3. 確率

このセクションでは偶然性を示す**確率**のあり方とその計算の仕方を扱います。はじめに簡単な二項分布確率を見て，それを正規分布確率にまで一般化します。次に，乱数実験から得られた確率と正規分布確率の分布が近似することを確認し，連続量の確率を累積正規分布から計算する準備をします。

3.1. 確率の分布

3.1.1. 確率変数

次の表は実際に 1 つのサイコロを 10 回投げて,それぞれの目が出た硬貨の数(X)とその頻度数(F)を示します。

サイコロの目(X)	1	2	3	4	5	6	和(S)
実際の頻度(F)	2	2	1	0	3	2	10

このような表は**度数分布表**(frequency distribution)とよばれます。一方，頻度数ではなくて，それぞれの頻度(F)を和(S)で割ると，それぞれの実験の確率(P)が計算されます。次のようにそれぞれを確率で示す表は**確率分布表**(probability distribution)とよばれます。

サイコロの目(X)	1	2	3	4	5	6	和(S)
実際の確率(P)	2/10	2/10	1/10	0/10	3/10	2/10	1

この確率を理論的に求めるならば次のようになるはずです。たとえば X=1 のときの確率は P(X=1)のように書かれます。

$$P(X=1) = P(X=2) = \ldots = P(X=6) = 1/6$$

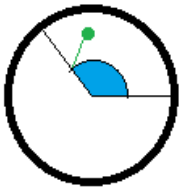
よって，実際の生起数に基づく確率ではなく，理論的な確率分布表は次になります。

サイコロの目(X)	1	2	3	4	5	6	和(S)
理論的な確率(P)	1/6	1/6	1/6	1/6	1/6	1/6	1

このように，実際と理論の微妙な違いはよくあることですが，それでも 2 枚の硬貨を投げる回数を多くすれば，その実験の実際の確率分布表は，

理論的な確率分布表に近づいていくはずです。なお，サイコロの{1, 2, 3, 4, 5, 6}の目のように，数字が区切れて並ぶような変数の確率は**離散的確率変数**とよばれます。

さらに，次のような円盤の上に一本の針（ピン）を落とし，その針先の位置が示す角度（円盤のある点をゼロ(0)としておきます）を測り，それぞれの角度が示す値(X)の確率を求めることを考えましょう。



ルーレットの円周は 40 個ぐらいの升目に区切られていますが，ここでは角度を正確に測ることを考えます。この角度の値は 0（0 を含める)から 360(360 を含めない)まで連続的であり，小数点以下まで求めれば，その精度は無限にあります。このような**連続的な範囲**は[0, 360)のように書かれます。角括弧「[」は「含める」，丸括弧「)」は「含めない」という意味です。

このような連続的な変数の場合，特定の 1 つの数値に対応する確率を計ることは，それぞれが必ず 1 回の度数になり，全体の範囲内にある数値の数は無限（∞）ですから，その確率は P = 1 / ∞ = 0 になってしまいます。しかし，たとえば[0 ~ 60)の範囲にある確率ならば，離散的な確率とおなじように想定できます。このような**連続的確率変数**の確率は P(0≤X<b)のように書かれます。次が連続的確率変数の確率分布表です。

X	[0, 60)	[60, 120)	[120, 180)	[180, 240)	[240, 300)	[300, 360)	和
P	1/6	1/6	1/6	1/6	1/6	1/6	1

3.1.2. 平均と分散

データの中心を示す**平均**と，データの散らばり具合を示す**分散**は数値データを統計的に扱うときに重要な指標です。このことは頻度分布のデータだけでなく，確率分布のデータでも同じです。このセクションでは，頻度分布の平均・分散から出発して確率分布の平均・分散を理解し，その重要な性質を確認します。

はじめに次のような簡単な**数値分布**の平均と分散を求めます。

d	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x	1	1	5	5	5	3	3	3	3	3	4	4	4	4	6

この平均(m)は

$$m = (\Sigma_i x_i) / n$$
$$= (1 + 1 + 5 + \ldots + 6) / 15 = 54 / 15 = 3.6$$

そして分散(v)は

$$v = \Sigma_i (x_i - m)^2 / n$$
$$= (1 - 3.6)^2 + (1 - 3.6)^2 + (5 - 3.6)^2 + \ldots + (6 - 3.6)^2] / 15$$
$$= 1.844$$

なお，次の**分散の別式**もよく使われます。

$$v = \Sigma_i x_i^2 / n - m^2$$

証明：

$$v = \Sigma_i (x_i - m)^2 / n$$
$$= \Sigma_i (x_i^2 - 2 m x_i + m^2) / n$$
$$= (\Sigma_i x_i^2 - \Sigma_i 2 m x_i + \Sigma_i m^2) / n$$
$$= (\Sigma_i x_i^2 - 2 m \Sigma_i x_i + \Sigma_i m^2) / n$$
$$= \Sigma_i x_i^2 / n - 2 m \Sigma_i x_i / n + \Sigma_i m^2 / n$$
$$= \Sigma_i x_i^2 / n - 2 m^2 + \Sigma_i m^2 / n$$
$$= \Sigma_i x_i^2 / n - 2 m^2 + n m^2 / n$$
$$= \Sigma_i x_i^2 / n - 2 m^2 + m^2$$
$$= \Sigma_i x_i^2 / n - m^2$$

この別式 $v = \Sigma_i x_i^2 / n - m^2$ は「分散＝2 乗の平均－平均の 2 乗」であることを示しています。

次に，それぞれのデータに頻度(f)があることを示す**度数分布表**のデータの平均(m)と偏差(v)を求めましょう。

data	1	2	3	4	5	和
x	1	5	3	4	6	19
f	2	3	5	4	1	15

$$m = (\Sigma_i x_i * f_i) / \Sigma_i f_i$$
$$= [(1 * 2) + (5 * 3) + (3 * 5) + (4 * 4) + (6 * 1) / 15$$
$$= (2 + 15 + 15 + 16 + 6) / 15$$
$$= 54 / 15 = 3.6$$

$$v = [\Sigma_i (x_i - m) * f_i] / \Sigma_i f_i$$
$$= [(1 - 3.6)^2 * 2 + (5 - 3.6)^2 * 3 + (3 - 3.6)^2 * 5 + (4 - 3.6)^2 * 4$$
$$+ (6 - 3.6)^2 * 1] / 15$$
$$= [(2.6)^2 * 2 + (1.4)^2 * 3 + (0.6)^2 * 5 + (0.4)^2 + (2.4)^2 * 1] / 15$$
$$= [6.76 * 2 + 1.96 * 3 + 0.36 * 5 + 0.16*4 + 5.76 * 1] / 15$$
$$= [13.52 + 5.88 + 1.8 + 0.64 + 5.76] / 15$$

= 27.6 / 15 = 1.84

さらに，次は同じデータの確率分布表です。

data	1	2	3	4	5	和
x	1	5	3	4	6	19
p	2/15	3/15	5/15	4/15	1/15	1

$m = \Sigma_i x_i * p_i$
$= [(1 * 2 / 15) + (5 * 3 / 15) + (3 * 5 / 15) + (4 * 4 / 15) + (6 * 1 / 15)$
$= (2 + 15 + 15 + 16 + 6) / 15$
$= 54 / 15 = 3.6$

$v = \Sigma_i (x_i - m)^2 * p_i$
$= [(1 - 3.6)^2 * 2 / 15 + (5 - 3.6)^2 * 3 / 15 + (3 - 3.6)^2 * 5 / 15$
$+ (4 - 3.6)^2 * 4 / 15 + (6 - 3.6)^2 * 1] / 15$
$= [(2.6)^2 * 2 + (1.4)^2 * 3 + (0.6)^2 * 5 + (0.4)^2 + (2.4)^2 * 1] / 15$
$= [6.76 * 2 + 1.96 * 3 + 0.36 * 5 + 0.16*4 + 5.76 * 1] / 15$
$= [13.52 + 5.88 + 1.8 + 0.64 + 5.76] / 15$
$= 27.6 / 15 = 1.84$

このように，同じ原データは度数分布でも確率分布でも平均と分散が同じ結果になることを確認して，以下では確率分布の式を使います。

確率分布の平均と分散を計算するときに便利な次の E(X), V(X)の式が使われます。平均(m)は**期待値**(expectation, expected value: E)とも呼ばれます。

$$m = E(X) = \Sigma_i x_i * p_i = 1/n \ \Sigma_i x_i$$
$$v = V(X) = \Sigma_i (x_i - m)^2 * p_i = 1/n \ \Sigma_i (x_i - m)^2$$

ここで $p_i = P(X = x_i)$は確率変数(X)が x_i のときの確率，m はデータの平均，n はデータの個数を示します。V(X)を期待値(E)で示すと

$$V(X) = E[(X - m)^2] = \Sigma_i (x_i - m)^2 * p_i$$

になることは，先の $E(X) = \Sigma_i x_i * p_i$ の x_i を$(x_i - m)^2$ に置き換えれば理解できます。$V(X) = E[(X - m)^2]$を言葉で表現すると「分散 V(X)は偏差の 2 乗 $(X - m)^2$ の平均(期待値)である」ということになります。

また，先に見た分散の別式 $v = \Sigma_i x_i^2 / n - m^2$ 「分散＝2 乗の平均－平均の 2 乗」を V を使って示すと次のようになります。

$$V(X) = V(X) = E(X^2) - [E(X)]^2 \quad \text{（分散の別式）}$$

確率分布の平均 E には次の性質があります。これらは重要な性質です。

E の性質(1)：$E(X + Y) = E(X) + E(Y)$, $E(X - Y) = E(X) - E(Y)$

E の性質(2)：$E(aX + b) = a\,E(X) + b$

$(a = 0)$：$E(b) = b$

$(b = 0)$：$E(a\,X) = a\,E(X)$

E の性質(3)：$E[E(X)] = E(X)$

E の性質(4)：$E(XY) = E(X)\,E(Y)$　[X, Y が独立のとき]

証明-1a：E の性質(1a)：$E(X + Y) = E(X) + E(Y)$

$E(X + Y) = \Sigma_i (x_i + y_i)\, p_i$　←$E(X) = \Sigma_i x_i\, p_i$

$= \Sigma_i (x_i\, p_i + y_i\, p_i)$　←p_i を分配

$= \Sigma_i x_i\, p_i + \Sigma_i y_i\, p_i$　←Σ_i を分配

$= E(X) + E(Y)$　←$E(X) = \Sigma_i x_i\, p_i$

証明-1b：E の性質(1b)：$E(X - Y) = E(X) - E(Y)$

$E(X - Y) = \Sigma_i (x_i - y_i)\, p_i$　←$E(X) = \Sigma_i x_i\, p_i$

$= \Sigma_i (x_i\, p_i - y_i\, p_i)$　←p_i を分配

$= \Sigma_i x_i\, p_i - \Sigma_i y_i\, p_i$　←Σ_i を分配

$= E(X) - E(Y)$　←$E(X) = \Sigma_i x_i\, p_i$

証明-2：E の性質(2)：$E(aX + b) = a\,E\,(X) + b$

$E(aX + b) = \Sigma_i (a\, x_i + b)\, p_i$　←$E(X) = \Sigma_i x_i\, p_i$

$= \Sigma_i (a\, x_i\, p_i + b\, p_i)$　←p_i を分配

$= \Sigma_i a\, x_i\, p_i + \Sigma_i b\, p_i$　←Σ_i を分配

$= a\ \Sigma_i x_i\, p_i + b \Sigma_i p_i$　←a, b を前に

$= a\,E\,(X) + b$　←$E(X) = \Sigma_i x_i\, p_i$; [1]　$\Sigma_i p_i = 1$

証明-3：E の性質(3)：$E[E(X)] = E(X) = m$

$E[E(X)] = E(m)$　←$E(X) = m_i$

$= 1/n\ \ \Sigma_i m$　←$E(m) = 1/n\ \ \Sigma_i m$

$= 1/n\ n\ m$　←$\Sigma_i m = n\ m$

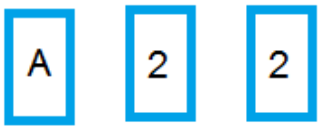
$= m$　←$1/n\ n = 1$

$= E(X)$　←$m = E(X)$

証明-4：E の性質(4)：$E(XY) = E(X)\,E(Y)$ [X, Y が独立のとき]

↓以下で詳しく見ます。

ここにトランプのエース(A)1 枚と，「2」のカード 2 枚を次のように用意します。



これを裏返してランダムに 1 枚引いたときのカードを X とします。次にそのカードを戻して，もう 1 枚引いたときのカードを Y とします。そうすると X と Y は無関係（独立）になります（独立でない場合→後述：「非独立の確率変数」）。そのときの同時確率分布は次の表で示されます。

X:Y	Y =「A」	Y =「2」	和
X =「A」	1/3 * 1/3 = 1/9	1/3 * 2/3 = 2/9	1/9 + 2/9 = 3/9 = 1/3
X =「2」	2/3 * 1/3 = 2/9	2/3 * 2/3 = 4/9	2/9 + 4/9 = 6/9 = 2/3
和	1/9 + 2/9 = 3/9 = 1/3	2/9 + 4/9 = 6/9 = 2/3	1

