講義「確率・統計」
中心極限定理の例

担当
静岡大学工学部機械工学科　岡本正芳

中心極限定理

- 確率変数x_1、…、x_nが互いに独立に同一の確率分布に従うものとし、その平均・分散をμ、σ^2とする。このとき、x_1、…、x_nの平均$x=(x_1+\cdots+x_n)/n$の示す確率分布は、nが十分に大きければ、正規分布$N(\mu,\sigma^2/n)$となる。

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

２項分布

- さいころを６回振って、１の目が何回出るかといったベルヌーイの試行を実行する。
- １の目が出る確率　$p = \dfrac{1}{6}$
- 確率関数

$$p(x) = {}_6C_x p^x (1-p)^{6-x}$$

実際の試行結果の一部

１万回分の結果

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	x_{20}
1	**1**	**0**	**1**	**1**	**0**	**3**	**1**	**0**	**0**	**1**	**0**	**2**	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**0**
2	**1**	**1**	**0**	**0**	**1**	**2**	**2**	**1**	**4**	**2**	**0**	**0**	**2**	**0**	**1**	**2**	**0**	**0**	**1**	**2**
3	**0**	**0**	**0**	**2**	**1**	**0**	**2**	**0**	**0**	**2**	**2**	**0**	**1**	**1**	**1**	**1**	**3**	**0**	**0**	**1**
	⋮																			
498	**1**	**2**	**2**	**0**	**1**	**0**	**1**	**0**	**0**	**0**	**0**	**1**	**1**	**1**	**0**	**0**	**0**	**3**	**0**	**1**
499	**2**	**2**	**1**	**1**	**0**	**1**	**0**	**0**	**2**	**1**	**3**	**1**	**0**	**0**	**0**	**1**	**3**	**1**	**1**	**1**
500	**0**	**1**	**0**	**1**	**2**	**1**	**1**	**1**	**1**	**1**	**0**	**1**	**1**	**0**	**0**	**0**	**1**	**3**	**0**	**1**

