20230512 の講義メモ

統計学の初歩の講義に対する不満
受講側
　面白くない（不愉快）
　　勝手に前提条件が持ち込まれる
　　例　正規分布という前提条件
　　　Ｔ分布、Ｆ分布とは何か
　　どんなモデルを考えているのか説明がない
　　例　何故、二項分布を習うのか
　　　正規分布は二項分布の極限だからというのは説明になっていない
　　ロジックが丁寧に説明されない
　　Ｔ 検定が２群の比較で、Ｆ 検定は３群以上の比較というのは、論理的説明ではない。

もっと、不愉快なのは、一部の学生は、どこで習ったのか常識としてそれを受け入れている。
　　教員はしょせん、話がわかるそいつらしか相手にしていない。
　　劣等感が刺激される（私の個人的な体験）
講義する側の問題
　○　誰が講義するか問題
　実用的に統計学的手法を使っている人が、解析学や線形代数に詳しいとは限らない
　解析学・線形代数に詳しい人が、統計学の実務に詳しいとは限らない。
　　結果２つのタイプの講義になる。
　　１．解析数学の基礎（Ｔaylor 展開、微積）の講義になってしまう
　　２．単なるやり方の講義になってしまう。
　○　今、必要か問題
　　コンピュータの計算力を使って、ランダムサンプリングを繰り返せば、正規分布を仮定しなくても、群間に差があるかを論ずることはできる。我々が習った統計学は古いのではないか

　　講義する側も、あまりやりたくない授業（自信がない。上手くやれる気がしない。）

本講義では、
　　何を考えているのか（モデルやロジック）に重点を置いて説明する。

統計学の論理

有意性：ある結果が得られたのは、偶然ではなくて必然だ。→何か原因・理由がある

おなじ群の属する物でも、個々のデータは少しずつ違っていて、ある広がりを持って分散している。その中から個体が選ばれるかは偶然。

データの広がり方は、確率密度曲線で表現できる。確率密度曲線を知っていれば、どのくらいの確率で、そのデータが選ばれるかがわかる。

たとえば

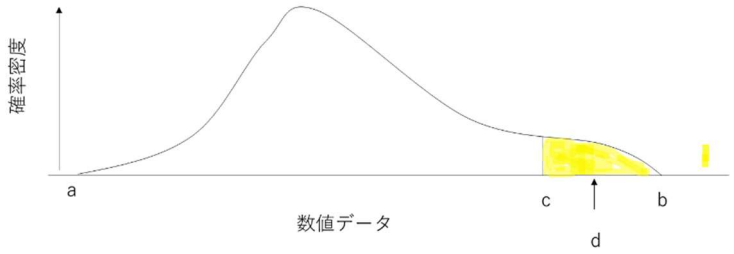


図１．確率密度分布の例

$P_{a\to b}=\int_a^b f_x dx=1$　　（$P_{-\infty\to\infty}=\int_{-\infty}^{\infty} f_x dx=1$）と書く方が普通）い

$P_{c\to b}=\int_c^b f_x dx=0.05$

ならば、ｄというデータが得られれる確率は 0.05 (5%)以下である。

めったにないこと（２０回やって１回しか起こらない偶然）だから、危険率５％以下で、有意だと判断できる。

確かにそうだが、あるべきデータの確率分布など知っているはずがない。知っていれば、統計学的判断など必要ない。

確率分布曲線を知っているとはどういうことか

位置、広がり、形

位置、広がりについては、もし、偶然の違いだけによって、サンプルされたデータの違いが出来ているのならば、こうなっているはずだという推測はできるものがある。

位置　おなじものならば位置は違わない。差が０

２群のあいだに違いがあるか→差の分布（ｔ分布）→ｔ検定

広がり　同じものならば広がりは同じ。

２群以上の間に違いがあるか→比の分布（Ｆ分布）→F 検定

つまり、群からサンプルした、サンプルの差あるいはサンプルされた群の、分散比の確率分布の図を描くことが出来れば、有意性の議論が出来る。

たとえば、もし、差の確率分布が図２のように矩形であったら

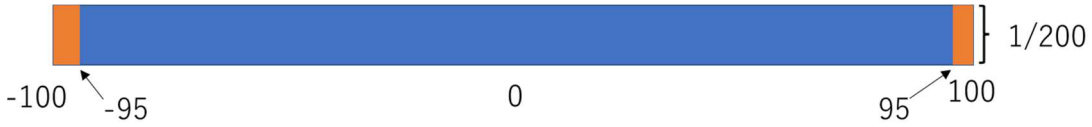


図２。もし差の確率分布の形が矩形だったら、
赤の部分の面積は$\left(-95-(-100)\right)\times\frac{1}{200}+(100-95)\times\frac{1}{200}=0.05$も

となりますから、二つの群れの平均値の差が、-100～-95 $(-100\leq x<95)$あるいは

95～$100(95<x\leq 100)$となる確率は５％以下で、二つの群には差があると結論する方が、

はるかに確率的に妥当だということになります。

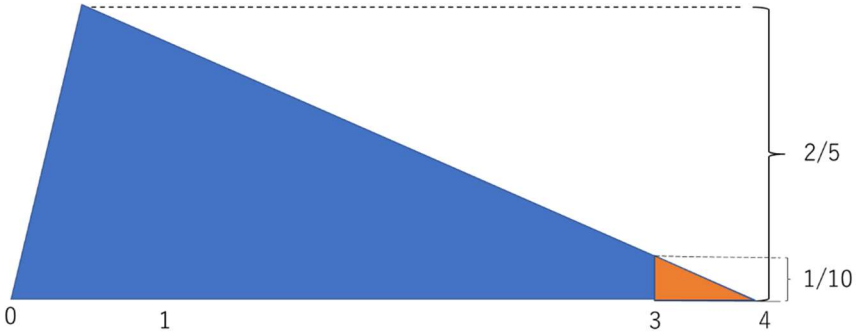


図３．もし、広がりの比の分布が３角形だったら
差の広がりが、$-\infty$から∞に広がっているのに対して、比の広がりは０から∞にひろがっていますが、ここでは、高さ 2/5 の三角形で、０から４の間に広がっているとします。ｋの分布であれば、3 から４の赤い直角三角形の面積は、$1\times\frac{1}{10}\times\frac{1}{2}=0.05$で、比が３倍以上になる確率は、$0.05(5\%)$で、めったに起こらない現象だとわかります。したがって、比較した２つの群のデータの広がりは違っていると５％の危険率で判断します。

差のところで取り上げた、全くフラットな分布をするデータとしては、例えば、順位がありますが、順位の差を論ずることはあまりないでしょう。比のところで取り上げた三角形の分布はかなり出鱈目です。こういう分布をするものを思いつきません。しかし、でたらめでも何でも、分布密度の図が出来れば面積計算に持っていけるということが、ここでの重要な情報です。

普通、確率密度曲線（ｘ軸を除く分布の輪郭線）は曲線です。この輪郭線を積分すれば面積計算が出来ます。

確率密度曲線を$P_{(x)}$とすると。

差の確率の場合　$\int_{-\infty}^{\infty}P_{(x)}\,dx=1$

（差データが存在する確率は無限小から無限大まであると考える。存在

確率の総和は１）

この条件で、曲線の係数をあたえて

$\int_x^\infty P_{(x)}\,dx$を計算すれば、 x 以上の値が得られる確率を計算できる。

ここまでは、数学の話。以下は分析者の視点の話（両側検定・片側検定の話）

もし、xが与えられる可能性がある確率水準から予測される分布範囲を外れているかと問う（違っているのかと問う）場合、例えば、５％以下の危険率の範囲に存在するかと問うのであれば、図２の説明のように、確率分布の両端の赤い部分の面積の両方の面積を合計したものが５％危険水準の範囲とした方が良い方が良い。どちらが A、B どちらが大きいかというのは、あらかじめわかっているわけではなくて、たまたま、分析者がどちらかが大きいと判断したのであって、そうなる確率は２分の１だから、両側検定にする。もし、A より B が大きいということがあらかじめわかる場合には、片側検定で良いので、片方の赤い領域の面積を確率水準とする。

比の確率の場合　$\int_0^\infty P_{(x)}\,dx = 1$

（比のデータの存在は０から無限大までである。存在確率の総和は１）

これを条件に$P_{(x)}$の係数を決定する。

$\int_x^\infty P_{(x)}\,dx$を計算すれば、比が$x$以上となる確率の値となる

ちょっと回り道

今、ここでは、深く取り上げませんが、統計的判断にはいつも危険率が伴います。統計的判断を続けていると、いつかは間違えるということです。一つの現象について、５％危険率の統計的判断を２０回繰り返すと、何処かで判断の誤りをしている可能性があります。そんな時は、危険率をもっと小さくした統計的判断をします（ボンフェローニ補正)。ただこの話は、今、語ろうとしている話のメインストーリーではないので、統計的判断とはそういうものだと理解して記憶にとどめてください。別のところでお話しします。

本題に戻る

今、重要な話は、差の分布、比の分布の形を表す式を作れば、その式を積分して、面積比を計算して、その現象が起こ確率を統計的に議論できるということです。

この先の見通し

目標

群間の差の分布の確率曲線を作る　ｔ分布曲線

群間のデータの広がり（分散:平均値まわりの２次の積率）の比の確率曲線を作る

Ｆ分布曲線

その先：検定：ｔ分布、Ｆ分布の使い方

ｔ検定：データからｔ値を作る

Ｆ検定：Ｆ比を作る。

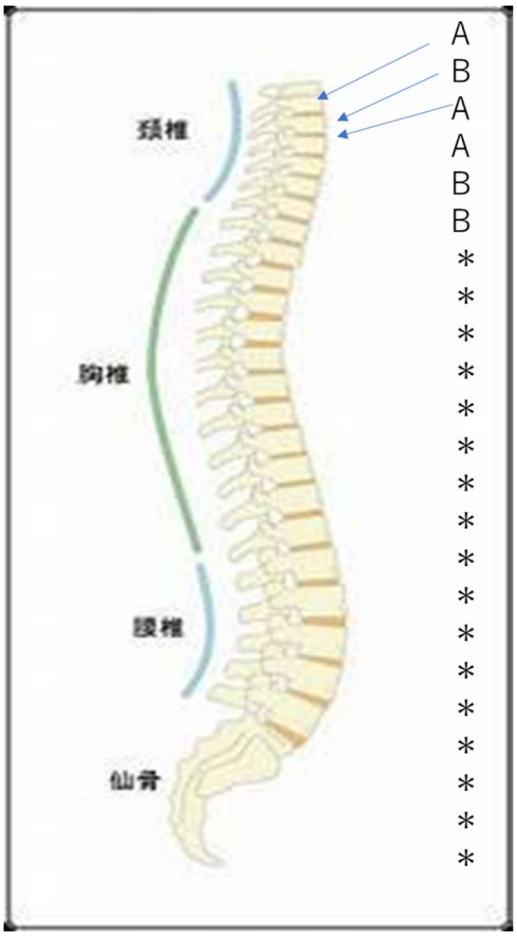
分散を各要素に分ける→二要因、多要因分散分析へ。

相関による分散を取り分ける→相関分析

ここまでで、基礎統計は終わり、次の段階に進む

線形代数を復習して、重相関、主成分分析、因子分析等々

数学的にに二項分布から正規分布を導出する前に、データの分布と確率について、理解する必要がありまが、データは様々な形で分布しているので、無条件でデータの分布関数を論ずることはできない。何か、有りそうなモデル（説得力のあるモデル）を考えて、分布関数をつくる。これが、二項分布から正規分布を作るという話です。
この話は、かなり身勝手な仮説から出発しますが、何故、二項分布から話が始まるのかを納得するには、これしかなさそうに思います。



どう考えているのかというと、データを構成する同じような要素がたくさんあるのだと考えているのです。個々の要素の働きは同じ効果を持つと考えます。実際には、個々の要素の働きは違うから、そんなことはありません。これはたとえ話です。個々の要素 A という状態がＢという状態か、それが相加的に重なり合って、全体の大きさが決まるのでと考えます。背骨には頸椎が７個、胸椎が 12 個、腰椎が５個、合わせて２４個の椎骨があります。これらが身長の構成要素だとします。それぞれの椎骨がＡ（例えば短い）であるかＢ（例えば長い）であるかによってその組み合わせは多様ですが、その組み合わせの違いによって身長が決まると考えます。

図４．二項分布のイメージ

このABの並び方を、頸椎の一番上から順番に横書きすると

AABAB BBAAA BABAB BAAAB ABBB

のようになります。この例では、A が１２個、B が１２個になっています。もし、１つの椎骨がAならば０、Bならば１点、score が入るとすると、この場合１得点のスコアがもらえて、その分だけ背が高くなります。すべてが A であった場合に比べて、12score 分だけ背が高いことになります。すべて、B ならば、24score 分だけ背が高くなります。

AAAAA AAAAA AABBB BBBBB BBBB

という並び方でも、B が１２個で score は 12 で、図示した例と同じだけの背の高さになります。このように、Score が１２になる組み合わせはほかにもたくさんあって、その数は全部で

$$\frac{24\times23\times22\times21\times20\times19\times18\times17\times16\times15\times14\times13}{12\times11\times10\times9\times8\times7\times6\times5\times4\times3\times2\times1}=2704156$$

式１

個あります。この計算を、２項係数の計算といいます。

もし、椎骨ひとつについて、B になる確率が 1/2 であれば、排反事象である A になる確率も 1/2 ですから、２４個の椎骨で、B が 12 個になる確率は、

$$2704156\times\left(\frac{1}{2}\right)^{12}\times\left(\frac{1}{2}\right)^{12}=0.16118$$

式２

となります。これが二項分布の確率密度の計算です。

式１の二項係数は、多分、高校の組み合わせ数学で学習した。n 個の中から k 個のものを取り出す場合の数の計算、コンビネーション ${}_nC_k$ の計算です。二項係数として表現する場合は$\binom{n}{k}$と書く方が普通かもしれません。式２の$\left(\frac{1}{2}\right)^{12}\times\left(\frac{1}{2}\right)^{12}$のところは、確率 1/2 の事象が同時に１２回、その反対事象が１２回が同時に１２回起こることの確率で、その2704156回分の総和という意味です。

知っていると思いますが、念のため、二項分布について、具体例を使って詳しく説明します。よく使われるのは、コインを数回投げた時の、表が出る確率の計算です。

コインを一つ、一回だけ投げた場合には

表　または　裏

この２つの事象があります。

２回投げた場合は

表|表　または　表|裏　または　裏|表　または　裏|裏

となります。おこる事象の総数としては４つですが、その内、表が 2 回現れる組み合わせ１つ、表と裏が１回ずつ現れる組み合わせは２つ、裏が２回現れる組み合わせは１つです。

表が２回出て、裏が０回出る確率は

$$1\times\left(\frac{1}{2}\right)^2\times\left(\frac{1}{2}\right)^0=\frac{1}{4}$$

表が１回出て、裏が１回出る確率は

$$2\times\left(\frac{1}{2}\right)^1\times\left(\frac{1}{2}\right)^1=\frac{1}{2}$$

表が０回出て、裏が２回出る確率は

$$1\times\left(\frac{1}{2}\right)^0\times\left(\frac{1}{2}\right)^2=\frac{1}{4}$$