たとえば 1 枚目が「A」であり，2 枚目が「2」であるときの確率 P(X=「A」, Y=「2」)が 1/3 * 2/3 =2/9 になることは理解できます。そして，この確率が X=「A」の確率 P(X=「A」)を示す（横）和 1/3 と，Y=「2」の確率 P(Y=「2」)を示す（縦）和 2/3 の積になることを確認します。ほかのマスに対応する P(X=「A」, Y=「A」)=1/3 * 1/3 =1/9，P(X=「2」, Y=「A」)2/3 * 1/3 =2/9，P(X=「2」, Y=「2」)2/3 * 2/3 = 4/9 についても同様です。

よって

$$E(XY) = E(X)\,E(Y) \quad [X, Y：独立]$$

ここで X, Y はそれぞれの確率変数を示します。積 E(XY)はそれぞれのマスにある積算を示し，E(X) E(Y)は確率の行和と列和の積を示します。

以上はカードの種類が「3」，「4」，…のように増えても同じです。そこで一般化して，次の(X, Y)の確率分布を見ます（X, Y：独立）。

X:Y	y_1	y_2	…, y_j	和
x_1	p_{11}	p_{12}	…	$p_{1\cdot}$
x_2	p_{21}	p_{22}	…	$p_{2\cdot}$
…, x_i	…	…	…	…
和	$p_{\cdot 1}$	$p_{\cdot 2}$	…	1

$$E(X\,Y) = \Sigma_i\ \Sigma_j\ x_i\ y_j\ p_{ij} \quad \leftarrow \text{表の } p_{11}, p_{12}, \ldots, p_{np} \text{ を個別に足す}$$
$$= \Sigma_i\ \Sigma_j\ x_i\ y_j\ p_{i\cdot}\ p_{\cdot j} \quad \leftarrow \text{表の行と列をまとめて全部足す}$$
$$= \Sigma_i\ x_i\ p_{i\cdot}\ \ \Sigma_j\ y_j\ p_{\cdot j} \quad \leftarrow \text{上の 2 つの表を参照}$$
$$= E(X)\,E(Y)$$

実験-1：E の性質(1a)：E(X + Y) = E(X) + E(Y)

data: X	1	2	3	4	5	和
x	1	5	3	4	6	19
f	2	3	5	4	1	15
x*f	2	15	15	16	6	54
p	0.133	0.200	0.333	0.267	0.067	1.000
x*p	0.133	1.000	1.000	1.067	0.400	3.600

data: Y	1	2	3	4	5	和
y	3	2	5	1	2	13
f	2	3	5	4	1	15
y*f	6	6	25	4	2	43
p	0.133	0.200	0.333	0.267	0.067	1.000
y*p	0.400	0.400	1.667	0.267	0.133	2.867

data: X+Y	1	2	3	4	5	和
x + y	4	7	8	5	8	32
f	2	3	5	4	1	15
(x+y)*f	8	21	40	20	8	97
p	0.133	0.200	0.333	0.267	0.067	1.000
(x+y)*p	0.533	1.400	2.667	1.333	0.533	6.467

Σ (x*p) + Σ (y*p) = 3.600 + 2.867 = 6.467

Σ [(x+y)]*p = 6.467

実験-2：E の性質(4)：E(XY) = E(X) E(Y) [X, Y が独立のとき]

X↓ Y→	2	3	和
4	1	4	5
5	2	6	8
和	3	10	13

p(X,Y)	2	3	和
4	0.067	0.320	0.385
5	0.167	0.450	0.615
和	0.231	0.769	1.000

X*Y*p(X,Y)	2	3	和
4	0.5	3.8	4.4
5	1.7	6.8	8.4
和	2.2	10.6	12.8

E(XY) = 4*2*p11 + +4*3*p12 + 5*2*p21 + 5*3*p22 = 12.8

E(X) * E(Y) = (4*.385 + 5 * .615) * (2*.231 + 3*.769) = 12.781

確率分布の分散 $V(X) = E[(X - m)^2]$ には次の性質があります。この分散の性質も重要です。

V の性質(1)：	$V(X) = E(X^2) - [E(X)]^2$
V の性質(2)：	$V(aX + b) = a^2 V(X)$　　←a^2 と b=0 に注意
V の性質(3)：	$V(X + Y) = V(X) + V(Y)$　　[X, Y：独立]
V の性質(4)：	$V(X - Y) = V(X) + V(Y)$　　[X, Y：独立] ←「+」に注意

それぞれを以下のように導きます。

証明-1：V の性質(1)：$V(X)= E(X^2) - [E(X)]^2$

$V(X) = E[(X - m)^2]$　　←分散の定義
$= E(X^2 - 2 m X + m^2)$　　←かっこ(...)内を展開
$= E(X^2) - 2 m E(X) + m^2$　　←$E(X + Y) = E (X) + E (Y)$
$= E(X^2) - 2 m^2 + m^2$　　←$m = E(X)$
$= E(X^2) - m^2$　　← $- 2 m^2 + m^2 = - m^2$
$= E(X^2) - [E(X)]^2$　　←$m = E(X)$

証明-2：V の性質(2)：$V(aX + b) = a^2 V(X)$

$V(aX + b)$
$= E\{ [aX + b - E(a X + b)]^2 \}$ ←$V(X) = E[(X - E(X))^2]$の X に aX+b を代入
$= E\{ [aX + b - (a E(X) + b)]^2 \}$　　←$E(aX) = a E(X)$ (E の性質)
$= E\{ [aX + b - a E(X) - b]^2 \}$ ←(...)を外す
$= E\{ [aX - a E(X)]^2 \}$　　←b を消去
$= E(aX - a m)^2$　　←$E(X) = m$
$= E[a^2(X - m)^2]$　　←a を 2 乗して前にくくる
$= a^2 E(X - m)^2$　　←$E(aX) = a E(X)$ (E の性質)
$= a^2 V(X)$　　←$V(X) = E(X - m)^2$ (定義)

証明-3：V の性質(3)：$V(X + Y) = V(X) + V(Y)$[X, Y：独立]

$V(X + Y)$
$= E[(X + Y)^2] - [E(X + Y)]^2$
　　←$V(X) = E(X^2) - [E(X)]^2$ (分散の別式)の X に X+Y を代入
$= E(X^2 + 2XY + Y^2) - [E(X + Y)]^2$　　←展開
$= E(X^2) + 2E(XY) + E(Y^2) - [E(X) + E(Y)]^2$　←E を配分
$= E(X^2) + 2E(XY) + E(Y^2) - \{[E(X)]^2 + 2 E(X) E(Y) + [E(Y)]^2\}$　　←展開
$= E(X^2) - [E(X)]^2 + 2[E(XY) - E(X)E(Y)] + E(Y^2) - [E(Y)]^2\}$　　←整理
$= V(X) + 2[E(XY) - E(X)E(Y)] + V(Y)$　　←$V(X) = E(X^2) - [E(X)]^2$
$= V(X) + 2[E(X)E(Y) - E(X)E(Y)] + V(Y)$　　←$E(XY) = E(X)E(Y)$
$= V(X) + V(Y)$　　← (X, Y: 独立)

(4) $V(X - Y) = V[X + (-1)\,Y]$

$= V(X) + V[(-1)\,Y]$ ←$V(X+Y) = V(X) + V(Y)$ [X,Y 独立]

$= V(X) + (-1)^2\,V(Y)$ ←$V(aX) = a^2\,V(X)$

$= V(X) + V(Y)$

実験：V の性質(2)：$V(aX + b) = a^2\,V(X)$

data:X	1	2	3	4	5	和	平均 m
x	1	5	3	4	6	19	3.8
f	2	3	5	4	1	15	
x*f	2	15	15	16	6	54	
p	0.133	0.200	0.333	0.267	0.067	1.000	
x-m	-2.800	1.200	-0.800	0.200	2.200	0.000	
(x-m)^2	7.840	1.440	0.640	0.040	4.840		
(x-m)^2*p	1.0453	0.2880	0.2133	0.0107	0.3227	1.8800	
y=2x+3	5	13	9	11	15	53	10.6
y-m	-5.600	2.400	-1.600	0.400	4.400		
(y-m)^2	31.36	5.760	2.560	0.160	19.360		
(y-m)^2*p	4.181	1.152	0.853	0.043	1.291	7.520	

V(2X + 3) = 2^2 + V(X) = 4 * 1.88 = 7.52

＊平均 E の性質と分散 V の性質については小寺(2002:97-111)を参照しました。

●非独立の確率変数

確率変数が独立していないときは，

$$E(XY) = E(X)\,E(Y),\ V(X+Y) = V(X) + V(Y)$$

は成立しません。たとえばトランプの「エース(A)」のカード 1 枚と，「２」のカード 2 枚を次のように用意し

これを裏返してランダムに 1 枚引いたときのカードを X とします。次にそのカードを戻さないで，残る 2 枚の中からもう 1 枚引いたときのカードを Y とします。そうすると，2 回目に引くカードの確率は 1 回目に引かれたカードの種類に依存するので，X と Y は無関係（独立）ではなくなります。

たとえば，1 枚目が「A」のときの同時確率分布は次の表で示されます。

X:Y	Y =「A」	Y =「2」	和
X =「A」	1/3 * 0 = 0	1/3 * 1 =1/3	0 + 1/3= 1/3.
X =「2」	2/3 * 1/2 = 2/6 =1/3	2/3 * 1/2 = 2/6 = 1/3	1/3 + 1/3 = 2/3.
和	0 + 1/3 = 1/3	1/3 + 1/3= 2/3	1

たとえば，1 枚目(X)が「A」のときの確率 P(X=A)は 1/3 であり，そのカードを戻さないで 2 枚目が「2」であるときの確率は「A」がなくなっているので，2 枚目(Y)が「A」になる確率 P(Y=A)は 0 です。そして，「2」のカード 2 枚の中から 1 枚をとるのでそれが「2」になる確率 P(Y=2)は 1(必ず「2」になる)，よって X=「A」，Y =「2」の確率は 1/3 * 1 =1/3 になります。確率の行和も列和も先に見た独立の場合と同じになりますが，この確率は，X=「A」の確率 P(X=「A」)を示す（横）和 1/3 と，Y=「2」の確率 P(Y=「2」)を示す（縦）和 2/3 の積にはなっていないことを確認できます(1/3≠1/3 * 1/3)。ほかのマスに対応する P(X=「A」, Y=「A」)，P(X=「2」, Y=「A」)，P(X=「2」, Y=「2」)についても同様です。そこで

$$E(XY) \neq E(X)\,E(Y) \quad \text{... X, Y：非独立}$$

また，先に見たとおり

$$V(X + Y) = V(X) + V(Y) \quad \text{... X, Y：独立}$$

を証明するときの途中の式で

$$E(XY) = E(X)\,E(Y) \quad \text{... X, Y：独立}$$

を使っているので，X, Y が非独立であれば，次のようになります。

$$V(X + Y) \neq V(X) + V(Y) \quad \text{... X, Y：非独立}$$

3.1.3. 二項分布

ある事象が起こる確率にはさまざまなものがあります。たとえば，サイコロには{1, 2, 3, 4, 5, 6}という目があるので，1 回サイコロを投げるとき(「試行」trial と言います)，それぞれの目「1」「2」...が出る確率はそれぞれ 1/6 ずつです。これらの目の中の 1 つ，たとえば「1」が出る確率は 1/6 なので，逆に「1」が出ない確率は 1 - 1/6 = 5/6 です。次の表の T (True)は「1」が出ることを示し，F (False)は「1」が出ないことを示しています。確率の総和が 1 になることを確認してください(5/6 + 1/6 = 1)。

「1」	T の数	確率
T	1	1/6 ≒ 0.167
F	0	5/6 ≒ 0.833

次にサイコロを 2 回投げる場合(試行回数 n=2)を考えましょう。たとえば 1 回目が F で，2 回目が T とすると，これを F, T と書きます。4 つの場合のそれぞれの確率は，2 つのサイコロで T/F の確率の積になります[1]。この場合も確率の総和は 1 です(25/36 + 5/36 + 5/36 + 1/36 = 1)。

「1」	T の数	確率
T, T	2	(1/6) * (1/6) = 1/36 ≒ 0.028
T, F	1	(1/6) * (5/6) = 5/36 ≒ 0.139
F, T	1	(5/6) * (1/6) = 5/36 ≒ 0.139
F, F	0	(5/6) * (5/6) = 25/36 ≒ 0.694

さらに，サイコロを 3 回投げる場合(試行回数 n=3)を考えます。この場合も確率の総和は 1 になります。

「1」	T の数	確率
T, T, T	3	(1/6) * (1/6) * (1/6) = 1/216≒0.005
T, T, F	2	(1/6) * (1/6) * (5/6) = 5/216≒0.023
T, F, T	2	(1/6) * (5/6) * (1/6) = 5/216≒0.023
T, F, F	1	(1/6) * (5/6) * (5/6) = 25/216 ≒ 0.116
F, T, T	2	(5/6) * (1/6) * (1/6) = 5/216 ≒ 0.023
F, T, F	1	(5/6) * (1/6) * (5/6) = 25/216 ≒ 0.116
F, F, T	1	(5/6) * (5/6) * (1/6) = 25/216 ≒ 0.116
F, F, F	0	(5/6) * (5/6) * (5/6) = 125/216 ≒ 0.579

ここで，たとえばサイコロを 3 回投げて順番を問題にせずに，全部で 2 回「1」が出る場合(T の数=2)の確率を求めると，上の表から

「1」	T の数	確率
T, T, F	2	(1/6) * (1/6) * (5/6) = 5/216 ≒ 0.023
T, F, T	2	(1/6) * (5/6) * (1/6) = 5/216 ≒ 0.023
F, T, T	2	(5/6) * (1/6) * (1/6) = 5/216 ≒ 0.023

