放射線の計算や測定における統計誤差

- 「平均の誤差」とその応用（1H)
- 2項分布、ポアソン分布、ガウス分布（1H）
- 最小二乗法（1H）

1.1 分散と標準偏差

ある量xをn回だけ繰り返し測定・計算し、i番目のxの値をx_iとする。
平均は

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

cm, cGy （単位の例）

分散s^2:1個1個のxのばらつきの程度を表す。
個々のxの値の平均x_{av}からの差の2乗の平均

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

cm^2, cGy^2

標準偏差s:分散の平方根。
xと同じ次元であるので、xのばらつきを表すのに多用される。

$$s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

cm, cGy

合計の誤差と平均の誤差の直感的な説明

- 100 cpmを1分間測定 100±10
  - N=100 x 1 min =100, σ=100$^{1/2}$=10
- 100 cpmを1分測定、4回繰り返し 400±20
  - N=100 cpm x 1 min x 4 =400,
  - σ=400$^{1/2}$ = 20
  - この”20”は10 x 4$^{1/2}$でも計算ができる。

n個のx_iの合計の誤差s_yは、x_iの誤差のn$^{1/2}$倍である。

- 上記を4 で除して1分当たりに戻す 100±5
  - この”5”は、10/4$^{1/2}$でも計算できる。

n個のx_iの平均の誤差s_{x_bar}は、x_iの誤差の1/n$^{1/2}$である。

1.2 合計(y)の誤差

$$
\begin{aligned}
y &= \sum_{i=1}^{n} x_i \\
&= x_1 + x_2 + \cdots + x_N \\
&= n\bar{x}
\end{aligned}
$$

yの分散は、x_iの偏差Δx_iを用いて次式で計算される。(誤差の伝搬)

$$
\begin{aligned}
\Delta y^2 &= \left(\frac{\partial y}{\partial x_1}\right)^2 \Delta x_1^2 + \left(\frac{\partial y}{\partial x_2}\right)^2 \Delta x_2^2 + \cdots + \left(\frac{\partial y}{\partial x_n}\right)^2 \Delta x_n^2 \\
&= 1 \cdot s^2 + 1 \cdot s^2 + 1 \cdots \cdot s^2 \\
&= ns^2
\end{aligned}
$$

ここで、偏微分の式を書く都合上、yの分散s_y^2をΔy^2と書いた。
yの標準偏差s_yはyの分散s_y^2の平方根である。

$$
\boxed{
\begin{aligned}
s_y &= \sqrt{s_y^2} \\
&= \sqrt{ns^2} \\
&= \sqrt{n}s
\end{aligned}
}
$$

n個のx_iの合計の誤差s_yは、x_iの誤差の$n^{1/2}$倍である。

1.3 x_iの平均x_{av}の誤差

$$
\begin{aligned}
\bar{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i \\
&= \frac{1}{n}\{x_1 + x_2 + \quad \cdots \quad + x_N\}
\end{aligned}
$$

x_{av}の分散は、x_iの偏差Δx_iを用いて次式で計算される。（誤差の伝搬）

$$
\begin{aligned}
\Delta \bar{x}^2 &= \left(\frac{\partial \bar{x}}{\partial x_1}\right)^2 \Delta x_1^2 + \left(\frac{\partial \bar{x}}{\partial x_2}\right)^2 \Delta x_2^2 + \quad \cdots \quad + \left(\frac{\partial \bar{x}}{\partial x_n}\right)^2 \Delta x_n^2 \\
&= \left(\frac{1}{n}\right)^2 \cdot s^2 + \left(\frac{1}{n}\right)^2 \cdot s^2 + \quad \cdots \quad + \left(\frac{1}{n}\right)^2 \cdot s^2 \\
&= n\left\{\left(\frac{1}{n}\right)^2 \cdot s^2\right\} \\
&= \frac{1}{n}s^2
\end{aligned}
$$

ここで、偏微分の式を書く都合上、x_{av}の分散$s_{xav}{}^2$を$\Delta x_{av}{}^2$と書いた。
x_{av}の標準偏差s_{xav}はx_{av}の分散$s_{xav}{}^2$の平方根である。

$$
\boxed{s_{\bar{x}} = \sqrt{s_{\bar{x}}^2} = \sqrt{\frac{1}{n}s^2} = \sqrt{\frac{1}{n}}s}
$$

n個のx_iの平均の誤差s_{x_bar}は、x_iの誤差の$1/n^{1/2}$である。

中心極限定理

- 分布がどのようなものであっても、平均μ、分散σ^2をもつ母集団からとられた大きさnの標本の平均値x_{av}の分布は、nが大きくなるとき正規分布$N(\mu、\sigma^2/n)$に近づく。

宮川公男、「基本統計学」

- モンテカルロ法の数学的裏付け

1.4 $N \pm \sqrt{N}$ の導出

x の分散　放射線計測において、各時間帯での計数 x_i の値が 0 または 1 であるよう時間帯を狭くして x_i を記録する場合を考える。あるいは、モンテカルロ計算の結果が、各入射粒子に対して 0 または 1 である場合を考える。時間帯の数 $n = 10$ とした場合の、x_i の分布の例を図 1 および図 2 に示す。このような値が 0 または 1 の 2 項分布を「単位 2 項分布」あるいは「ベルヌーイ分布」という。当たり率を a（応

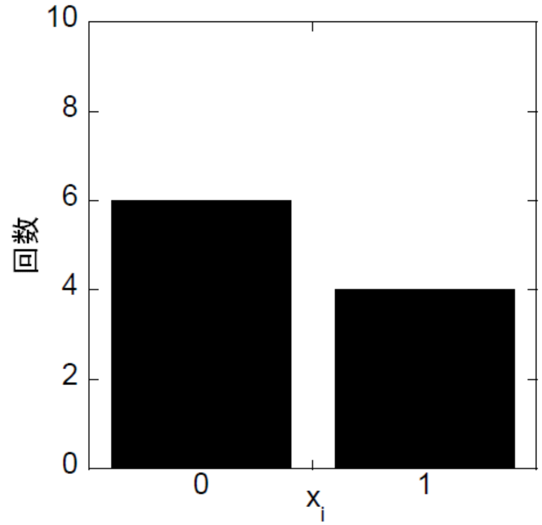


Figure 1: x_i の分布 時間帯の数 n=10、計数ありの回数=4 の場合。

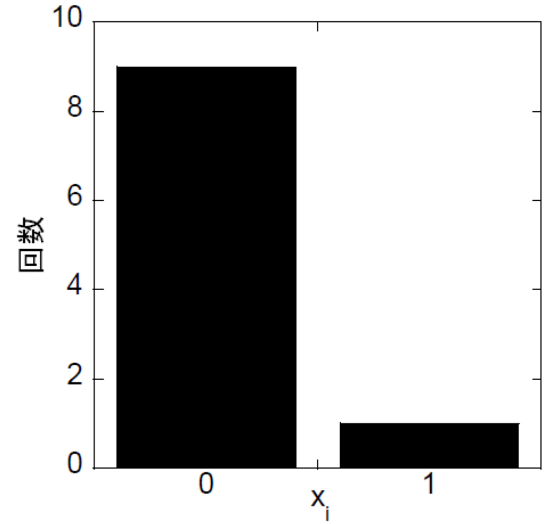


Figure 2: x_i の分布 時間帯の数 n=10、計数ありの回数=1 の場合。

答が1である確率）として x_i の平均 $\bar{x}$ を求めると、

$$\begin{aligned}\bar{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i \\ &= \frac{1}{n}(na\cdot 1 + n(1-a)\cdot 0) = a.\end{aligned}$$

次に、na 回だけ $x_i = 1$、$n(1-a)$ 回だけ x_i=0 であることを用いて分散 s_x^2 を求めると、

$$\begin{aligned}s^2 &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}(x_i - a)^2 \\ &= \frac{1}{n}\left[na(1-a)^2 + n(1-a)(0-a)^2\right] \\ &= \frac{1}{n}na(1-a)\left[(1-a)+a\right] \\ &= a(1-a) \qquad (8)\end{aligned}$$

例えば、図1および図2において、それぞれ $a = 0.4$ および $a = 0.1$ であるので、$s^2 = 0.24$ および $s^2 = 0.09$ である。このように s^2 の値は a のみに依存し、n には依存しない。標準偏差 s は、次式で計算することができる。

$$s = \sqrt{a(1-a)} \qquad (9)$$

s と a の関係を調べると、s は図3に示すように $a = 0$ と $a = 1$ を直径とする半円上にある。

合計の誤差　前節の (6) 式で示したように、x_i の合計 y の誤差は、x の標準偏差の $\sqrt{n}$ 倍である。これと上記 (8) 式を組み合わせると、

$$\begin{aligned} s_y &= \sqrt{n}s \\ &= \sqrt{n}\sqrt{a(1-a)} \\ &= \sqrt{na(1-a)} \\ &= \sqrt{y(1-a)} \end{aligned}$$

