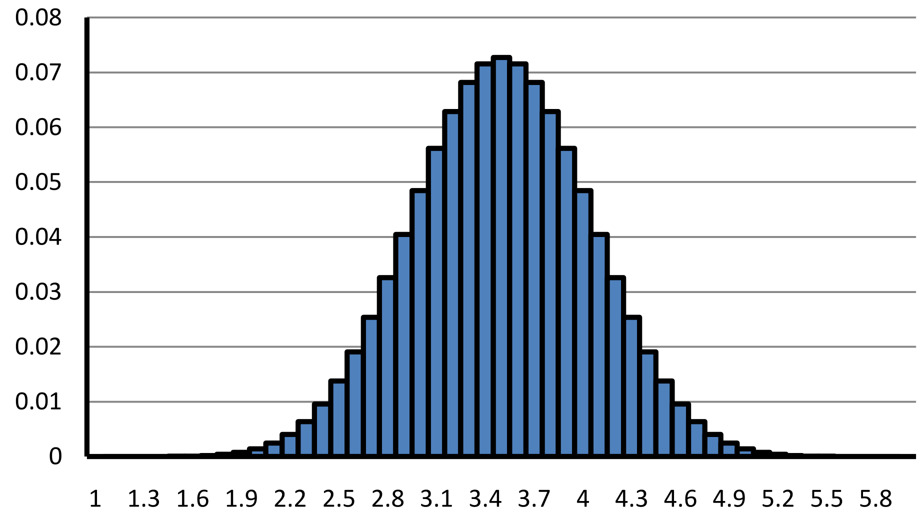
$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_5}{5}$$



$$E\left[\overline{X}\right] = 3.5, V\left[\overline{X}\right] = \frac{2.92}{5}$$

振った回数を増やすとき、左右対称の釣鐘型の分布（正規分布）に近づく。

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_{10}}{10}$$



$$E\left[\overline{X}\right] = 3.5, V\left[\overline{X}\right] = \frac{2.92}{10}$$

正規分布とは、

- 数多くの、互いに影響を受けない、小さな変動が積み重なった結果は正規分布に従う。
- うまく場合分けを行う（異質なグループを混ぜない）こと、あるいは、大きな影響を与える要素を取り除くことにより、正規分布になるようにする。
- 統計では最も重要で、最も利用される分布である。特に、物理学、工学、心理学。
- 身長、知能指数、株価収益率などは正規分布に従うと想定される。

