

放射線の計算や測定における統計誤差¹

放射線計測やモンテカルロ法による放射線輸送では統計誤差の理解が重要である。本稿は統計誤差とそれに関連するトピックスについての3時間の講義のためのものである。最初の一時間ではまず、平均と分散について復習し、合計の分散および平均の分散を導出する。そして、モンテカルロ法による放射線輸送計算で通常算出される統計誤差が、平均の分散の定義通りのものであることを見る。そして、放射線計測で普段使用している統計誤差の式が平均の分散の定義式から導出できることを示す。次の2時間目では、二項分布、ポアソン分布、ガウス分布の平均と分散について復習する。最後の3時間目では最小二乗法の基礎的な式を学ぶ。

1 平均の誤差

1. 「平均の誤差」の意味と計算方法はどのようなものか？
2. 放射線計測などで、計数 N の統計誤差を \sqrt{N} で見積もることが多い。例えば、アイソトープ手帳 [1] の統計的取扱の節では、総計数を

$$N \pm \sqrt{N} \quad (1)$$

で求めている。これはどのように導出されるのだろうか？

3. データ数に依存しない分散とは何か？

本章では、「合計の誤差」と「平均の誤差」をキーワードとしてこれらの疑問を考えていく。まず、分散と標準偏差を1.1節で導入し、続いて合計の誤差と平均の誤差について1.2節と1.3節で述べる。そして、合計の誤差を基礎として「 $N \pm \sqrt{N}$ 」の式を1.3節で導出し、平均の誤差を利用して、データ数に依存しない分散を1.4節で導出する。モンテカルロ計算でもちいる統計誤差の式 [例えば、(28) 式] は、これらの疑問に答えたら、直ちに理解できる量である。

1.1 分散と標準偏差

ある量 x を n 回だけ繰り返して測定・計算することを考える。 i 番目の x の値を x_i とすると、 x の平均は、

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

1個1個の x のばらつきの程度を表すため、次の式で個々の x の値の平均からの差の二乗の平均を求め、これを分散と呼ぶ。

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (3)$$

また、分散の平方根を標準偏差 s と呼ぶ。 x と標準偏差が同じ次元であるため、 x のばらつきを表すために、標準偏差 s が多く用いられる。

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (4)$$

¹2020-10-15 版

tex source: texfile¥2012¥計算放射線物理学 ¥comp_rad_phys.tex

1.2 合計の誤差

n 個の x_i の合計 y を考える。

$$\begin{aligned} y &= \sum_{i=1}^n x_i \\ &= x_1 + x_2 + \cdots + x_N \\ &= n\bar{x} \end{aligned} \tag{5}$$

次に y の誤差を少し丁寧に見てみよう。誤差の伝搬を考えると、 y の分散は、 x_i の偏差 Δx_i を用いて次のように書くことができる。

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

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 の分散の平方根である。

$$\begin{aligned} s_y &= \sqrt{s_y^2} \\ &= \sqrt{ns^2} \\ &= \sqrt{n}s \end{aligned} \tag{6}$$

n 個の x_i の合計の誤差 s_y は、 x_i の誤差の \sqrt{n} 倍である。

1.3 平均の誤差

n 個の x_i の平均 \bar{x} を考える。

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

平均の誤差を、誤差伝搬の式から導く。

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

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

$$s_{\bar{x}} = \sqrt{s_{\bar{x}}^2} = \sqrt{\frac{1}{n} s^2} = \sqrt{\frac{1}{n}} s \tag{7}$$

n 個の x_i の平均の誤差 $s_{\bar{x}}$ は、 x_i の誤差の $1/\sqrt{n}$ である²。(7) 式は、測定を繰り返して平均を得た場合にその誤差を評価するための重要な式であり、統計学の多くの教科書で言及されている³。

