Hand out

確率論 (*Probability*) 第 16 週

due	24 juillet 2014
cur	24 juillet 2014
ver	1
rev	2

w.yamamoto

1 確率分布間の関係

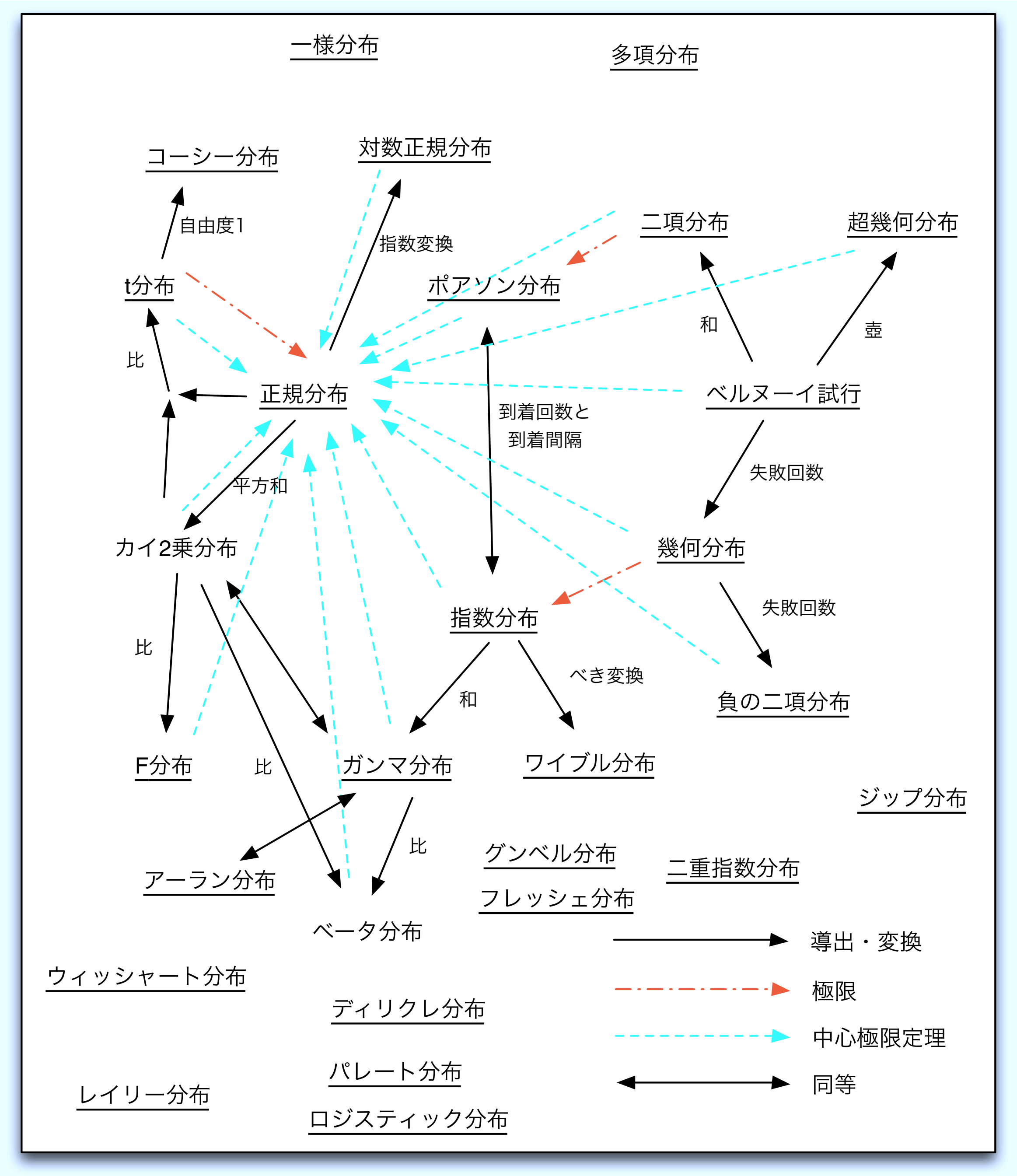

図 1: 確率分布間の関係

確率論や統計学の講義で登場する主な確率分布の間には、互いの導出や変換に関して、図 1 のような関係がある。この図には、講義で紹介した関係も、紹介できなかった関係も含まれている。主な関係を次に掲げておく。

和 $\sum_{i=1}^{n} X_i$ (平均 $\sum_{i=1}^{n} X_i/n$ は、和の $1/n$ 倍)

平方和 $\sum_{i=1}^{n} X_i^2$ (偏差平方和 $\sum_{i=1}^{n} (X_i - \mu)^2$ は平均からの差の平方和、分散は偏差平方和の $1/n$ 倍)

比 X_1/X_2 の分布、構成比 $X_1/(X_1+X_2)$ の分布

最小値 $\max_i\{X_i\}$ の分布、最大値 $\min_i\{X_i\}$ の分布

離散分布はベルヌーイ試行から出発して様々な確率分布が導出される。連続分布は、正規分布から出発することが多いが、正規分布以外にも多くの確率分布が提案されている。

正規分布は、多くの確率分布が中心極限定理を経て辿り着く先として、重要である。講義の範囲からは外れるが、グンベル分布とフレッシェ分布とワイブル分布も、多くの確率分布が極地理論によって辿り着く先として、重要である。

また一様分布と二項分布と多項分布は、どのような確率分布からも変換を用いて導出できる。

一様分布 累積分布関数の逆関数による変換 $U=F^{-1}(X)$ の分布

二項分布 確率変数がある値を超えるか否かの二値変換 $Y=I(X\geq a), a\in\mathcal{X}$

多項分布 標本空間の有限個の分割を $\mathcal{X}=A_1\cup A_n\cup\cdots\cup A_K, A_k\cap A_{k'}=\phi$ と与えたときの、確率変数が入った部分空間の番号を与える多値変換 $Y=\sum_{k=1}^{K}kI(X\in A_k)$

2 確率分布表

確率分布について、幾つかの表がよく用いられる。参考書の p.196 にある表 A.1 は、標準正規分布 $N\left(0,1^2\right)$ の確率表である。この表には、表頭と表側から小数点以下 3 桁の数値 α を、表の中身から標準正規分布の標本空間上の点 u を、読み取ることができる。