分布が釣鐘型をしていたら（正規分布と見てよいならば）

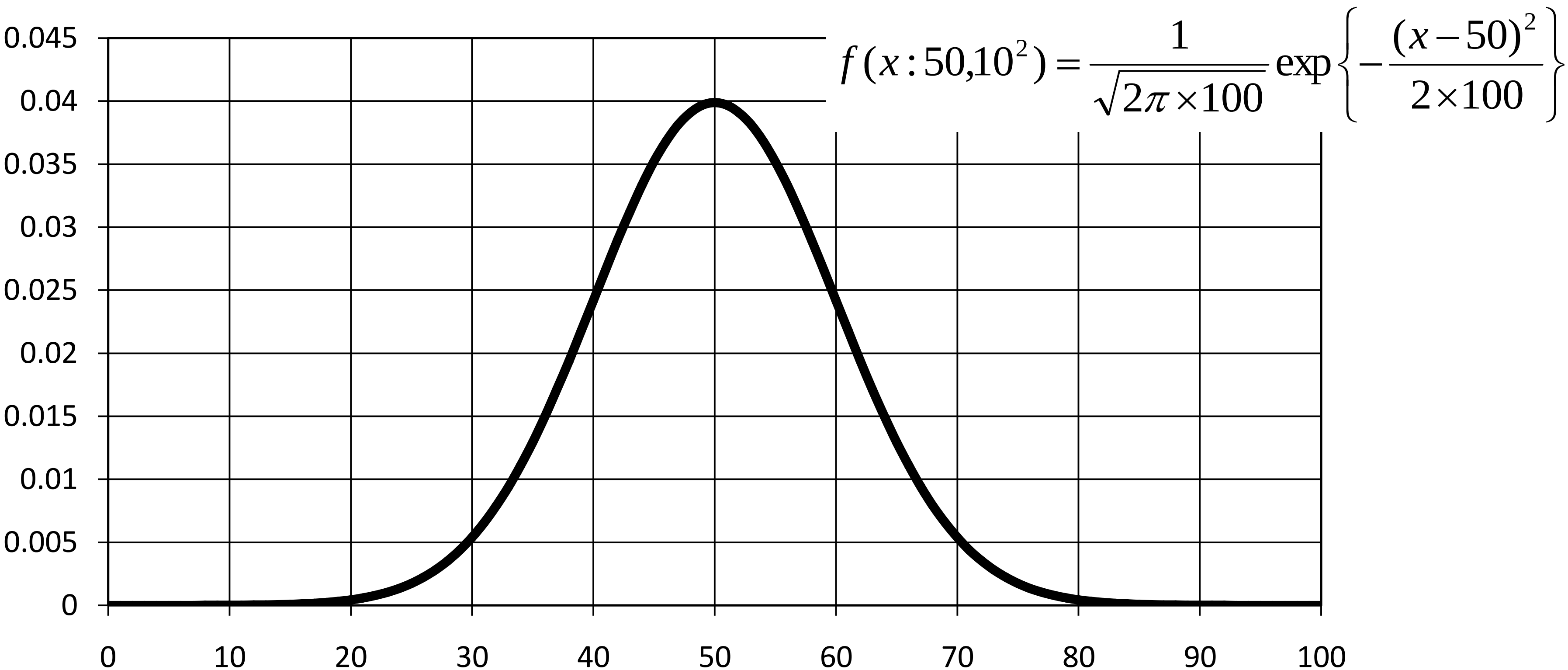
経験的に、以下の法則が成り立つ。

- 平均±標準偏差の範囲には、
  - 全体の約70%のデータが含まれる。
- 平均±２×標準偏差の範囲には、
  - 全体の約95%のデータが含まれる。
- 平均±３×標準偏差の範囲には、
  - 全体の99.7%のデータが含まれる。