となります。

コインだとこうなりますが、サイコロで、１が出る事象と、その排反事象として１以外の数が出る事象について考えます。サイコロはイカサマではないので１が出る確率を 1/6、１以外が出る確率を 5/6 とします。

１が２回出て、１以外が０回出る確率は

$$1\times\left(\frac{1}{6}\right)^2\times\left(\frac{5}{6}\right)^0=\frac{1}{36}$$

１が１回出て、１以外が１回出る確率は

$$2\times\left(\frac{1}{6}\right)^1\times\left(\frac{5}{6}\right)^1=\frac{10}{36}$$

１が０回出て、１以外が２回出る確率は

$$1\times\left(\frac{1}{6}\right)^0\times\left(\frac{5}{6}\right)^2=\frac{25}{36}$$

となります。念のため、全事象の確率の総和が１となることを確認してください。

コインとサイコロでは一回の試行で個々の事象が起こる確率がちがっているので、確率の計算結果は違いますが、２項係数のところは同じで、上から

$_2C_2$、$_2C_1$、$_2C_0$ コンビネーションとしての記法

$\binom{2}{2}$、$\binom{2}{1}$、$\binom{2}{0}$ 二項係数としての記法

コインやサイコロを何回投げるかによって、この値がどのように変わるかを考えます

１回　　$\binom{1}{1}\ \binom{1}{0}$

２回　　$\binom{2}{2}\ \binom{2}{1}\ \binom{2}{0}$

3 回　　$\binom{3}{3}\binom{3}{2}\binom{3}{1}\binom{3}{0}$

4 回　　$\binom{4}{4}\binom{4}{3}\binom{4}{2}\binom{4}{1}\binom{4}{0}$

5 回　　$\binom{5}{5}\binom{5}{4}\binom{5}{3}\binom{5}{2}\binom{5}{1}\binom{5}{0}$

起こり得る事象はこのようになっています。
このそれぞれの事象が起こる場合の数を計算すれば良いのですが、まず、普通にやってみます。１回の背反事象を A,B とします。
１回投げる場合は

A が起こるという事象の数は N(A)=1,
B が起こるという事象の数は N(B)=1,

$$\binom{1}{1}=1\,,\binom{1}{0}=1$$

２回投げると
N(A|A)=1,　N(B|A)=1,　N(A|B)=1,　N(B|B)=1,
N(B|A)は A が起きたという条件のもとに B という事象が起こること

$$\binom{2}{2}=N(A|A)1\,,\binom{2}{1}=\ \mathrm{N(A|B)+N(B|A)}=2,\binom{2}{2}=\ \mathrm{N(B|B)}=1$$

３回投げると
N(A|A|A)=1,N(B|A|A)=1,N(A|B|A)=1,N(B|A|A)=1,N(B|B|A)=1,N(B|A|B)=1,
N(A|B|B)=1, N(B|B|B)=1

$$\binom{3}{3}=\mathrm{N(A|A|A)}=1\,,\binom{3}{2}=\mathrm{N(B|A|A)+N(A|B|A)+N(B|A|A)}=3$$

$$\binom{3}{1}=\mathrm{N(B|B|A)+N(B|A|B)+N(A|B|B)}=3\,,\binom{3}{1}=\mathrm{N(B|B|B)}=1$$

４回投げると
N(A|A|A|A)=1, N(B|A|A|A)=1,N(A|B|A|A)=1,N(A|A|B|A)=1,N(A|A|A|B)=1,
N(B|B|A|A)=1,N(A|B|B|A)=1,N(B|A|B|A)=1,N(A||B|A|B)=1,N(B|A|A|B)=1,
N(A|A|B|B)=1,　N(A|B|B|B)=1, N(B|A|B|B)=1, N(B|B|A|B)=1, N(B||B|B|A)=1,
N(B|B|B|B)=1

$$\binom{4}{4}=\mathrm{N(A|A|A|A)}=1$$

$$\binom{4}{3}=\mathrm{N(B|A|A|A)+N(A|B|A|A)+N(A|A|B|A)+N(A|A|A|B)}=4$$

$$\binom{4}{2} = \mathrm{N(B|B|A|A)} + \mathrm{N(A|B|B|A)} + \mathrm{N(B|A|B|A)} + \mathrm{N\left(A\middle||B|A\middle|B\right)} + \mathrm{N(B|A|A|B)} +$$

$$\mathrm{N(A|A|B|B)} = 6$$

$$\binom{4}{1} = \mathrm{N(A|B|B|B)} + \mathrm{N(B|A|B|B)} + \mathrm{N(B|B|A|B)} + \mathrm{N(B||B|B|A)} = 4$$

$$\binom{4}{0} = \mathrm{N(B|B|B|B)} = 1$$

この計算結果は

1 回	1 1
2 回	1 2 1
3 回	1 3 3 1
4 回	1 4 6 4 1
5 回	1 5 10 10 5 1

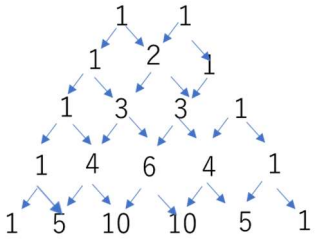


図５．パスカルの三角形

という形になっていて、試行回数が１回増えるごとに、矢印が向かってくる方向の数字をたせば、次の二項係数が計算できます。これにはパスカルの三角形という名前がついています。パスカルの三角形の計算を一般化して書くと、次のようになります。

$$\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1} \quad \text{ただし、} \binom{n}{-1} = 0$$

この計算は、結構、なじみのある計算で

$$p + q$$

という、２つの項からなる式のべき乗の計算で使っています。

$(p+q)^k$ *の結果が次のように表せるときに*

$$(p+q)^k = p^k + Ap^{k-1}q + Bp^{k-2}q^2 + \cdots\cdots + Z\,p^1 q^{k-1} + q^k$$

$$(p+q)^{k+1} = (p+q)^k (p+q)$$

$$= \left(p^k + Ap^{k-1}q + Bp^{k-2}q^2 + \cdots\cdots + Zp^1q^{k-1} + q^k\right)p$$

$$+ \left(p^k + Ap^{k-1}q + Bp^{k-2}q^2 + \cdots\cdots + Zp^1q^{k-1} + q^k\right)q$$

$$= \left(p^{k+1} + Ap^k q + Bp^{k-1}q^2 + \cdots\cdots + Z\,p^2 q^{k-1} + p^1 q^k\right)$$

$$+ (p^k\,q + Ap^{k-1}\,q^2 + Bp^{k-2}\,q^3 + \cdots\cdots + Zp^1\,q^k + q^{k+1})$$

$$= p^{k+1} + (A+1)p^k q + (A+B)p^{k-1}q^2 + \cdots + (Y+Z)\,p^2 q^{k-1} + (1+Z)p^1 q^k + q^{k+1}$$

となって、確かにパスカルの三角形の計算になっています。

２項係数という名前はこのことから来ています。この説明で２項係数という名前は覚えられますが、実際の計算のためには、一般化した公式を覚えた方が良いでしょう。

一般的な公式を導きます。

例えば、サイコロを５つ同時に投げるとします。一つ一つのサイコロには a,b,c,d,e と名前を付けておきます。このサイコロ一つの目が１となる事象を A　１以外のとなる排反事象を B とします。

サイコロの目を確認して、サイコロを回収します。サイコロを回収する順番はランダムに決まるものとしますが、その順番も記録します

たとえば、以下のような表が作れます

順番	1	2	3	4	5
サイコロ	b	c	e	a	d
事象	A	B	B	A	B

サイコロの順番がどのように並ぶかというのは、順列です。順列の数は、順番が n までとすれば、$n!$です。この場合は$5! = 120$

ところで、A の事象となっているのは、１番目に回収された b と４番目に回収された a で A の事象となる数が k 個だとすると、A の事象となったサイコロ（この場合は b と a）の中で、どのような順番で並ぶかもランダムで、その順列の数は$k!$ この場合は$2! = 2$　です。

B 事象となったのはこの場合は n- k 個、２番目に回収された c 、３番目に回収された e, 最後に回収された d の３つです。その中で、どのように並ぶかという順列の数は$(n-k)!$でこの場合は、3!=6 です。

その事象が何回起こるかは、２項係数で$\binom{n}{k}$個ですが、A、B の並び方はランダムです。A、B それぞれの並び方の中で、何番目に回収されるかについてもランダムに考えて、順列の数を数えたから、それを合計すれば、全体の順列の数に一致するはずです。つまり、以下の式が成立ちます。

$$k! \times (n-k)! \times \binom{n}{k} = n!$$

ここから、

$$\binom{n}{k} = \frac{n!}{k! \times (n-k)!}$$

という、公式が得られますが、

$$\frac{n!}{(n-k)!} = n \times (n-1) \times \cdots (n-k+1)$$

ですから、

$$\binom{n}{k} = \frac{n \times (n-1) \times \cdots (n-k+1)}{k!}$$

と記憶している人が多いと思います。
丁寧にやったので長くなりましたが
　　二項分布の確率密度関数は

$$P_{(k)} = \binom{n}{k} p^k q^{n-k} \quad \text{ただし} \quad p+q=1$$

となります。

二項分布は事象が何回起こるかという、個数（不連続数）の確率分布ですから、身長とか体重とか連続数の分布に拡張するとのが、二項分布から正規分布を作るという作業なのですが、従来、の講義や教科書はここのところの説明がとても下手で、乱暴です。中には、二項分布のｎの数を無限大に拡張したのが正規分布だと一言で済ませている例もあります。確かにｎを無現大に拡張していますが、それが正規分布の本質ではありません。試しに、二項分布のｎを大きくしていくという作業をしてみます。

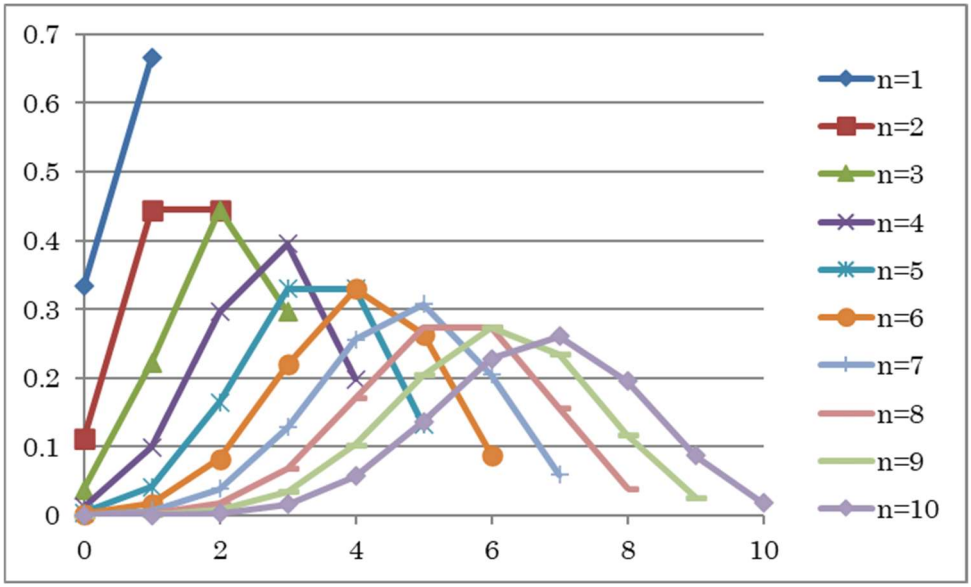


図６．二項分布のｎを大きくしていくとどうなるか。
縦軸が$W(k) =_n C_k p^k q^{(1-k)}$で、横軸がｋです。ｎが１から 10 まで変わるときに、そのｎごとに、ｋ の値に対して、$W(k) =_n C_k p^k q^{(1-k)}$の値を示したものです。この図は p=2/3 の時の図です。p+q=1 ですから、q は 1/3 です。コインのたとえでは、A（表）である確率が p が 2/3、B（裏）である確率 q が 1/3 という意味です。つまりイカサマのコインです。
この場合、ｎ回コインを投げて、A がｋ回、B が ｎ－ｋ 回出る確率が、W(k)です。ｎが大

きくなれば、当然、その確率の分布を表すW(k)の曲線は、右に広がっていきます。それにしたがって、頂上も右に移動していきます。分布の広がりも大きくなっていきます。形は左右対称のきれいな単峰形に近づいていきます。ただこれを無限回繰り返したら。山形もなくなって、ただのペッタンコの平地になるだけです。そんなことをしてくれと、誰も頼んでいません。やってもらいたいのは、位置と広がりを変えずに、形がどう変わるのか考得るということです。正規分布は二項分布の極限だという説明は、間違っているとは言いませんが、不十分な説明です。その説明で分るハズがありません。

二項分布から正規分布を作ってみます。
結構難しいので、まず、例によって、簡単に計算できる例について、考えてみますが、まず、その前に、図 6 の n=9 の曲線に注目してください。頂上は k=6 のところにあります、n=6 の時は、k=4、n=3 の時は k=2 のところに、頂上があります。k/n は 2/3 でｐの値になっています。n が他の値の場合、必ずしも頂上の位置がはっきりしませんが、だいたい k/n=p となるｋのところに頂上がありそうです。つまり、頂上となる k は k=np のように見えます。k は A となる回数のことですね。１回の試行でｐ＝が出る確率が 2/3 なのだから、２回投げればＡとなる確率が 2/3+2/3 のところで最大となるのは当然のことでしょう。「1 回の試行で A となる確率を知っている時、n 回試行を繰り返したら、何回Ａという事象が起こるか、もっとも確率が高い回数は何回か。」という問いに対する答えは np 回です。このような問いに対する答えを期待値と言います。頂上となるｋの値の期待値は、k=np です。山形の位置を決める時に、何か代表値を考えて、その代表値の位置を固定するという考え方が自然でしょう。それならば、頂上を代表値とするというのはありそうなことなので、頂点をデータのバラツキの位置を決める候補として使えそうだということはわかります。問題は、データの広がり方、バラツキをどのようにしてそろえるのかです。一試行について、背反事象 A、B が起きた時にそれぞれに与えられる得点を調整すれば良いのですが、具体的にイメージを描いた方がわかりやすいと思います。背の高さのつがいを作りだす要素として、背骨の椎骨の例を挙げました。人間の椎骨の数は 24 ですが、例えば、椎骨が１しかない想像上の動物を考えます。そのような動物では、椎骨が短い（A）=背が低い個体のグループ、椎骨が短い（B）=背が高い個体のグループ２つのグループに分かれます。データにはこの２種類しかありえません。椎骨が２つの生き物の場合、椎骨について、AA＝背が低い個体、AB＝背が普通の個体、BB＝背が高い個体、という３つのグループが出来ます。椎骨が３つの場合は、ＡＡＡ＝背が低い個体。 ＡＡＢ＝背がやや低い個体, ＡＢＢ＝差がやや高い個体, ＢBB＝背が高い個体。という４つのグループが出来ます。この時、A に対して０、B に対して１という score を与えると、椎骨が１の場合は、score:0 のグループと、score1 のグループが出来て、それぞれの確率が 1/2 ならば。これらのグループを無限個集めた母集団の最高値が１最低値が０　平均値が 0.5, 最高値が１です。椎骨が２つの場合は、最低値が０、最高値が２、平均値が１、椎骨が３つの場合は最低値が