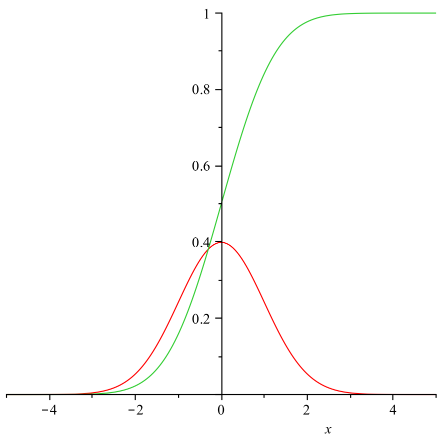

図 2: 標準正規分布の密度関数と累積分布関数

図 2 は、標準正規分布の密度関数と累積分布関数が示してある。表 A.1 の u は、図 2 の塗りつぶされた部分の端点であり、α は塗りつぶされた部分の面積 (確率) $Pr[X\geq u]$ である。上の方から塗りつぶしたので、$\alpha=1-F(u)$ となる。

$$\alpha=Pr[X>u]=1-Pr[X\leq u]=1-F(u)$$

この表は、正規分布が位置変換 (平均の移動) と尺度変換 (分散の拡大縮小) について共変となること、また正規分布が平均に関して対象であること、を利用して用いる。

- <u>$N(0,1)$ で X の上側 α %点を求めよ</u> とは、

  $$Pr[X>u]=1-F(x)=\alpha/100$$

  となる u を求めることである。例えば $\alpha=5\%$ のとき、表の中で $\alpha=0.050$ となる値を読み取れば、

  $$Pr[X>1.64485]=0.050$$

  と分かる。

- $N(0,1)$ で X の下側 α %点を求めよ とは、正規分布の平均に関する対称性を用いれば、常に

$$Pr[X \le u] = Pr[X > -u]$$

が成り立つ。これを用いて、$Pr[X > u] = \alpha/100$ となる u を求めてから、$-u$ を解答すれば良い。例えば $\alpha = 5\%$ のとき、表の中で $\alpha = 0.050$ となる値を読み取れば、

$$Pr[X > 1.64485] = 0.050$$

と分かり、下側 α % 点は

$$Pr[X \le -1.64485] = 0.050$$

より、-1.64485 と得る。

- $N(0,1)$ で $X > 3$ となる確率 は、表の中で 3 に最も近いセルを探すと、

$$Pr[X > 2.98916] = 0.002$$
$$Pr[X > 3.09023] = 0.001$$

が分かる。この 2 点から、線形補間をして $Pr[X > 3]$ を求める。

- $N(0,1)$ で $X \le -2$ となる確率 は、正規分布が平均について対象であること、連続確率分布であること、の 2 点から、$Pr[X > 2]$ と等しい。表の中で 2 に最も近いセルを探すと、

$$Pr[X > 2.01409] = 0.022$$
$$Pr[X > 1.99539] = 0.023$$

が分かる。この 2 点から、線形補間をして $Pr[X > 2]$ を求め、$Pr[X < -2]$ とする。

- 平均を跨ぐ区間については、例えば $N(0,1)$ に従う確率変数 X について、$X \le 2$ となる確率は

$$Pr[X \le 2] = 1 - Pr[X > 2]$$

また $X > -1$ となる確率は

$$Pr[X > -1] = 1 - Pr[X > 1]$$

などと求める。

平均が μ、分散が σ^2 の正規分布 $N(\mu, \sigma^2)$ についても同様だが、$N(\mu, \sigma^2)$ に従う確率変数 X について $X \le a$ となる確率 は、標準正規分布 $N(0,1])$ に従う確率変数 Z と X との関係 $X \sim Z\sigma + \mu$ [1] から

$$Pr[X \le a] = Pr[Z\sigma + \mu \le a] = Pr\left[Z \le \frac{a-\mu}{\sigma}\right]$$

となる、という関係を用いる。

- $N(\mu, \sigma^2)$ に従う確率変数 X について $X \le a$ となる確率 は、標準正規分布 $N(0,1])$ に従う確率変数 Z と X との関係 $X \sim Z\sigma + \mu$ から、$(a-\mu)/\sigma$ を計算して、標準正規分布表を読めばよい。

- $N(\mu, \sigma^2)$ に従う確率変数 X について $X > a$ となる確率 も同様に、$(a-\mu)/\sigma$ を計算して、標準正規分布表を読めばよい。

- 確率点を求める問題も同様である。

[1] この講義では $X \sim Y$ は確率変数 X と確率変数 Y の従う分布は等しい、という意味で用いている。$X \sim N(\mu, \sigma^2)$ ならば、確率変数 X の確率分布は $N(\mu, \sigma^2)$ である、となる。

他の表 (表 A.2〜A.5) は、t 分布、χ^2 分布、F 分布についてのものである。いずれも参考書の第 6 章 5 節「標本分布」に現れる。それぞれ

t 分布 正規分布の平均の、区間推定や検定をするときに用いる確率分布 (回帰分析などで現れる)

χ^2 分布 正規分布の分散の、区間推定や検定をするときに用いる確率分布 (尤度比検定などで現れる)

F 分布 正規分布の分散の比の区間推定や検定をするときに用いる確率分布 (分散分析などで現れる)

という意義を持つ、汎用性の高い確率分布である。しかし、それらの確率分布を導出するための確率変数の変換式がどれほど重要かを説明するには、仮説検定の説明が必要となり、「統計学」の範疇に踏み込んでしまうため、ここでは扱わない。

3 幾つかの確率不等式

確率について、確率分布に拠らず成り立つ不等式が幾つかある。これらは、確率の近似的な評価に用いられたり、様々な定理の証明に用いられたりする。

3.1 マルコフの不等式

非負の確率変数 X の期待値が有限の場合、任意の正数 $a > 0$ について

$$Pr[X \geq a] \quad \leq \quad \frac{1}{a}E[X] \tag{1}$$

が成り立つ。これを マルコフ (Markov) の不等式 という。[2]