を合計した確率，つまり，(5/216) + (5/216) + (5/216) = 15/216≒0.069 にな

[1] 先に見たように，互いに影響しない（独立な）複数の事象の確率はそれぞれの事象の確率の積になります。たとえば，ある趣味の会に，1/2 の確率で出席する A さんと 1/3 の確率で出席する B さんの 2 人が同時に出席する確率は(1/2) * (1/3) = 1/6 になります。もし，A さんと B さんが知り合いで誘いあってこの趣味の会に出席することがあるときは，互いに独立していないので，このような確率の積を使うことができません。

ります。これは「1」(T)が 2 回出る場合の確率(5/216)を 3 倍した数です。この倍数(=3)を求めるためには，このように少ない試行回数(3 回)ならばすぐ計算できますが，それが多くなると一般式を使わなければなりません。n 回の試行で T が r 回選ばれる場合の数は ${}_nC_r$ という「組み合わせ」(combination: ${}_nC_r$)の値になります[2]。ここでは，T が 2 個で F が 1 個の組み合わせになるので ${}_3C_2$ で計算します。そこで，3 回の試行で T が 2 回出る確率は

$$ {}_3C_2\,(1/6)^2\,(5/6) = (3 * 2) / (2 * 1)\,(1/6)^2\,(5/6) = 15/216 \fallingdotseq 0.069 $$

この確率（二項確率 binomial probability: Binom）を一般式で示すと

$$ \mathrm{Binom}(x, n, p) = {}_nC_x\, p^x\, (1 - p)^{n - x} $$

ここで n はサイコロを投げた総回数（試行数），x は成功回数（T の数），p は T の確率(成功確率：1/6)，1 － p は F の確率(失敗確率：5/6)を示します。次の表は Excel 関数 Binom を使って計算した二項確率です。x が 2(T の数が 2)のときの二項確率が先に見たように，15/216 (=0.06944…)になっています。

N	3
P	0.1667
M	0.5000
V	0.4167

x↓: n=3	BinPr
0	0.57870
1	0.34722
2	0.06944
3	0.00463

二項確率は，確率 p で起こる現象が，n 回の試行中で x 回出現する確率を求めたものです。ここで「確率」という言葉が異なる意味で 2 回使われているので，前提とする確率(p)を「期待確率」(Expected Probability: EP)と呼び[3]，出現する確率を「出現確率」(Occurrent Probability: OP)と呼んで区別しましょう[4]。また，この出現確率(OP)は n 回の中で x 回だけ出現する

[2] これは互いに区別のつく 3 個の物{a, b, c}の中から任意の 2 個(= T)を取り出す場合の数と同じです。もし，取り出す順番を考えるならば，ab, ac, ba, bc, ca, cb という 6 個の場合があります。これが「順列」(permutation: ${}_nP_r$)で，${}_nP_r = n\,(n - 1)(n - 2)\,\ldots\,(n - r + 1)$. ここで，順番を考慮しなければ（「組み合わせ」${}_3C_2$），ab と ba, ac と ca，bc と cb はそれぞれ同じなので場合の数を 2 で割らなければなりません。この 2 は ${}_2P_2$ の順列(2! = 2 x 1)です。よって ${}_3C_2 = (3 * 2) / (2 * 1)$. 組み合わせ ${}_nC_r$ の一般式は

$$ {}_nC_r = {}_nP_r / r! = [n\,(n - 1)\,(n - 2)\,\ldots\,(n - r + 1)] / r! = n! / [r!\,(n - r)!] $$

[3] 「期待確率」(expected probability)は「予想確率」という用語にしたほうがわかりやすいかもしれません。しかし統計学で expectation は「予想値」ではなく「期待値」と訳されているので，ここでも「期待確率」とします。

[4] ここで生起数(x)以下の確率を累積した確率を使い，生起数に対応する個

確率です。一方，n 回の試行中で 0 回から x 回まで出現する確率を「累積確率」(Cumulative Probability: CP)と呼んで区別します。

なお，上左表では個数(N)と確率(P)のほかに，平均(M=E(X))と分散(V=V(X))も示してあります。それぞれ次のように数理的に導出されます。

X	1	0	和
P	p	1 - p	1

n 回の試行での確率変数 $X_1, X_2, \ldots, X_n$ についてのそれぞれの平均 $E(X_i)$ と，それぞれの分散 $V(X_i)$を計算します。

$$
\begin{aligned}
E(X_i) &= \Sigma(i)\, x_i\, p_i \quad \leftarrow \text{平均（期待値）の定義 (3.1.2)} \\
&= 1 * p + 0 * (1 - p) \quad \leftarrow \text{上の表} \\
&= p
\end{aligned}
$$

$$
\begin{aligned}
V(X_i) &= \Sigma(i)\, (x_i - m)^2 * p_i \quad \leftarrow \text{分散の定義(3.1.2), m:平均} \\
&= (1 - m)^2 * p + (0 - m)^2 * (1 - p) \quad \leftarrow \text{上の表} \\
&= (1 - p)^2 * p + (0 - p)^2 * (1 - p) \quad \leftarrow m = p \\
&= (1 - p)^2 * p + p^2 * (1 - p) \\
&= p * (1 - p) * (1 - p + p) \\
&= p * (1 - p)
\end{aligned}
$$

この平均と分散が試行数 n 回の X について考えると

二項分布の平均(M)：$E(X) = n\, E(X_i) = n * p$
二項分布の分散(V)：$V(X) = n * V(X_i) = n * p * (1 - p)$

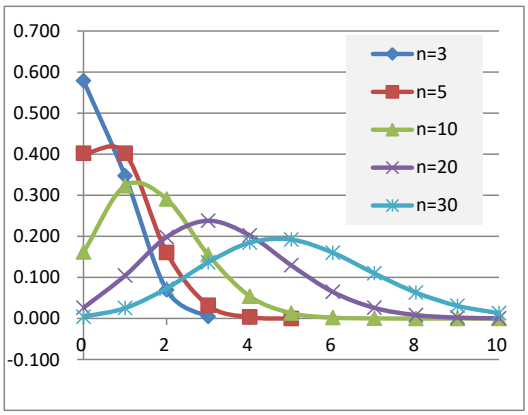
下左表は二項分布の試行数(N)，確率(P)，平均(M)，分散(V)を示します。平均は N*P，分散は N*P*(1-P)になります。下右表は試行回数を 3, 5, 10, 20, 30 にしたときの，それぞれの確率分布を示します。下図は，それらを結んだ平滑線です。試行数(N)が多くなるにつれ，左右対称の釣鐘状の分布（正規分布）に近づきます。

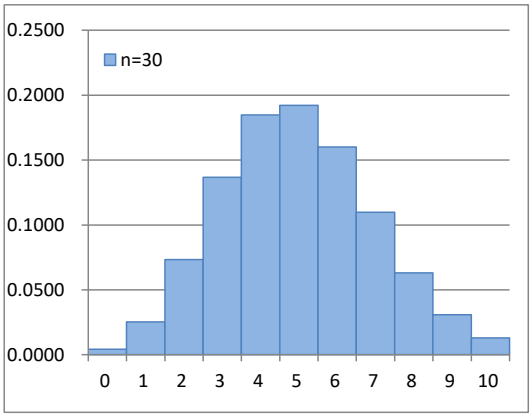
N	10
P	0.1667
M	1.6667
V	1.3889

x↓ : n→	n=3	n=5	n=10	n=20	n=30
0	0.5787	0.4019	0.1615	0.0261	0.0042
1	0.3472	0.4019	0.3230	0.1043	0.0253
2	0.0694	0.1608	0.2907	0.1982	0.0733
3	0.0046	0.0322	0.1550	0.2379	0.1368
4		0.0032	0.0543	0.2022	0.1847

別の確率そのものを使わない理由は，個別の確率はデータの個数が多くなるにつれて減少するので不都合だからです。累積生起確率は，当該の生起数以下すべてのケースの確率の総和です。つまり，その生起数以下の回数で起こる確率を示します。

5	0.0001	0.0130	0.1294	0.1921
6		0.0022	0.0647	0.1601
7		0.0002	0.0259	0.1098
8		0.0000	0.0084	0.0631
9		0.0000	0.0022	0.0309
10		0.0000	0.0005	0.0130





上左図はそれぞれの試行数(N)の確率分布を比較するために曲線で示しましたが[5]，二項確率は N が離散的なので，本来ならば上右図のようにそれぞれの N の確率を間隔のない棒グラフで示すべきです[6]。

●プログラム

次の二項分布確率の計算式には多くの階乗(!)があります。

$$Bin(x, n, p) = {}_nC_x\, p^x\, (1-p)^{n-x} \quad = n! / [x!\,(n-x)!]\, p^x\, (1-p)^{n-x}$$

階乗 x!の計算は単純ですが，x の値が大きくなると計算機はオーバーフローを起こします[7]。上式が示すように H の計算では巨大な数を巨大な数で割った結果，最終的には[0, 1]の範囲になるので，分子と分母の計算をす

[5] Excel：折れ線を選択→右クリック→データ系列の書式設定→線のスタイル→スムージング

[6] Excel：棒グラフを選択→右クリック→データ系列の書式設定→系列のオプション→要素の間隔：なし

[7] Excel の FACT 関数の上限は 170 です。JavaScript でもプログラム factorial(x)を作って実験すると x=170 (170!)が限界でした。

べて対数を使って行い，その結果得られた数を指数とします。対数を使えば，階乗と積算が加算になり割り算が減算になるので，巨大な数にはなりません。また，$(p)^x (1 - p)^{n-x}$ の部分にも指数があって，p も 1-p も確率なので 1 以下の数字ですから，指数部の数が大きくなると，p^x や $(1 - p)^{n-x}$ が非常に小さな数になって計算ができなくなります（アンダーフロー）。

そこで，プログラムでは上式の各項の積算・割り算・指数を対数の加算・減算・掛算にします。対数(log)によって指数の合計値を求め，最後にその合計値を関数 exp で自然対数の底 e の指数とします。その結果が個別の二項確率となります。次の式の対数(log)と指数(exp)の関係に注意してください。

$$a * b / c * d^{\wedge}e = \exp[\log(a * b / c * d^e)] = \exp[\log(a) + \log(b) - \log(c) + e*d]$$

以下の関数プログラムでは，はじめにこれらの対数(Log)を計算しておき，繰り返し演算(for)の中では，これらの対数を使って，i に従って次々に変化させ，下側累積確率または上側累積確率を足し上げていきます。

```
Function BinT(x, n, p, sel) '二項分布確率, sel=0(現確率); sel=1(下側); 2(上側)
  Dim i, Ar, s, t, o, q, r: ReDim Ar(n): Ar(0) = 0
  For i = 1 To n
    Ar(i) = Ar(i - 1) + Log(i) '階乗の対数の配列
  Next i
  If sel = 0 Then
    BinT = Exp(Ar(n) - Ar(x) - Ar(n - x) + x * Log(p) + (n - x) * Log(1 - p)) '現確率
  Else
    For i = 0 To n
      t = Exp(Ar(n) - Ar(i) - Ar(n - i) + i * Log(p) + (n - i) * Log(1 - p)) '当該確率
      If sel = 1 And i <= x Then BinT = BinT + t '下側累積確率
      If sel = 2 And i >= x Then BinT = BinT + t '上側累積確率
    Next
  End If
End Function
```

次は Excel 関数を使った二項検定プログラムです。それぞれの値の導出法はブラックボックスになります。

```
Function BinomT(x, n, p, sel) '二項分布確率, sel=0(現確率); sel=1(下側); 2(上側)
  If sel = 0 Then BinomT = Application.BinomDist(x, n, p, 0) '現確率
  If sel = 1 Then BinomT = Application.BinomDist(x, n, p, 1) '下側累積確率
  If sel = 2 And x = 0 Then BinomT = 1 '上側累積確率  x=0
  If sel = 2 And x > 0 Then BinomT = 1 - Application.BinomDist(x - 1, n, p, 1)
  '上側累積確率  x>0
```

```
End Function
```

●プログラム (VBA)

```
Function BinS(x, n, e, sel)
'二項分布有意率 s(x 出現数,n 試行数,e 期待確率,sel=0 密度:1 累積)
  If e = 1 Then e = 1 - 1 / (n * 10)
  If sel = 0 Then BinS = Application.BinomDist(x, n, e, 0): Exit Function
  If x = 0 Then BinS = 0: Exit Function
  BinS = Application.BinomDist(x - 1, n, e, 1)
End Function
```

●二項分布期待確率

はじめに，信頼係数(Significance: S)をエクセルの累積二項確率関数(BINOMDIST)を使って，次のように定義します[8]。

S = BINOMDIST(x-1, n, p, 1)

ここで，x:出現回数，n:試行回数，p:確率, 1:累積確率を示します。

たしかに，このように累積二項確率を使って，一定の出現数（以下）が現れる確率（信頼係数 S）が求められるのですが，そもそも前提とする確率（事前確率:p）がわからない場合には，上の式を使うことができません。

そこで，逆に考えて，出現回数(x)，試行回数(n)，信頼係数(S)を既知とし，未知の事前確率(p)を求めることを考えます。その事前確率(p)（「二項分布期待確率」Binomial expected probability: BinE とよびます）を次の関数を使って求めます。

p = BinE(x, n, S)