０、最高値が３、平均値が、１．５となります。つまり、分布が右に移動し、広がりが大きくなります。そうならないためには、平均値が０になるようにして、与える score を椎骨の数で割ればよいということになります。そこで、椎骨が１の時には、A ならば－１、B ならば１という sore にして、椎骨が２の時は A に対して－1/2、B に対して 1/2 の score を与えると言うように、椎骨が１の時に与える score を椎骨の数で割った数を score として与えることにします。

こうすると

椎骨が１の時

最低値（A）$=-1$　最高値（B）$=1$　平均値$=-1\times\frac{1}{2}+1\times\frac{1}{2}=0$

椎骨が２の時 (A,B の score は$-\frac{1}{2},\frac{1}{2}$　)

最低値:$S_{AA}=\left(-\frac{1}{2}\right)+\left(-\frac{1}{2}\right)=-1$　最高値:$S_{BB}=\left(\frac{1}{2}\right)+\left(\frac{1}{2}\right)=1$

平均値: $\frac{1}{4}S_{AA}+\frac{2}{4}S_{AB}+\frac{1}{4}S_{BB}=\frac{1}{4}(-1)+\frac{2}{4}(0)+\frac{1}{4}(1)=0$

椎骨が 3 の時(A,B の score は$-\frac{1}{3},\frac{1}{3}$　)

最低値:$S_{AAA}=\left(-\frac{1}{3}\right)+\left(-\frac{1}{3}\right)+\left(-\frac{1}{3}\right)=-1$　最高値:$S_{BBB}=\left(\frac{1}{3}\right)+\left(\frac{1}{3}\right)+\left(\frac{1}{3}\right)=1$

平均値: $\frac{1}{8}S_{AAA}+\frac{3}{8}S_{AAB}+\frac{3}{8}S_{ABB}+\frac{1}{8}S_{BBB}=\frac{1}{8}(-1)+\frac{3}{8}\left(-\frac{1}{3}\right)+\frac{3}{8}\left(\frac{1}{3}\right)+\frac{1}{8}(1)=0$

椎骨が 4 の時(A,B の score は$-\frac{1}{4},\frac{1}{4}$　)

最低値:$S_{AAAA}=\left(-\frac{1}{4}\right)+\left(-\frac{1}{4}\right)+\left(-\frac{1}{4}\right)+\left(-\frac{1}{4}\right)=-1$　最高値:$S_{BBBB}=\left(\frac{1}{4}\right)+\left(\frac{1}{4}\right)+\left(\frac{1}{4}\right)+\left(\frac{1}{4}\right)=1$

平均値: $\frac{1}{16}S_{AAAA}+\frac{4}{16}S_{AAAB}+\frac{6}{16}S_{AABB}+\frac{4}{16}S_{ABBB}+\frac{1}{16}S_{BBB}=\frac{1}{16}(-1)+\frac{4}{16}\left(-\frac{2}{4}\right)+\frac{6}{16}(0)+\frac{4}{16}\left(\frac{2}{4}\right)+\frac{1}{16}(-1)=0$

となり、グラフが、右に移動しなくなって、最低値、最高値を見る限り、椎骨の数によって、データの分布範囲の広がりも変わらなくなります。そう考えると、この椎骨に与えるスコアの絶対値をデータ広がりの基準とすればよいのではないかという気がしてきます。たとえば、左右対称で、絶対値が同じ正負の score がるから、正負の Score の二乗和の半分をデータの広がりを表す値とするという考え方です。もし、母集団の平均値がわかっていれば、その平均値と個々のデータとの差の二乗の平均値を広がりの基準にすればよいということです。それを分散と言い、その平方根を標準偏差と言います。これは、それなりに理屈の通った考え方ですが、何でデータをとるのかというと、対象となる集団の全体像が分からないからです。世の中に、データがとられた sample 集団しか存在しないのなら、その sample 集団こそ母集団ですから、sample 集団の平均値が母集団の平均値で、その平均値と個々のデータとの差の平方の平均が分散です。しかし、母集団のすべてを調べられないから、ある程度の数を抽出してさ sampling しているので、正しい母集団の平均値など知るわけがありません。我々が、問われているのは、標本データから、母集団の平均値や分散を知る方法です。ここで、問題になる分散は、広がり方を共通に表す値ですから、椎骨の数（sample するデータの数:data size）によって、数値が表すが変わっては困りま

す。例によって、具体例を作りながら、考えていきます。
作業を始める前に、必要な言葉を定義して、二項分布が持っている性質について解説します。二項分布が持っている性質の多くはそれを拡張した正規分布に引き継がれています。

確率分布の特性を表す値**(必要な言葉の定義)**

期待値：ある確率事象が起きたときに得られる値とその事象が起きる確率の積の、起きうる全事象についての総和

$$E(f(x))=\sum_{i=1}^{n} f(x_i)P_i \qquad \text{式 7}$$

平たく言えば、あることを無限回繰り返して、そのたびにある値を計算する。その値の無限回分の平均値のことです。

平均値

定義１．全データーの総和/データーの総数

定義２．$E(x)=\sum_{i=1}^{n} x_i P_i$　（期待値の表現を用いた新たな定義）

SS: Sum of square（平均値からの距離の２乗の和)、

２次の積率（２次のモーメント）　$E\left((x-\mu)^2\right)$

xはデーター、μは母集団の平均　標本手段の平均はｍ
nをデーターの総数とすれば標本集団については、２次の積率を SS/n として計算できます。２次の積率は、その標本集団のデーターが平均値からどのくらい隔たった値のデーターで構成されているかを平均的に表したもの。Ｅ（ｘ－ｍ）の平均でも良さそうですが、ｍは平均値だから、Ｅ（ｘ－ｍ）は 0 になってしまう。そこで、データーのばらつきに指標にはＥ（（ｘ－ｍ）^{2}）を使います。

感覚的な理解

二項分布するモデルB(n,p)について、そのモデル通りに理想的に標本（データ）が得られるものとして、それらのデータから平均や 2 次の積率（分散）を求めて、モデルが理論的に与える平均値や分散と一致するかどうかを考えます。１試行の中で n を変化させるということは、n 回の繰り返しを 1 セットとし、そのセットごとに平均や SS を求めて、それを無限回繰り返すというイメージです。この値を用いて、考えられる方法で、2 次の積率（母集団のばらつきの指標）を計算してみます。これを無限回、繰り返した時の期待値を計算して、それが既知の母集団の積率と一致するかどうかを検討してみます。

最も簡単なモデルで試してみます。

1/2 の確率で－1、1/2 の確率で１となる事象

例としてはコインを投げて表なら 1 円もらえて裏ならば 1 円払う。

この場合平均値は 0 であり 2 次の積率も 1 であることは自明なのですが、念のために、母集団の平均値を知っている場合と知らない場合について考えます。知らない場合は、標本集団の平均値を期待値としてが母集団の平均値と考えるしかないでしょう。

どちらの場合も、標本集団の積率（SS/n）が母集団の積率であると考えて良さそうですが、本当にそうでしょうか、試してみましょう。ついでに SS/(n-1)についても計算してみます。

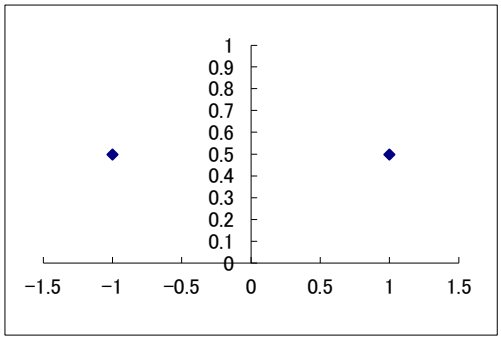


図 5, 繰り返し 1 の二項分布の例

１回の繰り返しの場合

n=1

母集団の分布中心（真の平均値μ）を知っている場合

実際のデーター	確率	真の平均値μ	SS	SS/n	SS/(n-1)
－1	1/2	0	1	1	-
1	1/2	0	1	1	-
期待値				1*	・

*: 1×1/2＋1×1/2

標本集団の平均値 m を母集団の平均値の期待値とする場合。

実際のデーター	確率	平均値 m	SS	SS/n	SS/(n-1)
－1	1/2	-1	0	0	-
1	1/2	1	0	0	-
期待値				0*	

2 回の繰り返しの場合

n= 2

母集団の分布中心（真の平均値μ）を知っている場合

実際のデーター	確率	真の平均値μ	SS	SS/n	SS/(n-1)
－1 －1	1/4	0	2**	1	2
－1 1	1/2*	0	2	1	2
1 1	1/4	0	2	1	2
期待値			2	1***	2****

*： ${}_2C_1(1/2)*(1/2)$

**： （－1－0）^{2}＋（1－0）^{2}＝2

***： 1×1/4＋1×1/2＋1×1/4 = 1

****： 2×1/4+2×1/2+2×1/4=2

標本集団の平均値 m を母集団の平均値の期待値とする場合

実際のデーター	確率	平均値 m	SS	SS/n	SS/(n-1)
－1 －1	1/4	－1	0	0	0
－1 1	1/2*	0	2**	1	2
1 1	1/4	1	0	0	0
期待値			1	1/2***	1****

*： ${}_2C_1(1/2)*(1/2)$

**： （－1－0）^{2}＋（1－0）^{2}＝2

***： 0×1/4＋1×1/2＋0×1/4 = 1/2

****： 0×1/4+2×1/2+0×1/4

実際に計算してみてイメージをつかみましょう

３回の繰り返しの場合

n= 3

母集団の分布中心（真の平均値μ）を知っている場合

実際のデーター	確率	真の 平均値μ	SS	SS/n	SS/(n-1)
－1 －1 －1	1/8	0	3	1	3/2
－1 －1 1	3/8	0	3	1	3/2
－1 1 1	3/8	0	3	1	3//2
1 1 1	1/8	0	3	1	3/2

期待値　　3　　1　　3/2

標本集団の平均値 m を母集団の平均値の期待値とする場合

実際のデーター	確率	平均値 m	SS	SS/n	SS/(n-1)
－1－1－1	1/8	－1	0	0	0
－1－1　1	3/8	－1/3	24/9*	8/9**	12/9***
－1　1　1	3/8	1/3	24/9	8/9	12/9
1　1　1	1/8	1	0	0	0
期待値			2	2/3****	1*****

*：　（－1－(－1/3)）^{2}＋（－1－(－1/3)）^{2}＋（1－(－1/3)）^{2}=24/9

**：　(24/9)/3

***：(24/9)/2

****：　0×1/8 +(8/9)×(3/8) +(8/9)×(3/8)＋0×(1/8)

*****：　0×1/8 +(12/9)×(3/8) +(12/9)×(3/8)＋0×(1/8)

４回の繰り返しの場合

n=４

実際のデーター	確率	平均値	SS	SS/n	SS/(n-1)
－1－1－1－1	1/16	－1	0	0	0
－1－1－1　1	4/16	－1/2	48/16	12/16	16/16
－1－1　1　1	6/16	0	4	1	4/3
－1　1　1　1	4/16	1/2	48/16	12/16	16/16
1　1　1　1	1/16	1	0	0	0
期待値			3	3/4*	1

*: 0×(1/16)+(12/16)×(4/16)+1×(6/16)+(12/16)×(4/16)+0×(1/16)

わかったこと、

この表を見ると SS/n は、n が変化すると変化してしまいますから、当然、母集団のデータの広がりを表現していません。つまり、標本を集団の平均まわりの積率）は母集団の 2 次の積率と一致しません、ただ、n が大きくなるにつ入れて、その差は小さくなっていきます。それに対して、SS/'(n-1)は、 n のおおきさにかかわらずいっていです。SS/'(n-1)は母集団のデータの広がりを示すものとして、確率分布曲線を描く単位に使えそうです。

何でこんなことになるのか

いくつかの、教科書では、どうしてそうなるのか説明しています。代表的なのは、「標本

集団の平均値はデータから計算されたものであり、それ自体データで、その平均値との差から計算された広がりの大きさは、標本集団のデータの広がりを表現するものとして、使える。そこから母集団の、広がりを推測するときに、標本集団の平均値との差から広がりを計算すると、外からデータを一つ加えたｋとになる。例えば、外から与えた場合、実際のデータを n-1 個のデータがわかると、残る一つのデータは自動的に決まってしまう。だからその SS は、n-1 個のデータの総和だと考えるべきである。平均値によって１つのデータが決まってしまうのだから、その分を差し引いた数を平均値から独立した自由に選べるデータの数として、自由度という。」という説明です。この文章、何を言っているのかわかりますか。正しいことを言っているのですが、その説明で、何故 n-1 で割るのか分かりますか？少なくとも私には説明になっていると思えません。私は次のように説明しています。

標本集団ではそれぞれの試行（繰り返しの 1 回分）の平均値と個々のデーターの差を求めています。この平均値を求めるときには、差を求めたデーターそのものも用いられています。したがって、その平均値は、その分だけ、母集団の平均値に近づいているのです。

2 回繰り返しのときは 1/2, 3 回繰り返しのときは 1/3、4 回のときは 1/4 近づいていることに注目しましょう。たとえば、3,4,5 というデータがあるときに、広がりの尺度として、平均値とデータの差を、

$$3-\frac{3+4+5}{3},\qquad 4-\frac{3+4+5}{3},5-\frac{3+4+5}{3}$$

と計算して、

ＳＳを計算すると

$$\left(3-\frac{3+4+5}{3}\right)^2+\left(4-\frac{3+4+5}{3}\right)+\left(5-\frac{3+4+5}{3}\right)^2$$

$$=(3-4)^2+(4-4)^2+(5-4)^2=2$$

平均を計算する分数の部分の赤いは、差し引こうとするデータです。自分自身が 1/n 分だけ加わってるのです。それをｎ個繰り返すと、１個分少なくなります。だから、広がり方を示す値として SS/3 と計算すると、2/3、　SS/2 と計算すると 1 でになるのです。これが、SS/n と計算すると 1/n だけ、広がりが小さくなる理由です。。

標本平均を母集団の平均値の推定値として使うことが問題ならば、母集団の平均値を使わずに、母集団の分散を計算することが出来ないかと考えてみます。問題は SS の計算ですね。

計算が煩瑣になるので、SS の定義式を変形しておきます。

定義式　$\sum_{k=1}^{n}(x_k-m)^2$

ｍは平均値で、

$$\sum_{k=1}^{n}(x_k-m)^2=\sum_{k=1}^{n}({x_k}^2-2mx_k-m^2)=\sum_{k-1}^{n}{x_k}^2+2m\sum_{k-1}^{n}x_k+nm^2$$

そもそも

$$m=\frac{1}{n}\sum_{k=}^{n}x_k$$

だから　定義式は

$$\sum_{k=1}^{n}(x_k-m)^2=\sum_{k-1}^{n}{x_k}^2-2\frac{(\sum_{k-1}^{n}x_k)^2}{n}+\frac{(\sum_{k-1}^{n}x_k)^2}{n}=\sum_{k-1}^{n}{x_k}^2-\frac{1}{n}\left(\sum_{k-1}^{n}x_k\right)^2$$