証明は単純で、期待値を求める定積分の区間が $[0, \infty)$ と非負の領域であることと、その定積分よりも区間 $[a, \infty)$ での定積分の方が小さい、という事実を用いる。連続確率分布の場合、

$$\begin{aligned} E[X] &= \int_0^\infty x f_X(x)\,dx \\ &\geq \int_a^\infty x f_X(x)\,dx \quad (\because \text{ 積分区間を狭めると非負の被積分関数の定積分は小さくなる}) \\ &\geq a\int_a^\infty f_X(x)\,dx \quad (\because\ a \text{ は区間 } [a,\infty) \text{ で最小値}) \\ &= aPr[X \geq a] \end{aligned} \tag{2}$$

と証明できる。最後の不等式は、

$$\begin{aligned} \int_a^\infty x f_X(x)\,dx - a\int_a^\infty f_X(x)\,dx &= \int_a^\infty (x-a) f_X(x)\,dx \\ &\geq 0 \end{aligned} \tag{3}$$

と証明しても良い。

離散確率分布でも

$$\begin{aligned} E[X] &= \sum_{k=0}^\infty k p_X(k) \\ &\geq \sum_{k=a}^\infty k p_X(k) \quad (\because \text{ 総和の区間を狭めると非負の関数の総和は小さくなる}) \\ &\geq a\sum_a^\infty f_X(x) \quad (\because\ a \text{ は区間 } [a,\infty) \text{ で最小値}) \\ &= aPr[X \geq a] \end{aligned} \tag{4}$$

となる。

[2] この不等式は、参考書にはない。

マルコフの不等式から、累積分布関数 $F(x)$ について

$$\begin{aligned} F(x) &= Pr[X \le a] = 1 - Pr[X \ge a] \\ &\ge 1 - \frac{1}{x}E[X] \end{aligned} \tag{5}$$

という関係が得られる。例えば期待値が $E[X] = 1$ のあらゆる非負の確率分布の累積分布関数は、図 1 の赤線より上になければならない、という強い不等式である。

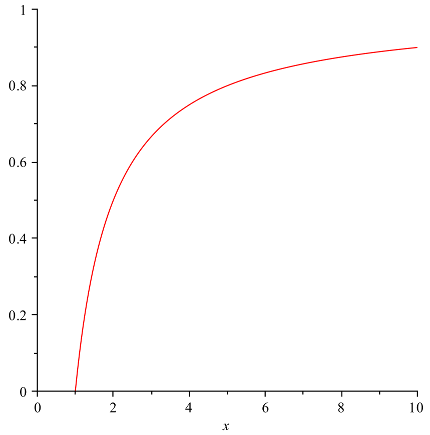

図 1 $E[X] = 1$ の場合のマルコフの不等式

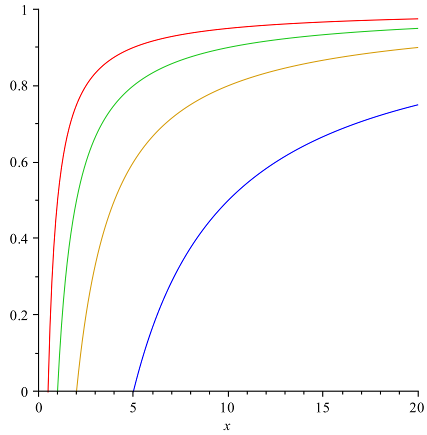

図 2 $E[X] = 0.5, 1, 2, 5$ の場合の下限

図 2 は、上から順に $E[X] = 1/2, 1, 2, 5$ の場合の下限である。いずれも期待値が変わると、累積分布関数の下限がそれに応じて変わり、これらの下限より下には累積分布関数は来ないことが分かる。

ところで図 1、図 2 からも分かるように、マルコフの不等式は $a > E[X]$ の範囲で、より精緻な区間を与える。$a \le E[X]$ では、確率は 0 以上 1 以下、という確率の公理のひとつの方がましである。そのことを確かめるために、マルコフの不等式の応用例を幾つか与えておく。

- 平均年収が 500 万円の世代の中で、年収が 5000 万円以上の人の割合は、1 割以下。

$$Pr[X \ge 5,000] \le \frac{E[X]}{5,0000} = \frac{500}{5,000} = \frac{1}{10} \tag{6}$$

- 平均年収が 500 万円の世代の中で、年収が 200 万円以下の人の割合は、0 以上。

$$Pr[X \le 200] \ge 1 - \frac{E[X]}{200} = 1 - \frac{500}{200} = -\frac{3}{2} \tag{7}$$

- 1 企業あたりの平均従業員数が 2 千人とすると、10 万人以上の従業員を抱える企業の割合は、2% 以下。

$$Pr[X \ge 2] \le \frac{E[X]}{100} = \frac{2}{100} \tag{8}$$

平均だけ分かれば、確率分布を知らずとも、上限や下限を与えられるのが、マルコフの不等式の実世界へのひとつの応用手段である。

マルコフの不等式は、すぐにチェビシェフの不等式の証明に使える。

3.2 チェビシェフの不等式

確率変数 X が平均 $E[X] = \mu$、分散 $V[X] = \sigma^2$ を持つとき、任意の正数 $a > 0$ に対して

$$Pr[|X - \mu| \ge a] \quad \le \quad \frac{\sigma^2}{a^2} \tag{9}$$

が成り立つ。これを チェビシェフ (Chevyshev) の不等式 という。[3]

[3]参考書では p.24 で定理 2.3 として紹介されていて、p.53 の定理 4.1 の証明にも現れる。

この不等式は、

$$Pr\left[|X-\mu| \geq a\right] \quad = \quad Pr\left[|X-\mu|^2 \geq a^2\right] \tag{10}$$

と左辺を変形すれば、$|X-\mu|^2$ が非負の確率変数であることから、あとはマルコフの不等式を少し変形するだけで証明できる。

$$\begin{aligned} Pr\left[|X-\mu|^2 \geq a^2\right] &\geq \frac{E\left[|X-\mu|^2\right]}{a^2} && (11)\\ &= \frac{E\left[(X-\mu)^2\right]}{a^2} && (12)\\ &= \frac{\sigma^2}{a^2} && (13) \end{aligned}$$