ここで，S として 0.95，0.99, 0.999 などを設定します。次が BinE 関数のプログラム(ExelVBA)です[9]。

```
Function BinE(x, n, s) '二項分布期待確率 e(x:出現数,n:試行数,s:信頼係数)
  Dim i, k, r, mx, mn, sc
  '中間値，繰り返し，精度，最大，最小，比較信頼係数
  If x = 0 Then BinE = 0: Exit Function 'x=0 ならば BinE=0
  r = 10 ^ 6: mx = r '期待確率精度：探索最大値
```

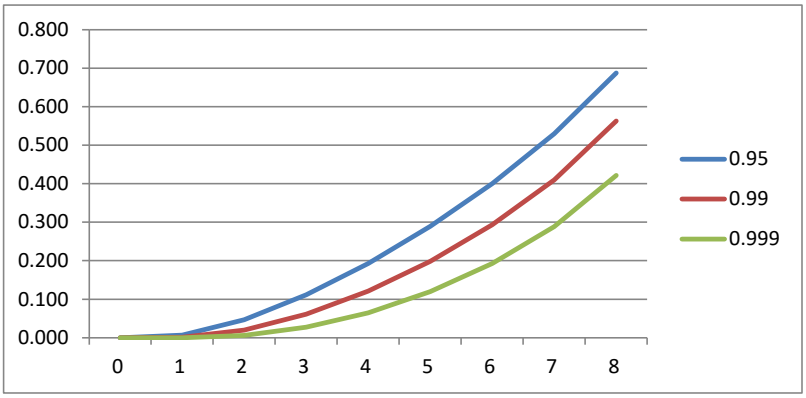
[8] S の 1 の補数(1-S)は二項検定に使う「危険率」になります。→検定

[9] このプログラムでは，x と，n と，次々に計算された確率 BinE から信頼係数 sc を求め，sc が先に設定された信頼係数 s に十分に近似するまで，二分探索法(binary search)を使って，繰り返します。十分に近似したときの確率 BinE が該当する確率 p になります。

```
  For k = 1 To 1000 '1000 回繰り返す
      i = (mx + mn) / 2:   BinE = i / r '中間値：期待確率候補
      sc = Application.BinomDist(x - 1, n, BinE, 1) '期待確率候補の有意率
      If sc < s - 1 / r Then 'sc が s-1/r に達しなければ
          mx = i '探索最大値を中間値まで下げる
      ElseIf sc > s + 1 / r Then 'sc が s-1/r を超えれば
          mn = i '探索最小値を中間値まで上げる
      Else
          Exit For 'sc が s±1/r の範囲に入ればループを脱出
      End If
  Next
End Function
```

次は，x, n を変えていくと二項分布期待確率がどのように変化するかを見た実験の結果です。信頼係数(S)を 0.95 (BinE.95)，0.99 (BinE.99), 0.999 (BinE.999)として，それぞれの期待確率を比較します。

x	n	BinE.95	BinE.99	BinE.999
0	8	.000	.000	.000
1	8	.006	.001	.000
2	8	.046	.020	.006
3	8	.111	.061	.027
4	8	.193	.121	.065
5	8	.289	.198	.120
6	8	.400	.293	.193
7	8	.529	.410	.289
8	8	.688	.562	.422



この結果を見ると，S=0.95 にすると結果がかなり甘く，S=0.999 にすると結果がかなり厳しくなっていることがわかります。S=0.99 が適切だと思います。

この「期待確率」とは，たとえば，8 回の試行で事象が 7 回出現したとき，99%の信頼性（1%の危険率）を条件として，そのようなことが起きるだろうと期待される確率は.410 と考えられる，ということです。つまり，8 回の試行で事象が 7 回出現したとき，99%の信頼性をもって，そのような事象が起こる確率を 0.410 と想定できる，ということになります[10]。

BinE(7, 8, 0.99) = 0.410

x	n	BinE: 0.95	0.99	0.999	Bin:0.95	0.99	0.999
0	8	.000	.000	.000	x	x	x
1	8	.006	.001	.000	.950	.990	.999
2	8	.046	.020	.006	.950	.990	.999
3	8	.111	.061	.027	.950	.990	.999
4	8	.193	.121	.065	.950	.990	.999
5	8	.289	.198	.120	.950	.990	.999
6	8	.400	.293	.193	.950	.990	.999
7	8	.529	.410	.289	.950	.990	.999
8	8	.688	.562	.422	.950	.990	.999

検算のために，次の式を使って，上表の右側(Bin)を計算しました。

Bin = BINOMDIST(x-1, n, BinE, 1)

x = 0 の場合を除いて[11], すべての信頼係数は表の左側と一致しています。

3.1.4. ポアソン分布

たとえば，国内で 1 日に起こる交通事故数や，多くの文字を含む 1 頁の中に見られる特定の文字の数などのように，ランダムに生起する現象ではその生起確率(p)が低くても母数(n)が大きければ, かなりの平均値(m = p*n)になることがあります。このような場合に二項分布をそのまま使うと多くの階乗を含むので計算が膨大になります。そこで n→∞(無限大), p→0 に近づけたときの二項分布の数式を理論的に導き，その式を使って二項分布の近似式とします。そのような近似式によって示される確率分布は発見者の名 Siméon Denis Poisson (1781-1840)から「ポアソン分布」(Poisson distribution)とよばれ，次の式(Po)で示されます(→**●ポアソン分布の導出**)。

$Po(X) = e^{-m} m^x / x!$, (x: 0, 1, 2, …; 平均 m = n * p)

[10] 逆に言えば，8 回の試行で事象が 7 回出現したとき，99%の信頼性を求めるならば，期待される生起確率は 41%しかない（半分以下），ということになります。
[11] x = 0 の場合は，BINOMDIST 関数の第 1 引数が x-1 であるために，計算できません。

次の表は，横軸に平均(Mean: m)を設定し，縦軸に生起回数(x)を設定して，平均と生起回数から，個別のポアソン確率を計算した結果です[12]。

Mean	1	2	3	4	5	6	7	8	9	10
x	m:1	m:2	m:3	m:4	m:5	m:6	m:7	m:8	m:9	m:10
0	0.368	0.135	0.050	0.018	0.007	0.002	0.001	0.000	0.000	0.000
1	0.368	0.271	0.149	0.073	0.034	0.015	0.006	0.003	0.001	0.000
2	0.184	0.271	0.224	0.147	0.084	0.045	0.022	0.011	0.005	0.002
3	0.061	0.180	0.224	0.195	0.140	0.089	0.052	0.029	0.015	0.008
4	0.015	0.090	0.168	0.195	0.175	0.134	0.091	0.057	0.034	0.019
5	0.003	0.036	0.101	0.156	0.175	0.161	0.128	0.092	0.061	0.038
6	0.001	0.012	0.050	0.104	0.146	0.161	0.149	0.122	0.091	0.063
7	0.000	0.003	0.022	0.060	0.104	0.138	0.149	0.140	0.117	0.090
8	0.000	0.001	0.008	0.030	0.065	0.103	0.130	0.140	0.132	0.113
9	0.000	0.000	0.003	0.013	0.036	0.069	0.101	0.124	0.132	0.125
10	0.000	0.000	0.001	0.005	0.018	0.041	0.071	0.099	0.119	0.125
11	0.000	0.000	0.000	0.002	0.008	0.023	0.045	0.072	0.097	0.114
12	0.000	0.000	0.000	0.001	0.003	0.011	0.026	0.048	0.073	0.095
13	0.000	0.000	0.000	0.000	0.001	0.005	0.014	0.030	0.050	0.073
14	0.000	0.000	0.000	0.000	0.000	0.002	0.007	0.017	0.032	0.052
15	0.000	0.000	0.000	0.000	0.000	0.001	0.003	0.009	0.019	0.035
16	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.005	0.011	0.022
17	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.006	0.013
18	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003	0.007
19	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.004
20	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002

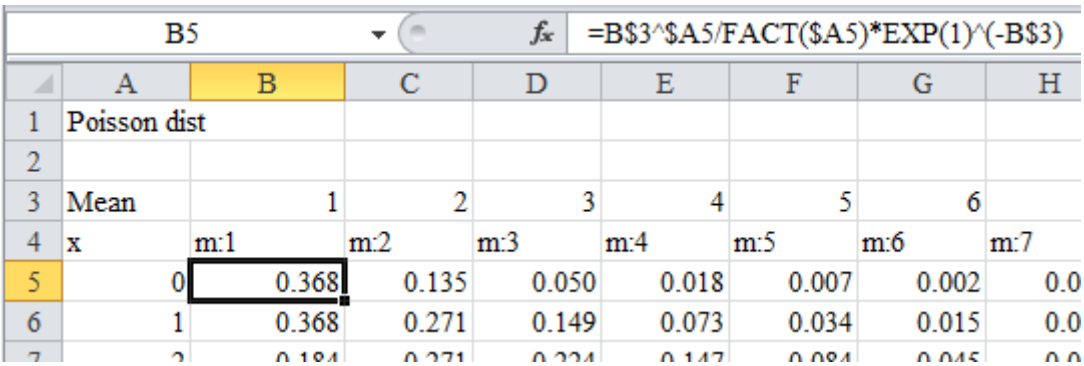
[12] たとえば，Po(x=0, m=1)はエクセル関数を使って，

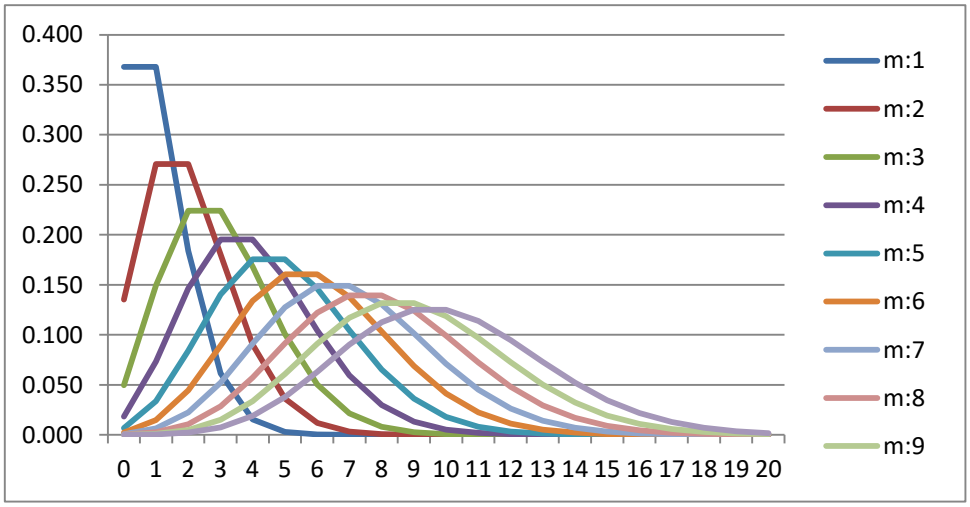
=EXP(1)^(-B$1)*B$1^$A5/FACT($A5)

または

=POISSON($A5,B$3, 0)

で計算します。





上の表によって,たとえば 60 分間に平均 3 回電話を受けている事務所が実際に 60 分間にランダムに電話を受ける回数が x=0, 1, 2, 3, 4, ...となるそれぞれの確率は，m:3 の列を見て，それぞれ 5.0% (x:0), 14.9% (x:1), 22.4% (x:2), 22.4% (x:3), 16.8% (x:4), ...になることが予想されます。このことは 1 時間を多くの区間，たとえば 20 個，30 個，60 個，...などの区間に分けて，各区間に起こる確率を 3/20, 3/30, 3/60, ...と考えます。そうすると，このようにnを増やしていくと確率はどんどん小さくなります。そして n*p = m を一定にしながら n→∞，p→0 という極限に近づけたときの確率の分布がポアソン分布になります。

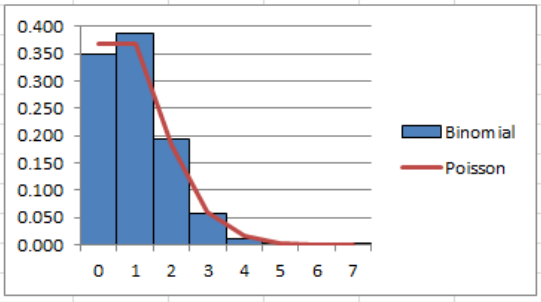
次の図は，n=10，20，30 にしたときの二項分布確率(Bin)とポアソン分布確率(Po)を比較したものです。ここで

Bin: B77=BINOMDIST(A77,10,1/10,0)

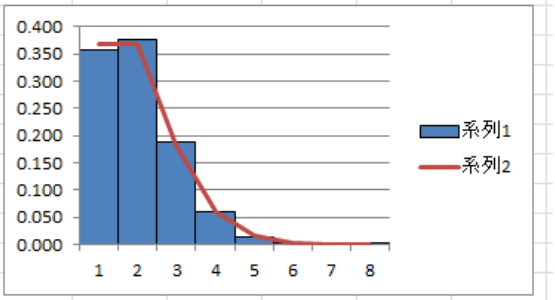
Po: C77=POISSON(A77,1,0)

試行回数 n が上昇するにしたがって二項分布の確率の変数 p を 1/10, 1/20, 1/50 と下げていきます。ポアソン分布の平均の変数 m = n*p は常に 1 とします。下の表とグラフによって，試行回数 n が上昇し，確率 p が下降するに従って，ポアソン分布が二項分布に近似する精度が高まっていることが確かめられます。