と簡略化できます。

「二乗の総和から総和の二乗を n で割ったものを引けば SS になる。」と覚えましょう。

a,b,c,d、４つのデータしかない母集団を考えます。母集団にには４つの個体しかない。だから、母集団の真の平均値μと分散 SS は計算できます。

$$\mu=\frac{a+b+c+d}{4}$$

$$\text{SS}=(a^2+b^2+c^2+d^2)-\frac{1}{4}(a+b+c+d)^2$$

母集団が４つしかないとはは知らないので、母集団から３つをサンプリングしたとする。そこから計算した SS は母集団の真の SS とは違うかもしれませんが、それを無限回繰り返せば、真の SS の限りなく使づくはずです。４つから３つを取り出す取り出し方は、

b,c,d　a,c,d　a,b,c　b,c,d　の３通りです。無限回繰り返せば、この４つの組み合わせがサンプルされる可能性は等しいので、４つの計算結果をたし合わせて平均化すれば、母集団の推定値として使えるはずです。どんな推定をするかということですが、サンプル集団の平均値を使ってはいけないというルールですから、やれることはただ一つ、サンプル集団の二つのサンプルの距離の二乗総和を SS の代理として計算してみることです。

母集団の個体のデータを　a,b,c,dとすると、最初の組はb,c,d、でその距離の２乗総和は

$$(b-c)^2+(b-d)^2+(c-d)^2=2b^2+2c^2+2d^2-2bc-2bd-2cd$$

以下

$$(a-c)^2+(a-d)^2+(c-d)^2=2a^2+2c^2+2d^2-2ac-2ad-2cd$$

$$(a-b)^2+(a-d)^2+(b-d)^2=2a^2+2b^2+2d^2-2ab-2ad-2bd$$

$$(a-b)^2+(a-c)^2+(b-c)^2=2a^2+2b^2+2c^2-2ab-2ac-2bc$$

是をつぃ合わせて＋

$$\text{式 A}=6a^2+6b^2+6c^2+6d^2-4ab-4ac-4ad-4bc-4bd-4cd$$

ところで

$(a+b+c+d)^2 = a^2+b^2+c^2+d^2+2ab+2ac+2ad+2bc+2bd$+2cd

移項して

$$2ab+2ac+2ad+2bc+\ +2bd+2\mathrm{cd} = -(a^2+b^2+c^2+d^2)+(a+b+c+d)^2$$

これを式 A に代入すると

式 A’ $6a^2+6b^2+6c^2+6d^2+2(a^2+b^2+c^2+d^2)-2(a+b+c+d)^2$

$$= 8(a^2+b^2+c^2+d^2)-2(a+b+c+d)^2$$

$$= 8((a^2+b^2+c^2+d^2)-\frac{1}{4}(a+b+c+d)^2)$$

これを、組み合わせの場合の数 4 で割れば

$$2((a^2+b^2+c^2+d^2)-\frac{1}{4}(a+b+c+d)^2)$$

これは、真の SS の 2 倍です。$(a^2+b^2+c^2+d^2)-\frac{1}{4}(a+b+c+d)^2$を母集団の個体数 4 で割れば、真の分散$\sigma^2$になります。

つまり、この方法でも、無限回繰り返せば、真の分散の 2 倍値が計算できるのですが、この計算は手間がかかりますから

$$SS = \sum_{k\ =1}^{n}(x_k-m)^2 = \sum_{k-1}^{n}{x_k}^2-\frac{1}{n}\left(\sum_{k-1}^{n}x_k\right)^2$$

標本集団の分散は SS/n、母集団の分散は SS/自由度と覚えます。この場合は、標本手段の平均値を母集団の平均値の期待値として、母集団の SS の計算の時に一回だけ外側から与えていますから、自由度は n-1 ですが、相関分析の時などは、説明変数と被説明変数の 2 つで平均値を期待値として外側から与えますから、自由度は n-2 です。自由度=n-1 と覚えない方が良いでしょう。

もう一つの問題として、母集団の平均値として、観察された標本集団の平均値を期待値とすることの確からしさの問題があります。

繰り返しの数が違うと繰り返しごとに求められる平均値の分布はどのように変化するでしょうか。

前に行った確率 1/2 の例から標本集団の平均値とその平均値が出る確率の図を作ってみます。横軸が求められる平均値の値、縦軸がその値となる確率です。

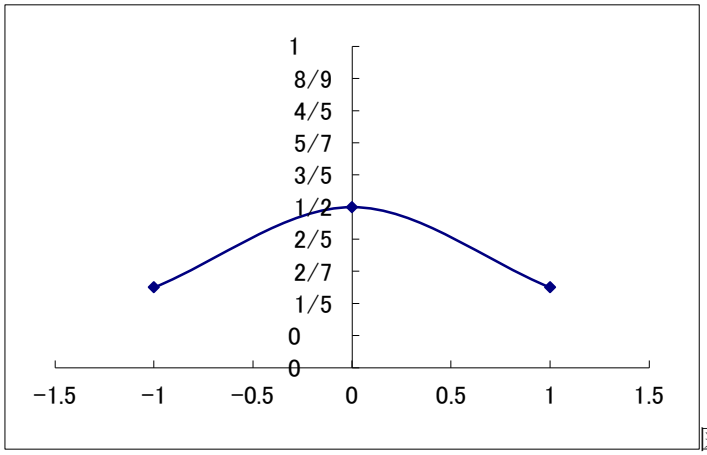


図 6.　n=2 の場合

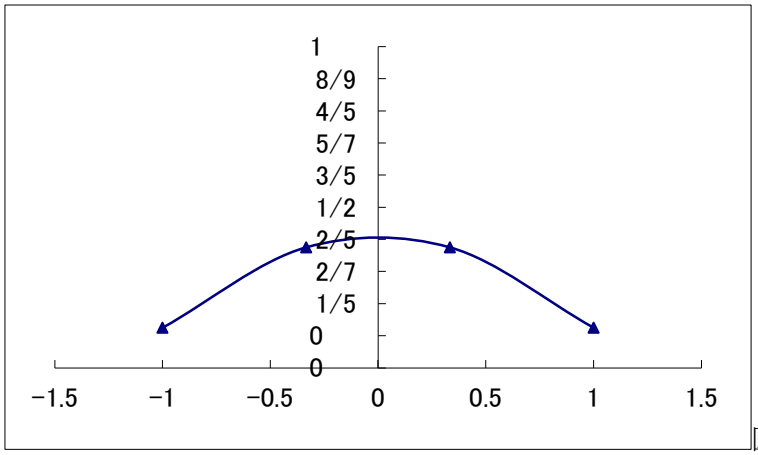


図 7.　n=3 の場合

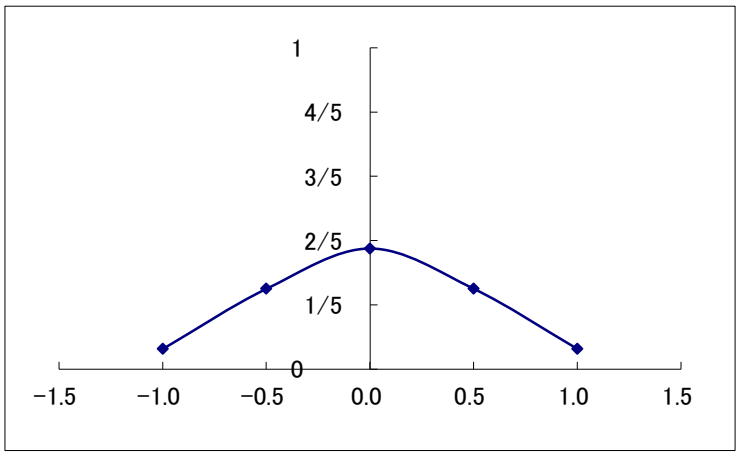


図 8.　n=4 の場合

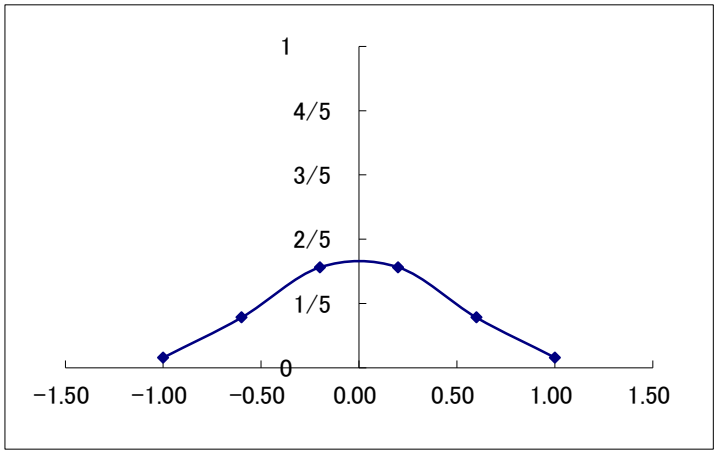


図 9 n=5 の場合

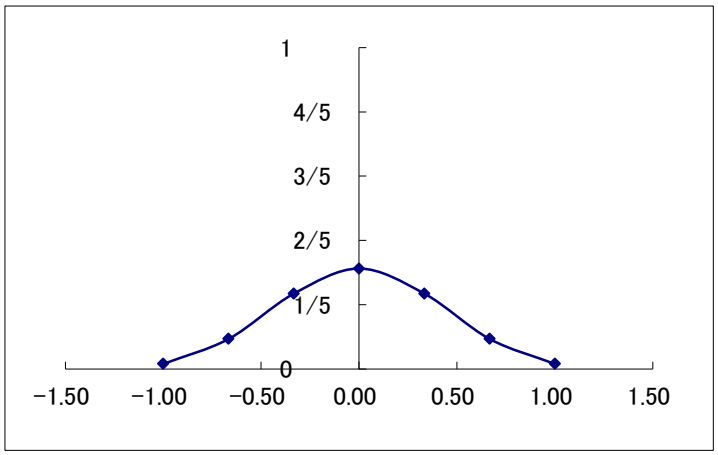


図 10 n=6 の場合

n が大きくなると両側の値が小さくなり、尖った形になり、正規分布に近づきます。n が無限大の時の二項分布を正規分布です。

ここまでは極めて単純なケースを用いて考察を行ってきました。一般的な確率事象はもう少し複雑です。そのような場合にも今まで考察してきた結果があてはまるか確認をしておきましょう。実際計算してみると感覚が身につきます。

母集団の確率分布がゆがんでいる場合
－１となる確率が１となる確率の２倍ある場合を考えます。

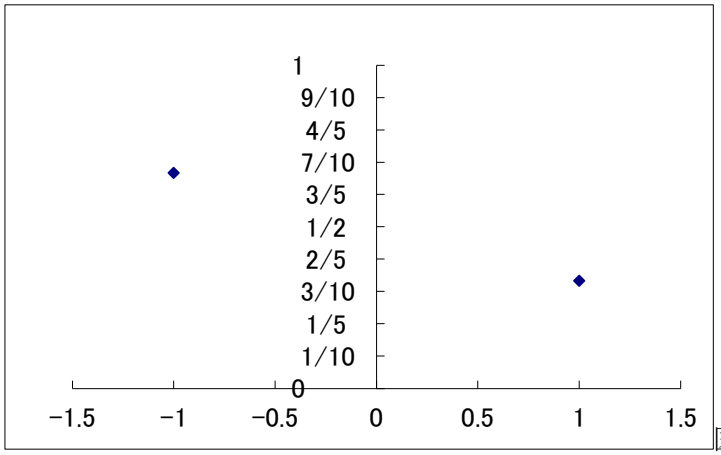


図 11. n=1 p=1/3 の二項分布

平均値は-1/3 2 次の積率は 8/9

１回の繰り返しの場合

n=1

実際のデーター	確率	平均値	SS	SS/n	SS/(n-1)
－１	2/3	－１	0	0	-
1	1/3	1	0	0	-
期待値				0	0

2 回の繰り返しの場合

n=２

実際のデーター	確率	平均値	SS	SS/n	SS/(n-1)
－１－１	4/9	－１	0	0	0
－１　1	4/9*	0	2**	1	2
1　1	1/9	1	0	0	0
期待値			8/9	4/9***	8/9****

*：　${}_2C_1(2/3)*(1/3)$

**：　（－１－０）^{2}＋（１－０）^{2}＝２

***：　0×4/9＋1×4/9＋０×1/9 = 4/9

****：0×4/9+2×4/9+0×1/9

３回の繰り返しの場合

n=３

実際のデーター	確率	平均値	SS	SS/n	SS/(n-1)
－１－１－１	8/27	－１	0	0	0
－１－１　１	12/27	－1/3	24/9*	8/9**	12/9***
－１　１　１	6/27	1/3	24/9	8/9	12/9
１　１　１	1/27	1	0	0	0
期待値			2	48/81****	8/9*****

*：　（－１－(－1/3)）^{2}＋（－１－(－1/3)）^{2}＋（１－(－1/3)）^{2}=24/9

**：　(24/9)/3

***：(24/9)/2

****：　0×8/27 +(8/9)×(12/27) +(8/9)×(6/27)＋0×(1/27)

*****：　0×8/27 +(12/9)×(12/27) +(12/9)×(6/27)＋0×(1/27)

４回の繰り返しの場合

n=４

実際のデーター	確率	平均値	SS	SS/n	SS/(n-1)
－１－１－１－１	16/81	－１	0	0	0
－１－１－１　１	32/81	－1/2	48/16	12/16	16/16
－１－１　１　１	24/81	0	4	1	4/3
－１　１　１　１	8/81	1/2	48/16	12/16	16/16
１　１　１　１	1/81	1	0	0	0
期待値			3	2/3	8/9