合計とその誤差は、

$$y \pm \sqrt{y(1-a)} \tag{10}$$

この式の導出により、0 と 1 からなる分布の分散が $y(1-a)$ であることが示された [2]。最後に $1-a \simeq 1$、当たり率が小さいと近似すると、

$$y \pm \sqrt{y} \tag{11}$$

が得られ、(1) 式と一致した。**ここまでの計算で、(1) 式を導出することができた**[4]。モンテカルロ計算の誤差の式も平均値の誤差の式に他ならないのであるから、モンテカルロ計算の誤差の式、(1) 式および 0 と 1 からなる分布の合計の誤差の 3 者は同じものであるとも言えるであろう。

最後に、式 (10) および式 (11) に示される誤差の値を計算してみる。試行回数 $n = 10000$ とし、当たり率 a の関数として s_y を計算すると図 4 が得られる。ここで、赤線と黒線はそれぞれ、$(1-a)$ の因子を考慮した場合と無視した場合の値である。例えば $a = 0.5$ の場合、計数とその誤差は 5000 ± 50 または 5000 ± 71 である。

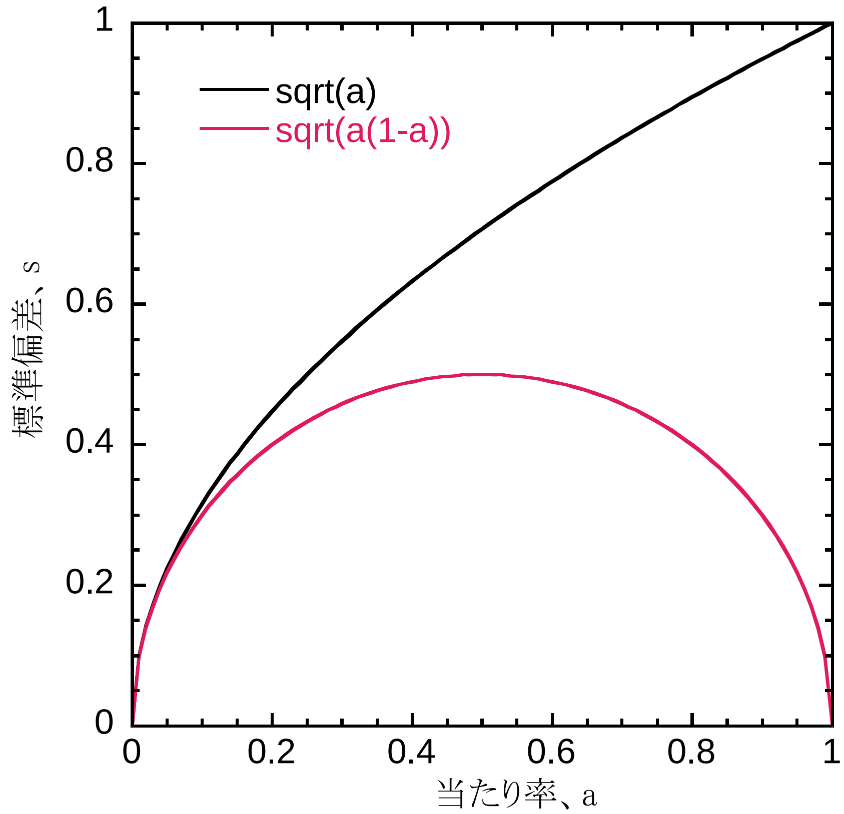


図3 xの標準偏差

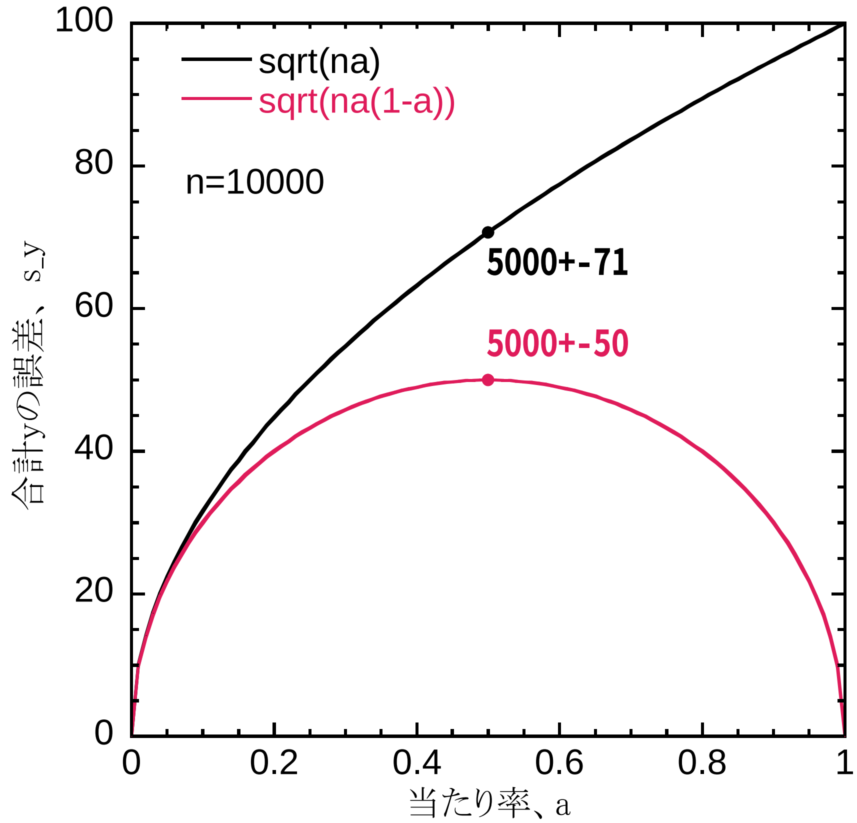


図4　合計yの誤差sy

平均の誤差　(13) 式を n で割ると、平均とその誤差が得られる。

$$a \pm \sqrt{\frac{1}{n}a(1-a)} \tag{14}$$

1.5　データ数に依存しない分散

(3) 式を用いて異なる n について分散 s^2 を計算すると、分散の値が n の値によって変化することが知られている。すなわち、s^2 の期待値は $(n-1)/n$ に比例する。$\bar{x}$ は x_i の平均であり、両者に関連があるためこのような現象が起こる。

　そこで平均の期待値 μ に対する分散 σ^2 を求めると、n に依存しないように分散が得られる。

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2 \tag{13}$$

(13) 式は分散の計算方法の一つであるものの、通常は μ が未知数であるため (13) 式を用いて分散を計算することはできない。n に依存しない分散を求めるもう一つの方法として、(3) 式の $1/n$ を $1/(n-1)$ とする方法がある。

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 \tag{14}$$

本当かな？→s^{2}を数値計算

分散s^2の数値計算

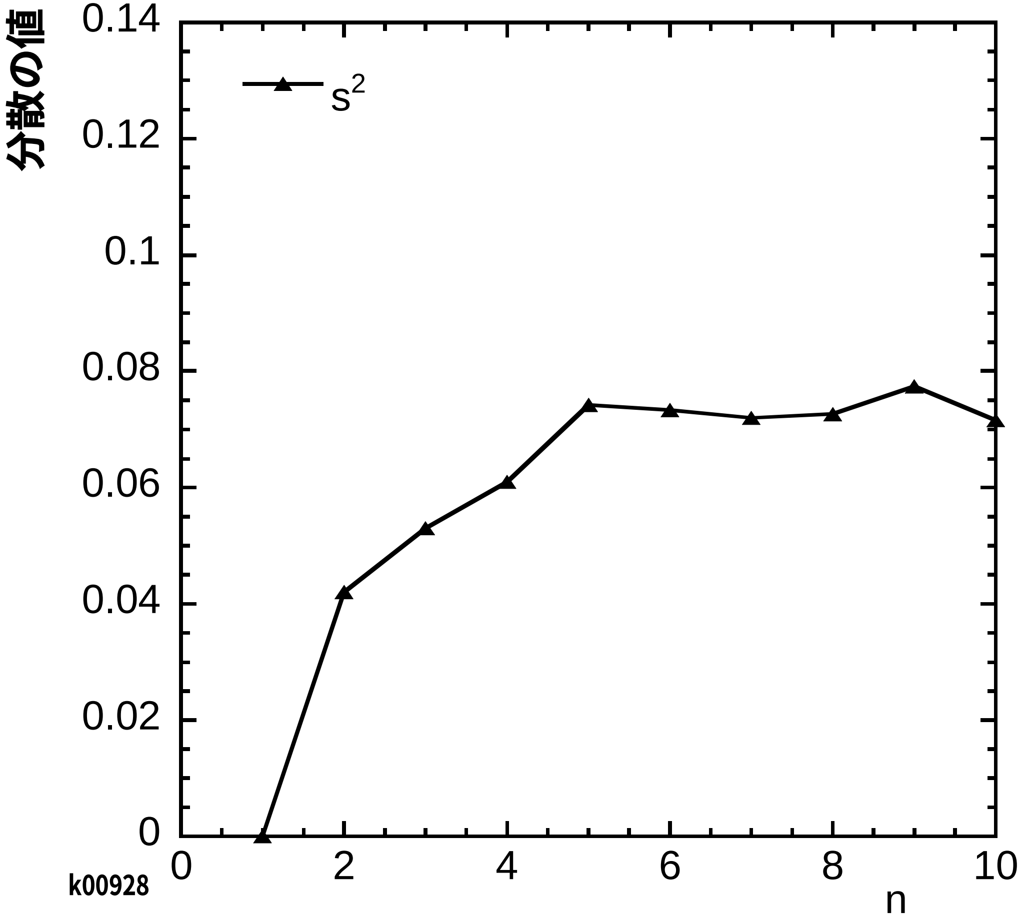
- ExcelのRAND()関数：(0,1)の乱数を発生
  - 10個の乱数の平均と分散s^2を計算。
  - 100組を計算し、s^2の平均を調べる。
  - 乱数の数n=9 ～ n=1に変えて繰り返す。