１万個のデータとしての処理

平均

$$\overline{x} = 0.9985$$

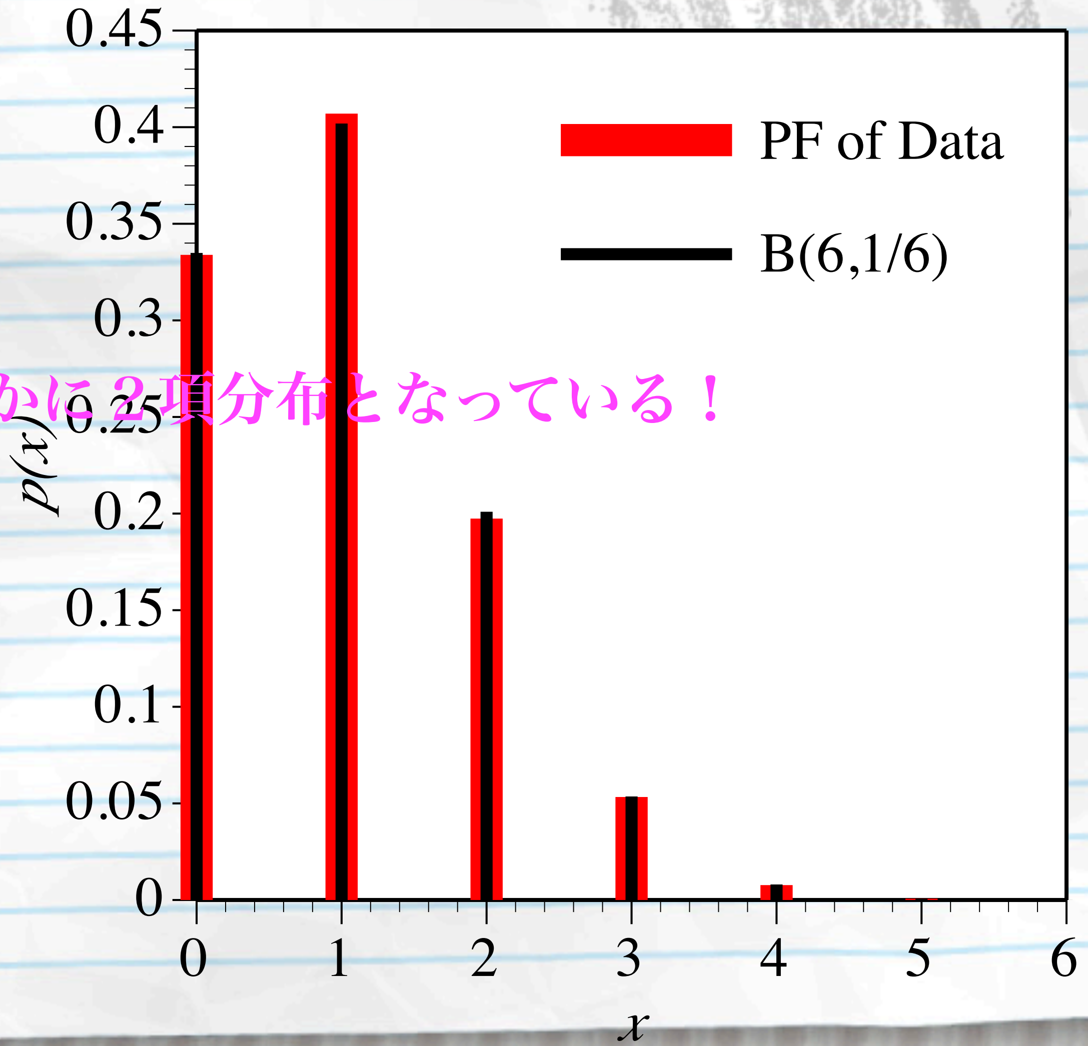
$$\mu = np = 1$$

分散

$$s^2 = 0.824984$$

$$\sigma^2 = np(1-p)$$

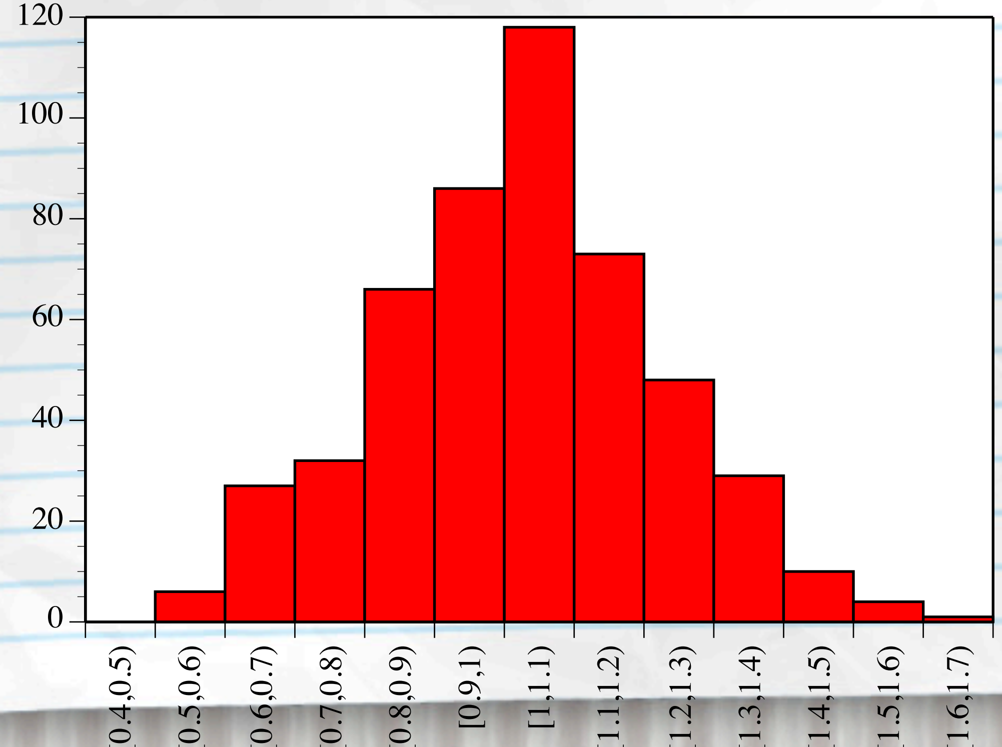
$$= \frac{5}{6} = 0.8333$$



データーは確かに２項分布となっている！

２０データ列としての評価

- 新たな確率変数 $X = \dfrac{1}{20}\sum_{i=1}^{20} x_i$
