26 juillet 2012
PROB #15

確率論 (*Probability Theory*) 第 15 週

w.yamamoto

1　一変量確率分布 (復習)

1.1　モーメント

モーメントは、確率分布の特徴量である。例えば人間の顔の特徴を、輪郭、目、瞳、鼻、口、耳、眉、髪、それぞれの形状、大きさ及びバランス、で捉えると次のようになる。

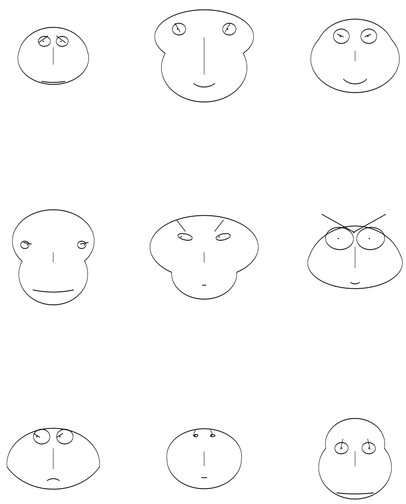

図 1: チャーノフの顔

これは 18 次元のデータのグラフ表現方法のひとつで、チャーノフの顔と呼ばれている。表示したデータは、ただの乱数で作成したデータだが、本資料の一番最後に掲げてある。元のデータが人間の顔のデータでなくとも良いので、あまり人間らしくない絵も含まれるが、目や鼻や口や耳などの部品が無くなることはない。

チャーノフの顔はデータの表現方法のひとつだが、確率分布もこれと同様にそれぞれに顔を持つ。確率分布は確率関数あるいは確率密度関数、それと累積分布関数で表現できるが、その確率分布の特徴は、モーメント母関数で表現できる。モーメントはモーメント母関数を原点まわりでテイラー展開したときの定数項であり、すべてのモーメント母関数を特定できれば、元の確率分布を特定できることは、既に述べた通り。

ラプラス変換前の関数とラプラス変換後の関数が、ラプラス変換とラプラス逆変換で一対一の関係を持つように、確率関数あるいは確率密度関数と、モーメント母関数との間にも、逆変換の公式集は無いものの、一対一の関係がある。またモーメント母関数のテイラー展開が収束するなら、その展開係数であるモーメントたちがあれば、モーメント母関数がなくても、元の確率関数あるいは確率密度関数に辿り着ける。その意味でモーメントは、元の関数の特徴をすべて有する、と言える。図 2 に、2 次までの中心モーメント $m_1 = E[X]$ 、$m_2 = E\left[X^2\right]$ とモーメント母関数 $M(t) = E\left[e^{tX}\right]$ の関係

$$M(t) \quad = \quad M(0) + M'(0)\,t + \frac{M''(0)}{2!}t^2 + O\left(t^3\right)$$

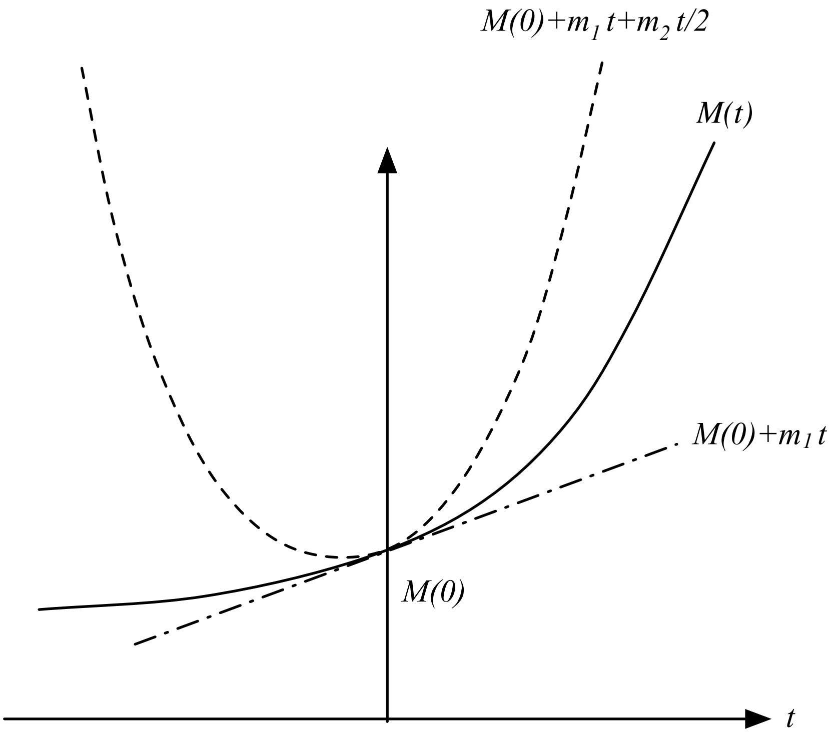

図 2: モーメント母関数とその原点まわりのテイラー近似

$$= \quad M(0) + m_1 t + \frac{m_2}{2!} t^2 + O\left(t^3\right) \tag{1}$$

を掲げた。

更に原点まわりのモーメント $m_k = E\left[X^k\right]$ と期待値まわりのモーメント μ_k

$$\mu_k = E\left[(X - m_1)^k\right] \tag{2}$$

の間には、次のような関係がある。

$$\mu_1 = m_1,\ \mu_2 = m_2 - {m_1}^2,\ \mu_3 = m_3 - 3m_1m_2 + 2{m_1}^3,\ \mu_4 = m_4 - 4m_1m_3 + 6{m_2}^2 - 3{m_1}^4,\ \ldots \tag{3}$$