Microsoft Excel - 917_統計.xls

N33

	A	B	C	D	E	F	G	H	I	J	K	L
1	乱数1	乱数2	乱数3	乱数4	乱数5	乱数6	乱数7	乱数8	乱数9	乱数10	平均	s2
2												
3	0.613968	0.447353	0.887473	0.411791	0.790167	0.46324	0.805012	0.372194	0.037351	0.908968	0.573752	0.695497
4	0.309133	0.009213	0.311154	0.737296	0.188814	0.171874	0.572172	0.863507	0.82629	0.842526	0.483198	0.93209
5	0.141664	0.740451	0.109383	0.835378	0.089547	0.769164	0.972715	0.903185	0.270545	0.602563	0.54346	1.122502
6	0.91265	0.644089	0.312647	0.8405	0.118923	0.941285	0.764885	0.73743	0.69444	0.532899	0.649975	0.622537
7	0.795863	0.47724	0.457544	0.703185	0.431988	0.014473	0.519322	0.851533	0.844248	0.848248	0.594364	0.646188
8	0.277756	0.85828	0.053229	0.441752	0.109082	0.441465	0.919993	0.765263	0.856533	0.681566	0.540492	0.927441
9	0.869281	0.002952	0.636036	0.438154	0.494728	0.177196	0.232511	0.705837	0.107985	0.893335	0.455802	0.912759
10	0.783704	0.770169	0.218915	0.494705	0.953385	0.921915	0.234182	0.854934	0.743638	0.885009	0.686056	0.67449

s^2のエクセルでの計算結果



確かに、nに依存してs^2が変化した！

サンプル数nに依存しない分散σ^2

- 平均の期待値μ(=0.5)との差の2乗和の平均

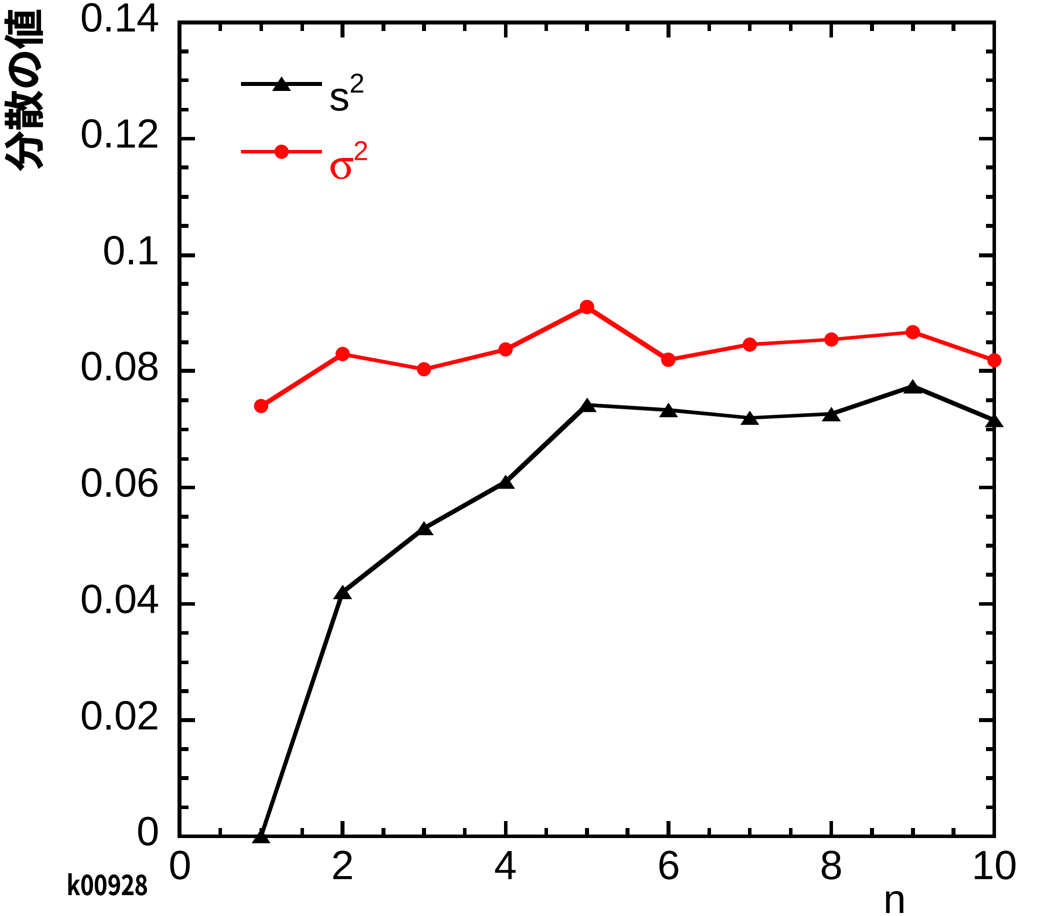
$$\boxed{\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2}$$

- エクセルでの数値計算
  - 10個の乱数の分散σ^2を計算。
  - 100組を計算し、σ^2の平均を調べる。
  - n=9 ～ n=1に変えて繰り返す。

σ^2
↓

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	乱数1	乱数2	乱数3	乱数4	乱数5	乱数6	乱数7	乱数8	乱数9	乱数10	平均	s2	sigma2
2													0.5
3	0.010153	0.038243	0.851855	0.44466	0.761285	0.595101	0.561725	0.157182	0.366967	0.752004	0.453918	0.838651	0.859887
4	0.714174	0.753715	0.033377	0.849182	0.76502	0.148712	0.500107	0.471678	0.6217	0.624779	0.548244	0.651452	0.674728
5	0.120007	0.064013	0.146619	0.905152	0.470914	0.004578	0.137465	0.815979	0.507815	0.94207	0.411461	1.218165	1.296557
6	0.6851	0.774129	0.776231	0.059662	0.519471	0.64841	0.133096	0.693669	0.118943	0.267886	0.46766	0.762764	0.773223
7	0.485525	0.139318	0.008247	0.000792	0.594535	0.485401	0.933304	0.817618	0.234627	0.910778	0.461015	1.143076	1.158274
8	0.70408	0.767963	0.228696	0.887579	0.67829	0.106101	0.459718	0.896207	0.802027	0.860475	0.639113	0.710459	0.903984
9	0.384619	0.910509	0.323403	0.033323	0.540377	0.44357	0.029983	0.437673	0.094165	0.423782	0.36214	0.640878	0.83093
10	0.236705	0.032559	0.39276	0.106378	0.184327	0.735616	0.798941	0.244476	0.453464	0.43319 7	0.361842	0.57984	0.770715