- 新たなデーター数は５００個ある。



中心極限定理の検証

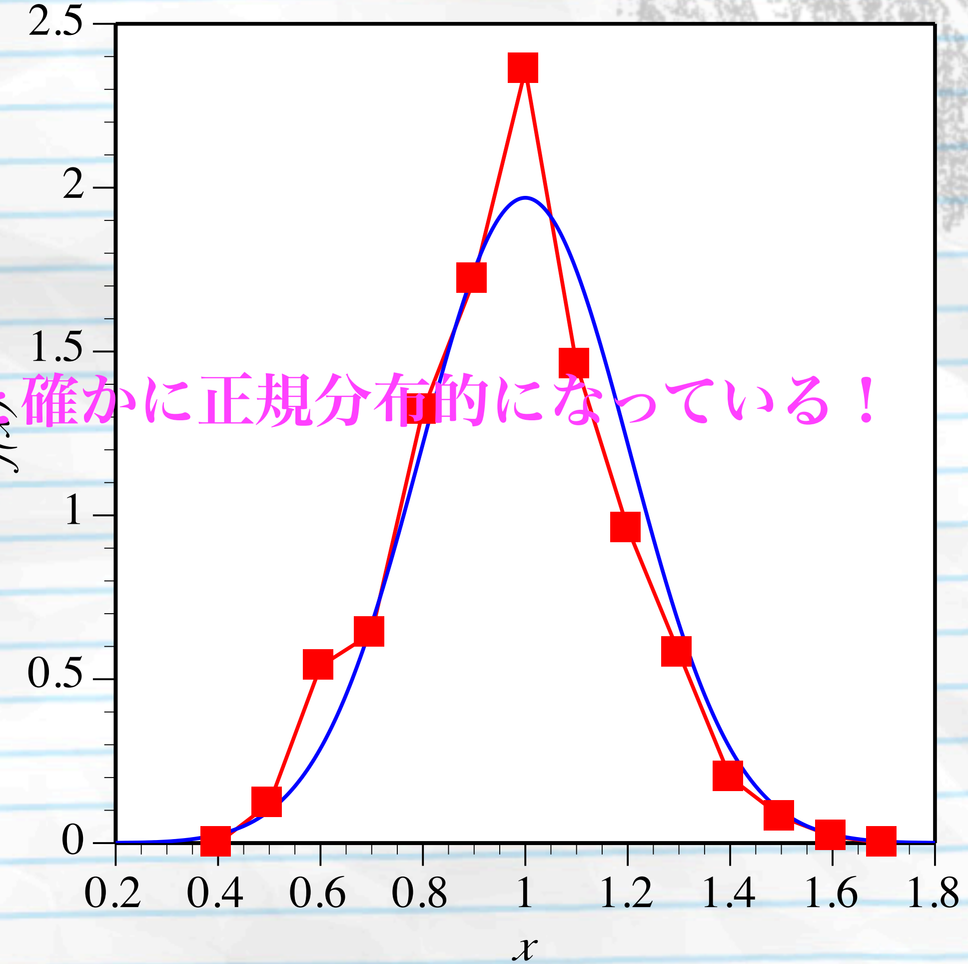
平均

$$\bar{X} = 0.9985$$
$$\mu = 1$$

２項分布が集合すると確かに正規分布的になっている！

分散

$$s^2 = 0.039682$$
$$\sigma^2 / 20 = 0.04166$$



ポアソン分布

- まれな現象を表す確率モデル
- λ= 3 のポアソン分布を考える。
- 確率関数

$$p(x) = \frac{3^x}{x!} e^{-3}$$

実際の試行結果の一部

１万回分の結果

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	x_{20}
1	0	0	3	5	2	4	3	4	2	3	5	3	2	4	2	1	5	3	4	1
2	4	3	4	2	6	0	1	3	2	2	3	5	5	3	4	2	4	2	4	1
3	3	3	3	5	2	5	6	3	0	2	5	3	3	3	4	4	7	6	3	6
													⋮							
498	5	1	3	4	2	4	9	3	5	3	1	0	3	3	1	2	3	3	4	4
499	5	4	2	4	3	5	5	2	3	3	5	3	4	5	2	1	2	2	3	3
500	2	4	1	6	5	0	2	2	2	5	0	2	1	3	1	4	2	3	1	2