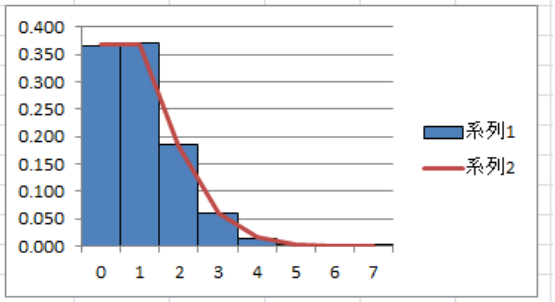
75	N:10			
76	x	Binomial	Poisson	B-P
77	0	0.349	0.368	-0.019
78	1	0.387	0.368	0.020
79	2	0.194	0.184	0.010
80	3	0.057	0.061	-0.004
81	4	0.011	0.015	-0.004
82	5	0.001	0.003	-0.002
83	6	0.000	0.001	0.000
84	7	0.000	0.000	0.000
85				



87	N:20			
88	x	Binomial	Poisson	B-P
89	0	0.358	0.368	-0.009
90	1	0.377	0.368	0.009
91	2	0.189	0.184	0.005
92	3	0.060	0.061	-0.002
93	4	0.013	0.015	-0.002
94	5	0.002	0.003	-0.001
95	6	0.000	0.001	0.000
96	7	0.000	0.000	0.000
97				

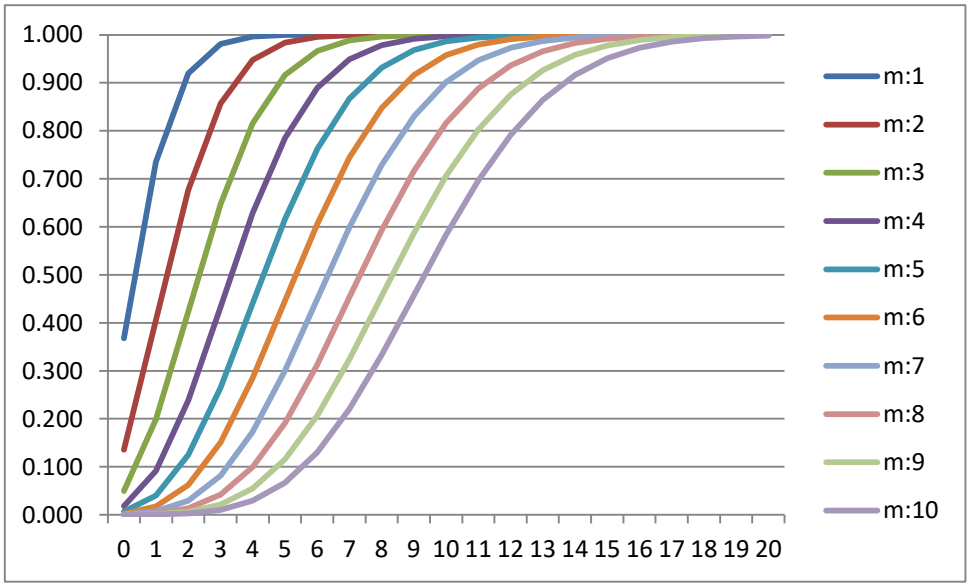


99	N:50			
100	x	Binomial	Poisson	B-P
101	0	0.364	0.368	-0.004
102	1	0.372	0.368	0.004
103	2	0.186	0.184	0.002
104	3	0.061	0.061	-0.001
105	4	0.015	0.015	-0.001
106	5	0.003	0.003	0.000
107	6	0.000	0.001	0.000
108	7	0.000	0.000	0.000
109				



次の表と図は累積ポアソン確率を示します。

Mean→	1	2	3	4	5	6	7	8	9	10
x ↓	m:1	m:2	m:3	m:4	m:5	m:6	m:7	m:8	m:9	m:10
0	0.368	0.135	0.050	0.018	0.007	0.002	0.001	0.000	0.000	0.000
1	0.736	0.406	0.199	0.092	0.040	0.017	0.007	0.003	0.001	0.000
2	0.920	0.677	0.423	0.238	0.125	0.062	0.030	0.014	0.006	0.003
3	0.981	0.857	0.647	0.433	0.265	0.151	0.082	0.042	0.021	0.010
4	0.996	0.947	0.815	0.629	0.440	0.285	0.173	0.100	0.055	0.029
5	0.999	0.983	0.916	0.785	0.616	0.446	0.301	0.191	0.116	0.067
6	1.000	0.995	0.966	0.889	0.762	0.606	0.450	0.313	0.207	0.130
7	1.000	0.999	0.988	0.949	0.867	0.744	0.599	0.453	0.324	0.220
8	1.000	1.000	0.996	0.979	0.932	0.847	0.729	0.593	0.456	0.333
9	1.000	1.000	0.999	0.992	0.968	0.916	0.830	0.717	0.587	0.458
10	1.000	1.000	1.000	0.997	0.986	0.957	0.901	0.816	0.706	0.583
11	1.000	1.000	1.000	0.999	0.995	0.980	0.947	0.888	0.803	0.697
12	1.000	1.000	1.000	1.000	0.998	0.991	0.973	0.936	0.876	0.792
13	1.000	1.000	1.000	1.000	0.999	0.996	0.987	0.966	0.926	0.864
14	1.000	1.000	1.000	1.000	1.000	0.999	0.994	0.983	0.959	0.917
15	1.000	1.000	1.000	1.000	1.000	0.999	0.998	0.992	0.978	0.951
16	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.996	0.989	0.973
17	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.995	0.986
18	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.998	0.993
19	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.997
20	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998



上の表と図で用いられた平均 m は整数を使っていますが，これは分数や小数になることもあります。たとえば，ある病気の全国の発生率が 1/2550 であって，ある地域の人口が 30000 人であれば，平均の発生数は 30000/2550=11.765...となります。このときたとえば個別の発生数，たとえば 10 件の確率や(第 3 引数=0)，10 件以下の発生数の累積確率(第 3 引数=1)を求めるときには，平均 m として 30000/2550 を先のそれぞれの式に代入します。

POISSON(10, 30000/2550, 0) = 0.109
POISSON(10, 30000/2550, 1) = 0.372

●ポアソン分布の数式の導出

次のポアソン分布の式(Po)を二項分布の式(Bin)から導出します。

$$Po(X) = e^{-m} * m^x / x!, (x: 0, 1, 2, \ldots;\ 平均\ m = np > 0)$$

上の式を次の二項分布の式から導きます。

$$Bin(X) = {}_nC_x * p^x * (1 - p)^{n-x} \quad (n \to \infty)$$

ここで

$$n * p = m \quad \rightarrow \quad p = m / n$$

とすると

$$Bin(X)$$
$$= {}_nC_x * p^x * (1 - p)^{n-x}$$

$= n! / [(n - x)!\ x!] * (m/n)^x * (1 - m/n)^{n-x}$ ←C の定義, p = m/n

$= n! / [(n - x)!\ x!] * m^x/n^x * (1 - m/n)^{n-x}$ ←指数 x を分配

$= n! / [(n - x)!\ n^x] * m^x/x! * (1 - m/n)^{n-x}$ ←x!と n^x の位置を交換

$= n! / [(n - x)!\ n^x] * m^x/x! * (1 - m/n)^n * (1 - m/n)^{-x}$ ←指数 n-x を分解

上の式を下線の(a), (b), (c), (d)に分け，n→∞とします。

(a) $n! / [(n - x)!\ n^x]$

$= [n * (n - 1) * \ldots * (n - x + 1)] / n^x$ ←階乗(!)を整理

$= n/n * (n - 1)/n * \ldots * (n - x + 1)/n$ ←各項に n を分配[13]

$= 1 * (1 - 1/n) * \ldots * [1 - (x + 1)/n]$ ←各項を計算

ここで n→∞とすると

$= 1 * 1 * \ldots * 1 = 1$ ←(n→∞)

(b) $m^x/x!$ ←このままにする

(c) $(1 - m/n)^n$

$= [1 - 1 / (n/m)]^n$ ←分子と分母に n/m を掛ける

$= \{[1 - 1 / (n/m)]^{-n/m}\}^{-m}$ ←指数部に-n/m を導入

$= \{[1 + 1/(-n/m)]^{-n/m}\}^{-m}$ ←分子と分母に-1 を掛ける

ここで x = - n/m とし，n→∞，よって，(x→-∞)とすると

$= [(1 + 1/x)^x]^{-m}$ (x→-∞)

$= e^{-m}$ ←ネイピア数 e の定義：$e = (1 + 1/x)^x$, x→±∞ (→後述)

(d) $= (1 - m/n)^{-x}$ ((n→∞))

$= 1$

よって

Po(X)

= Bin(X), n→∞

$= {}_nC_x * p^x * (1 - p)^{n-x}$ (n→∞)

= (a) * (b) * (c) * (d) (n→∞)

$= 1 * m^x / x! * e^{-m} * 1$

$= e^{-m} * m^x / x!$

[13] 分母 $n^x = n * n * \ldots$の各項(n)を，それぞれの分子に対応する分母にします。

●ポアソン分布の平均と分散

ポアソン分布の期待値 E(x)と分散 V(x)は二項分布の期待値 E(x)と分散 V(x)の式の中で p = m / n (←m = n * p)とし，n→∞とすることにより導かれます。

$$
\begin{aligned}
E(x) &= n * p \\
&= n * m / n \qquad \leftarrow\ p = m / n \\
&= m
\end{aligned}
$$

$$
\begin{aligned}
V(x) &= n * p * (1 - p) && (n \to \infty) \\
&= n * m / n\,(1 - m / n) && (n \to \infty) \leftarrow\ p = m / n \\
&= m * (1 - m / n) && (n \to \infty) \\
&= m
\end{aligned}
$$

このようにポアソン分布は期待値（平均）と分散が一致することが特徴です。

＊参考：一石(2004:60-61), 倉田・星野(2009:136-137).

●ネイピア数の定義

ネイピア数(e)は次のように定義されます（John Napier, 1550–1617: 高校数学 III）。

$$e = \lim_{h \to 0} (1 + h)^{1/h}$$

上の式は x = 1/h とすると次のように書き換えられます。

$$e = \lim_{x \to \pm\infty} (1 + 1/x)^{x} \qquad \leftarrow\ x = 1/h,\ h = 1/x$$

このとき，h→0 は x→±∞になります。h→0 の h はプラスでもマイナスでも成り立つので，x→±∞とします。このことを Excel で実験すると次のようになり，すべて同じ結果(2.71828...)に収束していくことがわかります。

h	$(1+h)^{1/h}$	h	$(1+h)^{1/h}$	x	$(1+1/x)^{x}$	x	$(1+1/x)^{x}$
0.1	2.59374	-0.1	2.86797	10	2.59374	-10	2.86797
0.01	2.70481	-0.01	2.73200	100	2.70481	-100	2.73200
0.001	2.71692	-0.001	2.71964	1000	2.71692	-1000	2.71964
0.0001	2.71815	-0.0001	2.71842	10000	2.71815	-10000	2.71842
0.00001	2.71827	-0.00001	2.71830	100000	2.71827	-100000	2.71830
0.000001	2.71828	-0.000001	2.71828	1000000	2.71828	-1000000	2.71828

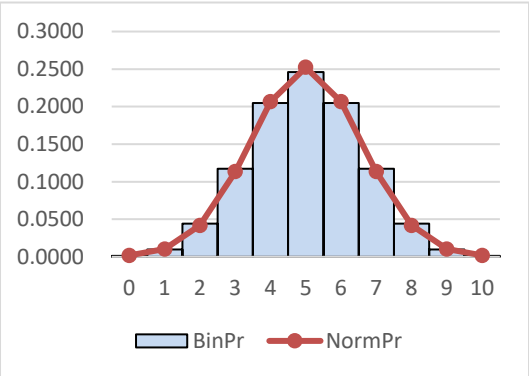
3.1.5. 正規分布

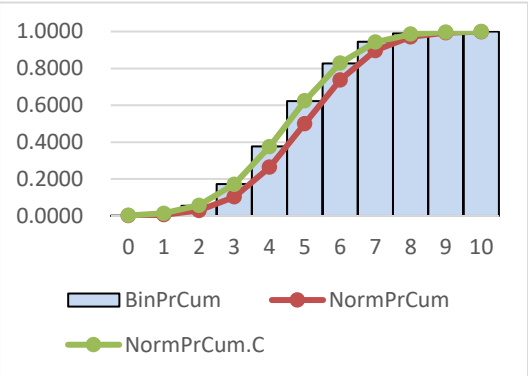
先に見たように, 二項分布のそれぞれの確率は出現回数(x), 試行回数(n), 出現確率(p)で求められます。Excel 関数では BINOMDIST(x, n, p, 0)を使います。一方，先述の正規分布の描く曲線は，平均(M)と分散(V)から得られる確率密度を使います。n が 30 ほどになると，次の表が示すように二項確率(BinPr)と正規確率密度(NormPr)の値は近似します[14]。