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

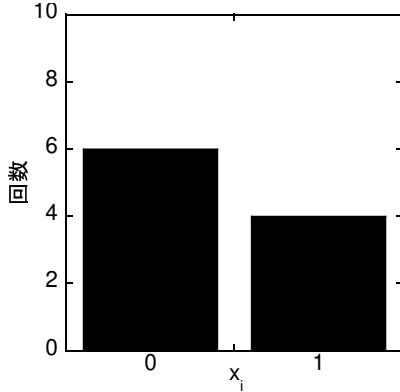


Figure 1: x_i の分布 時間帯の数 $n=10$ 、計数ありの回数=4 の場合。

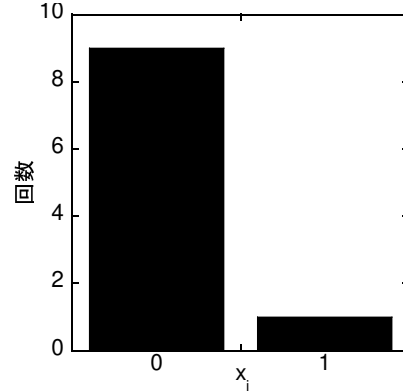


Figure 2: x_i の分布 時間帯の数 $n=10$ 、計数ありの回数=1 の場合。

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、

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

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

²ここまでで、EGS5 コードで計算している統計誤差の式と等価な式を導出することができた。EGS コードで求めている誤差は、平均値の誤差の定義に基づく値であることが分かった。モンテカルロコードの具体的な例では、EGS5 コードを取り上げた。しかし、本講義の内容は EGSnrc など他のコードの理解にも役立つはずである。ucnaicgv.f は EGS5 コードのサンプルユーザーコードの一つである。ucnaicgv の統計誤差の説明文および具体的な計算の内容をそれぞれ付録 A.1 と A.2 に示す。

³宮川 [2] は、「平均値 μ 、分散 σ^2 の母集団からとられた大きさ n の標本の平均値 \bar{x} の期待値は μ 、分散は σ^2/n である。」と述べている。ホーエル [3] は、「標本平均 \bar{x} は平均 μ 、標準偏差 $\sigma_{\bar{x}} = \sigma/\sqrt{n}$ の正規分布に従う。この標準偏差を平均 \bar{x} の標準誤差と呼ぶことがある。」と述べている。久我 [4] は、実験標準偏差 s と実験値 \bar{x} の標準不確かさ $\Delta\bar{x}$ の関係を、 $\Delta\bar{x} = s/\sqrt{n}$ と示している。

算することができる。

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

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、

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

合計とその誤差は、

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

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、

$$y \pm \sqrt{y} \quad (11)$$

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。

平均の誤差 (11) 式を n で割ると、平均とその誤差が得られる。

$$a \pm \sqrt{\frac{1}{n}a(1-a)} \quad (12)$$

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。

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

(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。

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

⁴ここまでのこの式の導出により、応答が 0 または 1 である場合の、平均の分散が $Na(1-a)$ であることが示された。これは、2 項分布の分散の導出の一つの方法である [2]。なお、2 項分布の分布の式からの分散の導出については後述する。2 項分布の分散は積率母関数 (Moment Generating Function) から求めることもできる。