1万個のデータとしての処理

平均

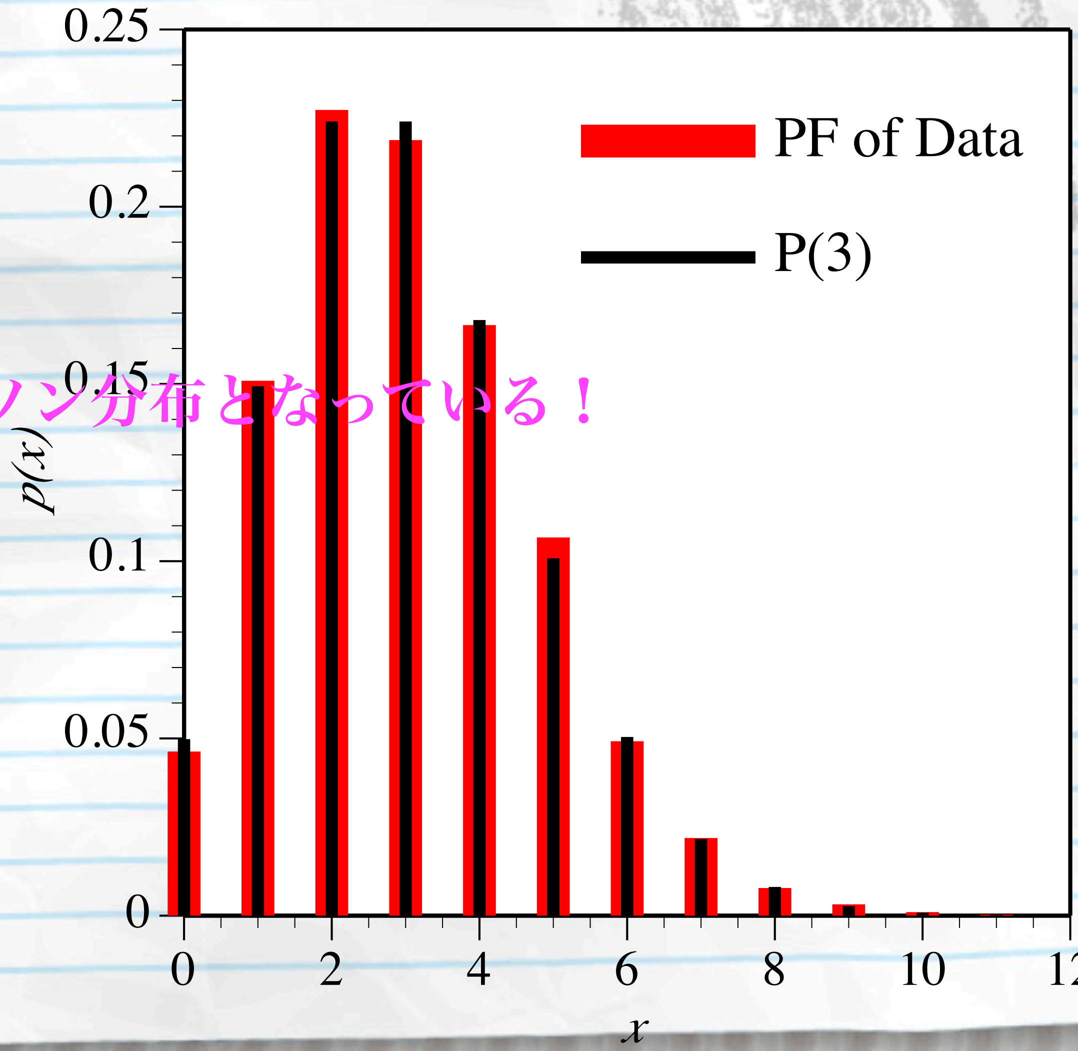
$$\bar{x} = 3.0148$$

$$\mu = \lambda = 3$$

分散

$$s^2 = 3.01238$$

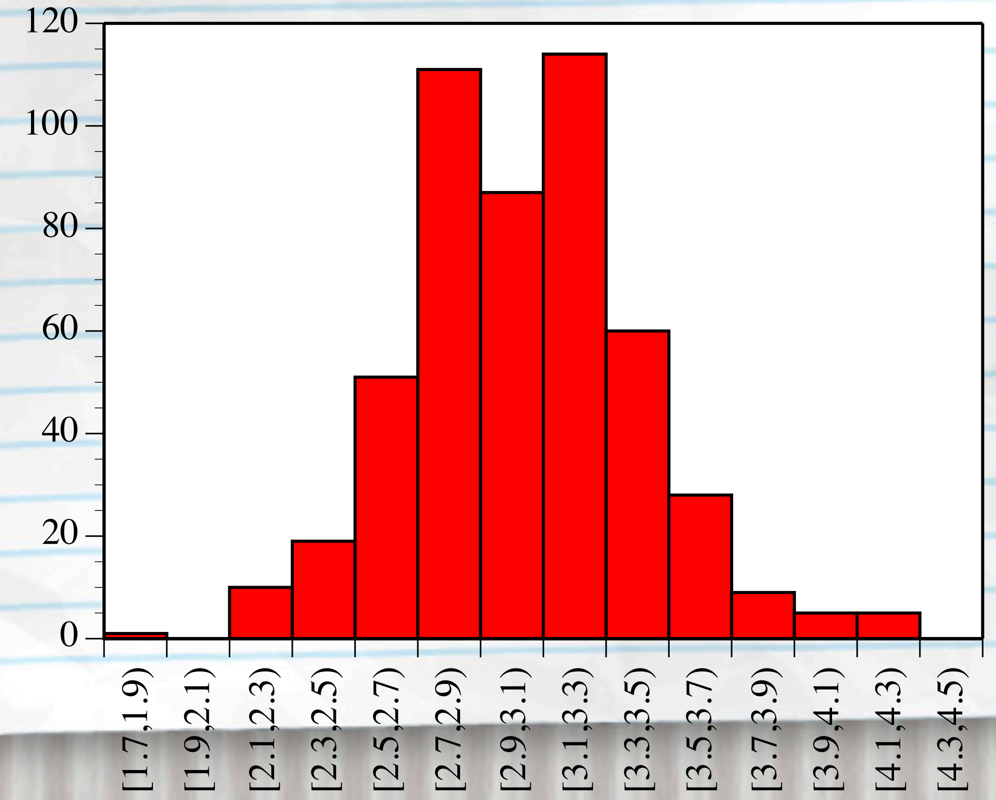
$$\sigma^2 = \lambda = 3$$



確かにポアソン分布となっている！

２０データー列としての評価

- 新たな確率変数　$X = \frac{1}{20}\sum_{i=1}^{20} x_i$
- 新たなデーター数は５００個ある。



中心極限定理の検証

平均

$$\bar{X} = 3.0148$$

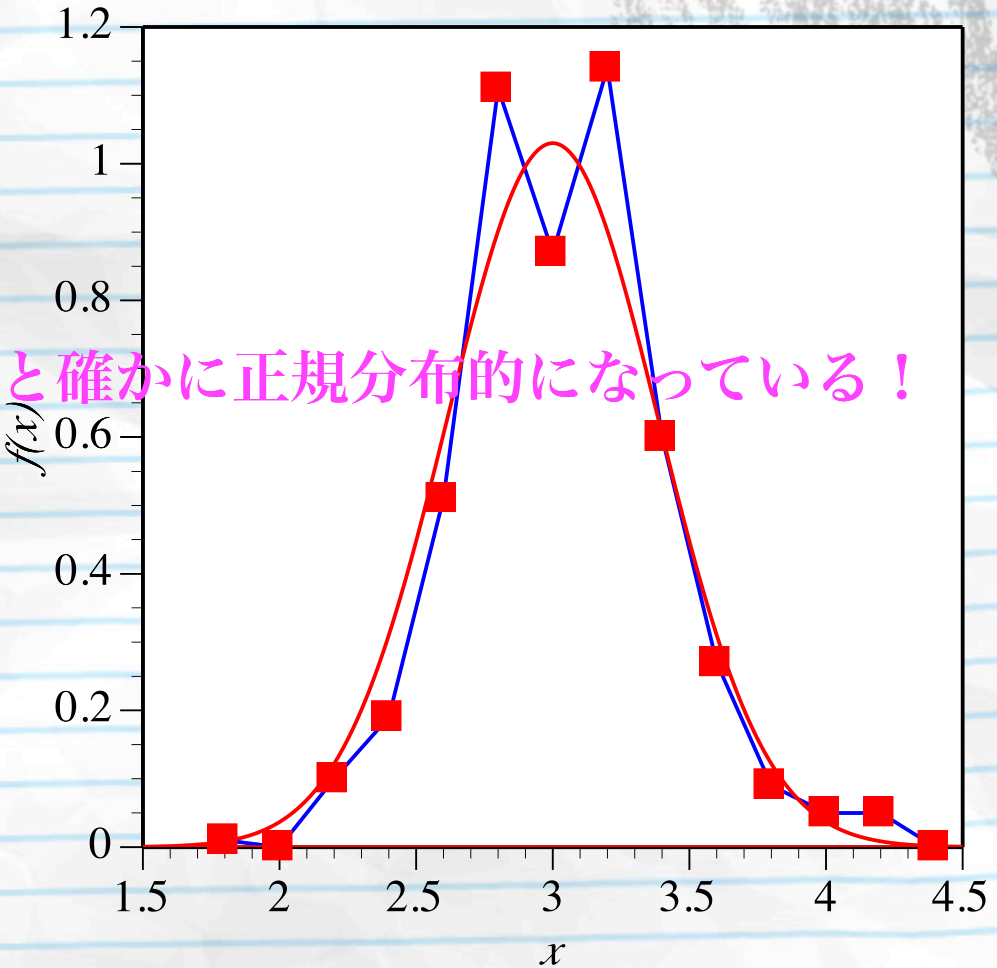
$$\mu = 3$$

ポアソン分布が集合すると確かに正規分布的になっている！

分散

$$s^2 = 0.135651$$

$$\sigma^2 / 20 = 3/20 = 0.15$$



一様分布

- ０から１の間の実数が発生する確率が一様である連続分布
- 確率密度関数

$$f(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & x < 0,\ x > 1 \end{cases}$$

実際の試行結果の一部

２万個の結果

0.2217003726400443
0.5078146234154 7 0
0.9089585021138 19
0.3610263243317 6 0
・
・
・
0.2175363376736 6 4
0.4879718795418 7 4