図を作ってみます。横軸が求められる平均値の値、縦軸がその値となる確率です。

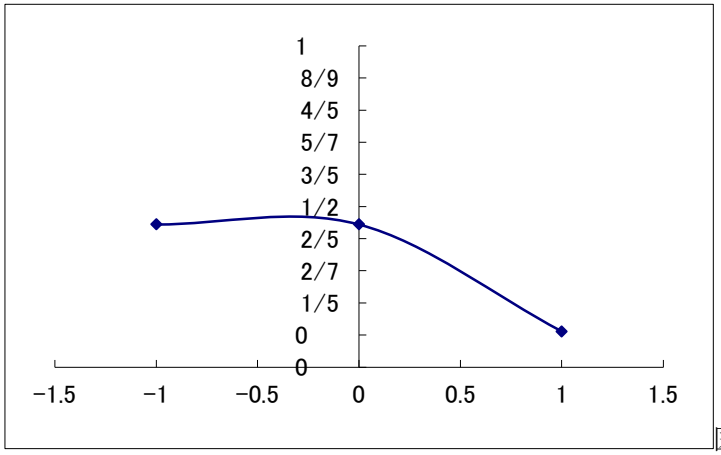


図 12. n=2 の場合

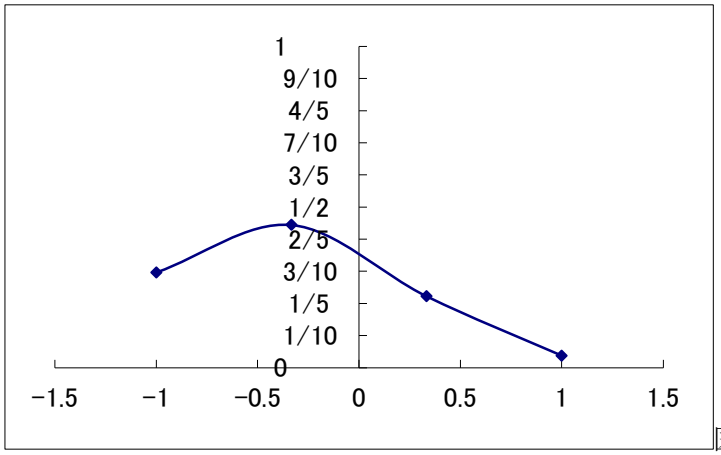


図 13. n=3 の場合

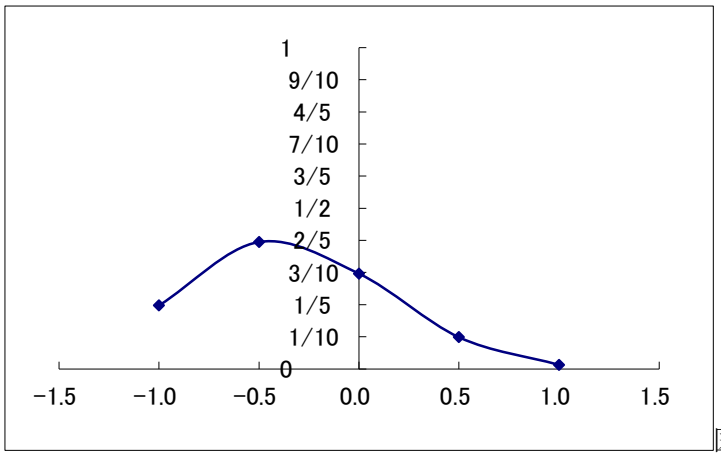


図 14. n=4 の場合

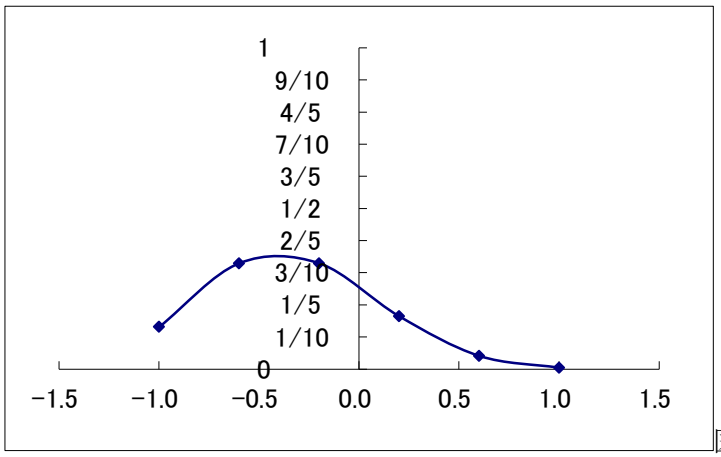


図 15　n＝5 の場合

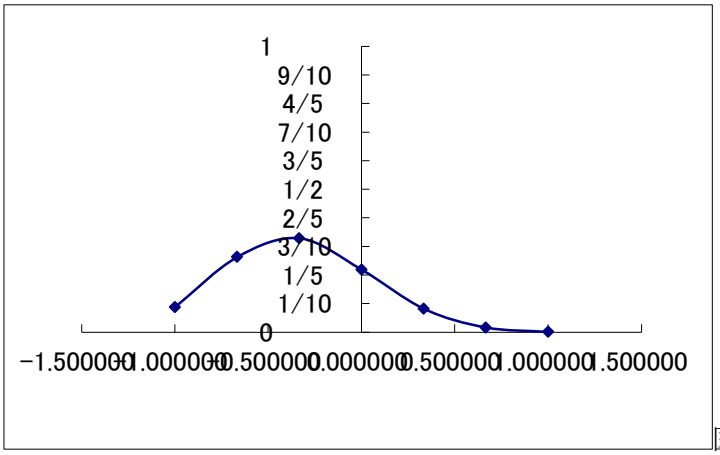


図 16. n=6 の時

わかったこと、

SS/'(n-1)が、母集団の 2 次の積率（母集団の原点まわりのバラツキの指標）表わしている。また、n が大きくなると正規分布に近づく

推定された母集団の平均値の確からしさを考える

n が大きくなると、母集団の平均値の推定値の確率分布が尖ってくる。つまり、予測値が母集団の平均値の周りの狭い範囲に集まってきます（これを中心極限定理といいます。無限大回繰り返せば幅が 0 となり、予測値は母集団の平均値そのものになるはずです。）。つまり、n が大きいほど予測値の確からしさは増します。このことは、よりたくさんのデーターから母集団の推定値を推測したほうが確かだという経験則にも一致します。

何をするのか

今までと同様に、母集団の分布がわかっている事例について、n 回の繰り返しで求めた平均値の推定値と母集団の平均値の差を求め、その 2 次の積率がどのように n によって変化するかを考えます。つまり本当の平均値と推定された平均値の差の 2 乗の平均が、n が大きくなることによってどのくらい小さくなっていくかを考えます。

数学的に表現すれば、母集団の真の平均値と推定された平均値の差の 2 乗の期待値（平均値周りの 2 次の積率：E（(M－μ）^{2})）と n の関係についての考察ということです。

（Mは標本集団から推定された平均値、μ は母集団の平均値をあらわす。）

今まで用いた考察のために用いてきた 1/2 確立で起こる事象のモデルをそのまま用いることにします。下には 1/2 確率 3 回の繰り返しの場合についての場合を示しました。この表の、下線引いた値（推定される平均値）と、母集団の平均値 0 の間の差を求め、その 2 乗にそれが起こる確率を乗じて、その総和を求め、それを母集団の平均値の分散として、その値と繰り返しの回数 n の関係を考察します。

今までと同じように、経験的にやるならば、真の平均値の値とデータから求められた平均値の差の計算を無限回繰り返して、平均を求めれば良いのだから

３回の繰り返しの場合

n=３

実際のデーター	確率	平均値	SS	SS/n	SS/(n-1)
－１－１－１	1/8	－１	0	0	0
－１－１　１	3/8	－1/3	24/9*	8/9**	12/9***
－１　１　１	3/8	1/3	24/9	8/9	12/9
１　１　１	1/8	１	0	0	0
期待値			2	2/3****	1*****

この場合求める値の計算は以下のとおり

（－１）^{2}×(1/8)+(－1/3)^2×(3/8)＋(1/3)^2×(3/8)+1^2×(1/8)=1/3

この計算結果は以下のとおりになります。（やってみることをお勧めします。）

ゆがみのない例の場合(σ^2=1)

n=1	1	σ^2/1
n=2	1/2	σ^2/2
n=3	1/3	σ^2/3
n=4	1/4	σ^2/4

ゆがみのある集団の例の場合（σ^2=8/9）

n=1	8/9	σ^2/1
n=2	4/9	σ^2/2
n=3	8/27	σ^2/3
n=4	8/36	σ^2/4

推測

母集団の真の平均値と推定された平均値の真の平均値周りの２次の積率Ｅ（(Ｍ－μ）^{2}）は、σ^2/ｎで求められそうです。

代数的な証明

母集団の 2 次積率の推定値であるσ^2を個々のデーターに基づく期待値として計算することを考えます。

$\mu = 0$ とすると

$$\sigma^2 = \sum_{j=1}^{\infty} \frac{1}{n} \sum_{i=1}^{n} (M_j + e_{ij})^2$$

$$= \sum_{j=1}^{\infty} M_j^2 + 2\frac{1}{n} \sum_{j=1}^{\infty} (M_j \sum_{i=1}^{n} e_i) + \frac{1}{n} \sum_{j}^{\infty} \sum_{i}^{n} e_{ij}^2$$

式 18

第一項はＥ（Ｍ^2）：つまり求める期待値Ｅ（（Ｍ－μ）^{2}）です。

第 2 項は０　（ここでもどんなMを選んだにしても e はその周りのばらつきなのだからその総和は０になります。）

第３項は SS の期待値の 1/n であることに気づきます。

SS=(n-1)σ^2だから

第３項は $\sigma^2 - \frac{1}{n}\sigma^2$

(式 18）にこれらを代入して

$$E(M^2) + \sigma^2 - \frac{1}{n}\sigma^2 = \sigma^2$$

$$E(M^2) = \frac{1}{n}\sigma^2$$

証明終わり：覚えておくこと、

標準偏差と標準誤差は違うものです。標本集団から推定した母集団の分散の平方根が標準偏差、標本集団から推定した母集団の平均値の信頼範囲を示すのが芳醇誤差です。グラフに誤差バーを入れるときに、母集団のデータのバラツキ方の推定値を示すのならば、標準誤差（例えばいくつかの標本群の母集団のデータの重なり合いを議論するときは標準偏差、それぞれの標本群の母集団の平均値の信頼範囲を示したいときは標準誤差を誤差バーとして書き入れます。

二項分布の特性

q=1－p ですから、$W(k) = {}_n C_k p^k q^{(1-k)}$という曲線は、n と p だけで決まります。n と p を知っていれば、W(k)という曲線が描けるのです。という理由で、２項分布することがわかっているデーターについて、いくつかのデーターから n と p を推測する方法を考えます。２項分布することがわかっているデータが、どんな分布するのかを理解すれば、実際のデーターの分布から n と p を推測する方法がわかるかもしれません。このことは、その n と p で二項分布の形の特性を表現することと同じです、

今までに示した２項分布の図から、確率分布の形は、頂上を一つ持つ山形で、n が大きくなるにつれて、その山が急峻に細くとがってくるということがわかります。１つしか頂上

がない曲線なのですから、その頂上はどこにあるのかということが曲線の形を表す重要な要素です。山が一つしかないのだから、１回微分して、f'(x)=0 となる x を求めれば、そこで極大値になりますから、その x のところで、頂上ということになります。しかし、どうも$W(k) =_n C_k p^k q^{(1-k)}$を k で微分するのはかなり難しそうです。あとでこれをやることになりますが、その前に、もう少し考えを整理しておきましょう。

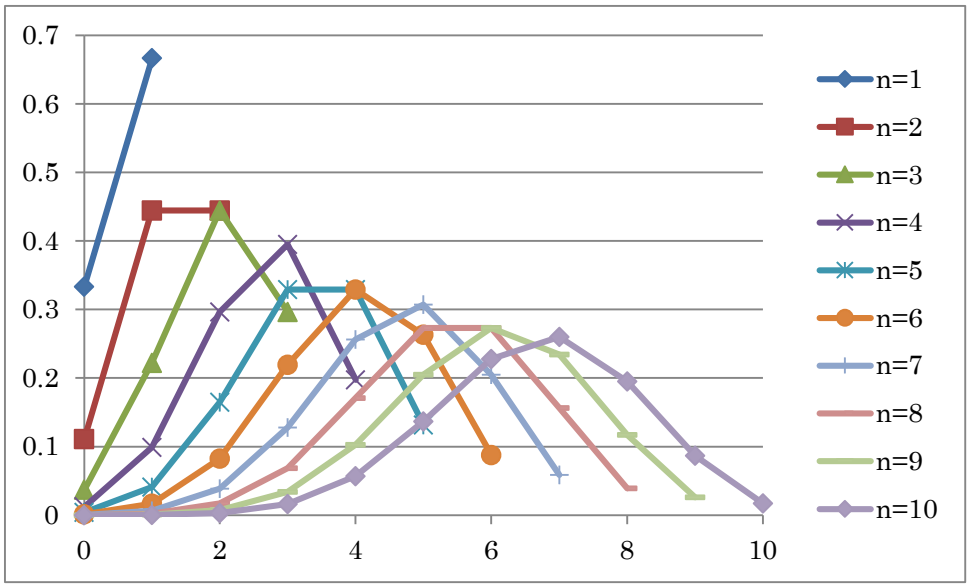


図 17. n の増加に伴う二項分布の変化

図 17 は、$W(k) =_n C_k p^k q^{(1-k)}$と n,k の関係を示したものです。縦軸が$W(k) =_n C_k p^k q^{(1-k)}$で、横軸が k です。n が 1 から 10 まで変わるときに、その n ごとに、k の値に対して、$W(k) =_n C_k p^k q^{(1-k)}$の値を示したものです。この図は p=2/3 の時の図です。p+q=1 ですから、q は 1/3 です。コインのたとえでは、A（表）である確率が p が 2/3、B（裏）である確率 q が 1/3 という意味です。この場合、n 回コインを投げて、A が k 回、B が n－k 回出る確率が、W(k)です。 n が大きくなれば、当然、その確率の分布を表す W(k)の曲線は、右に広がっていきます。それにしたがって、頂上も右に移動していきます。ここで、n=9 の曲線に注目してください。頂上は k=6 のところにあります、n=6 の時は、k=4、n=3 の時は k=2 のところに、頂上があります。k/n は 2/3 で p の値になっています。n が他の値の場合、必ずしも頂上の位置がはっきりしませんが、だいたい k/n=p となる k のところに頂上がありそうです。つまり、頂上となる k は k=np のように見えます。k は A となる回数のことですね。 １回の試行で p ＝が出る確率が 2/3 なのだから、２回投げればAとなる確率が k ＝2/3＋2/3 のところで最大となるのは当然のことでしょう。「1 回の試行で A となる確率を知っている時、n 回試行を繰り返したら、何回Aという事象が起こるか、もっとも確率が高い回数は何回か。」という問いに対する答えは np 回です。このような問いに対する答えを期待値と言います。頂上となる k の値の期待値（最も確率が高い、確率の頂

点となるｋ、最頻値）は、k=np です。
次に、実際のデーターから、期待値を推測する方法を考えます。直感的には、期待値≒平均値ですね。すでに前章の説明ではこのことを自明のこととして、母集団の平均値を推測する方法を考えました。これは直観的な方法です。ここではもう少し数学的に考えて、k の期待値（最頻値）＝平均値となるかどうかを検討します。
まず、２項分の曲線が表す分布通りの理想的なデーターが得られたときについて考えます。たとえば、n=9 のときに、頂上が k=6 のところにある、完璧な二項分布通りのデーターが試行を何回繰り返しても得られるという、理想的な世界を考えます。この時、k の平均値がどのような値になるかを考えます。図 17 の例でいえば、排反事象Ａ，Ｂがあって、その確率がそれぞれ p=2/3、q=1/3 となることを、9 回繰り返す。これを１試行として、無限回これを繰り返せば、２項分布が示す確率分布通りにデーターが出るはずです。そういう場合にＡという事象が１試行中に現れる回数の平均値がどうなっているのかという話です。
前の議論と同じことですが、m 回試行を繰り返して、Ａという事象が一つの試行で k 回現れる、試行回数の期待値は、m×W(k)ですね、ですから、その k の値については、m×W(k)×ｋ回、Ａという事象が起こることになります。これをすべてのｋについて足し合わせて、m 回という試行回数で割れば、平均的な k の値ということになります。k の平均を$\bar{k}$と表せば

$$\bar{k} = \frac{m \times W(k) \times k}{m} = W(k) \times k$$

ということです。
書き方としては面倒ですが、この式をきちんと計算できるように書きなおすと

$$\bar{k} = \sum_{k=0}^{n} \frac{m \times W(k) \times k}{m} = W(k) \times k$$

$$= \sum_{k=0}^{n} \binom{n}{k} p^k q^{(1-k)} \times k$$

$$= \sum_{k=0}^{n} \frac{kn!}{k!\,(n-k)!} p^k q^{(n-k)}$$

となりますが、この式はさらに簡略化して書くことができます。

$$\bar{k} = \sum_{k=0}^{n} \frac{kn!}{k!\,(n-k)!} p^k q^{(n-k)}$$

$$\bar{k} = \sum_{k=0}^{n} \frac{kn(n-1)!}{k!\,((n-1)-(k-1))!} p \times p^{(k-1)} q^{((n-1)-(k-1))}$$

$$\bar{k} = \sum_{k=0}^{n} \frac{n(n-1)!}{(k-1)!\,((n-1)-(k-1))!} p \times p^{(k-1)} q^{((n-1)-(k-1))}$$

$$\bar{k} = \mathrm{np} \sum_{k=1}^{n-1} \frac{(n-1)!}{(k-1)!\,\big((n-1)-(k-1)\big)!} p^{(k-1)} q^{((n-1)-(k-1))}$$

$$\frac{\bar{k}}{np} = \sum_{k=1}^{n-1} \binom{n-1}{k-1} p^{(k-1)} q^{((n-1)-(k-1))}$$