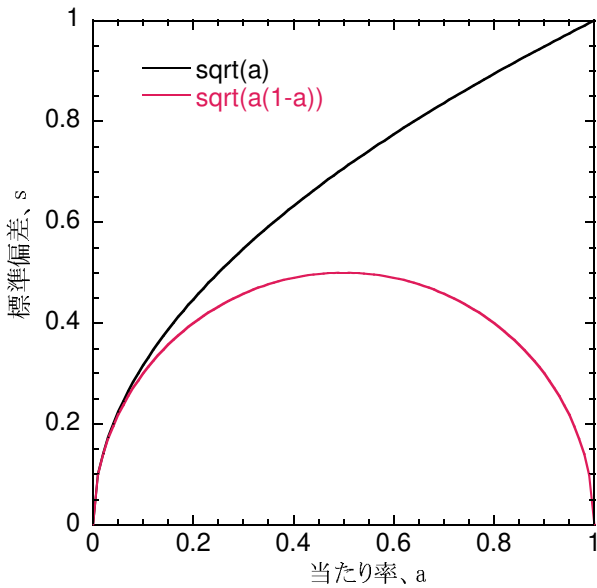


Figure 3: x の標準偏差 s

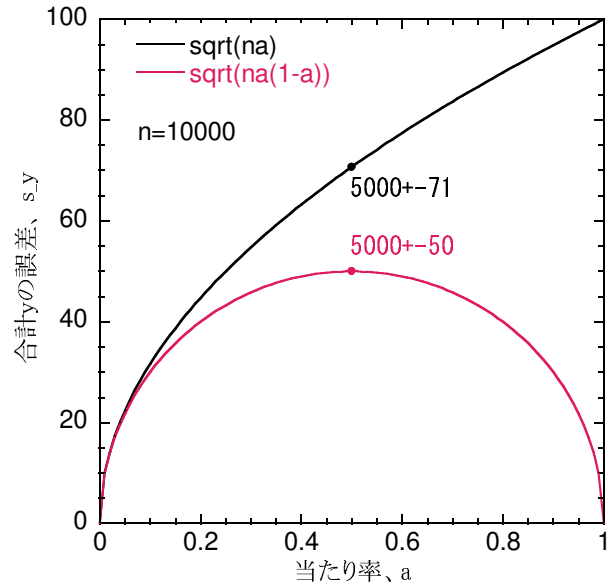


Figure 4: 合計 y の誤差 s_y

なお、教科書によっては、 $\hat{\sigma}$ を、標準分散 [2] あるいは、分散の不偏推定値 [3] と呼ぶことがあるものの、本プリントでは、単に分散と呼び、特に区別する必要がある場合には、記号で区別することとする。また、 n が十分に大きい場合には、 n と $n-1$ に有意差がなくなるため (3) 式と (4) 式を用いて分散と標準偏差を求める場合もある。

数値計算 エクセルの乱数を用いて、(3) 式、(13) 式および (14) 式の関連を調べてみる。まず、エクセルの RAND() 関数を用いて (0,1) の間に均一に分布する乱数を発生させ、10 個の乱数を 100 組入力する。続いて、10 個の乱数の平均を求め、(3) 式により s^2 を求める。 s^2 が 100 個得られるのでその平均を求める。また、平均の期待値 (この場合は 0.5) からのずれに基づく σ^2 を (13) 式で計算する。 σ^2 も 100 個得られるのでその平均を求める。この計算を $n=10$ から $n=1$ まで繰り返す。 σ^2 の期待値は、

$$\int_0^1 (x - 0.5)^2 dx = \int_{-0.5}^{0.5} y^2 dy = \left[\frac{y^3}{3} \right]_{-0.5}^{0.5} = \frac{1}{3} \left(\frac{1}{8} - \frac{-1}{8} \right) = \frac{1}{12} \quad (15)$$

図 5 に分散の値の比較を示す。(3) 式による s^2 は明らかに n の値に依存しており、 n が小さくなると s^2 も小さくなっている。これに対して、(13) 式による σ^2 は n によらず、期待値である $1/12$ にほぼ一致している。なお、 s^2 は $\frac{1}{12} \frac{n-1}{n}$ とほぼ一致している。 $\frac{n}{n-1} s^2$ [すなわち、(14) 式で計算される $\hat{\sigma}^2$] は n に依存しない。これらのことから、 s^2 は定義が直感的で簡便であるものの、 n が小さい場合にはその値が n に依存する実用上問題のある分散であることが分かる。また σ^2 および $\hat{\sigma}^2$ は、その値が n に依存しない良い分散であることがわかる。

(14) 式がデータ数に依存しないことの証明 (13) 式の右辺を次のように書き直してみる。

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

右辺第 3 項は、 $(x_i - \bar{x})$ に比例する量の i に関する総和なので 0 である。両辺の期待値は、

$$E \left[\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 \right] = E \left[\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right] + E \left[\frac{1}{n} \sum_i (\bar{x} - \mu)^2 \right] \quad (16)$$

左辺は (13) 式で表される平均の期待値 μ に対する分散 (期待値は σ^2)、右辺第 1 項は (3) 式で表されるデータの平均 \bar{x} からの分散、右辺第 2 項はデータの平均の分散 (期待値は σ^2/n) である。つまり、「 μ に対する分散 = \bar{x} に対する分散 + \bar{x} の分散」である。期待値を (16) 式に代入すると、(13) 式と (14) 式の期待値が一致することを示すことができる。

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

(証明終わり) 参考文献 [2] を参考にした。

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タ数 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、

$$f(x) = {}nC_x p^x (1-p)^{n-x} \quad (17)$$

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。

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 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、

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

$x=0$ 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、

$$E[x] = \sum_{z=0}^j np \frac{j!}{z!(j-z)!} p^z q^{j-z} = np \sum_{z=0}^j \frac{j!}{z!(j-z)!} p^z q^{j-z} \quad (18)$$

$$= np \quad (19)$$

ここで (18) 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。

分散 確率変数 x の分散 $V[x]$ は、

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

2項分布の分散を得るために、まず $E[x^2]$ を計算する。

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

上の式の最後の等号のところ、 $x=0$ 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

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

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、

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

一方、ポアソン分布の合計は、

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

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。

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 $y' = a + b(x - \bar{x})$ を引く。「なるべく近く」についてはいろいろな考え方があり、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、