２万個のデータとしての処理

平均

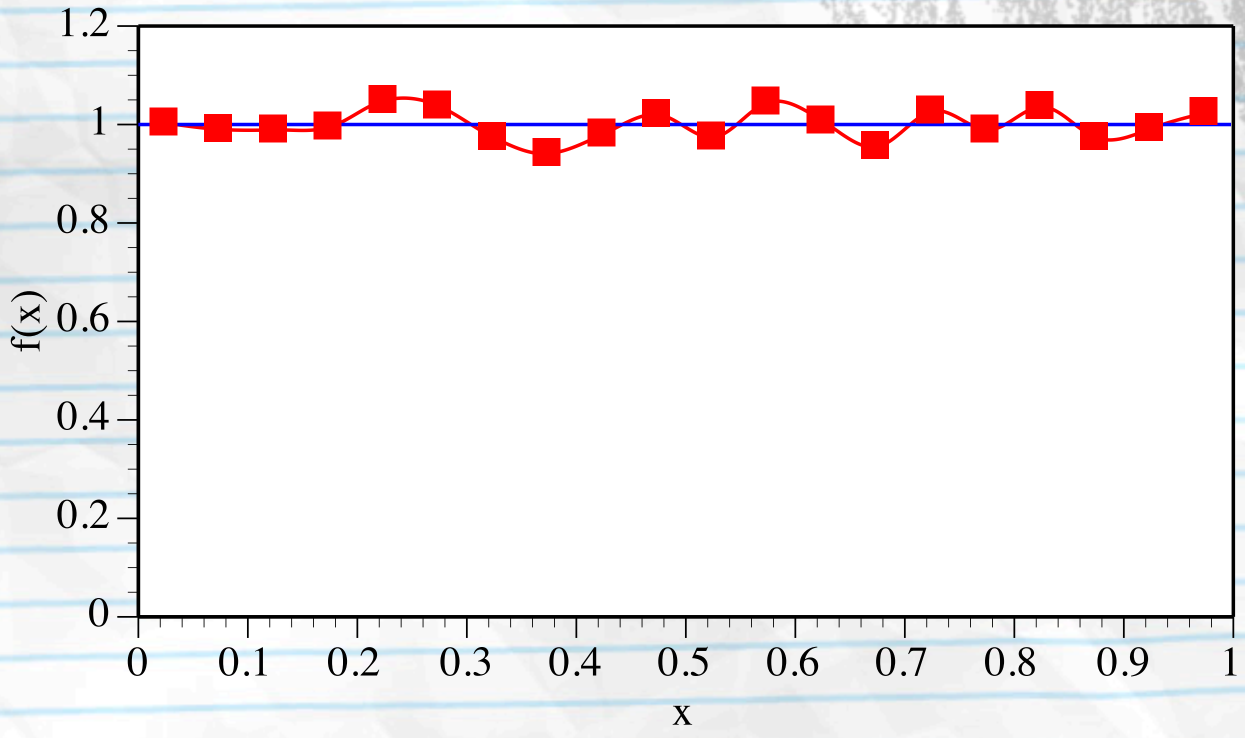
$$\bar{x} = 0.50022$$

$$\mu = 0.5$$

分散

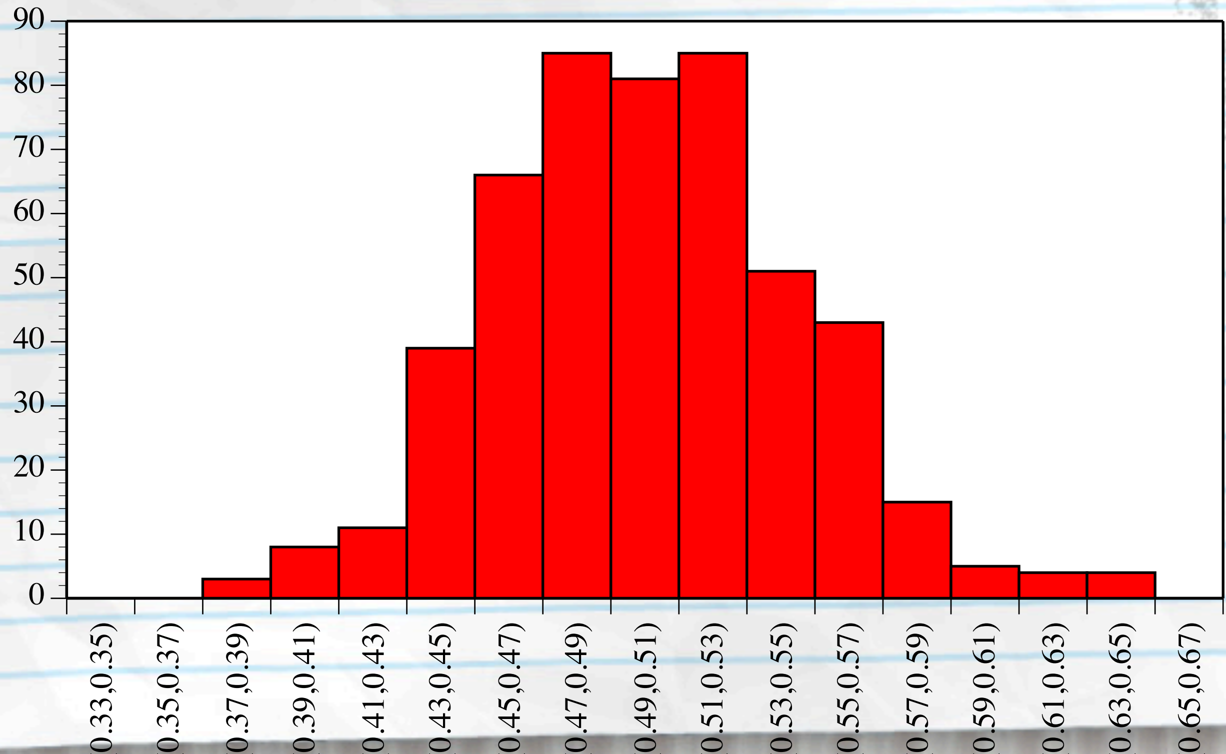
$$s^2 = 0.0835086$$

$$\sigma^2 = 0.08333$$



４０データー列としての評価

- 新たな確率変数 $X = \frac{1}{40}\sum_{i=1}^{40} x_i$
- 新たなデーター数は５００個ある。