となります。ところでΣ記号のところは、n=n-1 としたときの、全確率の総和になっています。全確率の総和は1に決まっています。したがって

$$\bar{k} = \mathrm{np}$$

となります。私たちがここでしたことは、確率分布通りに理想的にデーターが得られるのならば、平均値は期待値（最頻値）と一致して、np が予想できることを明らかにしました。

次に確率分布曲線の特性を表しているのは、分布曲線のふくらみ方ですね。分布曲線の水平方向の太さと言ってもよいかもしれません。これも n と p によって決まっているはずです。このふくらみ方をどのように表せばよいのかを考えます。とがり方は分布曲線の k 軸の値から一番とがったところ、つまり np までの距離の平均値（これも期待値として）を求めれば良いでしょう。

$$\sum_{k=0}^{n} |k - np| \, W_{(k)}$$

ですね。絶対値の記号をつけたのは、k が np よりも大きい時と小さい時があるからです。

$$\sum_{k=0}^{n} (k - np) \, W_{(k)}$$

とすると、k=np を中心として左右対称なのですから、$\sum_{k-1}^{n}(k-np)=0$となってしまいます。こういう場合は$(k-np)^2$の期待値を求めて、その平方根を求めれば良いでしょう。そこで、期待値の式

$$\sum_{k=0}^{n} (k - np)^2 \, W_{(k)}$$

式 9

これは2次の積率（モーメント）です。

$$= \sum_{k=0}^{n} (k^2 - 2knp + n^2p^2) \, W_{(k)}$$

$$= \sum_{k=0}^{n} k^2 W_{(k)} - \sum_{k=0}^{n} 2knpW_{(k)} + \sum_{k=0}^{n} n^2p^2 W_{(k)}$$

$$= \sum_{k=0}^{n} k^2 W_{(k)} - 2np \sum_{k=0}^{n} kW_{(k)} + n^2p^2 \sum_{k=0}^{n} W_{(k)}$$

第2項目のΣの値は、k の平均値を求めたときの式ですから、np です。第3項目のΣの値は、確率の総和だから1です。ということで

$$\sum_{k=0}^{n} k^2 W_{(k)} - 2n^2p^2 + n^2p^2$$

$$\sum_{k=0}^{n} k^2 W_{(k)} - n^2p^2$$

式 10

となります。つまり、$\sum_{k=0}^{n} k^2 W_{(k)}$について考えればよいことになりますが、ここで少しテクニックを使います。

式 10 をよく見ると、Σ記号のところは、k^2の期待値になっています。np は平均値ということは k の期待値です。つまり式 9 は、k の二乗の期待値から、k の期待値の二乗を引くという式になっているのです。ここで、x の分散を V(x)、x の期待値を E(x)と書くと,

$$V_{(x)} = E_{(x^2)} - E_{(x)}{}^2$$

式 11

前にやった、SS の計算を簡便化する式

$$\sum_{i=0}^{n} (x_i - \bar{x})^2 = \sum_{i=0}^{n} x_i{}^2 - \frac{1}{n}\left(\sum_{i}^{n} x_i\right)^2$$

式 12 は式 11 を変形したもので、式が意味する内容は同じです。

簡単に説明すると

$$V_{(x)} = E(x - \bar{x})^2$$

$$V_{(x)} = E(x^2) - 2Ex\bar{x} + E(\bar{x}^2)$$

$$= E(x^2) - 2E(x\bar{x}) + E(\bar{x}\bar{x})$$

$$= E(x^2) - 2\bar{x}E(x) + \bar{x}E(\bar{x})$$

$\bar{x} = E(x)$ですから

$$= E(x^2) - 2E(x)E(x) + E(x)E(x)$$

$$= E(x^2) - (\mathrm{E(x)})^2$$

となります。

さて、話を元に戻します。式 9

$$V_{(k)=} \sum_{k=0}^{n} k^2 W_{(k)} - n^2p^2$$

np をどのように求めるかはわかっているのですから、式 10 のΣのところ式、つまり E(k²)を n,p で表せば、この式を n,p で表すことができます

E(k²)を n,p で表すのは、結構難しくて、ちょっとした工夫が必要です。アイデアとしては、E(k)=np を求めたときのやり方、すなわち、n-1 の時の確率の総和を求めて、k と二項係数の式を消してしまうということが考えられますが、この場合 k² で k が二つ入っています。

そこが面倒なところです。2 回同じことを繰り返さなければならない。少しテクニックが必要です。k と k-1 が入っているのならば、同じことが繰り返せそうな気がします。
そこで、$\mathrm{E\{k(k-1)\}}$について考えることにします。

$$\mathrm{E\{k(k-1)\}} = E(k^2 - k) = E(k^2) - E(k) \qquad \text{式 13}$$

ですね。E(k)=np ですから、$\mathrm{E\{k(k-1)\}}$がわかれば $\mathrm{E(k^2)}$がわかるでしょう。
ということで、$\mathrm{E\{k(k-1)\}}$について考えます

$$\mathrm{E\{k(k-1)\}} = \sum_{k=0}^{n} k(k-1)\binom{n}{k} p^k q^{n-k}$$

$$= \sum_{k=0}^{n} \frac{k(k-1)n!}{(n-k)!\,k!} p^k q^{n-k}$$

$$\sum_{k=0}^{n} \frac{(k(k-1)n)(n-1)(n-2)!}{(n-k)!\,(k-2)!} p^k q^{n-k}$$

$$n(n-1)\sum_{k=0}^{n} \frac{(n-2)!\,k(k-1)}{(n-k)k!} p^2 p^k q^{n-k}$$

$$n(n-1)\sum_{k=0}^{n} \frac{(n-2)!}{((n-2)-(k-2))!\,(k-2)!} \cdot_n\ p^2 p^{k-2} q^{(n-2)-(k-2)}$$

$$n(n-1)p^2 \sum_{k=0}^{n} \frac{(n-2)!}{\big((n-k)\big)!\,(k-2)!} p^{k-2} q^{(n-2)-(k-2)}$$

$$= \ n(n-1)p^2 \sum_{k=0}^{n} \binom{n-2}{k} p^{k-2} q^{(n-2)-(k-2)}$$

Σ の値は、n=n-2 の時の確率の総和だから、1 です。
したがって、

$$= \mathrm{n(n-1)}p^2$$

これを式 12 にあてはめると

$$\mathrm{E\{k(k-1)\}} = E(k^2) - E(k)$$

$$\mathrm{n(n-1)}p^2 = \mathrm{E}(k^2) - np$$

となって、

$$\mathrm{E}(k^2) = \mathrm{n(n-1)}p^2 + np$$

これを、式 10 にあてはめると

$$V_{(k)} = E_{(k^2)} - {E_{(k)}}^2$$

$$= \mathrm{n(n-1)}p^2 + np - n^2p^2$$

$$= n^2p^2 - \mathrm{n}p^2 + \mathrm{np} - n^2p^2$$

$$= \mathrm{np(1-p)}$$

$$= \mathrm{npq}$$

分散 V(x)は、通常 σ^2 という記号で表します。分散の平方根$\sqrt{V_{(x)}}$が標準偏差といって、

データが平均値からどのくらい隔たっているか、その距離の平均値として使われ、それをσと表すからです。
上述の結論をσを使って表すと

$$\sigma^2 = \mathrm{npq} \qquad \mathrm{p}+\mathrm{q}=1$$

となります。
結論を示す

$$\mathrm{B}(n,p)$$

について

$$\mu = np, \sigma^2 = \mathrm{np}(1-p)$$

本日の課題
二項分布の正規分布への拡張。

二項分布$\mathrm{B}(n,p)$で確率pの現象が現れる回数 k の関数としてあらわすと。

$$W(k) =_n C_k p^k q^{(n-k)} \quad \text{あるいは} \quad \binom{n}{k} p^k q^{(n-k)}$$

コンビネーション記号を書き換えて分数で表すと

$$W(k) = \frac{n!}{k!(n-k)!} p^k q^{(n-k)} \qquad \mathrm{p+q=1}$$

これを対数にすると

$$\log \mathrm{W}(k) = \log(n!) - \log(k!) - \log(n-k)! + k\log(p) + (n-k)\log(q)$$

となります。
こうすると、複雑な式が対数の足し算に単純化できます。掛け算の形が足し算になったためにそれぞれの項を独立して考えることができます。ここでは、k を連続変数（整数にかかぎらず様々な値をとる実数）として、W(k)の形を考えるのですから、k=x と書き換えておきましょう。（一種の習慣です。 k は様々な値をとる不連続な整数のイメージです。これに対して x は連続して様々な値をとる実数のイメージです。）

$$\log \mathrm{W}(x) = \log(n!) - \log(x!) - \log(n-x)! + k\log(p) + (n-x)\log(q)$$

式 23

突然ですが、

$$\lim_{x\to\infty} \int_1^x log\, t\, dt = \log x!$$

です。つまり、xが十分に大きければ、$\int_1^x log\, t\, dt \fallingdotseq \log x!$です。
この式は単純な式で、慣れてくれば直感的にわからないこともないのですが、きちんと、これを証明するには手間がかかり、途中でいくつかのテクニックを使う必要があります。

この証明をしますが、長い退屈な証明なので、そういうことが嫌いな人は、ここは適当に読み飛ばしてください。それでも問題ないと思いますが、何をしているのか理解するためには、読んでおくと参考になるかもしれません。

$$\lim_{x\to\infty}\int_1^x log\,t\,dt = \log x!$$

式 24

の証明

この式は、

$$\lim_{x\to\infty}\frac{\int_1^x \log_e t dt}{\log_e x!} = 1$$

の変形ですが、式 24 の意味を考えながら、この形に持っていきます。

式 23 の極限記号の中

$$\int_1^x \log t dt$$

の意味は、曲線 log t と x 軸、直線 x=x に囲まれた次の図形の面積を求めるということです。

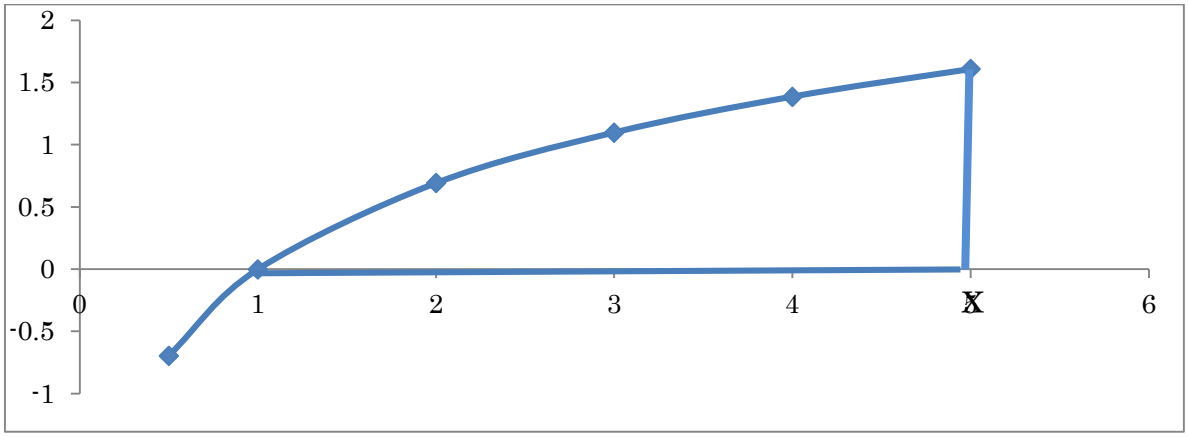


図 18-1, 対数の積分の極限の計算-1

この図に次のように、いくつかの長方形を書き加えます。

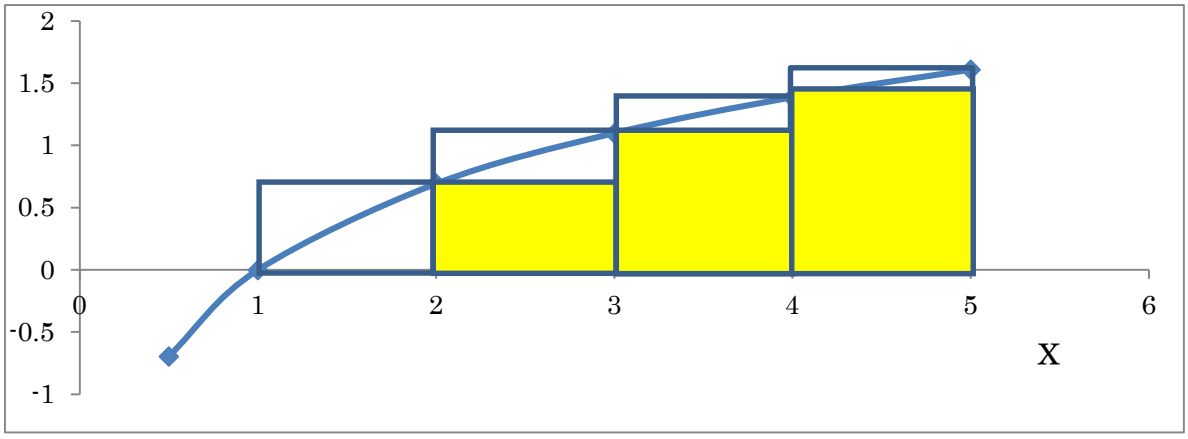


図 18-2, 対数の積分の極限式の計算-2

この図の意味は、以下の 2 つの図と log t の曲線を重ね合わせたものです。

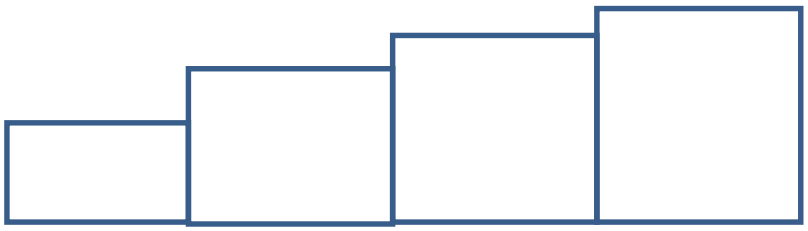


図 18-3, 対数の積分の極限式の計算-3

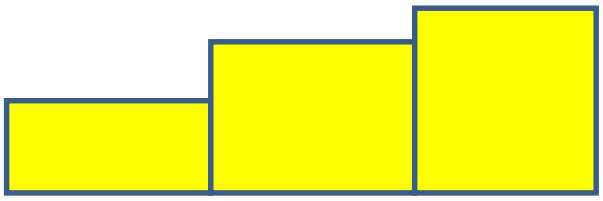


図 18-4, 対数の積分の極限式の計算-4

図 18-3 の 4 つの 4 角形の面積の合計を考えます。 1 つの 4 角形の幅は 1 です。そうすると、たとえば、一番左の四角形の面積は log2 ×1=log2 ですね。 4 つの四角形の面積の合計は log 2+log3+log4+log5 です。つまり log5!になります。
X がもっと大きくなった場合について一般化すると、面積の合計は logx!です。
同じようにして、図 18-4 の黄色い四角形の面積の合計は、log(x-1)!です。
ここで、面積の大きさを比べると、図 18-3 の四角形の面積の合計が一番大きくて、次が