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

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

$$S = A + 3a^2 + b^2B - 2aC - 2bD \quad (20)$$

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。

$$\begin{aligned} \frac{\partial S}{\partial a} &= 6a - 2C = 0 \\ a &= \frac{C}{3} \end{aligned} \quad (21)$$

$$\begin{aligned} \frac{\partial S}{\partial b} &= 2Bb - 2D = 0 \\ b &= \frac{D}{B} \end{aligned} \quad (22)$$

(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、

$$a = \bar{y} \quad (23)$$

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

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、10 個の入射粒子に対して、応答=1 が 3 回、応答=0 が 7 回得られたとする。
- ii. 単純化のため、最初の 3 回は 1、その後の 7 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~A3 セルに 1、A4~A10 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~A10 セルの var 関数、stdev 関数、stdevp 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。⁵

(d) エクセルのデータ分析の機能を利用して、この分布の統計量を求め、得られるパラメータの意味を考えよ。

⁵stdev 関数は、非バイアス法すなわち $(n-1)$ 法で標準偏差を求める。stdevp 関数はバイアス法すなわち n 法で標準偏差を求める。stdevp の p は母集団 (population) を示している。

- (e) 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、上の問題と同様のことを調べよ。
- (a) 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。単純化のため、最初の 3 回は 1、その後の 7 回は 0 という分布を考え、これを x とする
- (b) 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。 $0 < r < 1$ 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。
3. 分散の 3 種類の計算式のうち、(3) 式が n に依存し、(13) 式と (14) 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、(13) 式および (14) 式に基づいて分散を計算し、 n 依存の有無を検討せよ。
- (b) (14) 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 3 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) (21) 式および (22) 式、または (23) 式および (24) 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 $y' = a + b(x - \bar{x})$ 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).

付録 A モンテカルロ計算における統計誤差の算出方法

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、全効率、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:

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

- x_i の分散値を以下の式から求める。:

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

- \bar{x} の分散値は、

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

- 統計誤差として、

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

を用いる。

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。』

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
  pef2s=pef2s+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
pef2s=pef2s/ncount
sigpe=dsqrt((pef2s-avpe*avpe)/ncount)
```

と計算し、avpe と sigpe 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。

付録 B

B.1 平均値の誤差

ベルヌーイ分布の節で述べた平均の誤差は、平均のばらつきを表す量である。この量は重要な量であるにもかかわらず、次の理由から、言い表しにくい量である。ここでは、この量を「平均値の誤差」と呼ぶことにする。

- この量の名称が著者、時代や分野によって異なる。

Table 1: 「平均値の誤差」と同じ意味と推測される言葉の Google での検索結果

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

「平均値の誤差」と同じ意味と推測される言葉を Google で検索すると、表 1 に示すように日本語では「平均値の誤差」という呼び名が最も多いものの、「平均値の標準偏差」、「平均値の標準誤差」、「平均値の不確かさ」などの表現も多い。また、Wikipedia の「標準誤差」に関するページでは、「平均値の誤差」を「標準誤差」と呼んでいる。代表的なグラフ描画ソフトであるカレイダグラフでも、この量を「標準誤差」と呼んでいる。一方、総務省統計局の HP では、「標本誤差」と呼んでいる。初歩統計学の教科書の一つである [3] では、「平均 \bar{x} の標準誤差」と呼んでいる。⁶

- 放射線を取り扱う研究者は、通常、放射線計測の教科書で初歩的な統計学の概念を学ぶ。しかし、この量の記述がある教科書 [5] はまれであり、多くの教科書には、この量に関する記述がない [6, 7]。