N	10
P	0.5000
M	5.0000
V	2.5000

x↓: n=10	BinPr	NormPr	BinPrCum	NormPrCum	NormPrCum.C
0	0.0010	0.0017	0.0010	0.0008	0.0022
1	0.0098	0.0103	0.0107	0.0057	0.0134
2	0.0439	0.0417	0.0547	0.0289	0.0569
3	0.1172	0.1134	0.1719	0.1030	0.1714
4	0.2051	0.2066	0.3770	0.2635	0.3759
5	0.2461	0.2523	0.6230	0.5000	0.6241
6	0.2051	0.2066	0.8281	0.7365	0.8286
7	0.1172	0.1134	0.9453	0.8970	0.9431
8	0.0439	0.0417	0.9893	0.9711	0.9866
9	0.0098	0.0103	0.9990	0.9943	0.9978
10	0.0010	0.0017	1.0000	0.9992	0.9997

下左図は n=30 のときの確率分布を示します。棒グラフは二項確率分布を表し，折れ線は対応する正規分布を表します。ほぼ一致していることを確認してください。下右図は，それぞれの累積確率分布を示します[15]。





それぞれの Excel 関数は

二項確率（離散:BnPr）： =BINOMDIST(x,n,p,0)
正規確率（連続:NmPr）： =NORMDIST(x,m,sd,0)
累積二項確率（離散:BnCum）： =BINOMDIST(x,n,p,1)

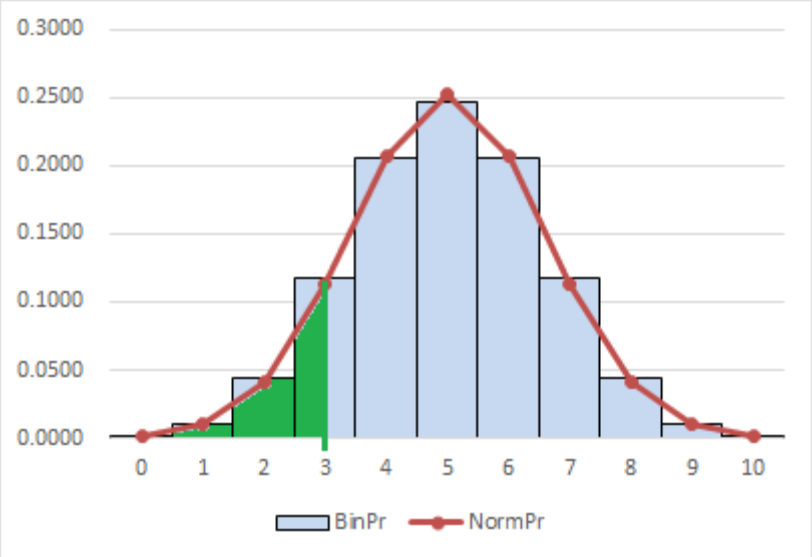
[14] N: 総和，P: 確率，M: 平均，V: 分散。
[15] Excel 関数 NORMDIST の第 3 引数は標準偏差(sd=$V^{1/2}$)です。

累積正規確率（連続:NmCum）：　　=NORMDIST(x,m,sd,1)

上右図には，正規累積確率 NmCum と補正正規累積確率 NmCumC を示しました。二項確率に近似させる正規累積確率 NmCum は，先に見たように(→「連続的確率変数」)，1 点での確率ではなく確率密度を示すので，累積二項確率と正確には一致しません。両者を正確に一致させるために次のように，x に 0.5 を足して「連続補正」(continuity correction)をします(市原 1990: 48)。

補正累積正規確率(Norm. L.:下側) = NORMDIST(x+0.5,m,sd,1)

連続補正をする理由は次の図を見るとわかります[16]。

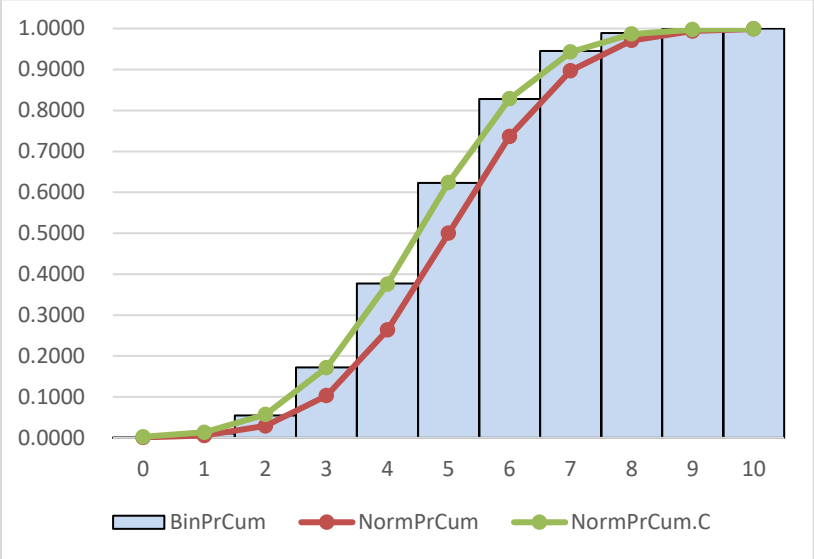


上の図を見ると，たとえば x = 3 までの累積確率を計算するとき，正規分布確率(Norm.p.:線グラフ)が二項分布確率(Norm.p.:棒グラフ)の x = 3 の部分の面積（確率）の左半分だけを含め（緑の部分），右半分を含めていないことがわかります。そこで，さらによく二項分布確率の累積に近似させるために，x に 0.5 を加え，x = 3.5 の確率までを含めることにします。x = 3 の左部分の上の三角形（青）と右部分の上の三角形（白）はほぼ同じ面積です。よって，x = 0 ~ 3.5 の折れ線（赤）で囲まれる面積は棒グラフの面積と非常に近似します。

補正累積正規確率(NormPr.Cum.C) = NORMDIST(x + 0.5,m,sd,1)

このことを累積確率の図を見て確かめしょう。

[16] N=10; 期待確率(e) = 0.5; 平均(m) = N*e = 10*0.5 = 5; 標準偏差(sd) = $[x*p*(1-p)]^{1/2} = 10 * 0.5*0.5)^{1/2}$.



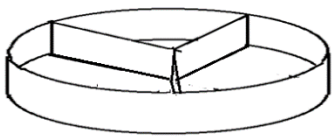
上図を見ると，x = 3 の正規分布確率の累積(NormPrCum：赤線)が二項分布確率の累積(BinPrCum)：青棒)の高さに達していないことがわかります。一方，0.5 の補正をした正規分布確率の累積(緑線：NormPrCum.C)は二項分布確率の累積(BinPrCum))の高さにほぼ達しています。

直接に二項分布確率を使わないで正規分布確率を使う理由は，二項分布確率の出現数(x)，試行数(n)が整数値だけに対応するためです。一方，さまざまな割合や平均値などの小数（小数点のある数: x）の確率を計算するときは平均値(m)と分散(v)を使った正規分布確率が役立ちます。

●プログラム(VBA)

```
Function NormT(x, m, v, sel) '正規分布確率, sel=0(現確率); sel=0(密度):1(累積)
   If sel = 0 Then NormT = Application.NormDist(x, m, Sqr(v), 0)
   If sel = 1 Then NormT = Application.NormDist(x + 0.5, m, Sqr(v), 1)
End Function
```

●小数点のあるデータ



上の例(中世スペイン語子音字の変異<u>, <b>, <v>)では，頻度を扱ったので，データはすべて整数です。次に，たとえば水量の配分(g)や平均時速(km/時間)などのように小数点があるデータの有意率を計算することを考えます。たとえば，12 グラムの水を左図のような 3 等分した容器にランダムに注いで，1 つの区分に 7.5 グラムが集まったときの統計的有意性を求めることを考えます。

このとき，2 項分布確率に従う数値データの平均(Av)と標準偏差(SD)は次になります。→「確率」

$Av = n * e$, $SD = [n * e * (1 - e)]^{1/2}$（n: 試行回数，e: 期待確率）

よって，このデータを正規分布に従うと想定すると，小数を含む数値 x の有意確率（安全率）は次の Excel 関数で計算されます。

Security = NORMDIST(x-0.5, Av, SD, 1)
= NORMDIST(x-0.5, $n * e$, $[n * e * (1 - e)]^{1/2}$, 1)

ここで NORMSDIST 関数の第 1 引数が 7.5 ではなく，7.5-0.5 = 7 とする理由は，連続補正(Continuity correction)で+0.5 とし（→「確率」），次に二項分布と揃えるために二項分布の第 1 引数からの 1 を引くためです。

全体(n)が 12 グラムで，その中の 7.5 グラム(x)が示す安全率と危険率は Excel の NORMSDIST 関数を使えば次のように計算されます。

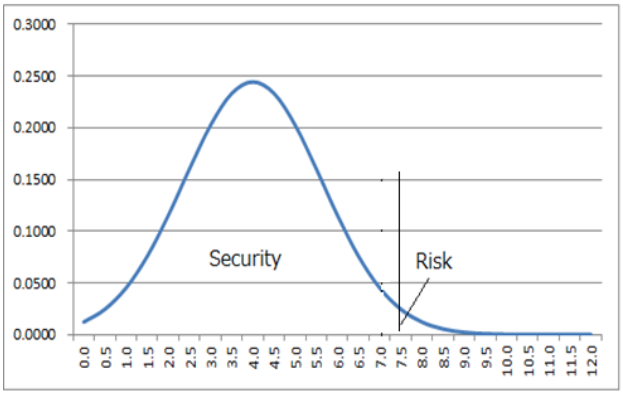
Security = NORMDIST(7.5-0.5, [12 * (1/3)], $[12 * (1/3) * (2/3)]^{1/2}$, 1)
= .967 (96.7%)

Risk = 1 - Security = .033 (3.3%)

次は x を変えたときの正規分布の確率密度(Norm)，危険率(Risk)，安全率(Security)を示します。

x	Norm.	Security	Risk
0.0	0.012	0.003	0.997
0.5	0.025	0.007	0.993
1.0	0.045	0.016	0.984
1.5	0.076	0.033	0.967
2.0	0.115	0.063	0.937
2.5	0.160	0.110	0.890
3.0	0.203	0.179	0.821
3.5	0.233	0.270	0.730
4.0	0.244	0.380	0.620
4.5	0.233	0.500	0.500
5.0	0.203	0.620	0.380
5.5	0.160	0.730	0.270
6.0	0.115	0.821	0.179
6.5	0.076	0.890	0.110
7.0	0.045	0.937	0.063
7.5	**0.025**	**0.967**	**0.033**
8.0	**0.012**	**0.984**	**0.016**
8.5	**0.005**	**0.993**	**0.007**
9.0	**0.002**	**0.997**	**0.003**

9.5	**0.001**	**0.999**	**0.001**
10.0	**0.000**	**1.000**	**0.000**
10.5	**0.000**	**1.000**	**0.000**
11.0	**0.000**	**1.000**	**0.000**
11.5	**0.000**	**1.000**	**0.000**
12.0	**0.000**	**1.000**	**0.000**



上の左表と図を見ると，x=7.5 の危険率(p 値)が.033 であり，これを p 値とすると 5%以下なので，7.5 以上の数値が統計的に有意であることがわかります。

■中世スペイン語古文献の平均語数の有意性

下左表は中世スペインで発行された各種文書の平均語数(Av)を文書の種類によってまとめたものです(C:王室，E:教会, J:法廷, M:役場, P:個人, Sum:計)。これらの数値はすべて文書数で割った値なので，このような数値がそのままで比較されることがありますが，それぞれの母数が年代によって大きく異なるので，厳密には比較することができません。そこで，下右表では安全率(Sec.: Security)を計算しました。

Av	**1200**	**1250**	**1300**	**1350**	**1400**	**1450**
C	53.8	**196.4**	**168.0**	150.3	177.7	**337.5**
E	**205.8**	**213.8**	**206.4**	**465.7**	**466.7**	**395.5**
J	5.2	29.4	24.9	50.6	46.6	43.9
M	2.7	13.2	15.6	53.3	27.8	43.6
P	**104.4**	76.8	**270.9**	198.4	**357.8**	224.6
Sum	371.9	529.5	685.8	918.3	1076.6	1045.0

Sec.	**1200**	**1250**	**1300**	**1350**	**1400**	**1450**
C	0.003	**1.000**	**0.998**	0.003	0.002	1.000
E	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	1.000
J	0.000	0.000	0.000	0.000	0.000	0.000
M	0.000	0.000	0.000	0.000	0.000	0.000
P	**1.000**	0.001	**1.000**	0.881	**1.000**	0.878

安全率が 95%を超えたデータは有意である，と考えて，太字で示しました。これらの安全率は，それぞれの年代の列の中で，偶然では起きない確率を示します。

●プログラム (VBA)

```
Function NormS(x, n, e, sel)
'正規分布有意率(x 出現量,n 全量,e 期待確率,sel=0 密度:1 累積)
 '(x: value, n: total, e: expected probability) H. Ueda (2017)
 If sel = 0 Then NormS = Application.NormDist(x, n * e, Sqr(n * e * (1 - e)), 0)
 If sel = 1 Then NormS = Application.NormDist(x - 0.5, n * e, Sqr(n * e * (1 - e)), 1)
 'Security in normal distribution (continuity correction +0.5, Binom -1 = -.5)
End Function
```