このため期待値回りのモーメントも、m_1 とセットで、確率分布の特徴を表すことができる。

確率分布のモーメントのうち、1 次のモーメントである平均 μ_1 は、分布の中心 (重心) を表し、2 次のモーメントである分散 $\mu_2 = \sigma^2$ は中心からの広がり (二乗距離の期待値) を表す。平均と分散以外にも、3 次のモーメント μ_3 はそのままではなく、標準偏差の 3 乗あるいは分散の 3/2 乗で割って、分散の影響を除いた量 $\mu_3/{\mu_2}^{3/2}$ が確率分布の歪度と呼ばれる。この値が、正規分布など対称な分布では 0 となる。4 次のモーメント μ_4 も同様に分散の 2 乗で割って、分散の影響を除いた量 $\mu_4/{\mu_2}^4$ が確率分布の尖り具合を表す。この量は正規分布では常に 3 となることから、正規分布と比べて尖っているか否かを見るために、 $\mu_4/{\mu_2}^4 - 3$ を尖度と定義するのが一般的である。

以上の 4 つの特徴量を表 1 にまとめた。

モーメントを計算したくなったり、最尤法を用いたくなるような確率分布は通常、1〜3 程度の少ない数の母数を持つ確率関数あるいは確率密度関数を持つ。例えば二項分布は母数は n と p の 2 つあるが未知のものは p のみ、正規分布は μ と σ^2 の 2 つ、ポアソン分布と指数分布は λ の 1 つ、など。つまり、確率分布族を限定できるなら、確率分布とモーメントたちとの間の関係から、幾つかの低次のモーメントのみで確率分布を一意に定めることができる。このことは、前回の資料の 1 ページにも表にまとめてある。

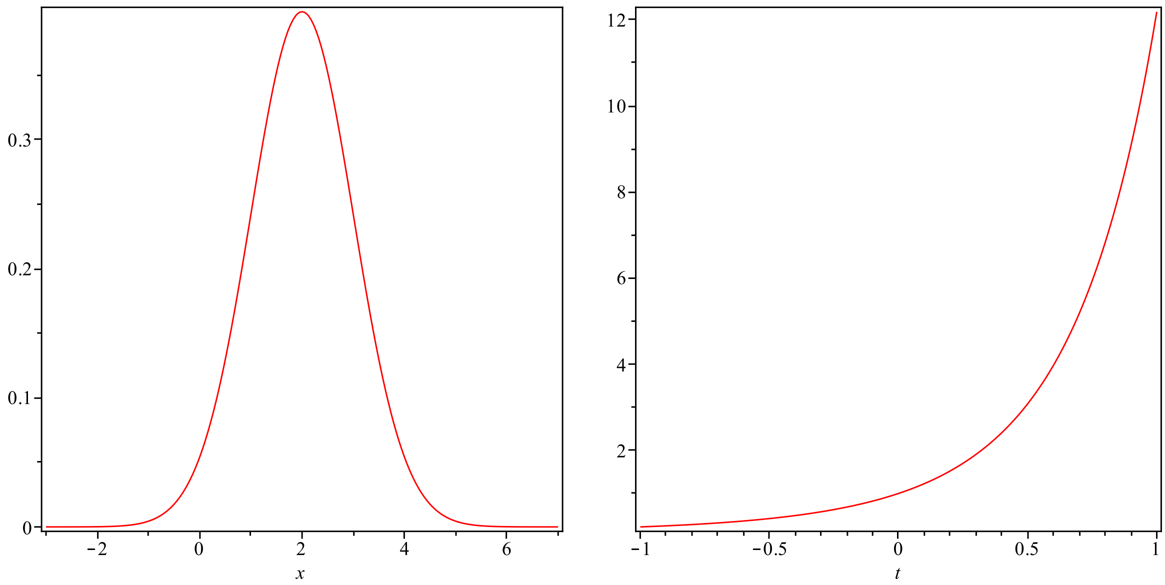

図 3: 正規分布 $N\left(\mu, \sigma^2\right)$ とそのモーメント母関数 $M(t) = \exp\left(\mu t + \sigma^2 t^2/2\right)$ の例 $(\mu = 2, \sigma^2 = 1)$

表 1: モーメントと確率分布の形状

モーメントの関数	表現	意味
平均	$m_1 = \mu_1 = \mu$	確率分布の重心
分散	$\mu_2 = \sigma^2 = m_2 - {m_1}^2$	確率分布の広がり
歪度	$\mu_3/{\mu_2}^{3/2}$	確率分布の対称性 (負ならば左に裾が長く右に偏っており、正ならば右に裾が長く左に偏っている、0 ならば対称)
尖度	$\mu_4/{\mu_2}^2 - 3$	確率分布のピークの鋭さと裾の重さ (正規分布の尖度は 0)

2 多変量確率分布

2.1 同時分布

複数の試行を同時は扱う時、それらの試行の結果が同時に従う確率分布を 同時確率分布 あるいは 同時分布 と言う。

2.1.1 同時分布

複数の試行を同時に行う、という問題を確率論で扱うとする。例えば、以下のような試行がある。

- コインを何枚か同時に投げる
- 人間の身長と体重と腹囲と体脂肪率を測る
- 朝に日経新聞を開いて東証 1 部の全銘柄の株価を眺める
- 実験に用いる複数の試薬を、瓶から取り出して、正確なつもりで量りながら用意する
- 参考書のランダムにページを選んで、教室に居る人全員で「の」の数を数える
- 実験結果を複数回測定する

これら以外にも、同時に行う試行は、私たちの身の回りに幾らでも溢れかえっている。通常は、同時に行われる試行のひとつひとつに確率変数を考え、それぞれを $X_1, X_2, \ldots X_n$ などと表す。

このとき $X_1, X_2, \ldots, X_n$ が、x_1 以下、x_2 以下、$\ldots$ x_n 以下、とそれぞれの範囲を同時に取る確率を定めると、それが 同時累積確率 となる。

$$F(x_1, x_2, \ldots, x_n) \quad = \quad Pr[X_1 \leq x_1,\, X_2 \leq x_2,\, \ldots,\, X_n \leq x_n] \tag{4}$$