こういったことは、二項分布、ポアソン分布、平均、分散、標準偏差など統計学の他の基本的な物事が放射線計測の教科書で記述され、それらの名称が長年変わっていないことと対照的である。

⁶このように表記が発散してしまった経緯を推測すると、元々「標本の平均 \bar{x} の標準誤差」と呼ばれていた量の名前から、「標本の」という部分と「平均値の」のどちらか片方または両方を省略することが慣習として行われていたのではないかと推測される。実際、エクセルのデータ分析の機能では、(n-1) 法で求めた不偏分散の平方根を「標準偏差」、それをデータ数で割った平均の不確かさを「標準誤差」として出力している。しかし、上記の呼称の表からも分かるように誤差と偏差は明確に区別されている量ではない。それらを用いて、個々の量のばらつきと、その平均のばらつきを「標準偏差」、「標準誤差」と呼び分けるのは万全とはいえない。

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 \bar{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に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。

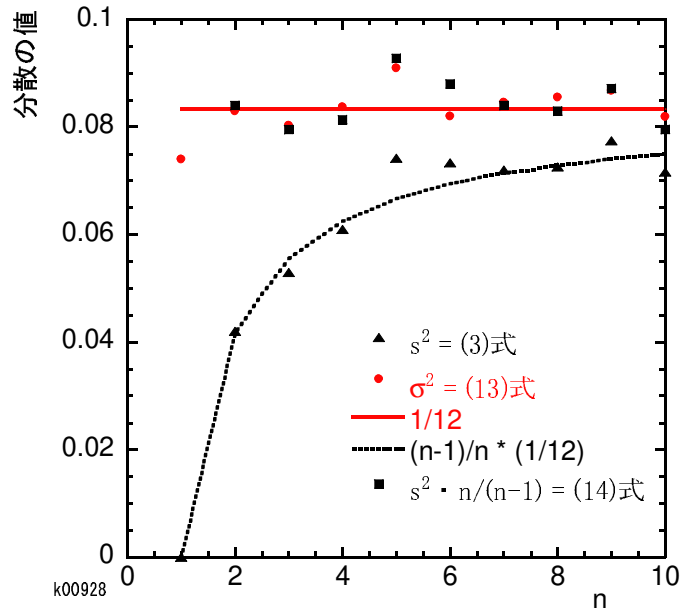


Figure 5: 分散の値の比較

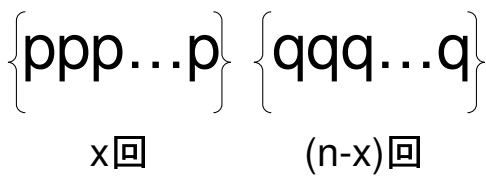


Figure 6: n個の入射粒子に対するモンテカルロ計算結果の概念図。光電ピークへの寄与のあり、なしをそれぞれpとqで示している。この図では、最初のx回続けて、光電ピークへの寄与があり、その後の(n-x)回続けて、光電ピークへの寄与がない場合を示している。

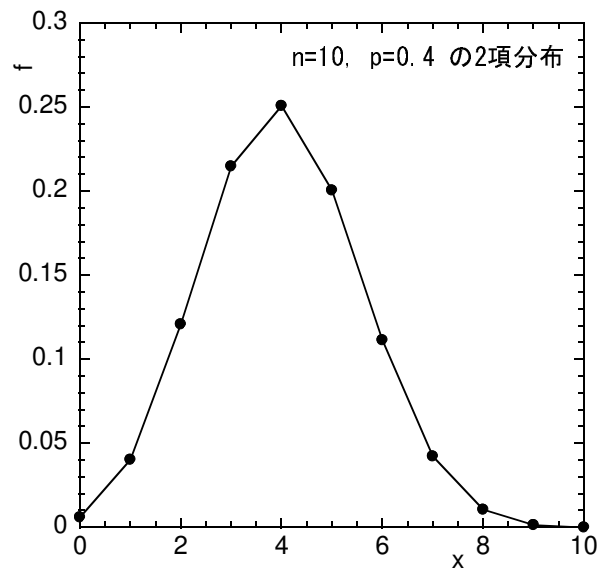


Figure 7: n=10, p=0.4 の2項分布

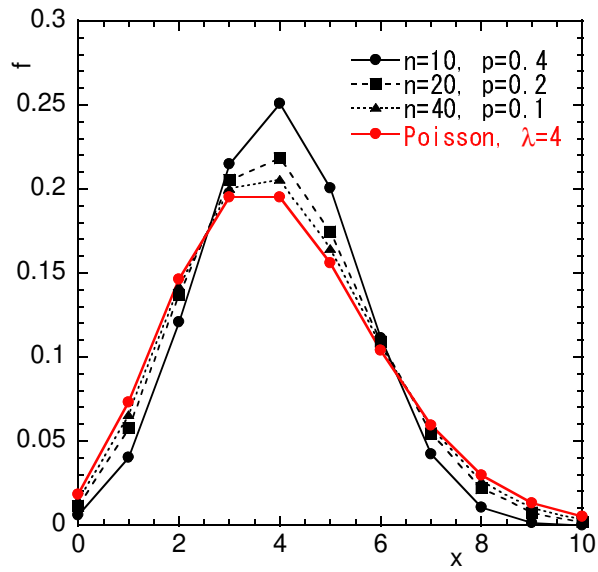


Figure 8: 2項分布とポアソン分布の比較

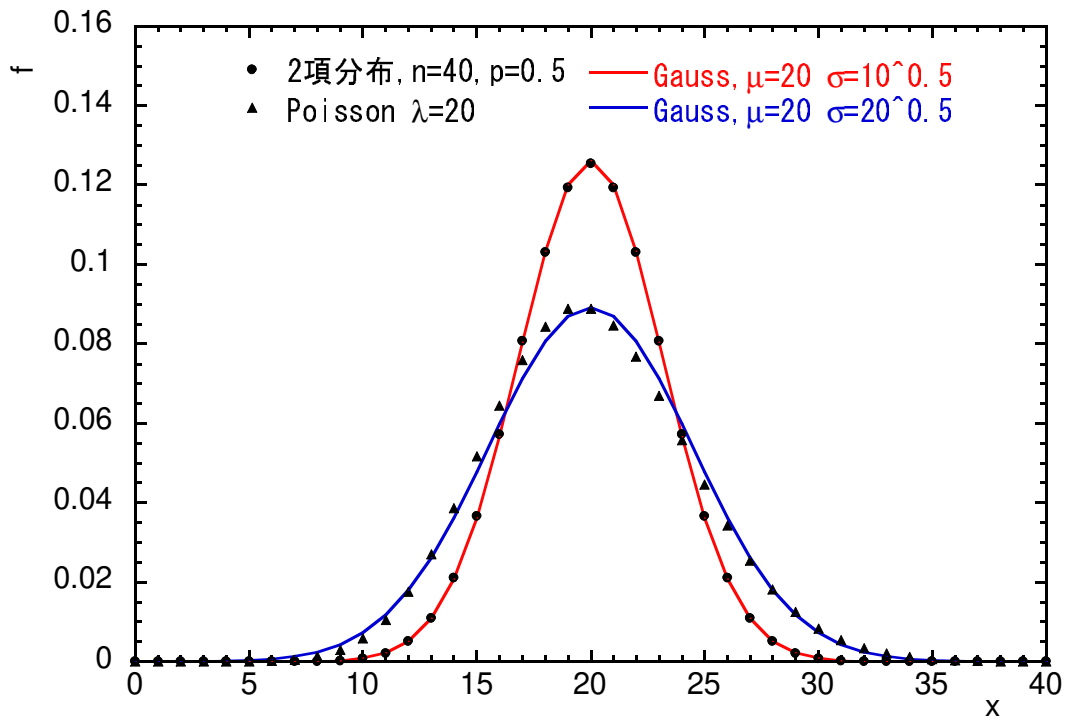


Figure 9: 2項分布とポアソン分布のガウス分布による近似

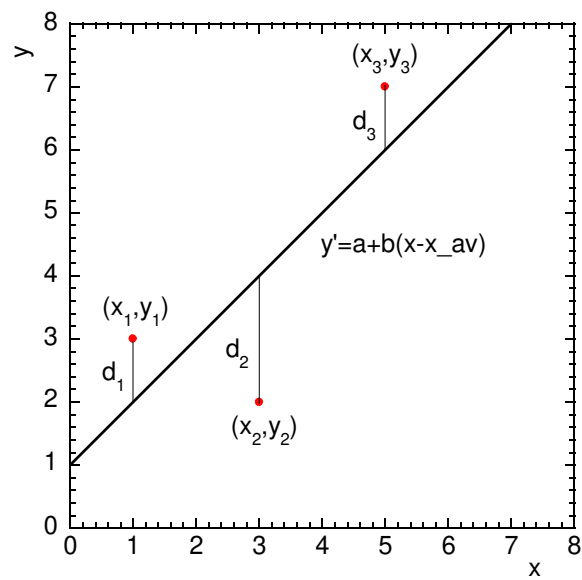


Figure 10: 最小二乗法の手順