この証明は、X が連続確率変数でも離散確率変数でも、マルコフの不等式を証明してあることから、問題なく成り立つ。

チェビシェフの不等式は、マルコフの不等式とは異なり、確率変数 X の定義域は非負でなくて良い。その代わりに、確率を評価する区間は、

$$X \geq a \ \text{あるいは} \ \ X \leq b \tag{14}$$

といった片側区間ではなく、確率分布の平均 μ に対称な

$$\begin{aligned} |X-\mu| \geq a \quad &\leftrightarrow \quad X \leq \mu - a \ \text{ or } \ X \geq \mu + a\\ |X-\mu| \leq b \quad &\leftrightarrow \quad \mu - b \leq X \leq \mu + b \end{aligned}$$

など、原点からではなく分布の平均 μ からの対称な区間 $[\mu - a, \mu + a]$ についての不等式である。

平均と分散が分かっている場合、チェビシェフの不等式から

- 年収は非負とすれば、平均年収が 500 万円、標準偏差が 200 万円の世代の中で、年収が 5000 万円以上の人の割合は、0.004% 未満。

  $$\begin{aligned} &Pr\left[X - 500 \geq 5,000 - 500 = 4,500 \ \text{ or } \ X - 500 \leq -5,000 + 500 = -4,500\right]\\ =&Pr\left[X - 500 \geq 4,500\right] \leq \frac{200^2}{4,500^2} = 0.0000309... < 0.00004 \end{aligned}$$

  最後の等式は、年収は負になり得ないことから。

- 年収は非負とすれば、平均年収が 500 万円、標準偏差が 500 万円の世代の中で、年収が 5000 万円以上の人の割合は、1.3% 未満。

  $$\begin{aligned} &Pr\left[X - 500 \geq 4,500 \ \text{ or } \ X - 500 \leq -4,500\right]\\ =&Pr\left[X - 500 \geq 4,500\right] \leq \frac{500^2}{4,500^2} = \frac{1}{81} = 0.012345679... \end{aligned}$$

  最後の等式は、年収は負になり得ないことから。

- 1 企業あたりの平均従業員数が 5 百人、標準偏差を 2 百人とすると、10 万人以上の従業員を抱える企業の割合は、0.0005% 以下。

  $$Pr\left[X - 5 \geq 995 \ \text{ or } \ X - 5 \leq -9995\right] = Pr\left[X \geq 1000\right] \leq \frac{2^2}{995^2} = 0.000004... < 0.000005 \tag{15}$$

など、マルコフの不等式よりは精緻な限界 (不等式) を得る。

チェビシェフの不等式は確率論の中では、確率の評価と大数の法則の証明など、幾つかの用法がある。

1. 確率の評価：上のような確率不等式による上限や下限を与えることができる
2. 大数の法則：証明に用いる

3.3 イェンセンの不等式

確率変数 X と、X の定義域を含む区間上の凸関数 $h(x)$ を考える。X の期待値と $h(X)$ の期待値がどちらも有限なとき、

$$E[h(X)] \quad \geq \quad h(E[X]) \tag{16}$$

が成り立つ。これを イェンセン (Jensen) の不等式 という。[4]

この不等式の証明は、関数 $h(x)$ が下に凸であることを利用する。$h(x)$ が下に凸であれば、任意の点 x_0 に対して、ある直線 $g(x) = ax + b$ が存在して、$h(x_0) = g(x_0)$ と

$$h(x) \geq g(x),\ \forall x \Leftrightarrow h(x) - g(x) \geq 0,\ \forall x \tag{17}$$

とが成り立つ。$x_0 = E[X]$ とおけば、

$$\begin{aligned} E[h(X) - g(X)] &= \int_{x\in X} \{h(x) - g(x)\} f(x)\, dx \\ &\geq 0 \end{aligned} \tag{18}$$

は式 (17) に示した $h(x)$ の凸性から明らか。一方、

$$\begin{aligned} E[g(X)] &= E[aX + b] \\ &= aE[X] + b \\ &= g(E[X]) \end{aligned} \tag{19}$$

再び式 (17) に示した凸性と h と g の関係より、

$$g(E[X]) \quad = \quad h(E[X]) \tag{20}$$

以上より、

$$E[h(X)] \geq E[g(X)] \tag{21}$$

が示される。離散分布についても同様。

イェンセンの不等式から、$E\left[X^2\right]$ が有限の任意の確率分布について

$$\{E[X]\}^2 \leq E\left[X^2\right] \tag{22}$$

が分かる。

$E\left[X^2\right] < \infty$ を仮定できると

$$\begin{aligned} E[|X|] &= \int_{-\infty}^{\infty} |x|\, f_X(x)\, dx \\ &= \int_{-\infty}^{-1} |x|\, f_X(x)\, dx + \int_{-1}^{1} |x|\, f_X(x)\, dx + \int_{1}^{\infty} |x|\, f_X(x)\, dx \\ &\leq \int_{-\infty}^{-1} x^2 f_X(x)\, dx + \int_{-}^{1} f_X(x)\, dx + \int_{1}^{\infty} x^2 f_X(x)\, dx \\ &\leq \int_{-\infty}^{\infty} x^2 f_X(x)\, dx + \int_{-\infty}^{\infty} f_X(x)\, dx = E\left[X^2\right] + 1 \leq \infty \end{aligned} \tag{23}$$

が成り立つのはおまけ。

イェンセンの不等式から、次のことが示せる。

- 算術平均 $(X_1 + X_2 + \cdots + X_n)/n$ と幾何平均 $(X_1X_2\cdots X_n)^{1/n}$ と調和平均 $n/\{1/X_1 + 1/X_2 + \cdots 1/X_n\}$ の間に、次の大小関係が成り立つ。(通常、別のイェンセンの不等式からも示せるが、この確率版の不等式からも示せる)

$$\frac{1}{\frac{1}{n}\left(\frac{1}{X_1} + \cdots + \frac{1}{X_n}\right)} \leq (X_1X_2\cdots X_n)^{1/n} \leq (X_1 + X_2 + \cdots + X_n)/n \tag{24}$$