σ^2のエクセルでの計算結果



確かに、σ^2はnに依存しない。

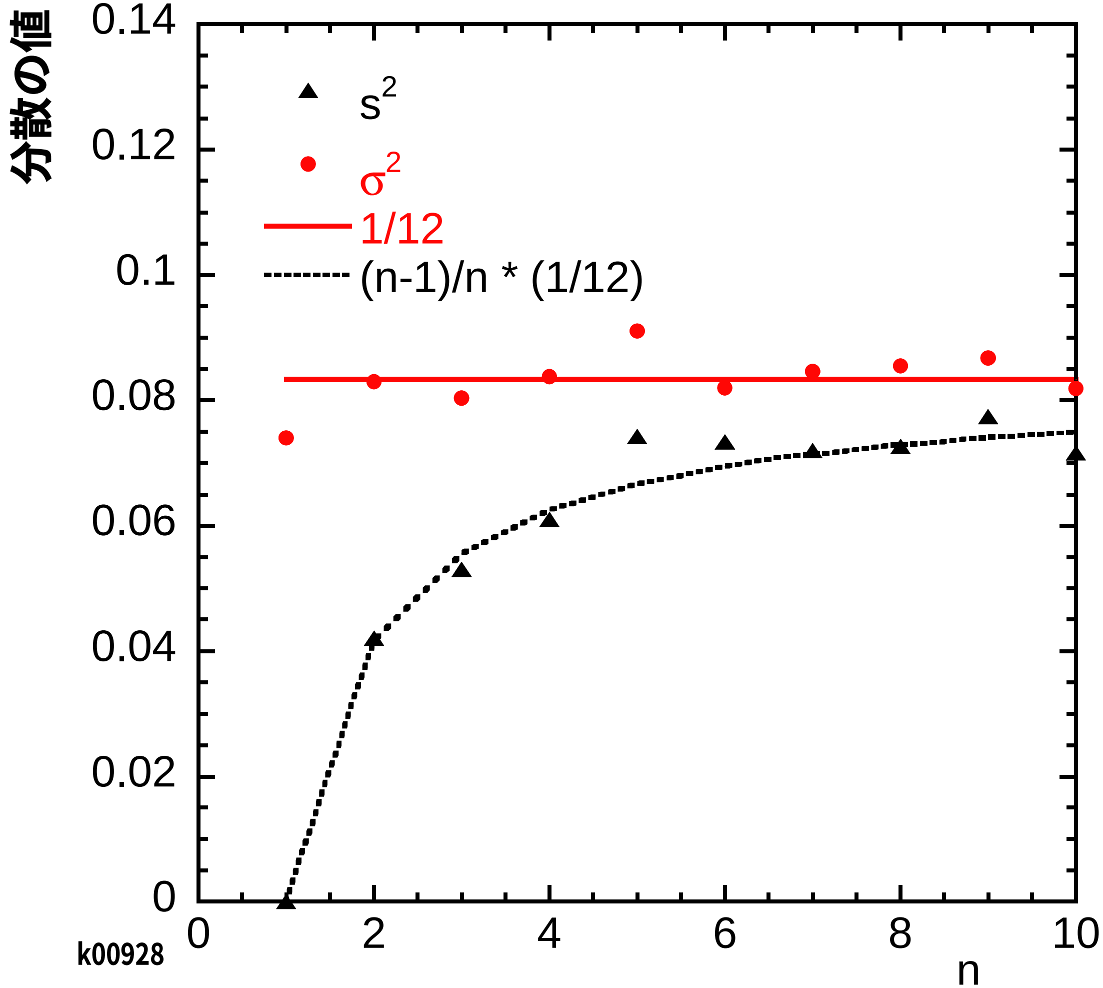
μ は通常、未知数なので、σ^2は分散の計算に使用できない
→困った！

とりあえず、σ²の期待値は？

$$\int_0^1 (x-0.5)^2 dx = \int_{-0.5}^{0.5} y^2 dy = \left[\frac{y^3}{3}\right]_{-0.5}^{0.5} = \frac{1}{3}\left(\frac{1}{8} - \frac{-1}{8}\right) = \frac{1}{12}$$

s^2の期待値は？

$E(s^2)=E(\sigma^2) \times (n-1)/n$



サンプル数nに依存しない分散（2）

s^2の1/nを1/(n-1)に変更した次式で、サンプル数nに依存しない分散を求める。

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

これを、標本分散あるいは分散の不偏推定値と呼ぶ教科書もある。

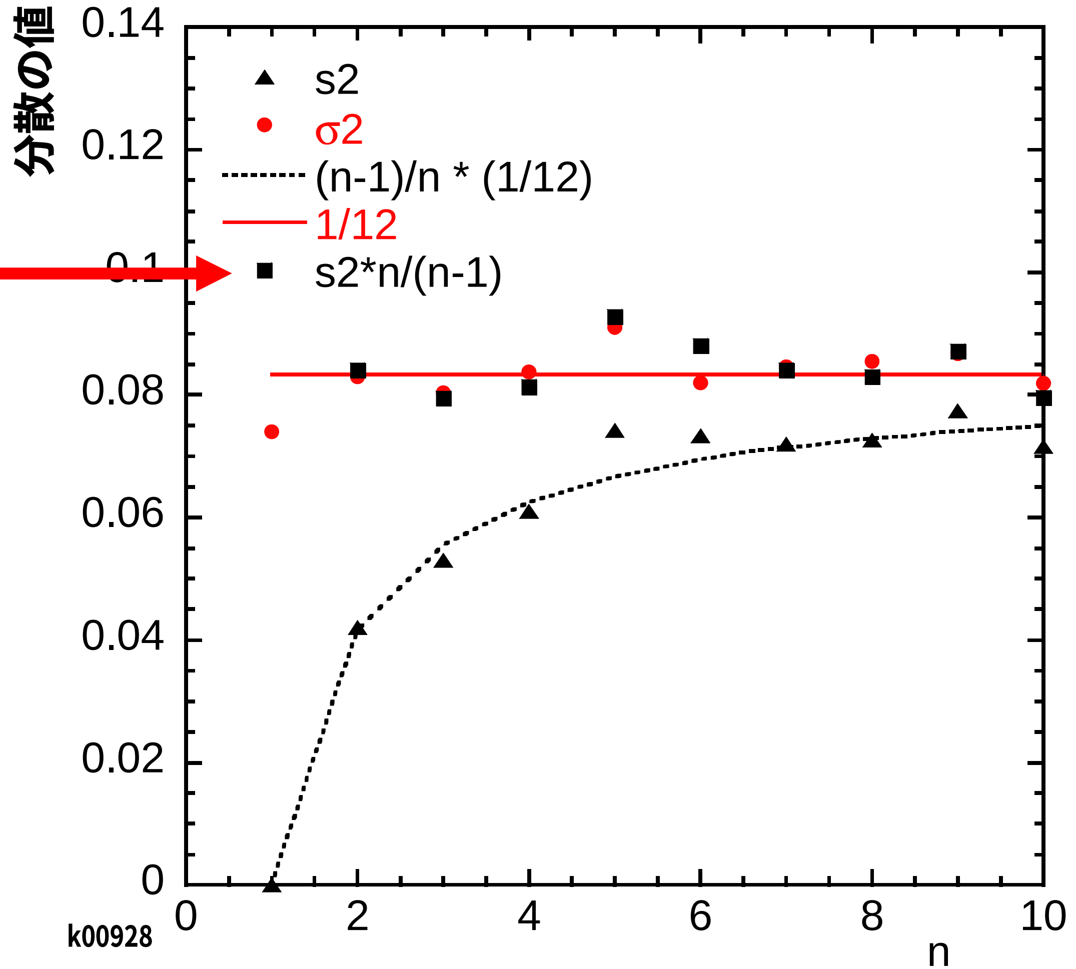


図1 分散の値の比較

データ数に依存しない分散の証明

平均の期待値μに対する分散の式を変形する

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2 &= \frac{1}{n}\sum_{i=1}^{n}\{(x_i-\bar{x})+(\bar{x}-\mu)\}^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2+\frac{1}{n}\sum_{i}(\bar{x}-\mu)^2+\frac{2}{n}\sum_{i=1}^{n}(x_i-\bar{x})(\bar{x}-\mu)
\end{aligned}
$$

期待値に書き直す。(右辺第3項は0)

$$
\boxed{E\left[\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2\right] = E\left[\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right]+E\left[\frac{1}{n}\sum_{i}(\bar{x}-\mu)^2\right]}
$$

μに対する分散 = x_{av}に対する分散 + x_{av}の分散

$$
\begin{aligned}
\sigma^2 &= E\left[\frac{1}{n}\sum_{i}(x_i-\bar{x})^2\right]+\frac{\sigma^2}{n} \\
\sigma^2 &= E\left[\frac{1}{n-1}\sum_{i}(x_i-\bar{x})^2\right] \\
&= E\left[\hat{\sigma}^2\right]
\end{aligned}
$$

1章「平均の誤差」のまとめ

A.1 ucnaicgv.f の統計誤差の説明

EGS5 コードのサンプルユーザーコードとして ucnaicgv.f を KEK で作成し、講習会などに使用している。このユーザーコードでは、NaI 検出器に光子を入射させ、ピーク効率、全講率、吸収エネルギー分布などを計算する。このコードの説明書 [9] の「統計誤差」の節では次のような説明をしている。

『x をモンテカルロ計算で計算したい量（スコアする量）とする。モンテカルロ計算結果には、その統計誤差が必要である。`ucnaicgv.f` では、次のような MCNP で使用している方法を採用している。

- ヒストリー数を N とする。
- x_i を i 番目のヒストリーの結果とする。
- x の平均値を計算する:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad (25)$$

データ数に依存しない分散(*)

分散の定義

- x_i の分散値を以下の式から求める。:

$$s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2 \simeq \bar{x^2} - \bar{x}^2 \quad (\bar{x^2} = \frac{1}{N}\sum_{i=1}^{N} x_i^2). \qquad (26)$$

- $\bar{x}$ の分散値は、

$$s_{\bar{x}}^2 = \frac{1}{N}s^2 \simeq \frac{1}{N}[\bar{x^2} - \bar{x}^2] \qquad (27)$$

- 統計誤差として、

平均の分散

$$s_{\bar{x}} \simeq [\frac{1}{N}(\bar{x^2} - \bar{x}^2)]^{1/2} \qquad (28)$$

を用いる。

N±N$^{1/2}$の誤差と等価(*)

先の計算した結果とその自乗の和は、上記の処理に用いる。』

*平均の誤差から導出

付録B.1「平均値の誤差」の表記について

- 初等統計学：平均の標準誤差
- 放射線計測の理論と演習：平均値の標準誤差
- 放射線計測ハンドブック：記述なし
- 放射線計測（プライス）：記述なし
- 総務省統計局HP：標本誤差
- Wikipedia：標準誤差
- カレイダグラフ：標準誤差

「標本平均の標準誤差」の表記の省略の組み合わせ表

標本	平均(値)の	標準誤差
(省略)	(省略)	標準偏差
		誤差
		不確かさ

「平均値の誤差」と同じ意味でありそうな言語の検索結果

検索語	検索結果の件数
Error for the sample	3,520,000
Standard error of mean	780,000
平均値の誤差	656,000
平均値の標準偏差	416,000
標準誤差	297,000
Error for the sample mean	118,000
平均値の標準誤差	70,400
平均値の不確かさ	45,600
標本平均の標準偏差	8,500
標本誤差	7,200
標本平均の標準誤差	4,500