これの導関数が存在すれば、それは同時密度関数と呼ばれる。

$$\frac{\partial^p}{\partial x_1 \partial x_2 \cdots \partial x_n} F(x_1, x_2, \ldots, x_n) \quad = \quad f(x_1, x_2, \ldots, x_n) \tag{5}$$

この関係は、密度関数が存在すれば、

$$Pr[X_1 \leq x_1,\, X_2 \leq x_2,\, \ldots,\, X_n \leq x_n] \quad = \quad \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_n} f(u_1, u_2, \ldots, u_n)\, du_1 du_2 \cdots du_n \tag{6}$$

より明らか。

$X_1, X_2, \ldots, X_n$ がすべて離散確率変数のときには、それぞれが $k_1, k_2, \ldots, k_n$ の値を同時にとる確率が定義できて、

$$Pr[X_1 = k_1,\, X_2 = k_2,\, \ldots,\, X_n = k_n] \quad = \quad p(k_1, k_2, \ldots, k_n) \tag{7}$$

を 同時確率関数 と呼ぶ。

以上のような、同時密度関数あるいは同時確率関数、そして同時累積分布関数で表される確率分布を 同時確率分布(略して同時分布) と呼ぶ。

2.1.2　同時分布を考える理由

同時分布を考えたくなるのは、$X_1, X_2, \ldots, X_p$ が互いに独立でない時、に決まっている。独立ならば、例えば同時累積分布関数は

$$F(x_1, x_2) = F_1(x_1)\, F_2(x_2) \tag{8}$$

と 1 次元の累積分布関数の積で表せる。それぞれの累積分布関数は

$$F_1(x_1) = \int_{-\infty}^{x_1} f_1(u)\, du,\ \ F_2(x_2) = \int_{-\infty}^{x_2} f_2(u)\, du \tag{9}$$

と、1 次元の積分で求まる。しかし独立でなければ、

$$F(x_1, x_2) = \int_{\infty}^{x_1} \int_{\infty}^{x_2} f(u_1, u_2)\, du_1 du_2 \tag{10}$$

と重積分 ($\int \int$) が必要になる。離散分布でも、複数の総和演算 (Σ) が必要となる。

同時累積分布関数だけでなく、各モーメントも同様に重積分 ($|int \int$) あるいは複数の総和演算 (Σ) が必要となる。

$$E[X_1] = \int_{\infty}^{x_1} \int_{\infty}^{x_2} u_1 f(u_1, u_2)\, du_1 du_2 \tag{11}$$

2.1.3　共分散：二つの確率変数のモーメント

同時分布を考えるときには、例えば X_1 と X_2 の間にどのような関係があるか、も考えることになる。その理由は簡単で、詳細は本講義の予定を超えるが、

1. 同時分布のモーメント母関数 $M(t_1,\dots,t_n)=E\left[e^{X_1t_1}\cdots e^{X_nt_n}\right]$ は同時確率の多重和もしくは同時確率密度の多重積分

2. 同時分布のモーメント母関数 $M(t_1,\dots,t_n)$ は n 変数関数

3. n 変数関数のテイラー展開は、例えば 2 変数関数で

$$
\begin{aligned}
g(t_1,t_2) &= g(u_1,u_2) \\
&\quad +(t_1-u_1)\left.\frac{\partial g(t_1,t_2)}{\partial t_1}\right|_{t_1=u_1,t_2=u_2}+(t_2-u_2)\left.\frac{\partial g(t_1,t_2)}{\partial t_2}\right|_{t_1=u_1,t_2=u_2} \\
&\quad +(t_1-u_1)^2\frac{1}{2}\left.\frac{\partial^2 g(t_1,t_2)}{\partial {t_1}^2}\right|_{t_1=u_1,t_2=u_2}+(t_2-u_2)^2\frac{1}{2}\left.\frac{\partial^2 g(t_1,t_2)}{\partial {t_2}^2}\right|_{t_1=u_1,t_2=u_2} \\
&\quad +(t_1-u_1)(t_2-u_2)\frac{1}{2}\left.\frac{\partial^2 g(t_1,t_2)}{\partial t_1\partial t_2}\right|_{t_1=u_1,t_2=u_2} \\
&\quad +(t_1-u_1)^3\frac{1}{3!}\left.\frac{\partial^3 g(t_1,t_2)}{\partial {t_1}^3}\right|_{t_1=u_1,t_2=u_2}+\cdots
\end{aligned}
$$

となるように、各変数のべき乗だけでなく、変数同士の積の項も含まれる

4. モーメントは、モーメント母関数のテイラー展開の展開係数。

$$
M(t_1) = M(0)+t_1E[X_1]+\frac{{t_1}^2}{2}E\left[{X_1}^2\right]+\frac{{t_1}^3}{3!}E\left[{X_1}^3\right]+\cdots \tag{12}
$$

これは 2 変数以上の場合にも同じこと。上のテイラー展開の式で $u_1=u_2=0$ と置けば、例えば 2 変数の場合には

$$
\begin{aligned}
M(t_1,t_2) &= M(0,0) \\
&\quad +t_1E[X_1]+t_2E[X_2] \\
&\quad +\frac{{t_1}^2}{2}E\left[{X_1}^2\right]+\frac{{t_2}^2}{2}E\left[{X_2}^2\right]+2\frac{t_1t_2}{2}E[X_1X_2] \\
&\quad +\frac{{t_1}^3}{3!}E\left[{X_1}^3\right]+3\frac{{t_1}^2t_2}{3!}E\left[{X_1}^2X_2\right]+3\frac{t_1{t_2}^2}{3!}E\left[X_1{X_2}^2\right]+\frac{{t_2}^3}{3!}E\left[{X_2}^3\right] \\
&\quad +\cdots
\end{aligned} \tag{13}
$$