$\int_1^x \log t dt$、で、図 18-4 の四角形の面積の合計が一番小さいということに気が付きます。

不等号で表すと次のようになります。

$$\log(x-1)! < \int_1^x \log t dt < \log x!$$

X は 1 以上の整数なのだから、$\log x!$は正の値になります。したがって、$\log x!$で各辺を割っても不等号の向きは変わらないでしょう。ですから、

$$\frac{\log(x-1)!}{\log x!} < \frac{\int_1^x \log t dt}{\log x!} < \frac{\log x!}{\log x!}$$

式 25

右辺が 1 であることは明らかです（分母と分子が同じだから）。
そこで一番左の辺について、その極限を考えます。

$$\frac{\log(x-1)!}{\log x!} = \frac{\log(x-1) + \log(x-2) + \cdots + \log 1}{\log x!}$$

$$= \frac{\log x + \log(x-1) + \log(x-2) + \cdots + \log 1 - \log x}{\log x!}$$

$$= \frac{\log x! - \log x}{\log x!}$$

$$= 1 - \frac{\log x}{\log x!}$$

この式で、x が無限大に大きくなれば、式の 2 項目は 0 に近づくでしょう。
このことは、私には自明のように思えますが、この辺の感覚は人によって違うかもせれません。
念のために、手数をかけて証明しておきましょう。ありそうなのは、以下の証明です。
証明したい内容は

$$\lim_{x\to\infty} \frac{\log x}{\log x!} = 0$$

式 26

です。
まず、以下の式がなりたつことを示します

$$\log k! > \log k + \log(k-1) + \cdots + \log\left[\frac{k}{2}\right] > \left(\frac{k}{2} - 1\right)\log\left[\frac{k}{2}\right]$$

式 27

この式の意味は、以下の通りです。

まず、$\left[\frac{k}{2}\right]$ですが、ここでは、k 半分を超えない整数の意味で使っています。たとえば、

k=5 の時は、$\left[\frac{k}{2}\right] = 2$、k=4 の時も$\left[\frac{k}{2}\right] = 2$

左辺と真ん中の辺の$\log k! > \log k + \log(k-1) + \cdots + \log\left[\frac{k}{2}\right]$のところは、部分が全体を超えることはないと言っているだけです。

$$\log k! = \log k + \log(k-1)! + \cdots + \log 1$$

で、これは正の数を足し合わせただけのものです。

これの値が、それよりも少ない工数を足し合わせた

$$\log\log k + \log(k-1) + \cdots + \log\left[\frac{k}{2}\right]$$

よりも大きいのは当然です。

分かりにくいのは、式 27 の真ん中の辺と右辺の関係です。

これは、次の図に示したことを言っているのです。

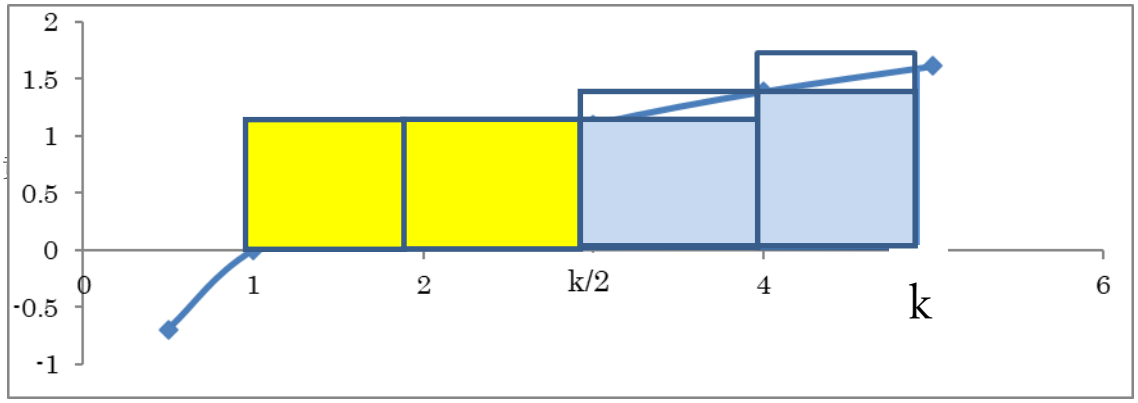


図 19 大小関係

これを$\frac{\log x}{\log x!}$に戻って考えると、x が 1 より大きいのでこの値は正ですから

次のようになります。なお、分子が変わらずに分母により小さなものが入るので、不等号の向きは反対になります。

$$0 < \frac{\log x}{\log x!} < \frac{\log x}{\left(\frac{x}{2}-1\right)\log\frac{x}{2}} = \frac{\log x}{\left(\frac{x}{2}-1\right)(\log x - \log 2)} = \frac{1}{\left(\frac{x}{2}-1\right)\left(1-\frac{2}{\log x}\right)}$$

$$\lim_{x\to\infty} \frac{1}{\left(\frac{x}{2}-1\right)\left(1-\frac{2}{\log x}\right)} = 0$$

ですから、挟み撃ちの原理で、はさんでいる両側が 0 なのだからはさまれているものも 0 です。

$$\lim_{x\to\infty} \frac{\log x}{\log x!} = 0$$

ということは、

$$\lim_{x\to\infty}(1-\frac{\log x}{\log x!})=1$$

$$\lim_{x\to\infty}\frac{\log(x-1)!}{\log x!}=1$$

式 24 に戻ると

$$\frac{\log(x-1)!}{\log x!}<\frac{\int_1^x \log t dt}{\log x!}<\frac{\log x!}{\log x!}=1$$

で左辺も x を無限大にしたときの極限は 1 ですから、これも挟み撃ちの原理で。

当然、$\frac{\int_1^x \log t dt}{\log x!}$についても

$$\lim_{x\to\infty}\frac{\int_1^x \log t dt}{\log x!}=1$$

です。分数の値が 1 ということは、分母分子が同じということですから、

$$\lim_{x\to\infty}\int_1^x \log t dt=\log x!$$

というか、x が十分大きい時

$$\log x! \fallingdotseq \int_1^x \log t dt$$

です。長かったけど、証明終わり。

ということで、式 26 に戻ります。

$$\log \mathrm{W}(x) \fallingdotseq \log(\mathrm{n}!)-\log(x!)-\log(n-x)!+x\log(p)+(n-x)\log(q)$$

$$\fallingdotseq \log(n!)-\int_1^x \log t dt-\int_1^{n-x}\log t dt+x\log p+(n-x)\log q$$

この両辺を微分します。

$$\{\log W(x)' \fallingdotseq -[\log t]_1^x+[\log t]_1^{n-x}+\log p-\log q$$

p+q=1、log1=0 ですから

$$\{\log W(x)' \fallingdotseq -[\log t]_1^x+[\log t]_1^{n-x}+\log p-\log 1-p$$

$$\fallingdotseq -\log x+\log(n-x)+\log p-\log(1-p)$$

$$\fallingdotseq \log\frac{(n-x)p}{x(1-p)}$$

式 28

$\log 1=0$ですから、この関数が 0 になるのは、

$$\frac{(n-x)p}{x(1-p)}=1$$

の時です。これを解いて

$$(n-x)p = x(1-p)$$

$$np - xp = x - xp$$

$$x = np$$

となります、

n、xに具体的な数字を入れてみるとわかりますが、$\{\log W(x)'$は減少関数ですから、logW(x)は、$x = np$で極大になります。ということは、W(x)も x=np で極大になるということです。この場合、極大値が一つしかありませんから、最大値になります。その値になる確率が最も高い。その値が出てくる頻度が最も高いということですね。そういう値を最頻値といいます。

ところで、np とはいったい何でしょうか、これはすでに二項分布のところでやりました。試行の回数にある現象が現れる確率を掛けたものですね。例を挙げると、「正確なサイコロを振った時に、もし 1 が出たら 1 円もらえると

します。サイコロを 3 回振ったらいくらもらえることが期待できますか。」というような問題の時に

$$3 \times \frac{1}{6} = \frac{1}{2}$$

と計算しますが、この例では、nが 3 で、n が$\frac{1}{6}$ npが$\frac{1}{2}$ということです。つまり、ある確率で起こる現象があって、それが現れるかどう n 回試した時に、何回現れるかを予想した値です。これを期待値と言います。普通、期待値は μ という記号で表します。実際のデーターから μ を予想するときは、データーの平均値$\bar{x}$をその予測値とします。

つまり、

$$\mu = np$$

です。

あることが起こるということとあることが起こらない、いいかえればお互いに同時に起こることがない事象ですが、これを反事象と言います。起こらない確率をqとすると、$p + q = 1$です。また、起こらない回数を z とすると、$n = x + z$

です。そこで、$p = 1 - q$、$x = n - z$ を式 28 にいれます。

$$\frac{(n-x)p}{x(1-p)} = 1$$

$$\frac{(n-(n-z))(1-q)}{(n-z)(1-(1-q))} = 1$$

$$\frac{z(1-q)}{(n-z)q} = 1$$

右辺が 1 だから、左辺の分母・分子を入れ替えて

$$\frac{(n-z)q}{z(1-q)} = 1$$

つまり、式 28 と同じ形になって、

$$z = nq$$
$$\mu = nq$$

となります。右から見ても左から見ても、式の形は同じということですね。また、もともと２項分布なのですから、p を一定にして n を大きくしていけば、左右対称に近づきます。N を無限大にすれば、その分布の形も左右平等です。
つまり、期待値＝最頻値＝中央値ということです。（これは２項分布の性質でもありますね。）

次にもう一回微分します。式 28 の 1 段階前の形で微分したほうが定数項が対数の外側に出ているので計算しやすいですね。

$$\{\log W(x)\}'' \fallingdotseq \{-\log x + \log(n-x) + \log p - \log(1-p)\}'$$
$$\fallingdotseq \{-\log x\}' + \{\log(n-x)\}'$$
$$\fallingdotseq -\frac{1}{x} - \frac{1}{n-x}$$

この１回と２回の微分式から、どこか一点の微分値を求めて、それを使って Taylor 展開をしたいのです。今、わかっているのは、x=np で$\{\log W(x)\}' = 0$ということです。
これを利用したいので、x=np の時の$\{\log W(x)\}''$を求めます。

$$\{\log W(x)\}'' \fallingdotseq -\frac{1}{x} - \frac{1}{n-x}$$

$$\{\log W(np)\}'' \fallingdotseq -\frac{1}{np} - \frac{1}{n-np}$$

$$\fallingdotseq -\frac{1}{n}\left(\frac{1}{p} + \frac{1}{1-p}\right)$$

$$\fallingdotseq -\frac{1}{np(1-p)}$$

ところで、式 14 で示したように、二項分布では$np(1-p) = \sigma^2$ですから

$$\{\log W(x)\}'' \fallingdotseq -\frac{1}{\sigma^2}$$

こでで、式 23 を Taylor 展開します。
Taylor 展開を知っていることは、全体を理解するために必ずしも必要ではないのですが、何をやっているのかわからないと、いきなり式が書き換えられたような気がして、話についていきにくくなります。Taylor 展開とは、複雑な式を分かりやすい多項式の式に近似的

に変換するテクニックです。与えられた式を何回か微分して、それらの微分式を別々の項として足し合わせる形に近似して式を扱いやすい形に変形します。Taylor 展開が何か知りたい人は、3-4-2.Taylor 展開を読んでください。
Taylor 展開の中身を知らなくても、Taylor 展開とは、式を何回か微分して、微分したものの和の形で式を近似的に簡略化することだと理解してください。ここでは２回まで微分します。

幸い私たちは、式 23、$\mathrm{logW}(x) = \log(\mathrm{n}!) - \log(x!) - \log(\mathrm{n}-x)! + \mathrm{k}\log(\mathrm{p}) + (\mathrm{n}-x)\log(\mathrm{q})$ の１回微分と、２回微分の結果を知っています。どこかのxの値の近傍で考えるならば、三回微分以降の式の値は十分小さいので無視できます。
ということで、二回微分の項まで、式 23 を Taylor 展開します。どの値の近傍で Taylor 展開するかが問題になりますが、もっともなだらかで、変化が少ないと考えられるところが良いでしょう。また、わかりやすいところの方が良いでしょう。そこで考えられるのは、期待値μの近傍で Taylor 展開することです。
$\log W(x)$の一回微分の x=μ における値、すなわち（logW(μ)'が 0 であることは確認ずみですね。

$$\mathrm{logW}(x) = \log(\mathrm{n}!) - \log(x!) - \log(\mathrm{n}-x)! + \mathrm{k}\log(\mathrm{p}) + (\mathrm{n}-x)\log(\mathrm{q})$$

$$\fallingdotseq \mathrm{logW}(\mu) + \frac{(\log(\mu))'}{1!}(x-\mu) + \frac{(\log(\mu))''}{2!}(x-\mu)^2$$

$$= \log W(\mu) + \frac{(\log(x))''}{2}(x-\mu)^2$$

$$= \log W(\mu) + \frac{-\frac{1}{\sigma^2}}{2}(x-\mu)^2$$

となるのですが、これを対数式でなく、もとの式に戻します。

$$\log_e e = 1$$

ですから、これをつかって次のように変形します。

著者注

ちなみに、ここで e は自然対数の底として知られるもので、数学的にはネイピア数と言います。高校の数学でネイピア数とは何かしっかりとした説明を受けていない人が多いということを最近知りました。そこで、ネイピア数についての解説を（3-4-3.ネイピア数）に書いておきました。参考にしてください。しっかり理解すると、以下の説明がわかりやすくなります。なお記号の約束事として、特に断らない限り対数$\log A$と書いたときの$\log$　は$\log_e A$のことで、対数はネイピア数を底とする自然対数だと理解してください。なお、対数の微分$\frac{d\log x}{dx} = \frac{1}{x}$、指数の微分$\frac{de^x}{dx} = e^x$は知っているものとして話を進めます。これがわからない人は（3-4-3.ネイピア数）を読んでください。

$$\log W(x) \fallingdotseq \log W(\mu) + \frac{-1}{2\sigma^2}(x-\mu)^2$$

$$= \log_e W(\mu) + \frac{-1}{2\sigma^2}(x-\mu)^2 \log_e e$$

$$= \log_e W(\mu) + \log_e e^{\frac{-1}{2\sigma^2}(x-\mu)^2}$$

$$= \log_e W(\mu) + \log_e e^{\frac{-1}{2\sigma^2}(x-\mu)^2}$$

$$= \log_e W(\mu) e^{\frac{-1}{2\sigma^2}(x-\mu)^2}$$

となるので、対数の中だけを考えれば、

$$\mathrm{W(x)} \fallingdotseq \mathrm{W(\mu)} e^{\frac{-1}{2\sigma^2}(x-\mu)^2}$$