2章 2項分布、ポアソン分布、ガウス分布

- 分布の概要
- 合計、平均、分散
- 相互の関連

2　2 項分布、ポアソン分布とガウス分布

2.1　2 項分布

放射線計測やモンテカルロ計算における計数は、一般に 2 項分布に従う。例えば、本稿で取り上げている NaI 検出器の応答関数のモンテカルロ計算を考え、全エネルギー吸収ピークの割合に注目する。そして、n 個の入射粒子について計算を行った後の、全エネルギー吸収ピークのカウント数 x の分布を考えてみる。ある入射粒子が、全エネルギー吸収ピークに寄与する確率を p とすると、図 6 のように 1 番目から x 番目までの x 個が寄与する確率は、p^x である。また、寄与しない確率は $(1-p)$ であるので、$(x+1)$ 番目から n 番目までの $(n-x)$ 個が寄与しない確率は、$(1-p)^{(n-x)}$ である。従って、1 番目から x 番目までの x 個が寄与し、$(x+1)$ 番目から n 番目までの $(n-x)$ 個が関与しない確率は、これらの確率を掛け合わせて、

$$p^x(1-p)^{(n-x)}$$

である。実際には、このように最初の部分だけ連続して寄与し、その後は連続して寄与しないなどということは極めてまれである。それは、寄与があったりなかったりを一見乱雑に、繰り返すからである。このような、異なる n 個の物から x 個を取る組合せの数は、

$${}_nC_x = \frac{n!}{x!(n-x)!}$$

である。結局、n 個の入射粒子のうち x 個が寄与する確率は、

$$f(x) \;=\; {}_nC_x p^x (1-p)^{n-x} \tag{17}$$

と書くことができる。この様な分布を 2 項分布と呼ぶ。$n=10$、$p=0.4$ の場合について、(17) 式をプロットすると、図 7 のようになる。

　放射線計測におけるある時間内の計数も 2 項分布である。ここでは、2 項分布の和、平均、分散を導出する。

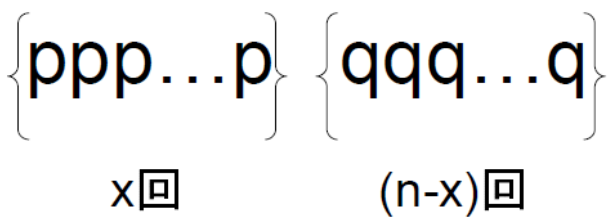


Figure 6: n 個の入射粒子に対するモンテカルロ計算結果の概念図。光電ピークへの寄与のあり、なしをそれぞれ p と q で示している。この図では、最初の x 回続けて、光電ピークへの寄与があり、その後の (n-x) 回続けて、光電ピークへの寄与がない場合を示している。

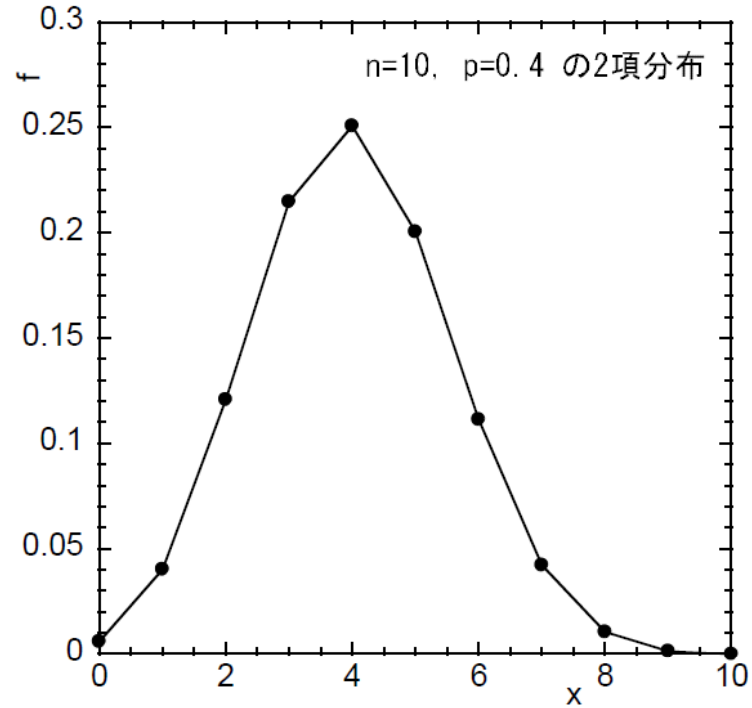


Figure 7: n=10, p=0.4 の 2 項分布

合計　2 項分布に従う確率密度関数 $f(x)$ の合計 $\sum_{x=0}^{n} f(x)$ を求める。次の式で表される「2 項定理」が知られている。

$$(a+b)^n = \sum_{x=0}^{n} \frac{n!}{x!(n-x)!} a^{n-x} b^x$$

ここで、a および b をそれぞれ p およびと q と書き換え、$p+q=1$ とすると、左辺=1、右辺=$\sum_{x=0}^{n} f(x)$ であるので、

$$\sum_{x=0}^{n} f(x) = 1$$

平均　離散型確率変数 x の平均 $E[x]$ を次式で与える。

$$E[x] = \sum_{x=0}^{n} x f(x)$$

２項分布の平均は、

$$\begin{aligned} E[x] &= \sum_{x=0}^{n} x C_x p^x q^{1-x} \\ &= \sum_{x=0}^{n} x \frac{n!}{x!(n-x)!} p^x q^{n-x} \end{aligned}$$

$x = 0$ の項は 0 であるので、この項を省略し、総和を $x = 1$ から始める。

$$E[x] = \sum_{x=1}^{n} x \frac{n!}{x!(n-x)!} p^x q^{n-x} = \sum_{x=1}^{n} x \frac{n!}{x(x-1)!(n-x)!} p^x q^{n-x} = \sum_{x=1}^{n} \frac{n!}{(x-1)!(n-x)!} p^x q^{n-x}$$

ここで $z = x - 1,\ j = n - 1$ とおくと、

$$\begin{aligned} E[x] &= \sum_{z=0}^{j} np \frac{j!}{z!(j-z)!} p^z q^{j-z} = np \sum_{z=0}^{j} \frac{j!}{z!(j-z)!} p^z q^{j-z} && (16) \\ &= np && (17) \end{aligned}$$

ここで (16) 式の最後の総和は、２項分布の総和であり、1 に等しいことを用いた。

Xの分散　V(x)

$$
\begin{aligned}
V(x) &= \frac{1}{n}\sum_{x=1}^{n}(x_i-\bar{x})^2 \\
&= \frac{1}{n}\sum_{x=1}^{n}x_i^2-\frac{2}{n}\bar{x}\sum_{x=1}^{n}x_i+\frac{1}{n}\sum_{x=1}^{n}\bar{x}^2 \\
&= \frac{1}{n}\sum_{x=1}^{n}x_i^2-\frac{2}{n}\bar{x}(n\bar{x})+\frac{1}{n}(n\bar{x}^2) \\
&= \frac{1}{n}\sum_{x=1}^{n}x_i^2-\bar{x}^2 \\
&= E[x^2]-E[x]^2
\end{aligned}
$$

分散　確率変数 x の分散 $V[x]$ を次式で与える。

$$V[x] = E[x^2] - E[x]^2$$

2 項分布の分散を得るために、まず $E[x^2]$ を計算する。

$$\begin{aligned}
E[x^2] &= \sum_{x=0}^{n} x^2 \, {}_nC_x p^x q^{n-x} \\
&= \sum_{x=0}^{n} x^2 \frac{n!}{x!(n-x)!} p^x q^{n-x} = \sum_{x=0}^{n} x^2 \frac{n!}{x(x-1)!(n-x)!} p^x q^{n-x} = \sum_{x=0}^{n} \frac{n!x}{(x-1)!(n-x)!} p^x q^{n-x} \\
&= \sum_{x=1}^{n} \frac{n!x}{(x-1)!(n-x)!} p^x q^{n-x}
\end{aligned}$$

上の式の最後の等号のところで、$x=0$ の項が 0 であるのでこの項を省略し、総和を $x=1$ から始めるという変更を行った。$z = x-1$, $j = n-1$ とおくと

$$\begin{aligned}
E[x^2] &= np \sum_{z=0}^{j} \frac{j!(z+1)}{z!(j-z)!} p^y q^{j-z} \\
&= np \sum_{z=0}^{j} \frac{j!z}{z!(j-z)!} p^z q^{j-z} + np \sum_{z=0}^{j} \frac{j!}{z!(j-z)!} p^z q^{j-z}
\end{aligned}$$

ここで、右辺の最初の総和は 2 項分布の平均 jp に等しく、右辺の 2 番目の総和は 2 項分布の総和 $(=1)$ に等しいので、

$$\begin{aligned}
E[x^2] &= np \cdot jp + np = n(n-1)p^2 + np \\
V[x] &= E[x^2] - E[x]^2 = n(n-1)p^2 + np - (np)^2 = np(1-p) = npq
\end{aligned}$$

