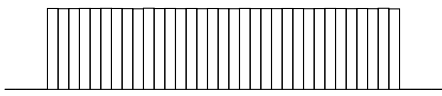


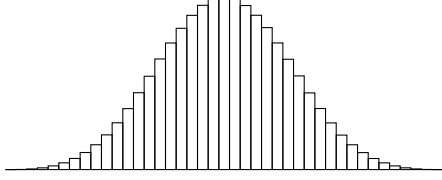
§ 4.3 中心極限定理

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る． 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 $\frac{X_1 + X_2}{2}$ 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る．

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 $\frac{X_1 + X_2 + X_3 + X_4}{4}$ 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る（図によって尺度は異なる）．

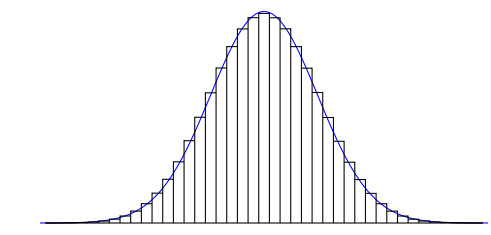
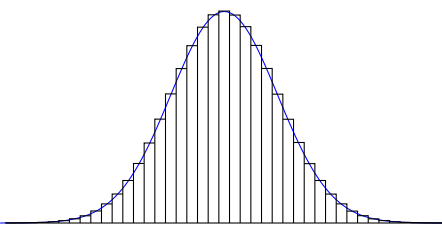


2 個の確率変数の平均の実現値の分布

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 $X_1, X_2, X_3, \dots, X_{10}$ の平均 $\frac{X_1 + X_2 + X_3 + \dots + X_{10}}{10}$ の実現値および 100 個の確率変数 $X_1, X_2, X_3, \dots, X_{100}$ の平均 $\frac{X_1 + X_2 + X_3 + \dots + X_{100}}{100}$ 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る（図によって尺度は異なる）． 青色の曲線

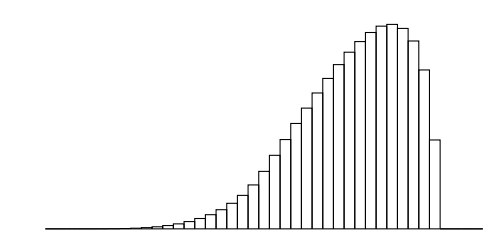
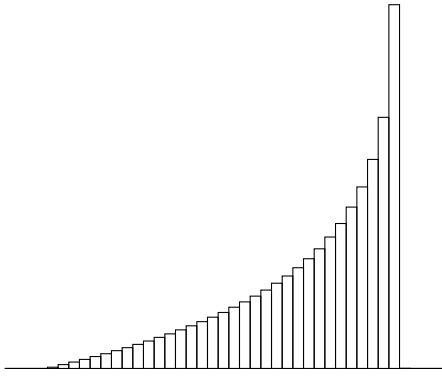
は正規分布の確率密度関数のグラフである．



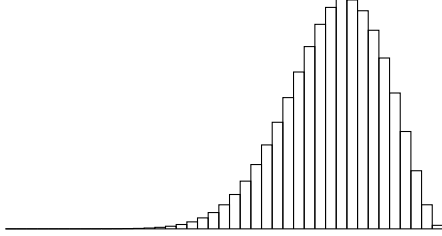
10 個の確率変数の平均の実現値の分布

100 個の確率変数の平均の実現値の分布

もう一つ例を示す． 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する． 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 $\frac{X_1 + X_2}{2}$ 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る． 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 $\frac{X_1 + X_2 + X_3 + X_4}{4}$ 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る（図によって尺度は異なる）．