式 29

となります。数学の答えとしてはこれで良いのかもしれませんが、これではあまりよく意味がわからないし、正規分布として私たちが知っている式とも表現の仕方が違います。
式に含まれている W(μ)はμが与えられば一定の値として定数になるはずですが、これがどのような値なのかは少なくとも知りたいところです。そこで、何らかの条件を与えて、
W(μ)
の値を求めることを考えます。すぐに気が付く条件は、この式は確率分布の式なのだから、その面積の総和は 1 ということです。つまり-∞から∞まで積分すれば、その値は 1 になるということです。
ですから、W(μ) = Aとして A について以下の式を解けばよいことになります。

$$1 = \int_{-\infty}^{\infty} W(\mu) e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

$$= \int_{-\infty}^{\infty} A\, e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

$$= \mathrm{A} \int_{-\infty}^{\infty} e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

ここではあまり関係がないのですが、指数のカッコの中の$\left(\frac{x-\mu}{\sigma}\right)$について考えておきます。

この値は、期待値（母集団の平均値・中央値）と実際に得られたデーターを、標準偏差で割ったものですね。つまり、μを起点(0)としたときに、μからデーターx までの距離を標準偏差σを 1 単位として表したものです。つまり、正規分布するデーターをそのばらつきの大きさにかかわらず、標準化して表すときの距離ということになります。

そういうことも意識しながら、

$$\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)^2 = X^2$$

と置いて、式を単純化します。

$$\mathrm{X} = \frac{x-\mu}{\sqrt{2}\sigma}$$

両辺を x で微分すると

$$\frac{dX}{dx} = \frac{1}{\sqrt{2}\sigma}$$

計算の便宜上、$\mathrm{dx} = \sqrt{2}\sigma\mathrm{dX}$と分離できるものとして、

式 28　$\mathrm{A}\int_{-\infty}^{\infty} e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$を

$$\mathrm{A}\int_{-\infty}^{\infty} e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \mathrm{A}\int_{-\infty}^{\infty} e^{-X^2}\sqrt{2}\sigma dX$$

$$= \sqrt{2}\sigma A\int_{-\infty}^{\infty} e^{-X^2} dX \qquad \text{式 30}$$

と変形します。これは分布の中心を 0 として σ を単位にした距離に変換する標準化のための作業です。

つまり、この問題は、

$$\int_{-\infty}^{\infty} e^{-X^2} dX$$

の答えを出す問題という問題に還元されます。

答えを先に言うと

$$\int_{-\infty}^{\infty} e^{-X^2} dX = \sqrt{\pi}$$

です。

一般の証明で、変数を X と書くのはあまり一般的でないので、変数をxと表して説明します。

$$I = \int_{-\infty}^{\infty} e^{-x^2} dx$$

原点を中心に左右対称なので、

$$\frac{I}{2} = \int_{0}^{\infty} e^{-x^2} dx$$

突然ですが、ここで、両辺を二乗します。

$$\frac{I^2}{4} = \int_{0}^{\infty} e^{-x^2} dx \times \int_{0}^{\infty} e^{-x^2} dx$$

右辺の１番目の定積分と２番目の定積分を区別して別々に計算するものとして、２番目の

定積分の変数をｙとして書き換えます

$$\frac{I^2}{4} = \int_0^\infty e^{-x^2}\,dx \times \int_0^\infty e^{-y^2}\,dy$$

これは積分したものの掛け算なのですが、ｘとｙが互いに独立で直行しているとすれば、積分したものを掛け合わせることと、重積分することは同じ結果になります。

$$\int_0^\infty e^{-x^2}\,dx \times \int_0^\infty e^{-y^2}\,dy = \int_0^\infty \int_0^\infty e^{-x^2} \times e^{-y^2} dxdy$$

ということです。

わざわざ２つの確率分布の掛け算の形にして複雑化しています。ここだけで考えると、この作業は正規分布の記述を簡略化するためなのですが、もう少し深く考えると、この作業によって確率分布同士の掛け算として分散を確率的に扱うことを可能にしていると言えます。作業の内容は、立体空間にできた確率分布を、新たに作った一つの軸で説明できるように書きなおすという作業です。作業のプロセスは立体空間の体積を記述する式を作ること、その立体に新たな座標軸を作って、その座標軸での積分で体積を表現できるようにすることです。座標変換については、別に項を設けて説明したのでそちらを参照してください（3-4-4. 畳み込み（重積分と座標変換）。

畳み込みを使って、式を変形していく流れだけを追います。

$$\int_0^\infty e^{-x^2}\,dx \times \int_0^\infty e^{-y^2}\,dy = \int_0^\infty \int_0^\infty e^{-x^2} \times e^{-y^2} dxdy$$

$$= \int_0^\infty \int_0^\infty e^{-(x^2+y^2)}\,dxdy$$

何をやっているのかというと、２つの変数の積分の積を、２つの変数の積の積分として表し、それを $x=r\cos\theta$、$y=r\sin\theta$という極座標に変換して、θで積分するため、無理やり、x^2+y^2を作っているのです。

それができたので、あらためて

$$x = r\cos\theta$$

$$y = r\sin\theta$$

で極座標変換すると

$$\frac{I^2}{4} = \int_0^\infty \int_0^\infty e^{-(x^2+y^2)}\,dxdy = \int_0^{\frac{\pi}{2}} \int_0^\infty e^{-r^2}\,rdrd\theta$$

式 31

まず、内側の積分$\int_0^\infty e^{r^2}\,rdr$について

$$r^2 = s$$

とおいて

$$2\mathrm{r} = \frac{ds}{dr}$$

$$rdr = \frac{1}{2}ds$$

$$\int_0^\infty e^{-r^2}\, rdr = \int_0^\infty e^{-s}\frac{1}{2}ds$$

$$= \frac{1}{2}\int_0^\infty e^{-s}\, ds$$

$$= \frac{1}{2}[-e^{-s}]_0^\infty$$

$$= \frac{1}{2}\left\{-\frac{1}{e^\infty} - \left(-\frac{1}{e^0}\right)\right\}$$

$$= \frac{1}{2}\{0 - (-1)\}$$

$$= \frac{1}{2}$$

式 31 に戻って

$$\frac{I^2}{4} = \int_0^{\frac{\pi}{2}}\int_0^\infty e^{-r^2}\, rdrd\theta$$

$$= \int_0^{\frac{\pi}{2}}\int_0^\infty e^{-s}\frac{1}{2}ds\, d\theta$$

$$= \int_0^{\frac{\pi}{2}}\frac{1}{2}d\theta$$

$$= \frac{1}{2}\int_0^{\frac{\pi}{2}}d\theta$$

$$= \frac{1}{2}[\theta]_0^{\frac{\pi}{2}}$$

$$= \frac{1}{2}\left(\frac{\pi}{2} - 0\right)$$

$$= \frac{\pi}{4}$$

したがって、

$$\frac{I^2}{4} = \frac{\pi}{4}$$

$$I^2 = \pi$$

$$\mathrm{I} = \sqrt{\pi}$$

ところで、制約条件は

$$\int_{-\infty}^{\infty} A\, e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = 1$$

です。

式 29, 30 に戻ると

$$\mathrm{A}\int_{-\infty}^{\infty} e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \sqrt{2}\sigma A \int_{-\infty}^{\infty} e^{-X^2}\, dX$$

$$= \sqrt{2}\sigma AI$$

$$= \sqrt{2}\sigma\sqrt{\pi}A$$

$$= \sqrt{2\pi}\sigma\mathrm{A}$$

したがって

$$\sqrt{2\pi}\sigma A = 1$$

$$\mathrm{A} = \frac{1}{\sqrt{2\pi}\sigma}$$

となって、A の値が決まります。

もう忘れてしまったかもしれませんが

$$\mathrm{W}(\mu) = \mathrm{A}$$

としたのでしたね。

そこで、式 28 にこの値を戻して

$$\mathrm{W}(x) \fallingdotseq \mathrm{W}(\mu) e^{\frac{-1}{2\sigma^2}(x-\mu)^2}$$

$$\mathrm{W}(x) \fallingdotseq \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-1}{2\sigma^2}(x-\mu)^2}$$

$$\mathrm{W}(x) \fallingdotseq \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

これで、一般ン知られている正規分布の式ができました。

平均がμ、分散がσ^2の正規分布を$\mathrm{N}(\mu, \sigma)$と書きあらわします。$\mathrm{N}(0, 1)$の正規分布を標準正規分布と言います。

xが$\mathrm{N}(\mu, \sigma)$に従うとき

$$\mathrm{P}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

式 32

確かにnが十分に大きければ、２項分布は正規分布に近づくことが示されました。正規分布は二項分布の極限だと説明されることが多いと思います。確かにそうなのですが、それ

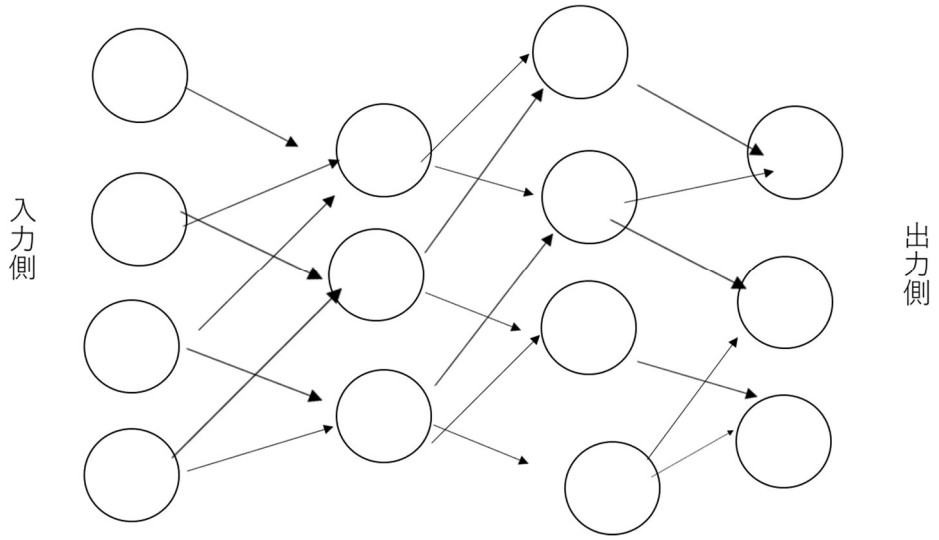
以上に、２個の同じ二項分布を重ね合わせて４つ折りすることによって、確率の式の中に、分散という中心からの距離の尺度を持ち込んだことが大きいと思います。畳込の操作を理解すると、この感覚が納得できると思います。

この過程では、第一段階で、対数化したうえで、最頻値（＝母集団の平均値μ）の近辺で、二次までの Taylor 展開を個なって変形します。この時、μを中心とした、左右相称の形になって、ネイピア数 e が持ち込まれます。第二段階では、その式を積分して、確率の総和が１という条件を使って、積分関数を求めます。その過程で、２つの積分の積の形にして、それぞれを極座標変換しているので、正規分布の定数項にπが入っているのです。いくつかの解析数学的なテクニックを使っているので、そのすべてを直ちに理解するのは難しいと思いますが、解説を読み直せば何をしているのか思い出せるという程度に記憶すればよいと思います。それが出来なければ、統計解析が理解できないということはありません（少なくとも私は、説明は理解できるけど、自分、一人でやってみろと言われるとできない。でも、それで困ったことはない。）

結構、重要かもしれない蛇足

ここで、使っている２項分布というモデルから、現実社会のデータ分布の形として正規分布を作るという考え方はもっともらしいのですが。落ち着いて考えると怪しげです。ある事象（表現型）にかかわるたくさんの要素があって、それらの要素が与える効果がほぼ等しいという仮定は現実的でしょうか。多くの要因がかかわる事象では、その要因のかかわり方はそれぞれに違うと思います。マイナーな要因は無視すればよいという言い逃れに対しても、それでは、メジャーな要因間でその影響は等しいのかと言われると、黙らざるを得ない。背骨のたとえ話を使うと、椎骨の大きさは、頸椎と胸椎と腰椎では違います。特に下の方の腰椎は大きいから、その影響は頸椎や腰椎より大きいでしょう。これはたとえ話だから、椎骨という単位で考えずに、別の単位で考えれば良いということもできますが、いつでもそれが可能とは限りません。もし、少数の要因がとても大きな影響を持って知多場合には、ヒストグラムは単峰形にならずに、双峰形、多峰形になると思います。データのひしとグラムを作って、単峰形になるかならないか、正規分布から大きく外れる分布になっていないかどうかを確かめることが、正規分布を使った統計処理をするために必要です。もし、正規分布的でないならば。大きな要因のそれぞれごとにデータを取り分けて、それぞれで、正規分布を前提とした統計処理をすべきです。多要因分散分析（ANOVA）は、正規分布を前提として、いくつかの要因を独立的（直交的）に統計分析する方法です。差の統計の後、F 検定を学んだ後に、その応用として多要因分散分析の方法を説明します。次に、二項分布の前提になっている考え方、個々の要因は相加的に機能するという考え方も、現実的かどうかわかりません。たとえば、生物学では表現型 P（背の高さとか、賢さ

とか、足の速さとか）は遺伝 G と環境 E の二つが関与しているとして、$P = G + E$と習います。G と E は相加的でしょうか。私は相乗的つまり$P = G \times E$だと思います。そう考えると、G の中のにも、言語理解力とか、空間認識とか、数理的理解力とか、好奇心、警戒心等々いろいろあって、こちらは相加的かもしれない。E の方にも、栄養、気候、教育、疾病等々あって、こちらは相加的なのか相乗的なのかわからない。単に相乗的に働く要因だけであれば、$P = G \times E$の対数をとって、$\log P = \log G + \log E$として、相加的に計算すれば良いかもしれない。しかし、相加的な要因と相乗的な要因が混合していたらどうすべきか。そこは、想像量も含めて様々な考え方がありそうです。とりあえず、独立していと考えられる要因を ANOVA で分析し、それぞれの、影響の程度を明らかにしたうえで、重相関ン持って行くとか、対数変換後の重相関に持って行くとか考えられそうです。もし、相加的なものと創造的なものが混合しているのならば、階層的に２段階のネットワークを作って、一段階目は相加的で、２段階目を相乗的にするとかありそうですし、コンピュータの計算力を最大限生かして、ニューラル・ネットワークモデルのような、多段階の階層的な相関関係を、同時に最適化化するようなことを考えても良いでしょう。



ニューラルネットワークのイメージ