2.2 ポアソン分布

当たり率が小さい場合に 2 項分布が、ポアソン分布と呼ばれる分布に漸近することが知られている。ポアソン分布の確率密度関数は次式で与えられる。

$$f(x) = e^{-\lambda}\frac{\lambda^x}{x!}$$

ここで、$\lambda = np$ である。ポアソン分布には次のような特徴がある。

- 2 項分布が試行回数と当たり率という 2 パラメータの分布であるのに対しポアソン分布は、平均値という 1 パラメータの分布であり簡便である。
- 2 項分布の変数が離散型（整数）であるのに対し、ポアソン分布の変数は連続型（実数）である。

図 8 に 2 項分布とポアソン分布との比較を示す。2 項分布は、(n=10,p=0.4), (n=20,p=0.2) および (n=40,p=0.1) の 3 種類。ポアソン分布は $\lambda = 4$。ここで $np = \lambda$ の関係が成り立つように、n と p の組合せを設定してある。p が減少するにつれて 2 項分布がポアソン分布に漸近することが分かる。

ポアソン分布の合計、平均、分散を導出してみる。

合計 e^λ をマクローリン展開すると、

$$e^\lambda = 1 + \lambda + \frac{1}{2!}\lambda^2 + \frac{1}{3!}\lambda^3 ... = \sum_{x=0}^{\infty}\frac{\lambda^x}{x!}$$

一方、ポアソン分布の合計は、

$$\sum_{x=0}^{\infty} e^{-\lambda}\frac{\lambda^x}{x!} = e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^x}{x!} = e^{-\lambda}e^\lambda = 1$$

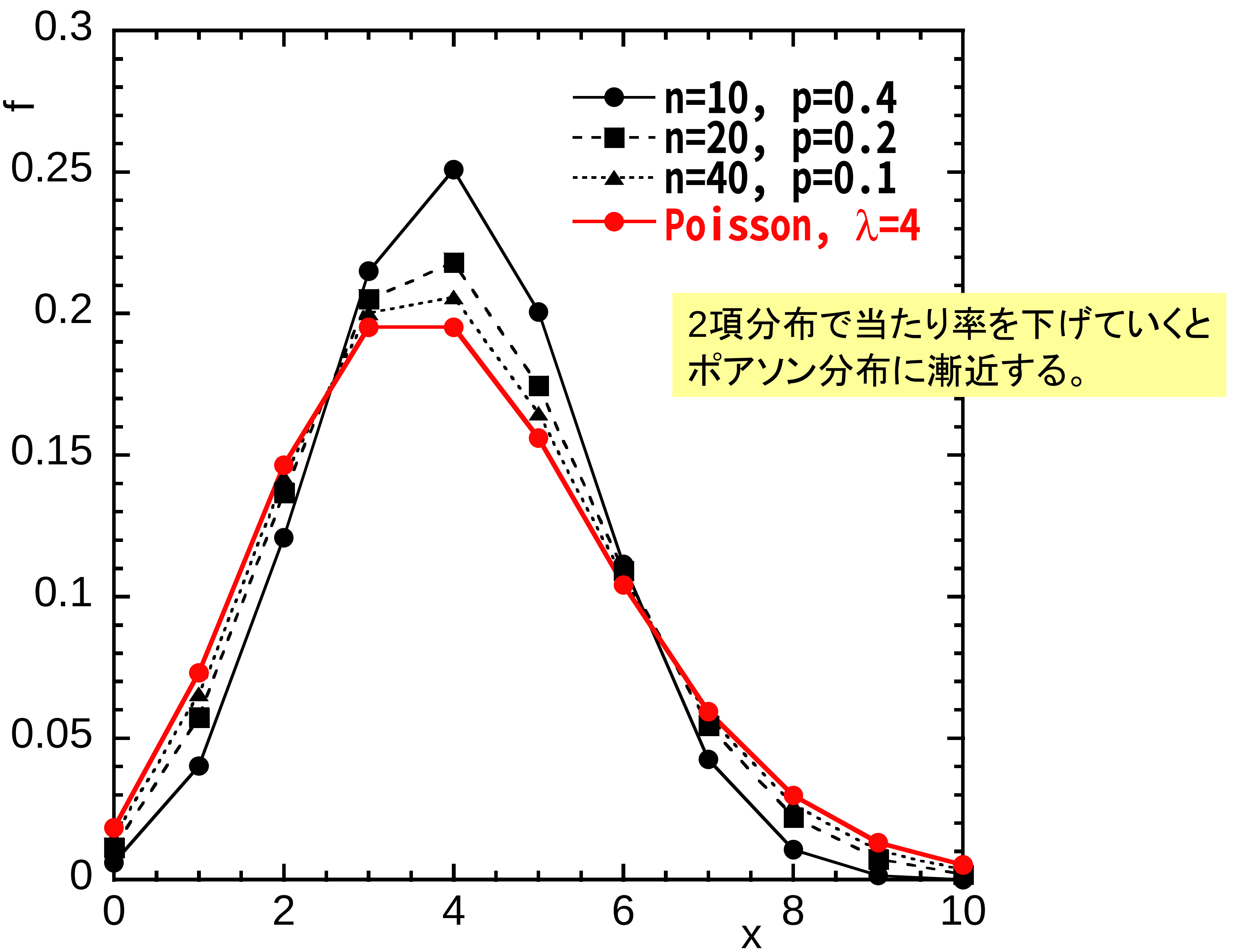


図8 2項分布とポアソン分布の比較

平均

$$E[x] = \sum_{x=0}^{\infty} x e^{-\lambda} \frac{\lambda^x}{x!} = \sum_{x=1}^{\infty} e^{-\lambda} \frac{\lambda^x}{(x-1)!}$$

ここで、$x = 0$ の項は 0 であるため消した。

$$E[x] = \sum_{x=1}^{\infty} \lambda e^{-\lambda} \frac{\lambda^{x-1}}{(x-1)!} = \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!}$$

ここで、$j = x - 1$ とおくと、

$$E[x] = \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!}$$

ここで、右辺の総和は e^{λ} に等しいので、

$$E[x] = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

分散

$$E[x^2] = \sum_{x=0}^{\infty} x^2 f(x) = \sum_{x=0}^{\infty} x^2 e^{-\lambda} \frac{\lambda^x}{x!}$$

ここで、x=0 の項が 0 であることから、これを消去すると、

$$E[x^2] = \sum_{x=1}^{\infty} x^2 e^{-\lambda} \frac{\lambda^x}{x!} = \sum_{x=1}^{\infty} x e^{-\lambda} \frac{\lambda^x}{(x-1)!}$$

ここで、$z = x - 1$ とおくと、

$$\begin{aligned} E[x^2] &= \sum_{z=0}^{\infty}(z+1)e^{-\lambda}\frac{\lambda^{z+1}}{z!} = \sum_{z=0}^{\infty}(z+1)\lambda e^{-\lambda}\frac{\lambda^z}{z!} \\ &= \lambda \sum_{z=0}^{\infty} z e^{-\lambda}\frac{\lambda^z}{z!} + \lambda \sum_{z=0}^{\infty} e^{-\lambda}\frac{\lambda^z}{z!} \end{aligned}$$

ここで、右辺の最初の総和がポアソン分布の平均 $(=\lambda)$ に等しく、右辺の 2 番目の総和がポアソン分布の総和 (=1) に等しいことから、

$$\begin{aligned} E[x^2] &= \lambda^2 + \lambda \\ V[x] &= E[x^2] - E[x]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda \end{aligned}$$

2.3　ガウス分布

試行回数が十分に大きい場合、2 項分布またはポアソン分布はガウス分布に近似できる。ガウス分布は、次の式で表される。

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

図 9 に 2 項分布とポアソン分布のガウス分布による近似を示す。すべて平均は 20 としている。2 項分布の当たり率は $p = 0.5$ としている。2 項分布およびポアソン分布を近似するガウス分布の広がりを $\sigma = \sqrt{10}$ および $\sigma = \sqrt{20}$ とした。ガウス分布により 2 項分布およびポアソン分布が精度良く近似されていることがわかる。