正規分布とは、

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

平均50、標準偏差10の正規分布



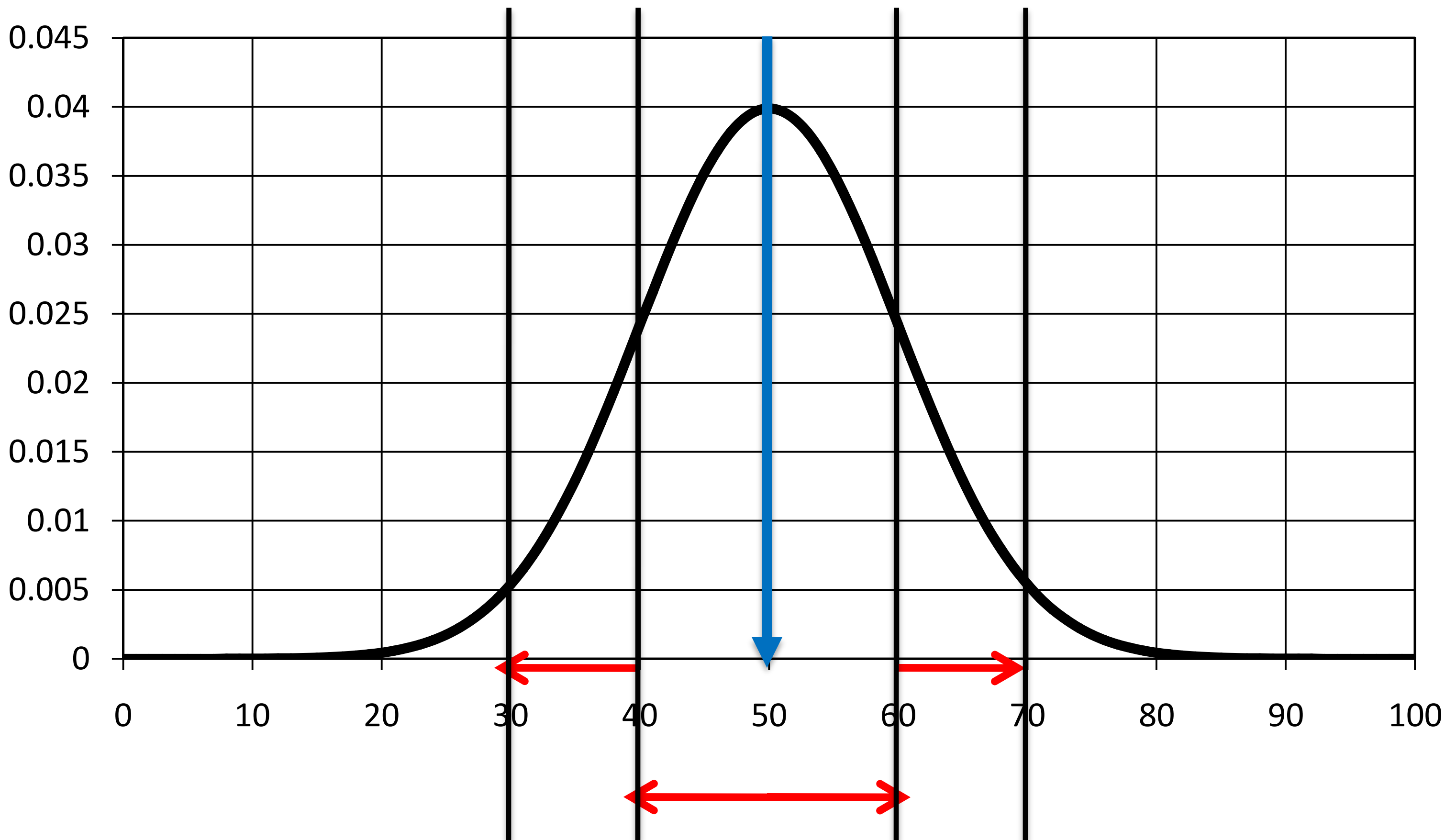
正規分布とは、

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

$$f(x:50,10^2)=\frac{1}{\sqrt{2\pi\times100}}\exp\left\{-\frac{(x-50)^2}{2\times100}\right\}$$