となる。

5. 中心モーメント $m_k=E\left[{X_1}^k\right]$ と期待値回りのモーメント $\mu_k=E\left[(X_1-m_1)^k\right]$ の間には関係式がある。最初の 4 次までの関係式を掲げる。

$$
\mu_1=m_1-m_1=0,\ \mu_2=m_2-{m_1}^2\ \mu_3=m_3-3m_2m_1+2{m_1}^2,\mu_4=m_4-4m_3m_1+6m_2{m_1}^2-3{m_1}^4,\dots \tag{14}
$$

このため $m_2,\dots,m_k$ と $\mu_2,\dots,\mu_k$ の間の変換は 1 対 1 の対応がある。さらに通常は、$\mu_1=m_1$ と置き、$\mu_k=E\left[(X_1-m_1)^k\right]$ という定義は $k\geq 2$ の場合のみに用いる。

これらを前提に、2 つの確率変数の間にも原点まわりの (k_1,k_2) 次のモーメント $m_{k_1k_2}$ を

$$
m_{k_1k_2} = E\left[{X_1}^{k_1}{X_2}^{k_2}\right]
$$

また期待値まわりのモーメント $\mu_{k_1k_2}$ を

$$
\mu_{k_1k_2} = E\left[(X_1-m_{10})^{k_1}(X_2-m_{01})^{k_2}\right]
$$

と、それぞれ定義する。これにより、式 (13) は

$$
\begin{aligned}
M(t_1, t_2) &= M(0,0) + t_1 m_{10} + t_2 m_{01} + \frac{{t_1}^2}{2} m_{20} + \frac{{t_2}^2}{2} m_{11} + 2\frac{t_1 t_2}{2} m_{02} \\
&\quad + \frac{{t_1}^3}{3!} m_{30} + 3\frac{{t_1}^2 t_2}{3!} m_{21} + 3\frac{t_1 {t_2}^2}{3!} m_{12} + \frac{{t_2}^3}{3!} m_{03} + \cdots
\end{aligned} \tag{15}
$$

と表せる。

3 つ以上でも同様に

$$
\begin{aligned}
m_{k_1 k_2 \cdots k_n} &= E\left[{X_1}^{k_1} {X_2}^{k_2} \cdots {X_n}^{k_n}\right] \\
\mu_{k_1 k_2 \cdots k_n} &= E\left[(X_1 - m_{10\cdots 0})^{k_1} (X_2 - m_{010\cdots 0})^{k_2} \cdots (X_n - m_{00\cdots 1})^{k_n}\right]
\end{aligned}
$$

とモーメントを定義して、モーメントとモーメント母関数の関係も得られる。しかしこの講義では、2 つの確率変数間の関係のみを扱うに留める。

2.1.4 共分散 $\mu_{11} = m_{11} - m_{10} m_{01}$

式 (13) に掲げたモーメント母関数のテイラー展開で、2 つの確率変数 X_1 と X_2 のモーメントのうち、両方に関係するもので一番最初に現れるのは、積の期待値 $m_{11} = E[X_1 X_2]$ である。

これに対応する、期待値まわりのモーメントは

$$
E[(X_1 - m_{10})(X_2 - m_{01})] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 f(x_1, x_2)\, dx_1 dx_2 \tag{16}
$$

である。これを X_1 と X_2 の共分散といい、μ_{11} もしくは $Cov[X_1, X_2]$ と表す。共分散は $(X_1 - m_{10})(X_2 - m_{01})$ の期待値であり、確率変数の関数としてのこの値が正をとる確率と負をとる確率を比較して、正をとる確率の方が大きい場合に正、負をとる確率の方が大きい場合に負、の値をとる。

共分散と原点まわりのモーメントの間には、次の関係がある。

$$
\begin{aligned}
E[(X_1 - m_{10})(X_2 - m_{01})] &= E[X_1 X_2 - m_{10} X_2 - X_1 m_{01} + m_{10} m_{01}] \\
&= E[X_1 X_2] - E[m_{10} X_2] - E[X_1 m_{01}] + E[m_{10} m_{01}] \\
&= m_{11} - m_{10} E[X_2] - m_{10} E[X_2] + m_{10} m_{01} \\
&= m_{11} - m_{10} m_{01}
\end{aligned}
$$

また共分散には

$$
-\sqrt{\mu_{20}\mu_{02}} \leq \mu_{11} \leq \sqrt{\mu_{20}\mu_{02}} \tag{17}
$$

という関係がある。

先週までの表記に戻せば、${\sigma_1}^2 = V[X_1]$、${\sigma_2}^2 = V[X_2]$、$\sigma_1 2 = Cov[X_1, X_2]$ と置いて

$$
-\sigma_1 \sigma_2 \leq \sigma_{12} \leq \sigma_1 \sigma_2 \tag{18}
$$

となる。この不等式は、2 つの確率変数 X_1, X_2 の関数 $tX_1 + X_2$ の分散について、コーシーシュワルツの不等式を用いると、証明できる。

$$
\begin{aligned}
Var[tX_1 + X_2] &= E\left[(tX_1 + X_2 - E[tX_1 + X_2])^2\right] \\
&= E\left[(tX_1 - E[tX_1] + X_2 - E[X_2])^2\right] \\
&= E\left[(tX_1 - E[tX_1])^2 - 2t(X_1 - E[X_1])(X_2 - E[X_2]) + (X_2 - E[X_2])^2\right] \\
&= t^2 V[X_1] + 2tCov[X_1, X_2] + V[X_2]
\end{aligned} \tag{19}
$$