●二項分布確率・ポアソン分布確率・正規分布確率の比較

二項分布の分散 V = np(1-p)と正規確率 p の値によって，次のような場合分けがされます[17]。

(1) V = n*p*(1-p) < 10

(1a) p ≧ 0.1 → 二項分布による検定

(1b) p < 0.1 → ポアソン分布による検定

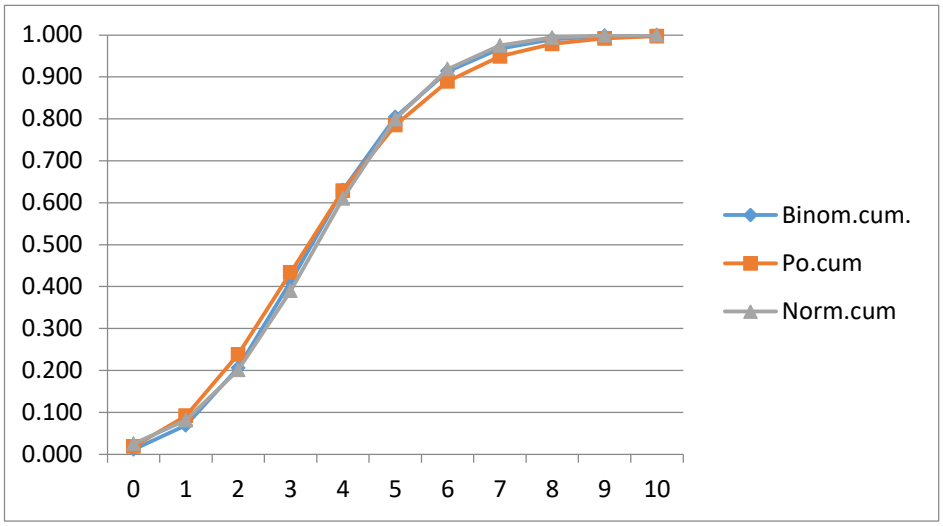
(2) V = n*p*(1-p) ≧ 10 → 正規分布による検定

次の表と図は(1a) V=3.2, p = 0.2 の場合の二項分布，ポアソン分布，正規分布の累積確率を示します[18]。

N	P	M	V
20	0.2	4	3.2
x	Binom.cum.	Po.cum	Norm.cum
0	0.012	0.018	0.025
1	0.069	0.092	0.081
2	0.206	0.238	0.201
3	0.411	0.433	0.390
4	0.630	0.629	0.610
5	0.804	0.785	0.799
6	0.913	0.889	0.919
7	0.968	0.949	0.975
8	0.990	0.979	0.994
9	0.997	0.992	0.999
10	0.999	0.997	1.000

[17] 正規分布の累積確率(Norm.cum.)は後述の連続補正をしてあります。

[18] 二項分布の累積確率: =BINOMDIST(x, N, P, 1)
ポアソン分布の累積確率: =POISSON(x, M,1)
正規分布の累積確率: =NORMDIST(x+0.5,M,SQRT(V),1)

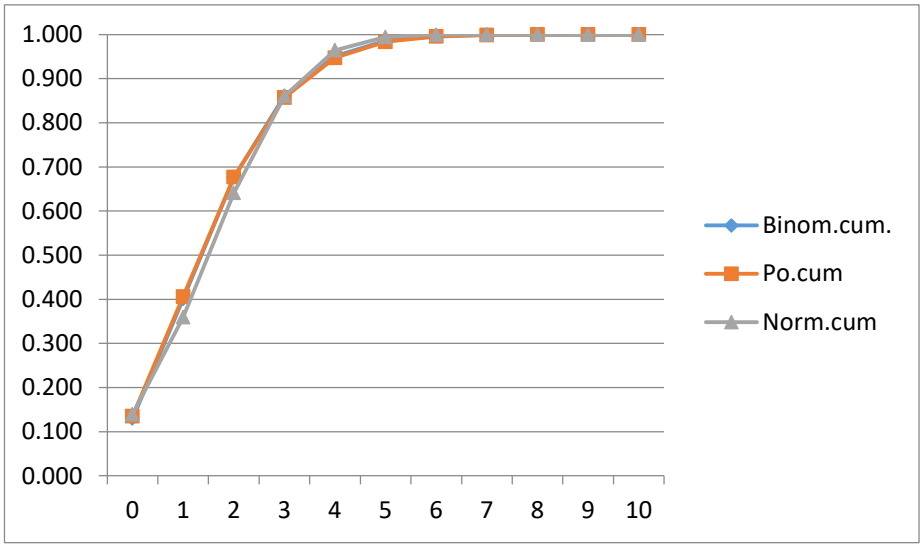


ポアソン分布も正規分布もかなり二項分布に近似していますが，すこし誤差があることがわかります（とくに正規分布の誤差が大きい）。

次の表と図は(1b) V=1.92 < 10, p = 0.04 < 0.1 の場合の二項分布，ポアソン分布，正規分布を累積確率を示します。

N	P	M	V
50	0.04	2	1.92

x	Binom.cum.	Po.cum	Norm.cum
0	0.130	0.135	0.140
1	0.400	0.406	0.359
2	0.677	0.677	0.641
3	0.861	0.857	0.860
4	0.951	0.947	0.964
5	0.986	0.983	0.994
6	0.996	0.995	0.999
7	0.999	0.999	1.000
8	1.000	1.000	1.000
9	1.000	1.000	1.000
10	1.000	1.000	1.000

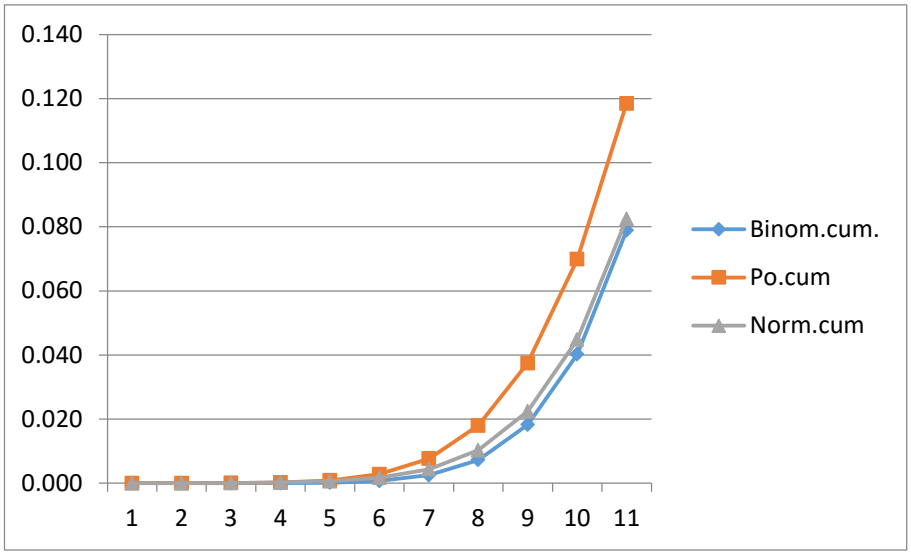


たしかにポアソン分布は二項分布によく近似していますが，正規分布の近似があまりよくありません。

最後に，次の表と図は(2) V=10.5 > 10, p = 0.3 > 0.1 の場合の二項分布，ポアソン分布，正規分布を確率を示します。

N	P	M	V
50	0.3	15	10.5

x	Binom.cum.	Po.cum	Norm.cum
0	0.000	0.000	0.000
1	0.000	0.000	0.000
2	0.000	0.000	0.000
3	0.000	0.000	0.000
4	0.000	0.001	0.001
5	0.001	0.003	0.002
6	0.002	0.008	0.004
7	0.007	0.018	0.010
8	0.018	0.037	0.022
9	0.040	0.070	0.045
10	0.079	0.118	0.082



このように，ポアソン確率分布も正規確率分布も二項確率分布の近似であるので，できるかぎり二項確率分布を使うべきです。そのうえ二項確率分布はわかりやすいのですが，ポアソン確率分布や正規確率分布を理解するにはかなりの数学的準備が必要です。従来，巨大なサンプル数(N)や微小の期待確率(P)のときに，二項確率の計算は困難であったのですが，Excel 関数 BINOMDIST は問題なく対応する確率を返します。また，対数・指数変換を使ったプログラムでも，この問題は解決されています。→**●プログラム**

参考：市原清志(1990)『バイオサイエンスの統計学』（南江堂）p.116-121.

3.2. 乱数

3.2.1. 乱数の確率

先に見た投げたサイコロの目や，円盤に投げた針が示す角度のように，それぞれの数値が次の数値を予測することができず，それぞれの数値や範囲に対応する度数が均等になるような数は**乱数**(ramdam numbers)と呼ばれます。乱数を生成するエクセル関数 Rnd()は使われる度に，[0, 1)の範囲内で，次のように小数点以下 15 桁まで出力されます。

0.288230019515856
0.569841439211386
0.616048897998326
(…)

はじめに次のプログラムで乱数を出力させます。

```
Sub Rnd1() '乱数実験 1
  Dim i&, Dn(10, 1)
```

```
   Cells(1, 1) = "Ex1": Cells(1, 2) = "Rnd"
   For i = 1 To 10
      Cells(i + 1, 1) = i
      Cells(i + 1, 2) = Rnd()
   Next
End Sub
```

出力：

Ex1	Rnd
1	.9276
2	.5495
3	.9850
4	.2122
5	.6167
6	.0829
7	.8321
8	.5783
9	.0458
10	.4628

次に乱数の確率が一定の階級の範囲内で等しいかどうかを確かめるプログラムを作ります。ここでは 10 個の階級を用意します。

```
Sub Rnd2() '乱数実験 2
   Dim i, j, r, p, Xnp(10, 2)
   Cells(1, 1) = "Ex.2": Cells(1, 2) = "Cnt"
   For i = 1 To 10 ^ 4
      p = Int(Rnd * 10) + 1 '一様分布乱数[1, 10)
      Xnp(p, 1) = p
      Xnp(p, 2) = Xnp(p, 2) + 1
   Next
   For i = 1 To 10
      For j = 1 To 2
         Cells(i + 1, j) = Xnp(i, j)
      Next j
   Next i
End Sub
```

上の P は乱数[0, 1)を 10 倍した数[0, 10)の整数部に 1 を足したもので，これを配列 Dn の位置とし，この配列位置の数値を全体で 1 万個分足しあげ

ます。

出力：

Ex.2	Cnt
1	980
2	966
3	1026
4	998
5	994
6	1073
7	974
8	978
9	1029
10	982

このように 1 万個の乱数がそれぞれの階級にほぼ均等に分配されていることがわかります。

●乱数の平均を求める実験

乱数[0, 1)を多数発生させ，その平均(=.5)を実験的に確かめます。プログラム（→後述）を使って，たとえば 10 万個の乱数を発生させると，それらの乱数の平均はおよそ 0.5 となり，その分散は 0.0843...になりました。乱数の範囲が[0, 1)ですから，この平均がおよそ 0.5 になることは想像できますが，分散がこの数値(0.0843....)になる理由は直ちにはわかりません。ここでは，はじめに具体的な例で実験的に平均と分散を求め，次に数理的にそれを一般化します。

次のような数値(x)と，その頻度(f)からなる頻度分布の例を見ましょう。

x	0	0.1	0.2	…	0.9	和
f	100	100	100	…	100	1000

この頻度分布表を使って平均(m)を求めると次のようになります。

m = [(0 * 100) + (0.1 * 100) + (0.2 * 100) + ... + (0.9 * 100)] / 1000

次に，この頻度(f)を確率(P)に変えて，次の確率分布にします。

X	0	0.1	0.2	…	0.9	和
P	1 / 10	1 / 10	1 / 10	…	1 / 10	1

上表のように確率(p)の和は必ず 1 になります。

[1]　　$\Sigma_i p_i = 1$

この確率分布を使って平均(m)を求めます。分数の分母のゼロの連続を避けるためにマイナスの指数を使います。下の第一式が先の頻度分布表と同じであることを確かめてください。

$$
\begin{aligned}
m &= \Sigma_i x_i p_i \qquad (i = 0, 1, 2, \ldots, 9) \\
&= (0 * 10^{-1}) + (10^{-1} * 10^{-1}) + (0.2 * 10^{-1}) + \ldots + (0.9 * 10^{-1}) \\
&= (0 + 0.1^1 + 0.2 + \ldots + 0.9) * 10^{-1} \qquad \leftarrow \text{各項の } 10^{-1} \text{ を外へ} \\
&= (0 + 1 + 2 + \ldots + 9) * 10^{-1} * 10^{-1} \qquad \leftarrow \text{(*)内の各項の } 10^{-1} \text{ を外絵} \\
&= (0 + 1 + 2 + \ldots + 9) * 10^{-2} \qquad \leftarrow \text{分母を整理} \\
&= (9 * 10 / 2) * 10^{-2} \qquad \leftarrow \text{脚注}^{19} \\
&= 45 * 10^{-2} = 0.45
\end{aligned}
$$

次に小数点以下 2 桁までの乱数の平均(m')は

$$
\begin{aligned}
m' &= \Sigma_i x_i p_i \qquad (i = 0, 1, 2, \ldots, 99) \\
&= (0 * 10^{-2}) + (0.01 * 10^{-2}) + (0.02 * 10^{-2}) + \ldots + (0.99 * 10^{-2}) \\
&= (0 + 0.01 + 0.02 + \ldots + 0.99) * 10^{-2} \\
&= (0 + 1 + 2 + \ldots + 99) * 10^{-2} * 10^{-2} \\
&= (0 + 1 + 2 + \ldots + 99) * 10^{-4} \\
&= (99 * 100 / 2) * 10^{-4} \\
&= 4950 * 10^{-4} = 0.495
\end{aligned}
$$