μ σ

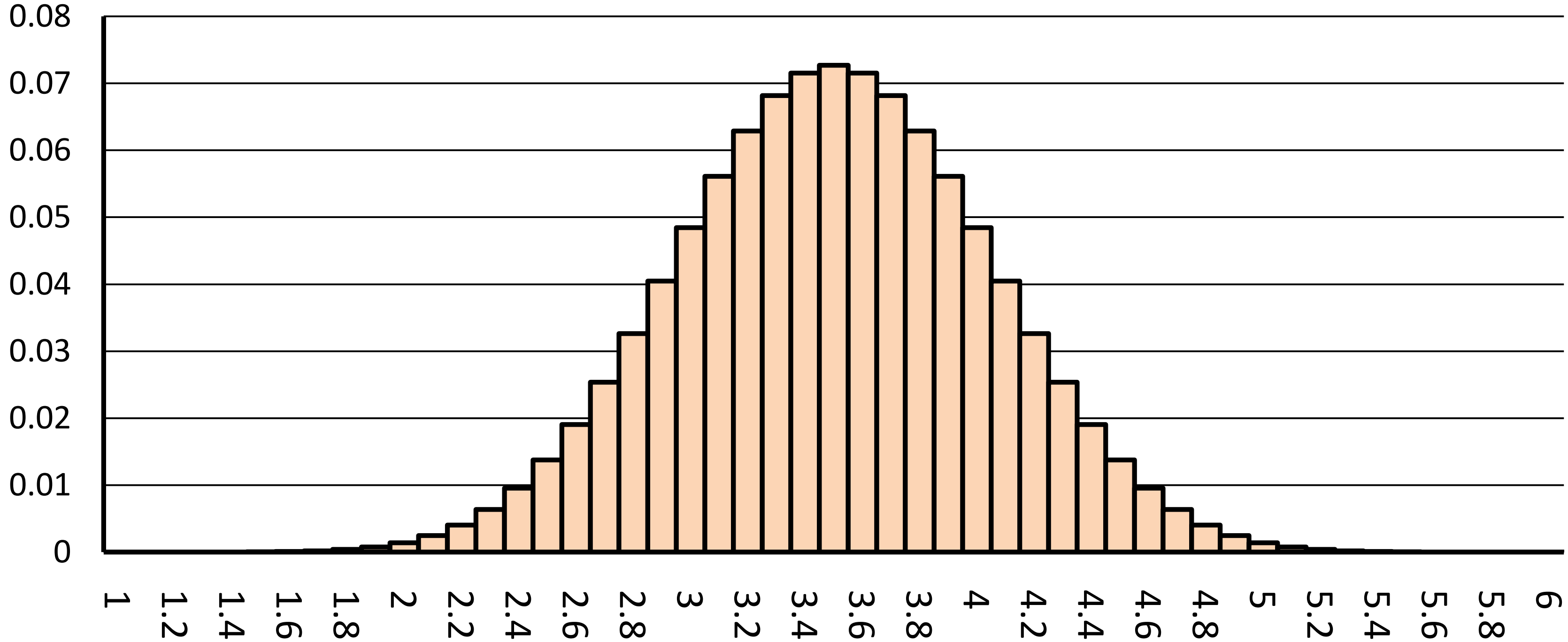
平均50、標準偏差10の正規分布



サイコロの目の平均の確率分布（n＝10）

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_{10}}{10} \qquad \mu_{\overline{X}} = E\left[\overline{X}\right] = 3.5,\ \sigma_{\overline{X}}^2 = V\left[\overline{X}\right] = \frac{2.92}{10},$$

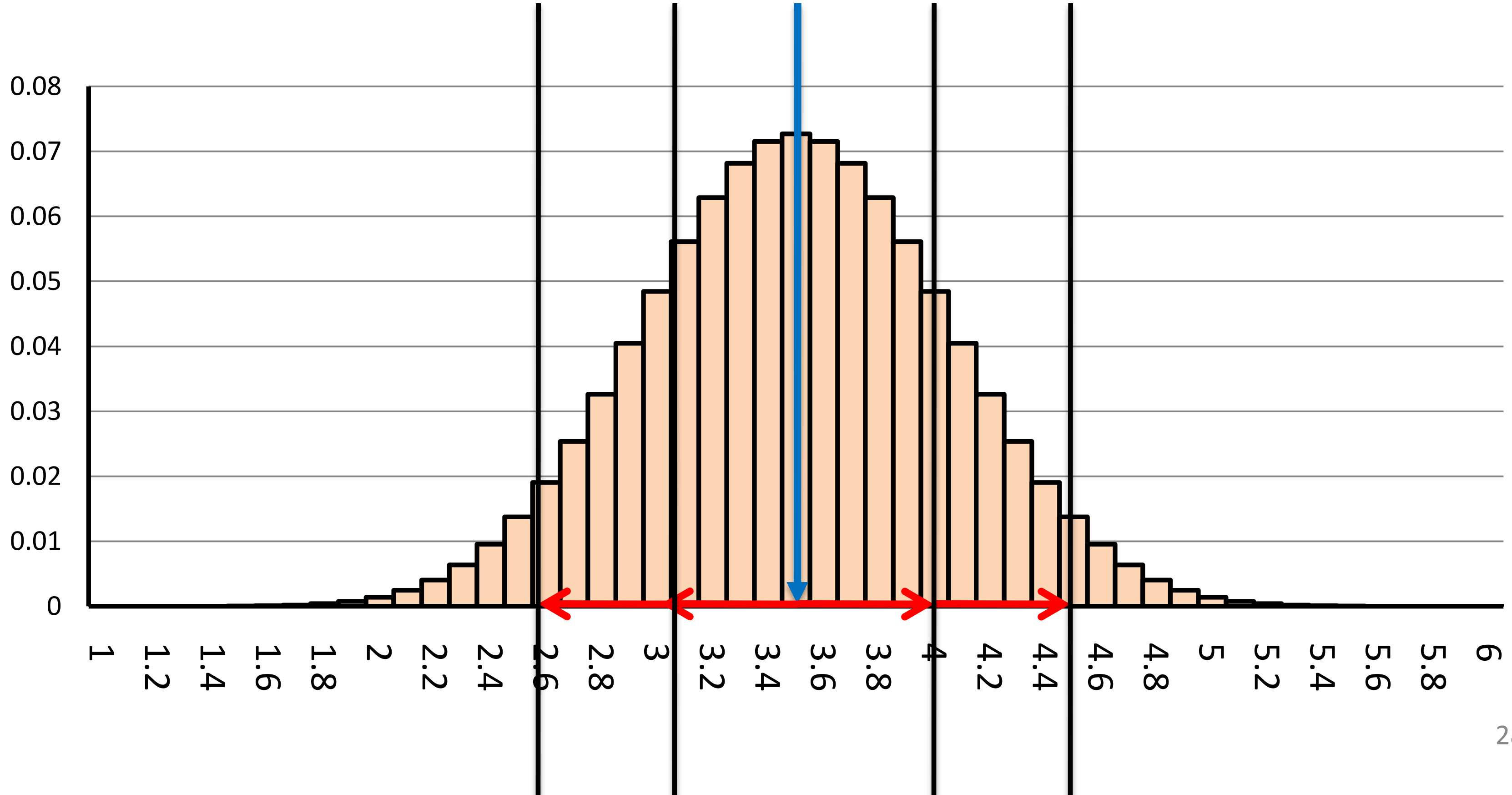
$$\sigma_{\overline{X}} = \sqrt{V\left[\overline{X}\right]} = \sqrt{\frac{2.92}{10}} = 0.540: \text{標準偏差}$$