中心極限定理の検証

平均

$$\bar{X} = 0.50022$$

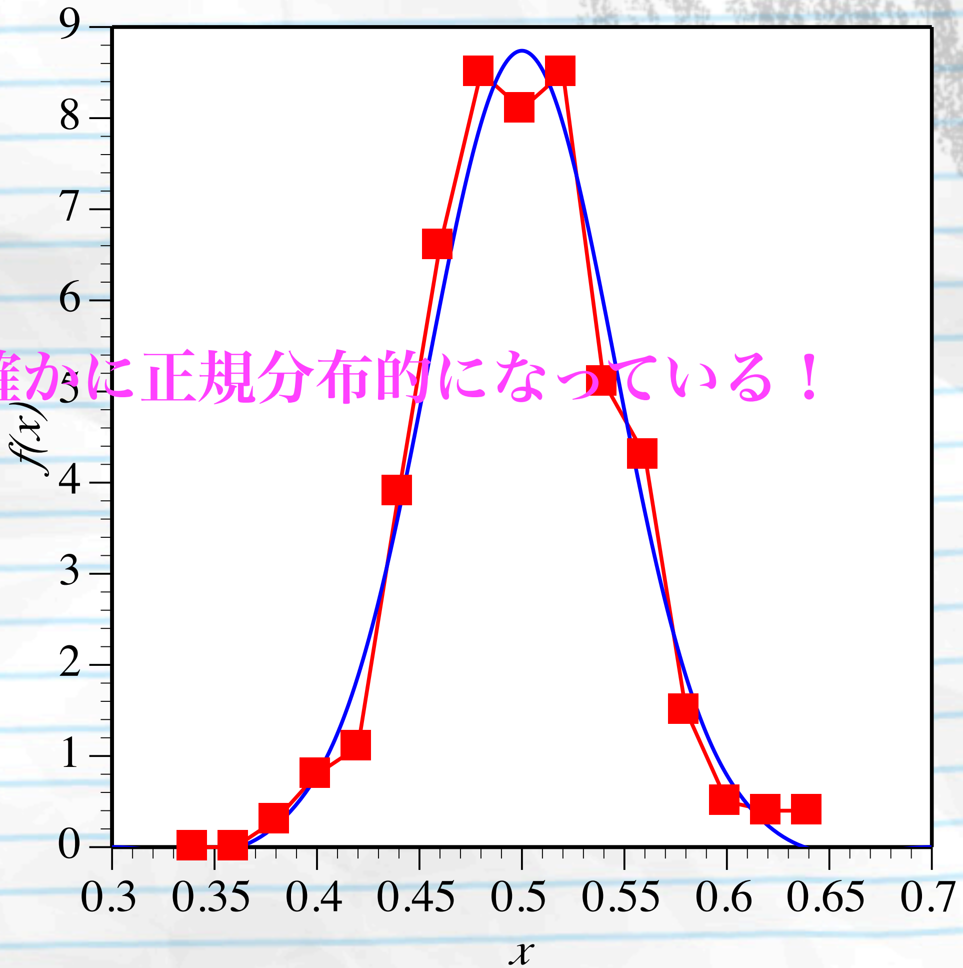
$$\mu = 0.5$$

分散

$$s^2 = 0.00198816$$

$$\sigma^2 / 40 = 0.002083$$

一様分布が集合すると確かに正規分布的になっている！



n に対する依存性

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

nが増加していくと分布幅が狭まって平均μの値を示すδ関数へと移行する