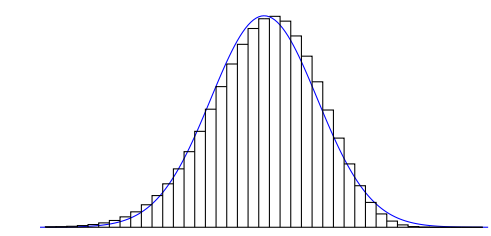
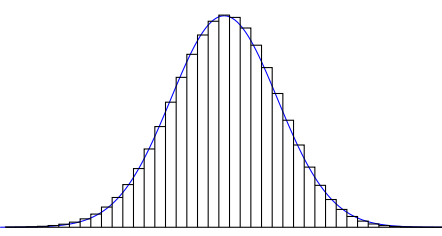
2 個の確率変数の平均の実現値の分布



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 $X_1, X_2, X_3, \dots, X_{10}$ の平均 $\frac{X_1 + X_2 + X_3 + \dots + X_{10}}{10}$ の実現値および 100 個の確率変数 $X_1, X_2, X_3, \dots, X_{100}$ の平均 $\frac{X_1 + X_2 + X_3 + \dots + X_{100}}{100}$ 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る（図によって尺度は異なる）． 青色の曲線

は正規分布の確率密度関数のグラフである．



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

このように、同じ分布に従う多数の互いに独立な確率変数の平均は近似的に正規分布に従うことが多い． このことを述べたのが次の**中心極限定理** (central limit theorem) である． その証明は省く．

[中心極限定理] 自然数 n は充分大きいとする． ある一つの分布に従う n 個の確率変数 $X_1, X_2, X_3, \dots, X_n$ が互いに独立でありそれらの平均値 μ と標準偏差 σ とがあるならば、 $X_1, X_2, X_3, \dots, X_n$ の平均を表す確率変数 $\frac{1}{n} \sum_{k=1}^n X_k$ は近似的に正規分布 $N\left(\mu, \frac{\sigma^2}{n}\right)$ に従う．

中心極限定理における前提は、充分大きい自然数 n に対して、互いに独立な n 個の確率変数 $X_1, X_2, X_3, \dots, X_n$ が同じ分布に従い、その分布の平均値と標準偏差とがあるということなので、その分布は平均値と標準偏差とがあればどのような分布でもよいことが重要である．

正の自然数 n に対して、ある一つの分布に従う n 個の確率変数 $X_1, X_2, X_3, \dots, X_n$ が互いに独立でありそれらの平均値 μ と標準偏差 σ とがあるとき、 $X_1, X_2, X_3, \dots, X_n$ の平均を表す確率変数 $\frac{1}{n} \sum_{k=1}^n X_k$ の分布は、 n が 100 以上あれば中心極限定理により正規分布 $N\left(\mu, \frac{\sigma^2}{n}\right)$ で近似することが多いが、 $X_1, X_2, X_3, \dots, X_n$ が従う分布が正規分布に近いときは n が 50 ぐらいでも中心極限定理によって正規分布で近似することがある．

【例題】 1 から 9 までの一様分布に従う 100 個の確率変数 $X_1, X_2, X_3, \dots, X_{100}$ が互いに独立であるとする． これらの平均を表す確率変数を $\overline{X}$ とおく：

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

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

確率変数 $X_1, X_2, X_3, \dots, X_{100}$ 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、

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

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

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

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

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

よって

$$P(\overline{X} \leq 5.2) = P\left(Z \leq \frac{\sqrt{3}}{2}\right) \doteq \Phi\left(\frac{\sqrt{3}}{2}\right) \doteq \Phi(0.866) \doteq 0.803 . \quad \boxed{\text{終}}$$

【問4.3】 ポアソン分布 $Po(6)$ に従う 150 個の確率変数 $X_1, X_2, X_3, \dots, X_{150}$ が互いに独立であるとする． これらの平均を表す確率変数を $\overline{X}$ とおく：

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

この確率変数 $\overline{X}$ について、確率 $P(\overline{X} \geq 5.7)$ を近似的に求めよ．