サイコロの目の平均の確率分布（n＝10）

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_{10}}{10} \qquad \mu_{\overline{X}} = E\left[\overline{X}\right] = 3.5,\ \sigma_{\overline{X}}^2 = V\left[\overline{X}\right] = \frac{2.92}{10},$$

$$\sigma_{\overline{X}} = \sqrt{V\left[\overline{X}\right]} = \sqrt{\frac{2.92}{10}} = 0.540: \text{標準偏差}$$



サイコロを10回投げ、目の平均を求める作業を $\overline{X}$ であらわすとき、

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_{10}}{10}$$

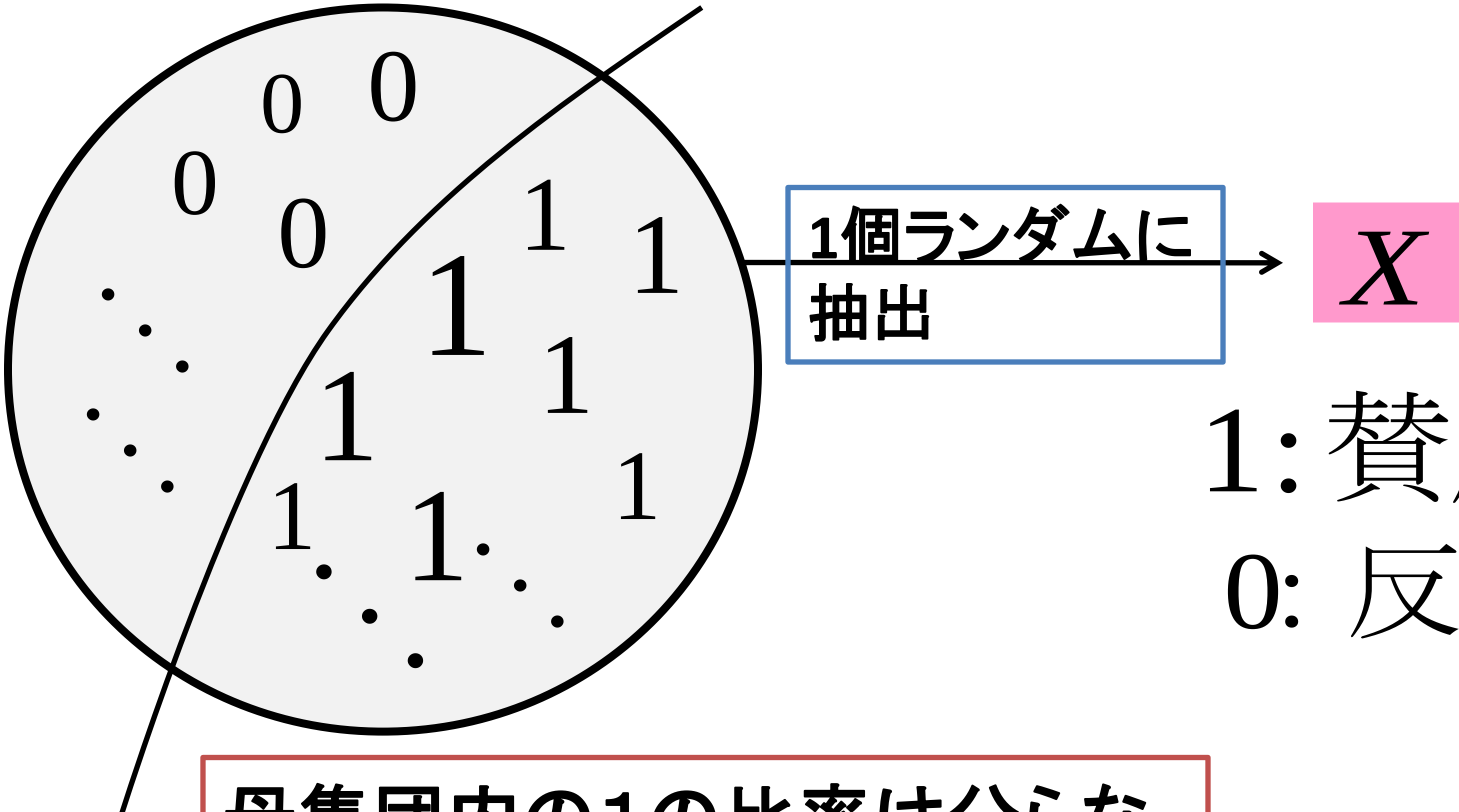
区間(記号)	区間(数値)	確率
$\mu - \sigma \sim \mu + \sigma$		68%
$\mu - 2\sigma \sim \mu + 2\sigma$		95%
$\mu - 3\sigma \sim \mu + 3\sigma$		99.7%