[4] この不等式も、参考書には無い。

- 算術平均 $(X_1+X_2+\cdots+X_n)/n$ と幾何平均 $(X_1X_2\cdots X_n)^{1/n}$ と調和平均 $n/\{1/X_1+1/X_2+\cdots 1/X_n\}$ の間に、次の大小関係が成り立つ。

$$E_X\left[\frac{1}{\frac{1}{n}\left(\frac{1}{X_1}+\cdots+\frac{1}{X_n}\right)}\right] \leq E_X\left[(X_1X_2\cdots X_n)^{1/n}\right] \leq E_X\left[(X_1+X_2+\cdots+X_n)/n\right] \tag{25}$$

3.4 ヘルダーの不等式

２つの正の数 p,q が $1/p+1/q=1$ を満たすとする。$E[|X|^p]$ が有限の確率変数 X と、$E[|Y|^p]$ が有限の確率変数 Y に対して、

$$E[|XY|] \quad\leq\quad \{E[|X|^p]\}^{1/p}\{E[|Y|^q]\}^{1/q} \tag{26}$$

が成り立つ。これを ヘルダー (Hölder) の不等式 という。[5]

ヘルダーの不等式で特に $p=2, q=2$ とした場合

$$E[|XY|] \quad\leq\quad \sqrt{E\left[|X|^2\right]}\sqrt{E\left[|Y|^2\right]} \tag{27}$$

を、コーシー・シュワルツ (Cauchy-Schwartz) の不等式 という。[6]

4 大数の法則

マルコフの不等式やチェビシェフの不等式と同じく、ここから先の議論は、特定の確率分布ではなく、条件を満たすすべての確率分布について成り立つ。

4.1 独立な確率変数列

ベルヌーイ試行を独立に繰り返すとき、n 回の試行中の成功回数 X は二項分布に従うことは以前に説明した。ここで $n\to\infty$ とするとき X/n はどのようなことになるか、を考える。その準備として、独立な確率変数の列 を定義する。

標本空間 $\mathcal{X}$ 上のある確率変数 X が確率分布 F に従うとする。この確率変数を n 個用意して並べ、それぞれに添え字をけたものを、

$$X_1, X_2, \ldots, X_n$$

と記す。これらを確率ベクトル $\boldsymbol{X}_n$ で表す。

$$\boldsymbol{X}_n = (X_1, X_2, \ldots, X_n)$$

同様に標本空間内の点もベクトルで

$$\boldsymbol{x}_n = (x_1, x_2, \ldots, x_n)$$

と記す。このとき、$X_1, X_2, \ldots, X_n$ が 互いに独立 であれば、$\boldsymbol{X}_n$ の確率分布は

$$\begin{aligned} F(\boldsymbol{x}_n) &= F(x_1)F(x_2)\cdots F(x_n) \\ &= \prod_{i=1}^{n} F(x_i) \end{aligned}$$

[5]この不等式も、参考書には記述がない。
[6]参考書では、p.48 の問題 3.6 になっている。

という確率分布関数を持つ。このことは

$$\begin{aligned} Pr\left[X_1 \le x_1 \quad \text{and} \quad X_2 \le x_2 \quad \cdots \quad \text{and} \quad X_n \le x_n\right] &= Pr\left[X_1 \le x_1\right] Pr\left[X_2 \le x_2\right] \cdots Pr\left[X_n \le x_n\right] \\ &= F(x_1) F(x_2) \cdots F(x_n) \\ &= \prod_{i=1}^{n} F(x_i) \end{aligned}$$

から確かめられる。

$X_1, X_2, \ldots, X_n$ 互いに独立であれば同様に、確率関数や密度関数も

$$\begin{aligned} f(\boldsymbol{x}_n) &= \prod_{i=1}^{n} f(x_i) \\ p(\boldsymbol{k}_n) &= \prod_{i=1}^{n} p(k_i) \end{aligned}$$

となる。

確率論では、互いに独立な確率変数の列はとても重要な位置を占めており、これを 独立な確率変数列 と呼ぶ。

4.2 標本平均

互いに独立な確率変数の列 $X_1, X_2, \ldots, X_n, \ldots$ に対して、最初の n 個の算術平均

$$\overline{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

を、それらの 標本平均 と呼ぶ。標本という言葉は、無数の確率変数の中から (無作為に)n 個を選んだ時の、それらの組み合わせを挿して言う。

4.3 大数の弱法則

互いに独立で同一な確率分布に従う確率変数列 $X_1, X_2, \ldots$ を考える。各 X_i は期待値 $E[X_i] = \mu$ を持つとする。[7] $n \to \infty$ とするとき、任意の $\epsilon > 0$ に対して、

$$Pr\left[\left|\overline{X}_n - \mu\right| \ge \epsilon\right]$$

は 0 に収束する。これを 大数の弱法則 と言う。

大数の弱法則は、「どんなに小さい正 ϵ を選んで固定したとしても、n を大きくしていけば、$\overline{X}_n$ が μ から ϵ 以上離れる確率 $Pr\left[\left|\overline{X}_n - \mu\right| \ge \epsilon\right]$ は、かならず 0 に収束する」という意味[8]を持つ。この意味で $\overline{X}_n$ は μ に収束するのである。

大数の弱法則の証明には幾つかのバリエーションがある。分散の存在 $V[X_i] = \sigma^2$ を仮定できれば[9]、

$$E\left[\overline{X}_n\right] = \mu, \quad V\left[\overline{X}_n\right] = \sigma^2 / n$$

という事実と、$\overline{X}_n$ が平均 $|mu$ から ϵ 以上離れている確率が、$\overline{X}_n$ の分散を ϵ^2 より小さい、というチェビシェフの不等式を用いる。

実際、チェビシェフの不等式より

$$Pr\left[\left|\overline{X}_n - \mu\right| \ge \epsilon\right] \le \frac{\sigma^2}{n\epsilon^2}$$

であり、$n \to \infty$ とするとき、この不等式の右辺が 0 に収束することから、左辺も 0 に収束する。(確率が非負であることは、今さら述べる必要もあるまい。)

