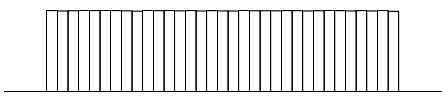
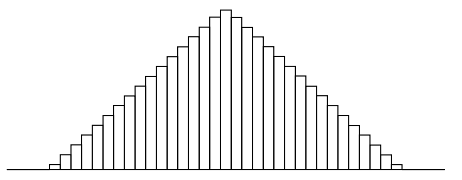
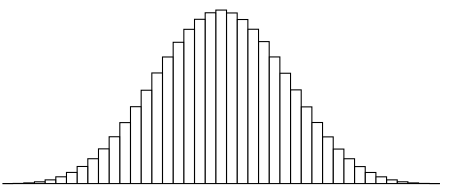
§4.3　中心極限定理

一様分布に従う確率変数の実現値を独立に何回も調べてヒストグラムを作ると例えば右図のようになります．この一様分布に従う独立な2個の確率変数 X_1, X_2 の平均 $\dfrac{X_1+X_2}{2}$ の実現値を独立に何回も調べてヒストグラムを作ると例えば下の左の図のようになります．また，この一様分布に従う独立な4個の確率変数 X_1, X_2, X_3, X_4 の平均 $\dfrac{X_1+X_2+X_3+X_4}{4}$ の実現値を独立に何回も調べてヒストグラムを作ると例えば下の右の図のようになります（図によって尺度は異なります）．



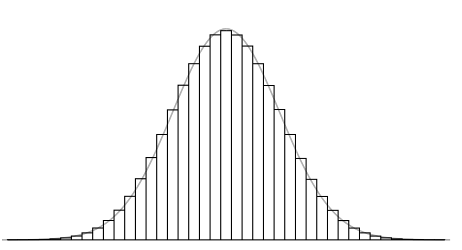


2個の確率変数の平均の実現値の分布

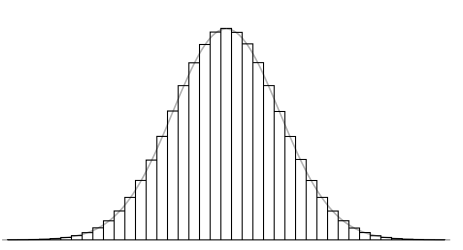


4個の確率変数の平均の実現値の分布

更に，この一様分布に従う独立な10個の確率変数 $X_1, X_2, X_3, \ldots, X_{10}$ の平均 $\dfrac{X_1+X_2+X_3+\cdots+X_{10}}{10}$ の実現値および100個の確率変数 $X_1, X_2, X_3, \ldots, X_{100}$ の平均 $\dfrac{X_1+X_2+X_3+\cdots+X_{100}}{100}$ の実現値の各々を独立に何回も調べてヒストグラムを作ると例えば下の左右の図のようになります（図によって尺度は異なります）．



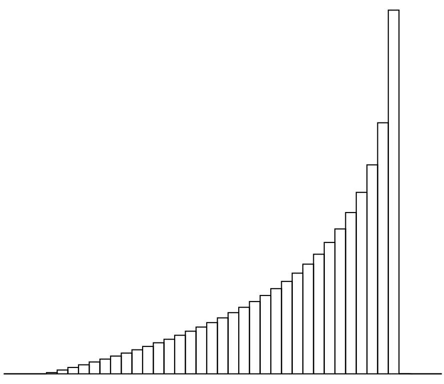
10個の確率変数の平均の実現値の分布

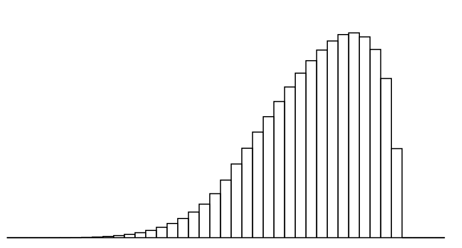


100個の確率変数の平均の実現値の分布

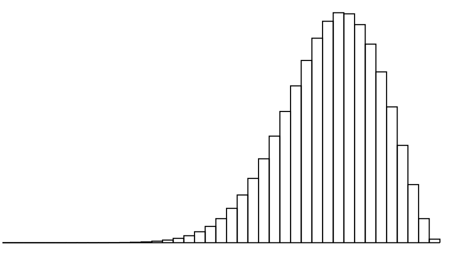
網目で描かれたグラフは正規分布に従う確率変数の実現値を独立に何回も調べてできる度数分布のグラフです．

もう一つ例を示します．ある確率変数の実現値を独立に何回も調べてヒストグラムを作ると例えば右図のような偏った分布になるとします．この確率分布に従う独立な2個の確率変数 X_1, X_2 の平均 $\dfrac{X_1+X_2}{2}$ の実現値を独立に何回も調べてヒストグラムを作ると例えば下の左の図のようになります．また，この確率分布に従う独立な4個の確率変数 X_1, X_2, X_3, X_4 の平均 $\dfrac{X_1+X_2+X_3+X_4}{4}$ の実現値を独立に何回も調べてヒストグラムを作ると例えば下の右の図のようになります（図によって尺度は異なります）．