この式は非負でなければならないので、t^2 に関する二次関数と見て、判別式が次の不等式を満たすべきでる。

$$(Cov\,[X_1, X_2])^2 - V\,[X_1]\,V\,[X_2] \leq 0 \tag{20}$$

この不等式を整理すると、元の不等式

$$-\sigma_1\sigma_2 \leq \sigma_{12} \leq \sigma_1\sigma_2 \tag{21}$$

を得る。等式が成り立つのは、

$$E\left[(tX_1 + X_2)^2\right] = 0 \tag{22}$$

が成り立つ時のみ。この時 X_1 と X_2 の間には

$$aX_1 + bX_2 = 0 \tag{23}$$

という関係が成り立っている。

2.1.5 相関係数 $\rho_{11} = \mu_{11}/\sqrt{\mu_{20}\mu_{02}}$

X_1 と X_2 の共分散 $Cov\,[X_1, X_2]$ を、それぞれの分散の積 $V\,[X_1]\,V\,[X_2]$ の平方根で割った量

$$\rho = \frac{Cov\,[X_1, X_2]}{\sqrt{V\,[X_1]\,V\,[X_2]}} = \frac{\sigma_{12}}{\sigma_1\sigma_2}$$

を 相関係数 という。不等式 (21) から、相関係数には

$$-1 \leq \rho \leq 1 \tag{24}$$

の不等式が成り立つ。また前節の結果から、相関係数 ρ が 0 となるのは、X_1 と X_2 が直線関係 $aX_1 + bX_2 = 0$ にある時のみとなる。

2.2 周辺分布

複数の試行を同時には行うが、一部の試行のみ、結果を観測できるとき、観測できる部分だけが従う確率分布を 周辺確率分布 あるいは 周辺分布 と言う。

2.2.1 同時試行と一部観測

まず二次元の確率変数の確率分布について、周辺分布は一次元の確率変数の確率分布になる、という事実を確認しておく。そのために、まず 試行 と 観測 の関係を述べる。

2 枚のコインを同時に投げる問題を考える。1 枚だけ投げればベルヌーイ試行だが、2 枚を同時に投げると、双方とも表になる確率が、1 枚ずつ投げたときの表が出る確率の積になるとは限らない。言い換えれば、2 枚分の試行が、互いに独立とは限らない。

表 2 枚のコインを同時に投げる試行

	2:表	2:裏
1:表	0.3	0.2
1:裏	0.2	0.3

さて、試行は同時に行われるが、1 枚目のコインの結果だけが観測される状況を考える。この時、1 枚目のコインについての観測は、表が出る確率も裏が出る確率も 0.5 のベルヌーイ試行と等しくなる 。

表 2 枚のコインを同時に投げるが、1 枚目のみが観測される試行

	周辺確率
1:表	0.5
1:裏	0.5

周辺分布を考える時は同時分布よりは情報が損なわれている、という。それは例えば、相関や共分散が計算できない、あるいは観測されない試行との同時分布が得られないなど、本来の確率的な変動の一部が観測されないことを意味する。

2.2.2 周辺分布関数、周辺確率関数

$$F_1(x_1) = F(x_1, \infty) = Pr[X_1 \leq x_1, X_2 < \infty] \tag{25}$$

$$p_1(k_1) = \sum_{k_2=0}^{\infty} p(k_1, k_2) = Pr[X_1 = x_1, X_2 < \infty] \tag{26}$$

2.2.3 周辺密度関数

$$f_1(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2)\, dx_2 \tag{27}$$

2.2.4 周辺モーメント

周辺分布 F_1 のモーメントを、周辺モーメント という。

2.3 条件付き分布

同時分布が $F(x_1, x_2)$ の二変量確率分布に関して、周辺分布は $F_1(x_1)$ と $F_2(x_2)$ の 2 つしかない。しかし条件付き分布は $F_{1|2}(x_1|X_2 \leq x_2), x_2 \in (-\infty, \infty)$ と $F_{2|1}(x_2|X_1 \leq x_1), x_1 \in (-\infty, \infty)$ があり、それぞれの条件の部分の値の場合を考えると、無数に存在することもある。先週の配布資料の 2.1 節の図に照らせば、確率分布を横から眺めるのは軸方向に限れば 2 通りしかないが、確率分布の断面は無数に考えうる、ということになる。

また工学では、条件付けはとても重要であり、図 4 にあるように、時点 0 において何の情報もなく時点 100 までの同時分布を考えなければならない状況と、時点 70 までの情報が得られていて、その条件の下でその先の同時分布を考えれば良い状況とでは、物事の不確実さの度合いが異なる。