[7] $\int_X x f(x)\, dx$ もしくは $\sum_X x p(x)$ が有限の値を持つ、という条件。

[8] 弱法則の「弱」は、収束の意味合いを反映している。

[9] $\int_X (x-\mu)^2 f(x)\, dx$ もしくは $\sum_X (k-\mu)^2 p(k)$ が有限である、という仮定。

4.4 大数の強法則

互いに独立で同一な確率分布に従う確率変数列 $X_1, X_2, \ldots$ を考える。各 X_i は期待値 $E[X_i] = \mu$ を持つとする。このとき

$$Pr\left[\lim_{n\to\infty} \overline{X}_n = \mu\right] = 1$$

が成り立つ。これを 大数の強法則 と言う。

大数の強法則は、「$n \to \infty$ とするとき $\overline{X}_n$ の確率分布が、標本空間が $\{\mu\}$ だけの 1 点分布になる」という意味[10]である。この意味で $\overline{X}_n$ は μ に収束するのである。

こちらの証明は少し厄介で、

$$\lim_{n\to\infty} \overline{X}_n = \mu$$

を示さなければならず、今日は省略する。

5 中心極限定理

5.1 中心極限定理

中心極限定理は、大数の法則と同じく、特定の確率分布の性質ではなく、多くの分布で成り立つ性質である。

確率変数列 $X_1, X_2, \ldots$ が互いに独立で、同じ確率分布に従うとする。その確率分布の平均 μ と分散 σ^2 が存在するとする。このとき、すべての実数 x に対して

$$Pr\left[\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sqrt{n}\sigma} \le x\right] \to \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \quad (n \to \infty) \tag{28}$$

が成り立つ。これは

$$Pr\left[\frac{(X_1 + X_2 + \cdots + X_n)/n - \mu}{\sigma} \sqrt{n} \le x\right] \to \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \quad (n \to \infty) \tag{29}$$

と書いても同じである。

証明は複雑で、参考書にも概略のみが記載されている。実は証明は何通りかあり、「中心極限定理」という名前の書籍もある程、この定理は奥深い。また、互いに独立、という仮定をどこまで緩められるか、という理論的な検討は、現在もなお続けられている。

証明の方針には、標本平均を基準化した量 $\left(\overline{X} - \mu\right) \sqrt{n}/\sigma$ について

1. その確率分布のモーメント母関数が、n が大きくなると、正規分布のモーメント母関数それに収束することを証明する
2. その確率分布の特性関数が、n が大きくなると、正規分布の特性関数に収束することを証明する
3. エントロピー概念を用いる証明
4. 漸近展開を用いる証明

などがある。

さて、不等式の右辺の量は

$$\begin{aligned} \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sqrt{n}\sigma} &= \frac{n}{\sqrt{n}\sigma} \frac{X_1 + X_2 + \cdots + X_n - n\mu}{n} \\ &= \frac{\sqrt{n}}{\sigma} \left(\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right) \end{aligned}$$

と変形できて、標本平均と母平均の差を $\sqrt{n}$ 倍しているようにも見える。実際、この $\sqrt{n}$ が無ければ、大数の法則により差は 0 に収束する。$\sqrt{n}$ 倍することで、差が正規分布に従うように収束する。この収束を、確率分布が正規分布に近づく、という意味で 法則収束 という。

[10] 強法則の「強」も、収束の意味合いを反映している。

5.2　例

ここでは、確率分布と大数の法則と中心極限定理の関係を、幾つかの確率分布を例に眺める。

図 3 から図 5 までは、ある確率分布に従う乱数を用いてデータ $X_1, X_2, \cdots, X_n$ を発生させたもののヒストグラム (上段)、それらの平均 $\overline{X}_n$ を求めて期待値との差分 $\overline{X}_n - \mu$ を計算したもののヒストグラム (中段)、それからその差分を $\sqrt{n}/\sigma$ 倍した $\sqrt{n}\left(\overline{X}_n - \mu\right)/\sigma$ のヒストグラム (下段)、の 3 つを一組にして掲げている。確率分布には、図 3 は標準正規分布 $N(0,1)$、図 4 はガンマ分布 $\Gamma(2,3)$、図 5 は二項分布 $Bin(10, 0.1)$、をそれぞれ用い、平均をとる単位 n は、$n = 1, 10, 100, 1000$ と変化させている。

図 3 は、X_i の確率分布も $\overline{X}_n$ の確率分布も、正規分布とならなければならない。実際、上段、中段、下段とも正規分布によく似たヒストグラムを見ることができる。中段のグラフを比較すると、標本平均と母平均の差 $\overline{X}_n - \mu$ の分布が、n が大きくなるにつれて 0 に集中していく様子が見て取れるだろうか。これが大数の法則である。下段のグラフを比較すると、$\sqrt{n}\left(\overline{X}_n - \mu\right)/\sigma$ の分布が、安定して正規分布に近いことが見て取れる。これはそもそも $\overline{X}_n$ が正規分布に従うこと、その分散が $\sigma/\sqrt{n}$ であることから、理論的にも裏付けられる。

次に図 4 だが、X_i の確率分布はガンマ分布であり、その確率分布に対称ではなく、偏りがある。標本平均 $\overline{X}_n$ の従う分布は、形状母数 α が整数の場合には、形状母数が異なる $n\alpha$ のガンマ分布に従うはずである。中段のグラフどうしを比較していくと、標本平均と母平均の差 $\overline{X}_n - \mu$ の分布が、n が大きくなるにつれて 0 に集中していく様子が見て取れる。これが大数の法則である。更に下段のグラフどうしを比較すれば、$\sqrt{n}\left(\overline{X}_n - \mu\right)/\sigma$ の分布が、n を大きくするにつれてえ正規分布に近づくことが見て取れる。これが中心極限定理の意味である。

最後に、図 5 は二項分布 $Bin(10, 0.1)$ の例である。これは離散分布だが、ここでも大数の法則と中心極限定理を確認することができる。

いずれの場合も、期待値と分散が存在していることから、大数の法則と中心極限定理の証明の前提は保証されている。

参考文献

[1] 微分積分学の教科書.

[2] 永田靖 (2005)「統計学のための数学入門 30 講」朝倉書店.

[3] 宮川雅巳 (1988)「統計技法」共立出版.

[4] 稲垣宣生 (2003)「数理統計学」改訂版, 裳華房.