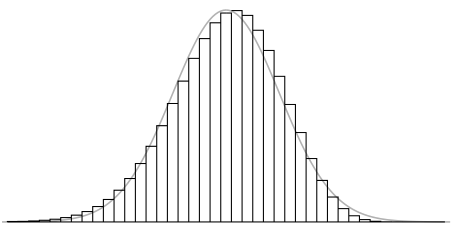


2個の確率変数の平均の実現値の分布

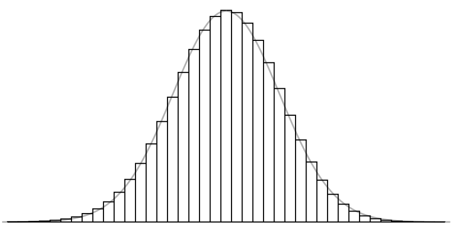


4個の確率変数の平均の実現値の分布

更に，この偏った分布に従う独立な10個の確率変数 $X_1, X_2, X_3, \ldots, X_{10}$ の平均 $\dfrac{X_1+X_2+X_3+\cdots+X_{10}}{10}$ の実現値および100個の確率変数 $X_1, X_2, X_3, \ldots, X_{100}$ の平均 $\dfrac{X_1+X_2+X_3+\cdots+X_{100}}{100}$ の実現値の各々を独立に何回も調べてヒストグラムを作ると例えば下の左右の図のようになります（図によって尺度は異なります）．



10個の確率変数の平均の実現値の分布



100個の確率変数の平均の実現値の分布

網目で描かれたグラフは正規分布に従う確率変数の実現値を独立に何回も調べてできる度数分布のグラフです．

これらの例を見て分かるように一般的に次のことが成り立ちます：

同じ分布に従う多数の独立な確率変数の平均は近似的に正規分布に従う．

このことをはっきり述べたのが次の中心極限定理 (central limit theorem) です．その証明は面倒なので省略します．

定理 4.3（中心極限定理）　自然数 n は充分大きいとする．互いに独立な n 個の確率変数 $X_1, X_2, X_3, \cdots, X_n$ が総て同じ確率分布に従い，その確率分布の平均値 μ と標準偏差 σ とがあるとき，$X_1, X_2, X_3, \cdots, X_n$ の平均 $\dfrac{X_1+X_2+X_3+\cdots+X_n}{n}$ は近似的に正規分布 $\mathrm{N}\left(\mu, \dfrac{\sigma^2}{n}\right)$ に従う．

$\boxed{\text{例題 4.3}}$　100個の確率変数 $X_1, X_2, X_3, \cdots, X_{100}$ が互いに独立で，総て1から9までの一様分布に従うものとする．これらの平均を表す確率変数を $\overline{X}$ とおく：

$$\overline{X} = \frac{X_1+X_2+X_3+\cdots+X_{100}}{100} .$$

この確率変数 $\overline{X}$ について，確率 $\mathrm{P}\left(\overline{X} \leq 5.2\right)$ を近似的に求める．

確率変数 $X_1, X_2, X_3, \cdots, X_{100}$ は互いに独立で，総て1から9までの一様分布に従うので，

$$\mathrm{E}[X_1] = \mathrm{E}[X_2] = \mathrm{E}[X_3] = \cdots = \mathrm{E}[X_{100}] = \frac{1+9}{2} = 5 ,$$
$$\mathrm{V}[X_1] = \mathrm{V}[X_2] = \mathrm{V}[X_3] = \cdots = \mathrm{V}[X_{100}] = \frac{(9-1)^2}{12} = \frac{16}{3} .$$

中心極限定理より，確率変数 $\overline{X}$ は近似的に正規分布 $\mathrm{N}\left(5, \dfrac{16}{300}\right) = \mathrm{N}\left(5, \dfrac{4}{75}\right)$ に従う．よって，確率変数

$$Z = \frac{\overline{X}-5}{\sqrt{\dfrac{4}{75}}} = \frac{5\sqrt{3}}{2}\left(\overline{X}-5\right)$$

は近似的に標準正規分布 $\mathrm{N}(0,1)$ に従う．標準正規分布の累積分布関数を $\Phi(x)$ とおく．

$$\overline{X} \leq 5.2 \iff \frac{5\sqrt{3}}{2}\left(\overline{X}-5\right) \leq \frac{5\sqrt{3}}{2}(5.2-5) \iff Z \leq \frac{\sqrt{3}}{2} .$$

よって

$$\mathrm{P}\left(\overline{X} \leq 5.2\right) = \mathrm{P}\left(Z \leq \frac{\sqrt{3}}{2}\right) \fallingdotseq \Phi\left(\frac{\sqrt{3}}{2}\right) \fallingdotseq \Phi(0.866)$$
$$\fallingdotseq 0.803 .$$

$\boxed{\text{問題 4.3}}$　150個の確率変数 $X_1, X_2, X_3, \cdots, X_{150}$ が互いに独立で，総てポアソン分布 $\mathrm{Po}(6)$ に従うものとします．これらの平均を表す確率変数を $\overline{X}$ とおきます：

$$\overline{X} = \frac{X_1+X_2+X_3+\cdots+X_{150}}{150} .$$

この確率変数 $\overline{X}$ について，確率 $\mathrm{P}\left(\overline{X} \geq 5.7\right)$ を近似的に求めなさい．