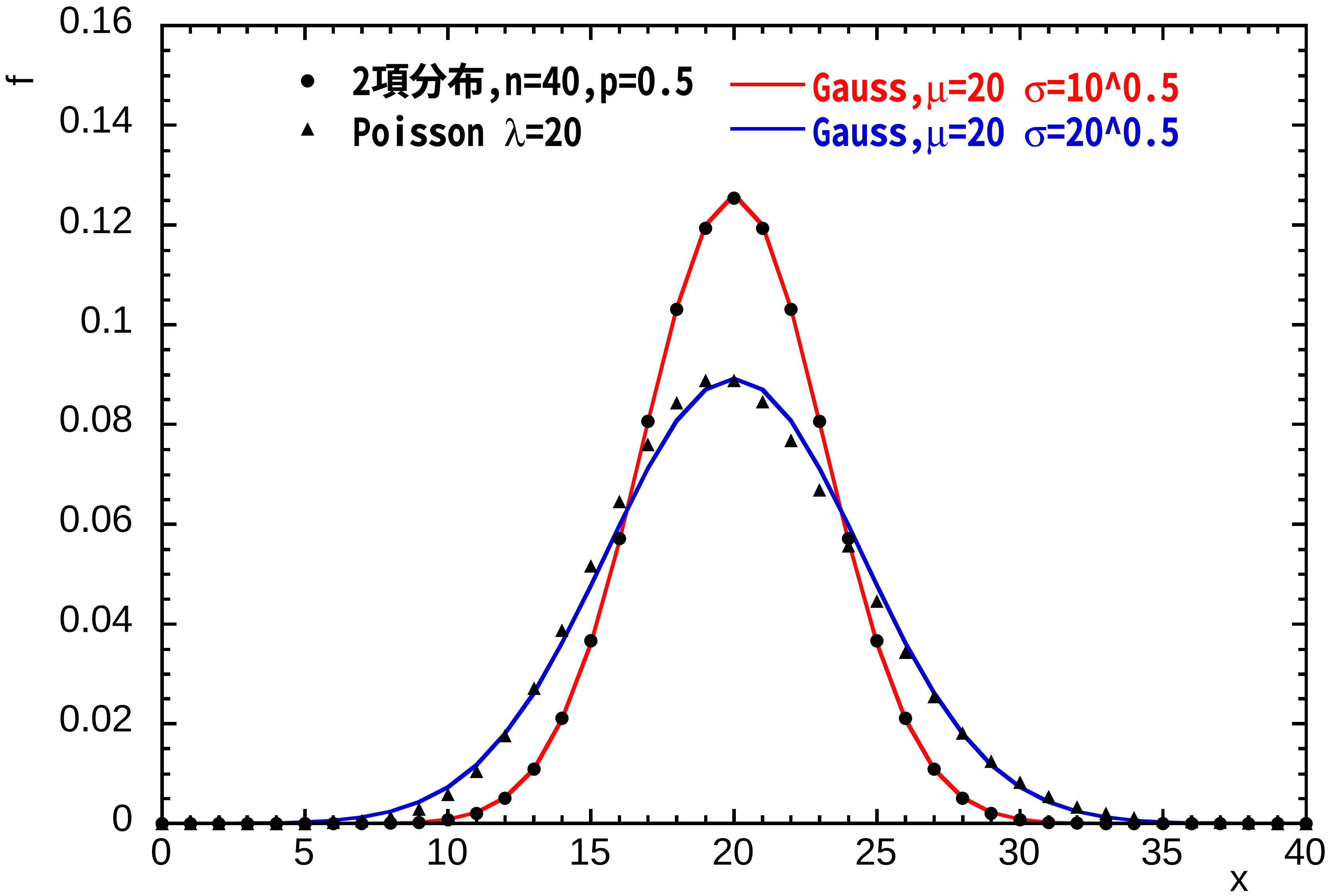


図9 2項分布とポアソン分布のガウス分布による近似

B.2 放射線計測の教科書での２項分布とポアソン分布の和、平均と分散の導出

「放射線計測の教科書で初歩的な統計学の概念を学ぶ」ことの有効性を知るために、放射線計測の数冊の教科書で２項分布とポアソン分布の和、平均と分散の導出が記述されているか調べてみた。もちろん、これは教科書の優劣を論じるための比較ではない。

ツルファニデス著・阪井英次訳「放射線計測の理論と演習」　２項分布の和、平均と分散およびポアソン分布の和、平均と分散の導出を記述している。

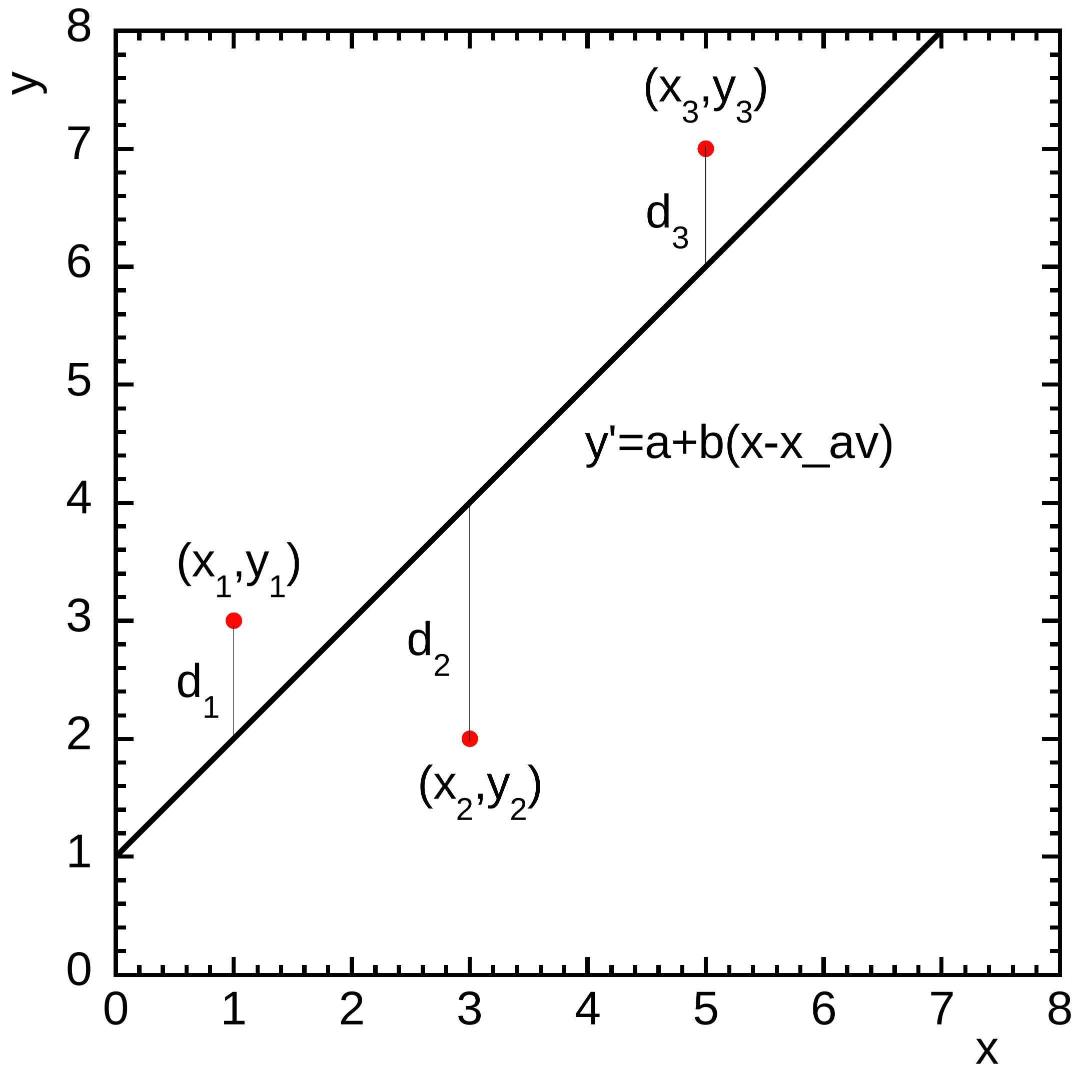
ノル著・木村逸郎、阪井英次訳「放射線計測ガイドブック」　ポアソン分布を簡略化した結果としてガウス分布を導入している。そして、ガウス分布の性質として「予想分散が $\overline{x}$」と述べている。しかし、これはポアソン分布を近似したガウス分布のみで成り立つことであり、誤解を招きやすい記述である。（一般のガウス分布では、分散と平均は独立である。）２項分布の和、平均と分散およびポアソン分布の総和、平均と分散の導出についてはふれていない。

プライス著・関口晃訳「放射線計測」　２項分布の分散の式を示しているものの、ポアソン分布の分散の式を示していない。２項分布の和、平均と分散およびポアソン分布の和、平均と分散の導出についてはふれていない。

２項分布とポアソン分布の和、平均および分散は、モンテカルロ計算や放射線計測で必要な統計学の理解に役立つ知識である。放射線計測の教科書では、これらの式自体は記述していることが多いものの、その導出の記述は少なく、放射線計測の教科書のみから、これらの分野に必要な統計学の知識を得ることは容易ではないことがわかる。一方、統計学の教科書は、大学１、２年生を対象にした平易なものから、専門的なものまでレベルが多岐にわたっている。モンテカルロ計算や放射線計測に関連する統計学の知識を体系的に得る必要がある場合には、放射線計測の教科書の統計の部分を読むとともに、基礎的な統計学の教科書を読むことが勧められる。

第3章　最小二乗法

最小2乗法の概念：$d_1^2+d_2^2+d_3^2$の最小値を求める。(図10)



3　最小二乗法の基本的な式の導出

図 10 に示すように (x,y) 平面上に 3 つの点 (x_1,y_1), (x_2,y_2), (x_3,y_3) があるとする。ここで 3 点の「なるべく近く」を通るように直線を $y' = a + b(x - \bar{x})$ 引く。「なるべく近く」についてはいろいろな考え方がある。ここでは、これらの点から、この直線に向けて y 軸に平行に線分を引き、線分の長さ d_1, d_2, d_3 の二乗和が最も小さくなるように係数 a と b を決める。$d_1 \sim d_3$ を式で書くと、

$$\begin{aligned}
d_1 &= y_1 - y'_1 = y_1 - \{a + b(x_1 - \bar{x})\} \\
d_2 &= y_2 - y'_2 = y_2 - \{a + b(x_2 - \bar{x})\} \\
d_3 &= y_3 - y'_3 = y_3 - \{a + b(x_3 - \bar{x})\}
\end{aligned}$$