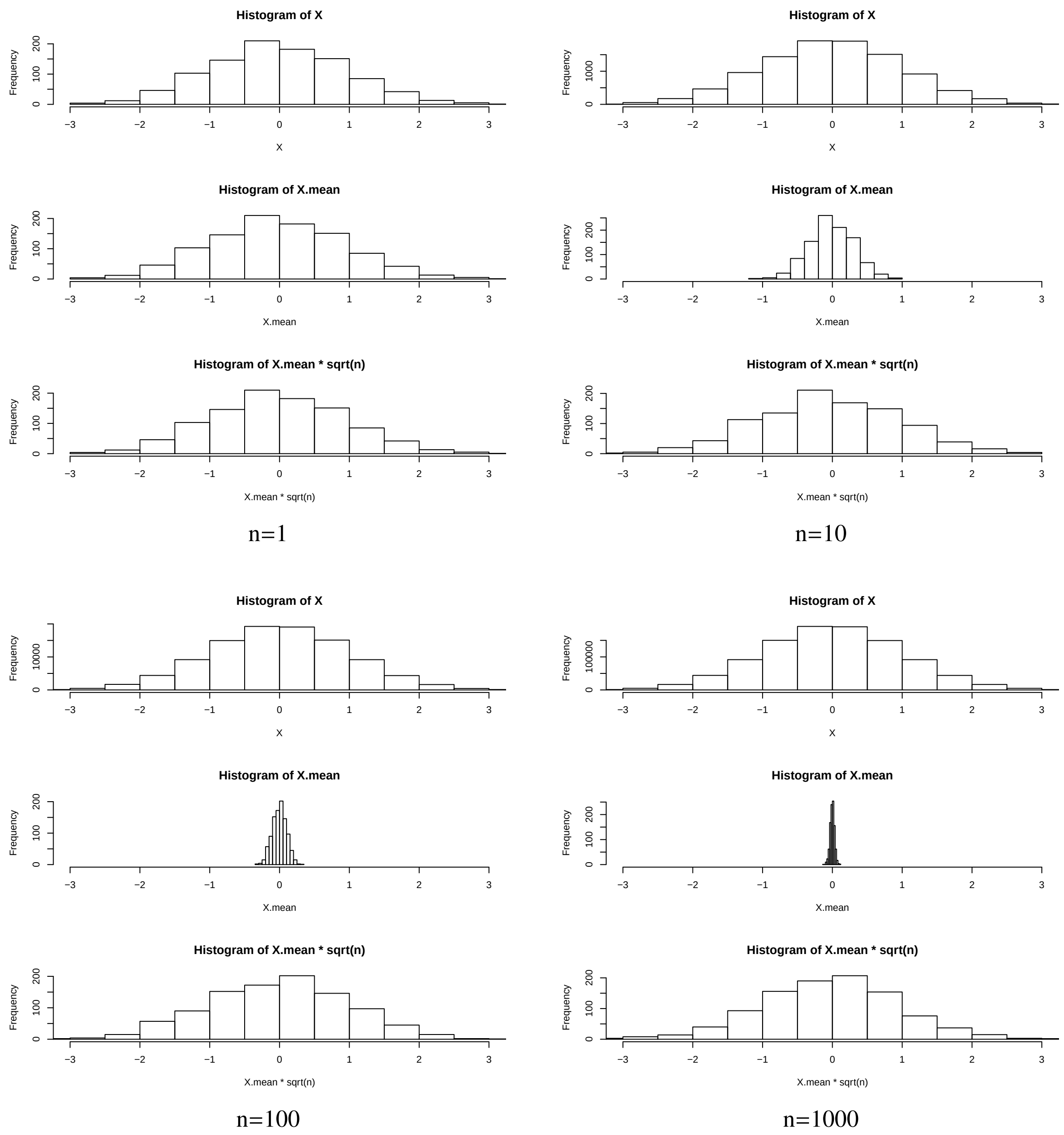

図 3: 中心極限定理の例 (正規分布の場合)
ヒストグラムはそれぞれ、上段：元のデータ、中段：(平均-真値)、下段：(平均-真値) $\sqrt{n}/\sigma$