2.3.1 条件付き分布関数、条件付き確率関数

$$F_{1|2}(x_1|X_2 \leq x_2) = \frac{F(x_1, x_2)}{F_2(x_2)} \tag{28}$$

$$p_{1|2}(k_1|X_2 = k_2) = \frac{p(k_1, k_2)}{p_2(k_2)} \tag{29}$$

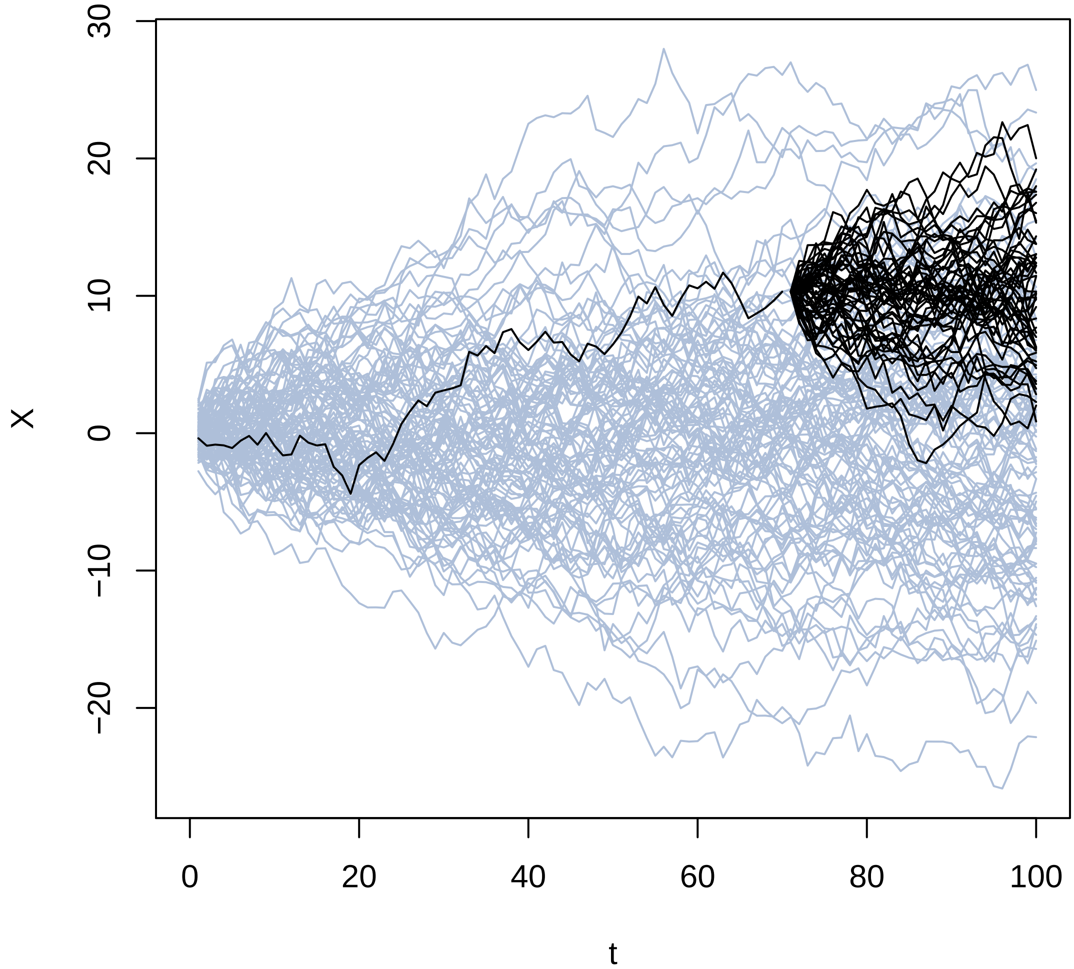

図 4: 条件付き分布と条件無し分布

2.3.2　条件付き密度関数

$$f_{1|2}(x_1|X_2 = k_2) = \frac{f(x_1, x_2)}{f_2(x_2)} = \frac{f(x_1, x_2)}{\int_{-\infty}^{\infty} f(x_1, x_2)\, dx_1} \tag{30}$$

2.3.3　条件付きモーメント

条件付き分布のモーメントを、条件付きモーメントという。

2.4　例

2.4.1　二変量正規分布

先週の資料の後半、あるいは参考書の 6 章を参照のこと。

2.4.2 三項分布

省略。

3 中心極限定理

3.1 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

$$Pr\left[\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sqrt{n}\sigma} \leq x\right] \to \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \quad (n \to \infty) \tag{31}$$

が成り立つ。

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。

3.2 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 5 から図 7 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 5 は標準正規分布 $N(0,1)$、図 6 はガンマ分布 $\Gamma(2,3)$、図 7 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 5 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 6 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 7 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。

付録

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	[,11]	[,12]	[,13]	[,14]	[,15]	[,16]	[,17]	[,18]
[1,]	-0.1	0.4	-1.2	-0.9	1.3	0.3	1.6	0.4	0.5	-1.0	-1.1	-0.6	0.3	1.0	0.4	-0.6	-0.2	-0.6
[2,]	0.5	0.3	1.2	1.0	0.6	1.0	-2.2	0.8	0.6	-1.0	0.6	-0.1	-0.1	-1.3	0.4	-0.4	-1.5	-1.3
[3,]	0.7	1.3	-0.1	-0.3	1.1	-0.8	-0.7	0.7	1.0	-0.4	-0.4	0.5	-0.5	0.4	0.4	-0.3	1.0	-1.8
[4,]	-0.6	-0.2	1.4	0.0	-1.1	-1.0	0.3	0.4	0.6	-1.5	1.9	1.2	0.0	-1.0	-1.0	0.3	2.0	0.5
[5,]	0.1	-1.0	0.6	0.9	0.9	-0.9	0.5	-0.4	-1.2	-0.6	-0.6	1.3	1.6	-0.7	0.9	0.8	-0.4	-0.3
[6,]	1.4	0.2	-0.8	-1.4	2.7	0.6	0.5	1.0	0.6	-0.7	0.1	0.4	0.5	2.2	0.2	1.7	0.5	1.4
[7,]	1.7	-0.3	-1.0	-0.1	-0.2	0.6	0.9	-0.2	0.5	0.0	-0.5	1.0	-0.2	1.1	1.2	-0.7	0.6	-2.1
[8,]	0.3	0.1	-1.0	0.2	0.8	-0.1	-0.4	0.8	-0.9	0.4	-0.8	-0.6	1.5	-1.2	0.7	0.8	-1.8	-0.9
[9,]	-0.6	1.0	0.4	-0.8	-0.6	-0.5	1.5	0.4	1.2	-1.8	0.1	-2.1	0.3	-0.1	0.1	0.4	-1.8	-0.4