この $d_1 \sim d_3$ の二乗和 S は、

$$\begin{aligned}
S &= d_1^2 + d_2^2 + d_3^2 \\
&= \{y_1 - a - b(x_1 - \overline{x})\}^2 + \{y_2 - a - b(x_2 - \overline{x})\}^2 + \{y_3 - a - b(x_3 - \overline{x})\}^2 \\
&= \{y_1^2 + y_2^3 + y_3^2\} + 3a^2 + b^2\{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + (x_3 - \overline{x})^2\} - 2a(y_1 + y_2 + y_3) \\
&\quad -2b\{y_1(x_1 - \overline{x}) + y_2(x_2 - \overline{x}) + y_3(x_3 - \overline{x})\} + 2ab\{(x_1 - \overline{x}) + (x_2 - \overline{x}) + (x_3 - \overline{x})\}
\end{aligned}$$

ここで $2ab$ の項は、それに係る { } 内が 0 であるので消える。S のその他の部分を書き直すと

$$S = A + 3a^2 + b^2B - 2aC - 2bD \tag{20}$$

ここで、

$$\begin{aligned} A &= y_1^2 + y_2^2 + y_3^2 \\ B &= (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 \\ C &= (y_1 + y_2 + y_3) \\ D &= y_1(x_1 - \bar{x}) + y_2(x_2 - \bar{x}) + y_3(x_3 - \bar{x}) \end{aligned}$$

(20) 式は、a の 2 次式と b の 2 次式の和である。S を a または b で微分して S を最小にする係数を求める。

$$\begin{aligned} \frac{\partial S}{\partial a} &= 6a - 2C = 0 \\ a &= \frac{C}{3} \qquad (21) \\ \frac{\partial S}{\partial b} &= 2Bb - 2D = 0 \\ b &= \frac{D}{B} \qquad (22) \end{aligned}$$

(21) 式および (22) 式は 3 点の場合である。これを一般化して、n 点について考えると、

$$a = \bar{y} \tag{23}$$

$$b = \frac{\sum_{i=1}^{n} y_i(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{24}$$

これらの式の導出には参考文献 [8] を参考にした。

4 演習問題

1. 次の手続きで 0 と 1 からなる乱数を生成し、その平均、分散、標準偏差および平均の誤差を求めよ。
   (a) 分布の作成
      i. モンテカルロ計算を行い、100 個の入射粒子に対して、応答=1 が 30 回、応答=0 が 70 回得られたとする。
      ii. 単純化のため、最初の 30 回は 1、その後の 70 回は 0 という分布を考え、これを x とする。
      iii. x をカレイダグラフまたはエクセルなどの表計算ソフトに入力せよ。具体的には、エクセルの A1 ～ A30 セルに 1、A31 ～ A100 セルに 0 を入力せよ。またそれをカレイダグラフにコピーせよ。
   (b) x の統計量を手計算で求めよ。
      i. x の分散と標準偏差をそれぞれ (8) 式と (9) 式により求めよ。
      ii. x の合計を求めよ。x の合計の誤差を (1) 式および (10) 式により求めよ。
      iii. x の平均を求めよ。x の平均の誤差を (12) 式により求めよ。
   (c) エクセルを利用して x の統計的性質を調べよ。A1 ～ A100 セルの var 関数、stdev 関数の値を読み取り、それらを手計算の値と比較して意味を考えよ。
   (d) カレイダグラフを利用して x の統計的性質を調べよ。平均、分散、標準偏差、平均の誤差をカレイダグラフの統計機能を用いて求めよ。

2. 次の手続きで 0 と 1 からなる乱数を生成し、その平均、分散、標準偏差および平均の誤差を求めよ。

   (a) 分布の作成

      i. 学籍番号の下二桁の数を 100 で割り、その数を当たり率 a の値とせよ。

      ii. $0 < r < 1$ を満たす乱数 r を 100 個生成し、$r < a$ の場合に 1、そうでない場合に 0 となる数 x を 100 個作成せよ。

      iii. x をカレイダグラフまたはエクセルなどの表計算ソフトに入力せよ。

   (b) x の統計量を手計算で求めよ。

      i. x の分散と標準偏差をそれぞれ (10) 式と (11) 式により求めよ。

      ii. x の合計を求めよ。x の合計の誤差を (1) 式および (12) 式により求めよ。

      iii. x の平均を求めよ。x の平均の誤差を (14) 式により求めよ。

   (c) エクセルを利用して x の統計的性質を調べよ。var 関数、stdev 関数の値を読み取り、その意味を考えよ。

   (d) カレイダグラフを利用して x の統計的性質を調べよ。平均、分散、標準偏差、平均の誤差をカレイダグラフの統計機能を用いて求めよ。

3. 分散の 3 種類の計算式のうち、(3) 式が n に依存し、(5) 式と (6) 式が n に依存しないことを調べよ。

   (a) 数値計算。エクセルなどのソフトを用い、乱数の n 個の組を 100 組程度生成せよ。$(1 < n < 101)$ それらについて (3) 式、(5) 式および (6) 式に基づいて分散を計算し、n 依存の有無を検討せよ。

   (b) (6) 式が n に依存しないことを証明せよ。ヒント：参考文献 [2] を参照。

4. 次の分布を計算せよ。エクセルなどの計算機プログラムを用いてもよい。

   (a) n=10, p=0.4 の場合の 2 項分布を計算し、図示して、図 7 と比較せよ。

   (b) 図 8 に示される、3 種類の 2 項分布とポアソン分布を計算し、図示せよ。ポアソン分布に最も近い 2 項分布はどれか？

   (c) 図 9 に示される 2 項分布、ポアソン分布、ガウス分布を計算し、図示せよ。ガウス分布は、2 項分布およびポアソン分布をよく近似しているか？

5. 次の手続きで、(x,y) 平面上に、点を 4 個置き、それらの近傍を通る直線を最小二乗法により求めよ。

   (a) 学籍番号を一桁の数に分解し、それらを小さい順に並べ、$(x_1, y_1) \sim (x_3, y_3)$ に当てはめよ。例えば学籍番号が 874913 ならば、$(x_1, y_1) = (1, 3)$, $(x_2, y_2) = (4, 7)$, $(x_3, y_3) = (8, 9)$.
   (b) (19) 式および (20) 式、または (21) 式および (22) 式を用いて、a および b を求めよ。
   (c) $(x_1, y_1) \sim (x_3, y_3)$ および $y' = a + b(x - \overline{x})$ をグラフ上に描き、最小二乗法により、点の近くを通る直線を引いたことを確認せよ。

6. 次の式を導出せよ。

   (a) 2項分布の和、平均および分散。分散については当たり率から考えても良い。
   (b) ポアソン分布の和、平均および分散。
   (c) 最小二乗法の係数の式。

References

[1] アイソトープ手帳 10 版 (2001) アイソトープ協会.

[2] 宮川公男、「基本統計学」第 3 版、有斐閣 (2007).

[3] P.G. ホーエル, 「初等統計学」 培風館 (1981).

[4] 久我隆弘、"物理学実験講座"測る" 第 3 回 精確さの指標", パリティ **25**, 54-60 (2010-06).

[5] ニコラス・ツリファァニディス、「放射線計測の理論と実習」2.8 節.

[6] ノル、「放射線計測ハンドブック」第 3 版、日刊工業新聞社 (1982), 第 3 章.

[7] プライス、放射線計測 コロナ社 (1979), 第 3 章.

[8] 最小 2 乗法に関する中部大学・鈴木肇氏の HP. http://szksrv.isc.chubu.ac.jp/lms/lms1.html

[9] 平山、波戸,EGS5 サンプルプログラム (ucnaicgv.f) NaI 検出器の応答計算 (Apr 16 2009).

A.2 コードの内容と具体的な数値

ここでは、上記の説明に出てくる式を理解し、疑問に答えるためコードの中身および具体的な数値を見ていく。ピーク効率の計算に着目して ucnaicgv.f の内部を見てみると、モンテカルロ計算を行うための Shower call loop 内で

```
if(depe.ge.ekein*0.999) then
  pefs=pefs+wtin
  pefs2=pefs2+wtin
end if
```

と計算し、ピーク効率に寄与する粒子の重みを積算している。ここで、ekein と depe はそれぞれ、入射エネルギーと、1 個の入射粒子に対する有感領域内での吸収エネルギーであり、これらを用いて最初の if 文で、注目している粒子が全エネルギー吸収ピークとして計数されるかどうかを判定している。pefs は、全エネルギー吸収ピークのカウント数である。

そして、モンテカルロ計算終了後に計算結果を整理する step 9 で

```
avpe=pefs/ncount
pefs2=pefs2/ncount
sigpe=dsqrt((pefs-avpe*avpe)/ncount)
```

と計算し、avpe と pefs2 をピーク効率とその誤差として出力している。

具体的な数値は、ncount=10000, wtin=1, pefs=3728, pefs2=3728 である。pefs2 は、ピーク効率に寄与する入射粒子の重みの二乗をスコアしており、このユーザーコードでは重みは常に 1 であるため、pefs と pefs2 の値は等しい。これらを用いて、avpe=0.3728, pefs2=0.3728, sigpe=4.8e-3 と値を求め、ピーク効率を $37.28 \pm 0.48\%$ と出力している。