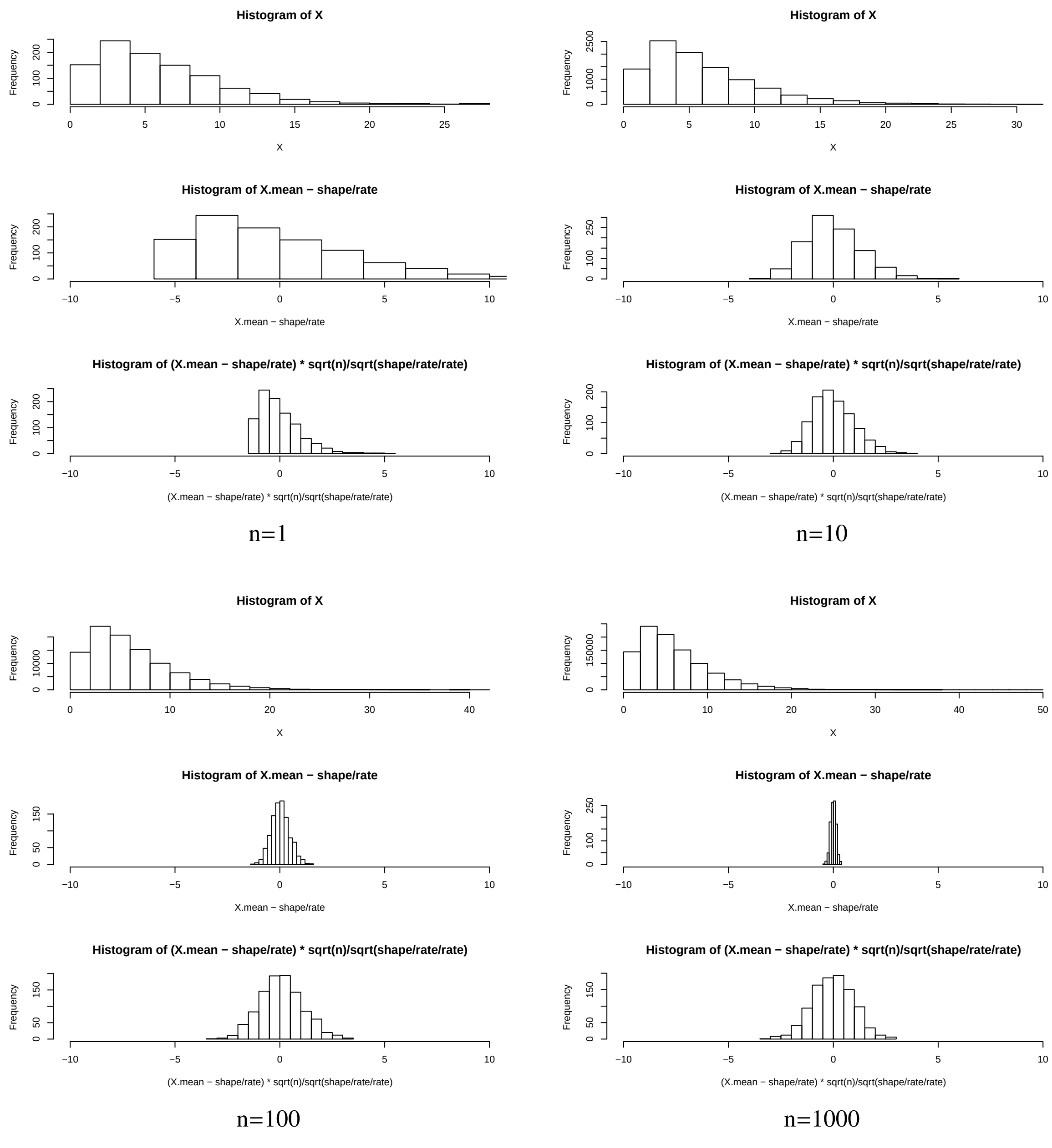

図 4: 中心極限定理の例 (ガンマ分布の場合)
ヒストグラムはそれぞれ、上段：元のデータ、中段：(平均-真値)、下段：(平均-真値) $\sqrt{n}/\sigma$

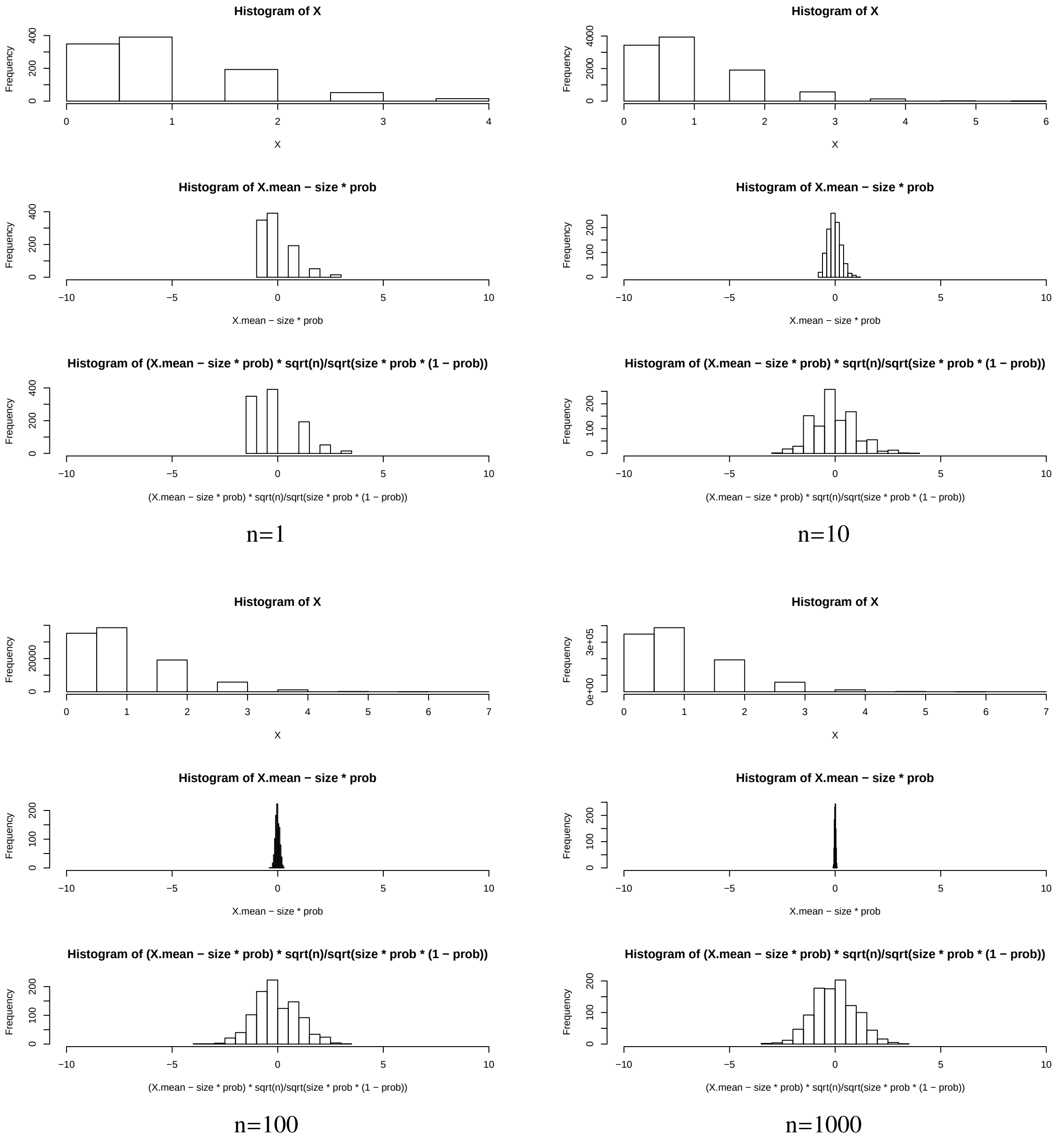

図 5: 中心極限定理の例 (二項分布の場合)
ヒストグラムはそれぞれ、上段：元のデータ、中段：(平均-真値)、下段：(平均-真値) $\sqrt{n}/\sigma$