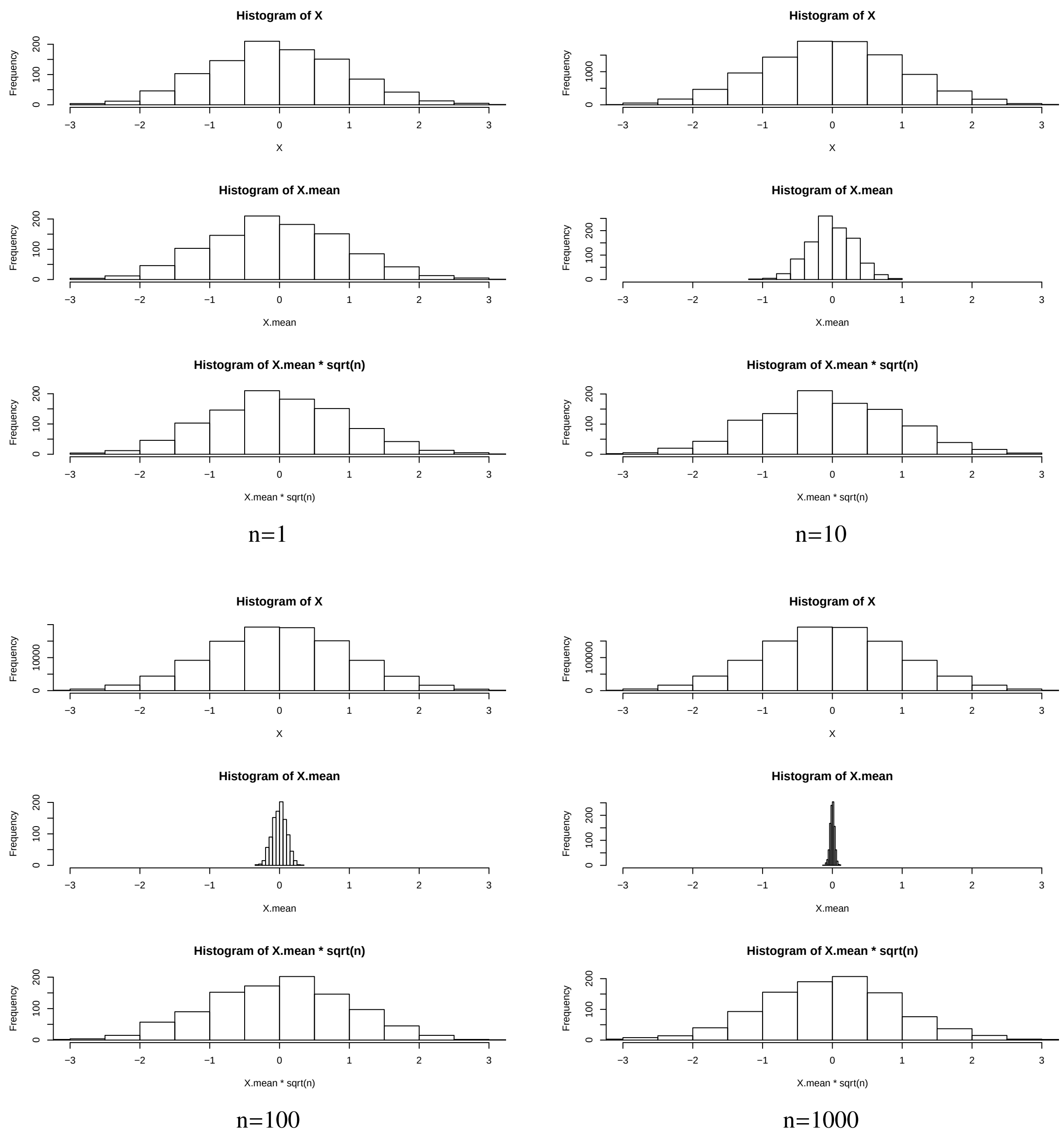

図 5: 中心極限定理の例 (正規分布の場合)
ヒストグラムはそれぞれ、上段：元のデータ、中段：(平均-真値)、下段：(平均-真値) $\sqrt{n}/\sigma$