サイコロを10回投げ、目の平均を求める作業を $\overline{X}$ であらわすとき、

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_{10}}{10}$$

区間(記号)	区間(数値)	確率
$\mu - \sigma \sim \mu + \sigma$	$3.5 - 0.54 \sim 3.5 + 0.54$ $2.96 \sim 4.04$	68%
$\mu - 2\sigma \sim \mu + 2\sigma$	$3.5 - 2 \times 0.54 \sim 3.5 + 2 \times 0.54$ $2.42 \sim 4.58$	95%
$\mu - 3\sigma \sim \mu + 3\sigma$	$3.5 - 3 \times 0.54 \sim 3.5 + 3 \times 0.54$ $1.88 \sim 5.12$	99.7%

世論調査に当てはめてみよう



1: 賛成

0: 反対

母集団内の1の比率は分らないので、πとしておく。

反対0賛成1（成功確率π）のときの平均と分散

k	a_k	p_k	$a_k p_k$	$(a_k-\mu)^2 p_k$
1	0	$1-\pi$	0	$(0-\pi)^2\times(1-\pi)$
2	1	π	π	$(1-\pi)^2\times\pi$
$\sum$		1.0	π	$\pi(1-\pi)$

$$\mu = E[X] = \sum_{k=1}^{2} a_k p_k$$

$$\sigma^2 = V[X] = E\left[(X-\mu)^2\right] = \sum_{k=1}^{2} (a_k-\mu)^2 p_k$$

母集団からn個取り出して平均を見る

母集団から1個取り出す作業をXとおくとき、n個取り出す作業は、

$$X_1, X_2, X_3, \cdots, X_n$$

これをもとに、n個の平均を調べる作業は、

$$\overline{X}_n = \frac{X_1 + X_2 + X_3 + \cdots + X_n}{n}$$

と表わされる。

$\overline{X}_n$ の平均と（分散）標準偏差は、

$$E\left[\overline{X}_n\right] = \pi$$

$$V\left[\overline{X}_n\right] = \frac{\pi(1-\pi)}{n}$$

$$\sqrt{V[\overline{X}_n]} = \sqrt{\frac{\pi(1-\pi)}{n}}$$

標準偏差の計算

課題

$\sqrt{\frac{\pi(1-\pi)}{n}}$		n: 標本数				
		500	1000	1500	2000	2500
π	0.3					
	0.4					
	0.5					
	0.6					
	0.7					