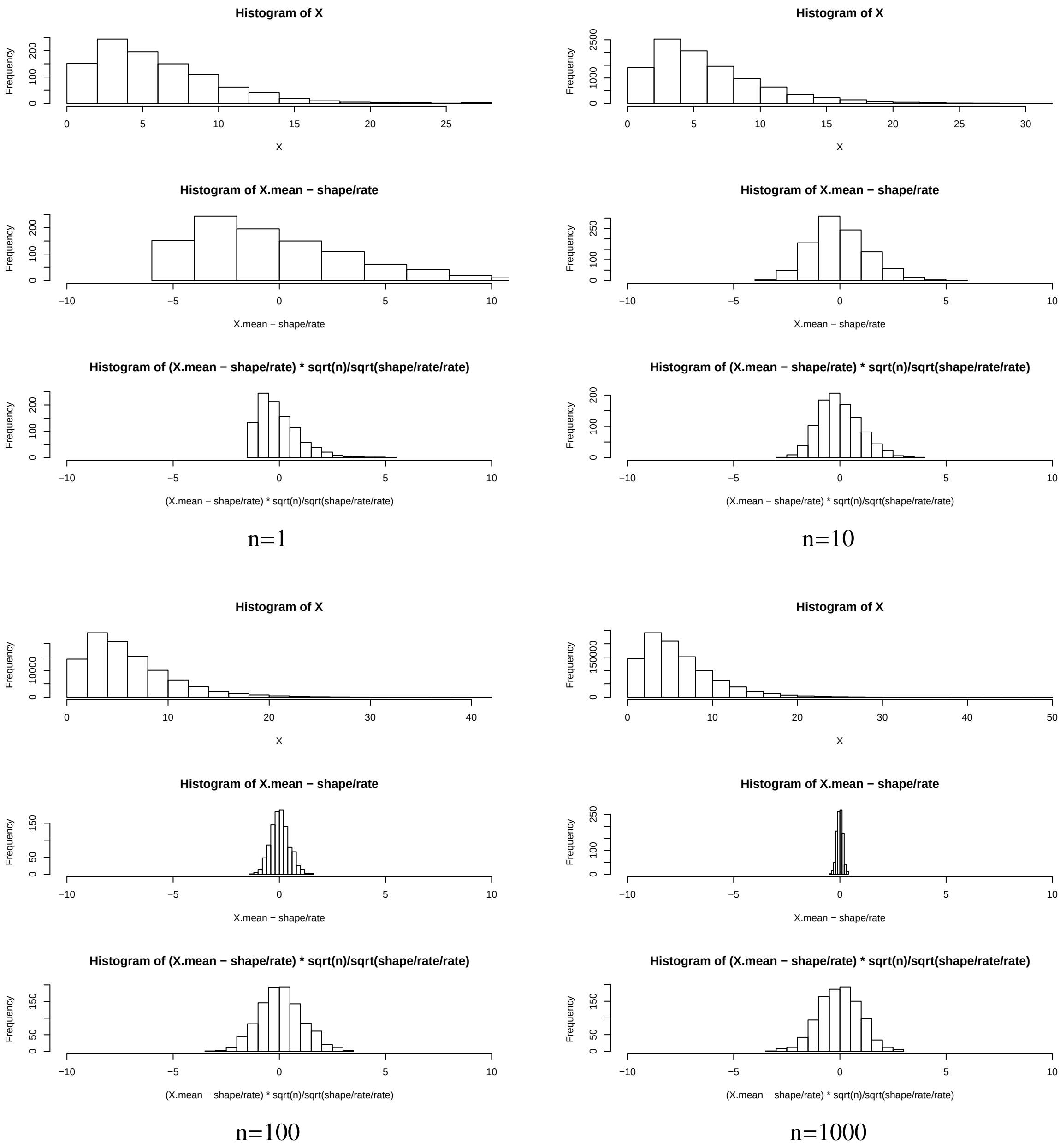

図 6: 中心極限定理の例 (ガンマ分布の場合)
ヒストグラムはそれぞれ、上段：元のデータ、中段：(平均-真値)、下段：(平均-真値) $\sqrt{n}/\sigma$

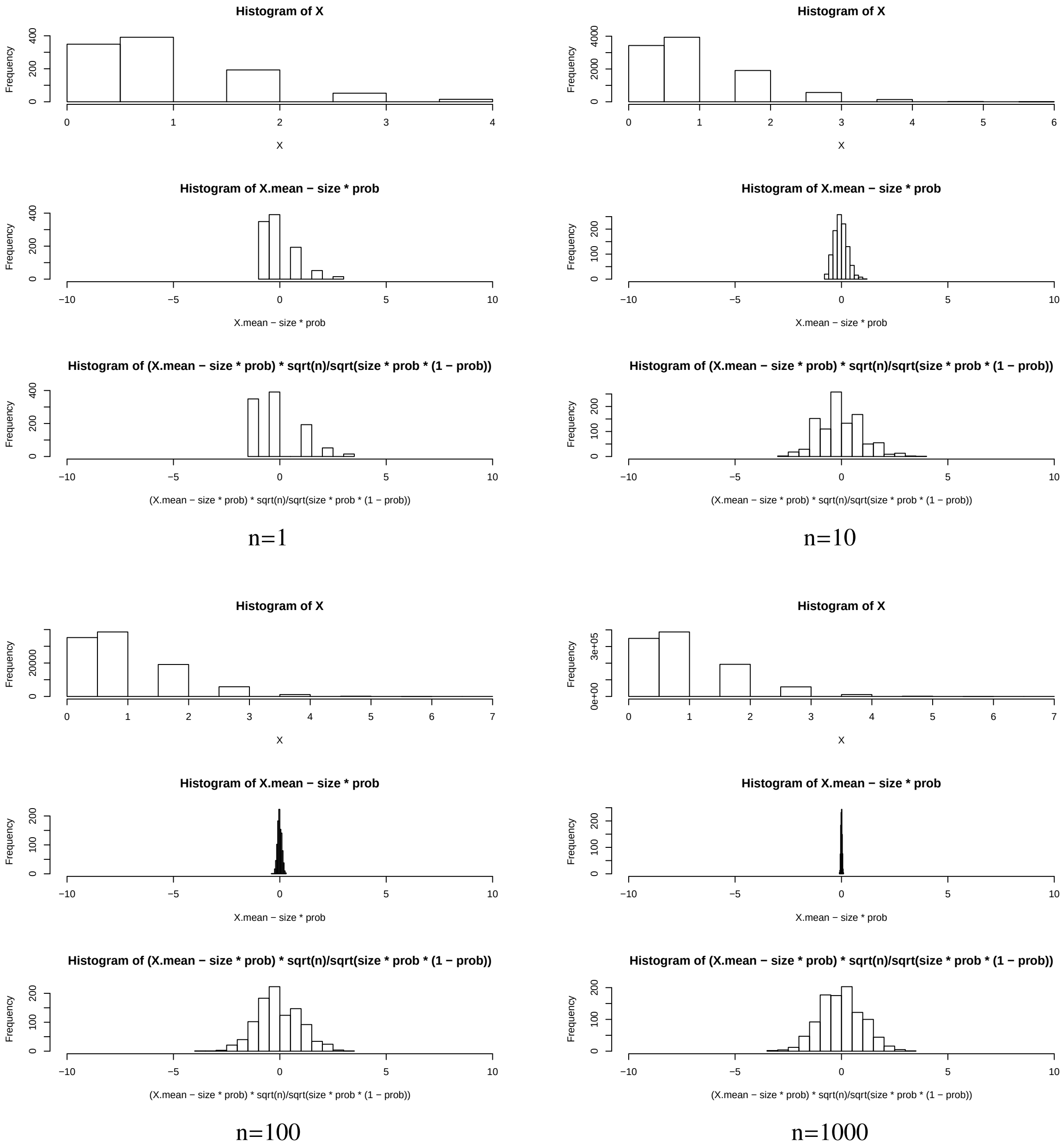

図 7: 中心極限定理の例 (二項分布の場合)
ヒストグラムはそれぞれ、上段：元のデータ、中段：(平均-真値)、下段：(平均-真値) $\sqrt{n}/\sigma$