さらに，小数点以下 3 桁までの乱数の平均(m")は

$$
\begin{aligned}
m'' &= \Sigma_i x_i p_i \qquad (i = 0, 1, 2, \ldots, 999) \\
&= (0 * 10^{-3}) + (0.001 * 10^{-3}) + (0.002 * 10^{-3}) + \ldots + (0.999 * 10^{-3}) \\
&= (0 + 0.01 + 0.02 + \ldots + 0.999) * 10^{-3} \\
&= (0 + 1 + 2 + \ldots + 999) * 10^{-3} * 10^{-3} \\
&= (0 + 1 + 2 + \ldots + 999) * 10^{-6} \\
&= (999 * 1000 / 2) * 10^{-6} \\
&= 499500 * 10^{-6} = 0.4995
\end{aligned}
$$

このように乱数の間隔を次第に小さくし,乱数の種類を多くしていくと，乱数の平均は次第に 0.5 に近づくことがわかります。後述するように，乱数の間隔を無限にゼロ(0)に近づければ，平均は無限に 0.5 に近づくことが予想できます。そして，範囲が[0, 1)の乱数の平均が 0.5 に近づくことは，私たちが直感で納得できることです。

---

[19] 数列(1, 2, …, *n*)の和＝*n* (*n* + 1) / 2，よって *n* = 9 のときの和は 45.←高校数学 B. わかりやすいようにこの部分を括弧(...)で囲みます。

●乱数の分散を求める実験

先に見たように，確率分布の分散は

$$V(X) = E(X^2) - [E(X)]^2$$

そこで,分散 V(X)を求めるには,先に平均 E(X) = 0.5 を求めてあるので,あとは $E(X^2)$がわかればよいことになります。

X	0	0.1	0.2	…	0.9	和
X^2	0^2	$(0.1)^2$	$(0.2)^2$		$(0.9)^2$	
P	0.1	0.1	0.1	…	0.1	1

$$E(X^2) = \Sigma_i X_i^2 p_i \quad (i = 0, 1, 2, \ldots, 9)$$
$$= 0^2*10^{-1} + (10^{-1})^2*10^{-1} + (0.2)^2*10^{-1} + (0.3)^2*10^{-1} + \ldots + (0.9)^2*10^{-1}$$
$$= [0^2 + (0.1)^2 + (0.2)^2 + (0.3)^2 + \ldots + (0.9)^2] * 10^{-1}$$
$$= [0 + 1^2 + 2^2 + 3^2 + \ldots + 9^2] * 10^{-2} * 10^{-1}$$
$$= [0 + 1^2 + 2^2 + 3^2 + \ldots + 9^2] * 10^{-3}$$
$$= [9 * (9 + 1) * (2 * 9 + 1) / 6] * 10^{-3}$$ ←注[20]
$$= 285 * 10^{-3} = 0.285$$

よって，分散 V(X)は

$$V(X) = E(X^2) - [E(X)]^2 = 0.285 - (0.5)^2 = 0.285 - 0.25 = 0.035$$

次に，小数点以下 2 桁までの乱数の分散 V(X)'を求めます。

$$E(X^2)' = \Sigma_i X_i^2 p_i \quad (i = 0, 1, 2, \ldots, 99)$$
$$= 0^2*10^{-2}+(0.01)^2*10^{-2}+(0.02)^2*10^{-2}+(0.03)^2*10^{-2}+ \ldots + (0.99)^2*10^{-2}$$
$$= [0 + (10^{-2})^2 + (0.02)^2 + (0.03)^2 + \ldots + (0.99)^2] * 10^{-2}$$
$$= [0 + 1^2 + 2^2 + 3^2 + \ldots + 99^2] * 10^{-4} * 10^{-2}$$
$$= [0 + 1^2 + 2^2 + 3^2 + \ldots + 99^2] * 10^{-6}$$
$$= [99 * (99 + 1) * (2 * 99 + 1) / 6] * 10^{-6}$$
$$= 328350 * 10^{-6} = 0.32835$$

$$V(X)' = E(x^2) - [E(x)]^2 = 0.328 - (0.5)^2 = 0.32835 - 0.25 = 0.07835$$

さらに，小数点以下 3 桁までの乱数の分散 V"(X)は

$$E(x^2)'' = \Sigma_i x_i^2 p_i \quad (i = 0, 1, 2, \ldots, 999)$$
$$= 0^2*10^{-3}+(0.001)^2*10^{-3}+(0.002)^2*10^{-3}+(0.003)^2*10^{-3}+\ldots+ (0.999)^2*10^{-3}$$
$$= [0 + (0.001)^2 + (0.002)^2 + (0.003)^2 + \ldots + (0.999)^2] * 10^{-3}$$

[20] 数列$(1^2, 2^2, \ldots, n^2)$の和＝$n*(n+1)*(2n+1) / 6$，よって $n = 9$ のときの和は 285.←高校数学 B．わかりやすいようにこの部分を括弧[...]で囲みます。

$= [0 + 1^2 + 2^2 + 3^2 + \ldots + 999^2] * 10^{-6} * 10^{-3}$

$= [0 + 1^2 + 2^2 + 3^2 + \ldots + 999^2] * 10^{-9}$

$= [999 * (999 + 1) * (2 * 999 + 1) / 6] * 10^{-12}$

$= 332833500 * 10^{-12} = 0.3328335$

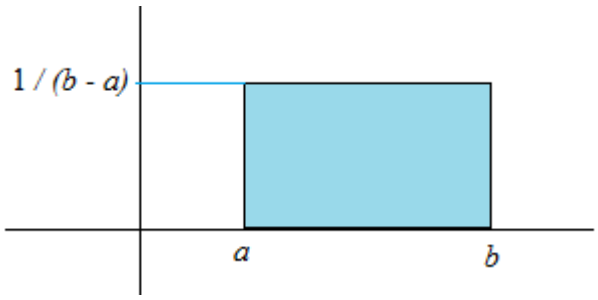
$V(X)'' = E(x^2)'' - [E(x)]^2 = 0.3328335 - (0.5)^2 = 0.3328335 - 0.25 = 0.0828335$

この段階まで求めた分散 0.0828335 が，先に実験的に確かめた乱数の分散 0.0843...に近似することがわかりました。以上で，それぞれの数値に対応する確率を個別に区切ってその平均と分散を求めました。そのような個別の確率は**離散的確率**(discrete probability)と呼ばれます。ここでは i = 9, 99, 999 まで増やしながら乱数の分散を求めましたが，さらに i = 9999, 99999, ... のように精度を高めていき，i → ∞のときの乱数の分散を求めることを，次に考えます。

3.2.2. 連続的確率

たとえば，{1, 2, 3, 4, 5, 6}という目をもつサイコロを次々に投げたとき，次に出る目は[1, 6]の範囲内でまったく予測できませんが，それぞれの確率はすべて 1/6 で同じになります。このように確率が等しく，次の数値が予測できない数値は**乱数**(random numbers)とよばれます。

先に離散的確率変数の平均と分散を求めましたが，実は乱数の小数点以下の桁数は非常に大きく理論的には無限にあると考えられるので，厳密に言えば，確率分布表の p ではなく，次のようなグラフと式で示される**一様分布**(uniform distribution)の**確率密度**(probability density)の関数 f(x)を使わなければなりません。



$f(x) = \begin{cases} 1 / (b - a) & [a \sim b] \\ 0 & [-\infty \sim a,\ b \sim +\infty] \end{cases}$

ここで，a, b はそれぞれ区間の下端（開始点）と上端（終了点）を示します。[0, 1)の区間にある乱数では，a = 0, b = 1 になります。x が 0 以下または 1 以上のときは f(x)はゼロ(0)になります。

$f(x)' = 1 / (1 - 0) = 1 \quad [0 \sim 1]$

はじめに，このような一様分布の確率密度関数の全体の値(総和：S)を積分を使って求めます。先の離散的な確率ではシグマ（Σ：和）を使って，個別の確率を掛けて足し合わせていきましたが，ここでは連続的な確率になるので，次のような定積分を使います(←高校数学 II)。

$$S = \int_0^1 f(x)\ dx = \int_0^1 1\ dx = [x]_0^1 = 1 - 0 = 1$$

上式では[0 ~ 1]の区間で計算していますが，特定の点での積分値はゼロになるので[21]，乱数の区間[0 ~ 1)でも同じです。また，[0 ~ 1]の区間以外のf(x)の値はゼロなので，区間[-∞ ~ +∞]にしても同じように結果は 1 になり，このことは一様分布の確率の総和が 1 になることを示しています。

さて，この f(x) = 1 を用いて連続的確率変数の平均を求めると

$$E(x) = \int_0^1 x\, f(x)\ dx = \int_0^1 x\, 1\ dx = \left[\frac{x^2}{2}\right]_0^1 = (1^2 / 2) - (0^2 / 2) = 1 / 2$$

よって，連続的な乱数の平均値は 1 / 2 = 0.5 になります。このことは先の実験で確かめました。

次に分散を求めるために，二乗の平均 $E(x^2)$を計算します。

$$E(x^2) = \int_0^1 x^2\, f(x)\ dx = \int_0^1 x^2\, 1\ dx = \left[\frac{x^3}{3}\right]_0^1 = (1^3 / 3) - (0^3 / 3) = 1 / 3$$

よって，分散 V(x)は

$$V(x) = E(x^2) - [E(x)]^2 = (1 / 3) - (1 / 2)^2 = (1 / 3) - (1 / 4)$$
$$= (4 / 12) - (3 / 12) = 1 / 12 = 0.0833...$$

以上で，先に乱数の実験で求めた分散 0.0843...と，小数点以下 3 桁の離散的確率変数で求めた分散 0.0828 が，連続的確率変数を使って数理的に求めた分散 1 / 12 = 0.0833...と近似することを確かめました。

＊一様分布の平均と分散については永田(2005: 61, 66)を参照しました。

●プログラム

```
Sub RndTest() '●乱数の和・平均・分散
  Dim i&, N&, Xn
  N = 5000: ReDim Xn(N)
  'Rnd (-1) '繰り返し数：シード値一定
  For i = 1 To N
    Xn(i) = Rnd '乱数[0, 1)
  Next i
  Cells(1, 1) = "和":    Cells(1, 2) = Application.Sum(Xn)   '和
```

[21] $\int_k^k f(x)\ dx = [F(x)]_k^k = F(k) - F(k) = 0$ (F(x)は f(x)の原始関数)

```
    Cells(2, 1) = "平均": Cells(2, 2) = Application.Average(Xn) '平均
    Cells(3, 1) = "分散": Cells(3, 2) = Application.VarP(Xn) '分散
End Sub
```

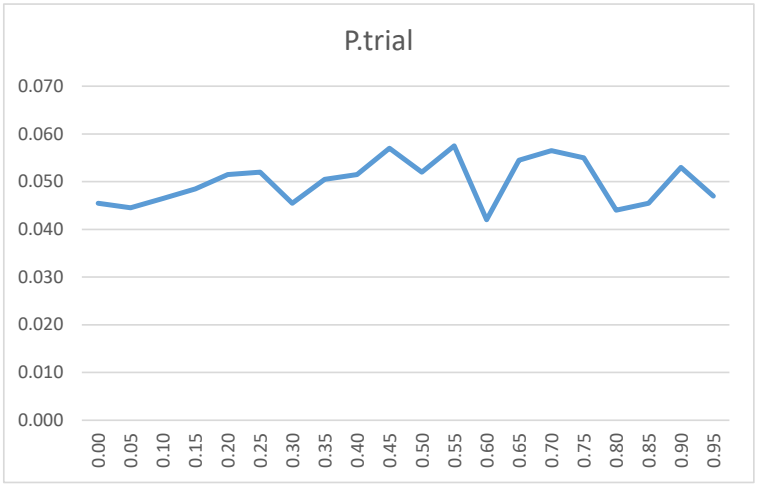
R 個の乱数を配列 Xn に格納し，Xn の和，平均，分散をそれぞれの Excel 関数で求め，該当するセルに出力します。

●一様分布の実験

一様分布の乱数[0, 1]を発生させ，20 個の階級(class)のそれぞれの確率が 1 / 20 = 0.05 になることを実験して確認します。

	A	B	C	D	E	F	G	H	I	J
1	U-dist			Trial	Z		Z	from	to	P.trial
2	Set			1	0.794		0.00	0.00	0.05	0.046
3	m	10		2	0.718		0.05	0.05	0.10	0.045
4	v	10		3	0.686		0.10	0.10	0.15	0.047
5	sd	3.162		4	0.407		0.15	0.15	0.20	0.049
6				5	0.361		0.20	0.20	0.25	0.052
7	Trial			6	0.684		0.25	0.25	0.30	0.052
8	M	0.506		7	0.118		0.30	0.30	0.35	0.046
9	V	0.080		8	0.496		0.35	0.35	0.40	0.051
10	SD	0.283		9	0.487		0.40	0.40	0.45	0.052
11				10	0.086		0.45	0.45	0.50	0.057
12	Renew: [F9]			11	0.170		0.50	0.50	0.55	0.052

	A	B	C	D	E	F	G	H	I	J
1	U-dist			Trial	Z		Z	from	to	P.trial
2	Set			1	=RAND()		0	=G2	=G2+0.05	=COUNTIFS($E$2:$E$2001,">"&H2,$E$2:$E$2001,"<="&I2)/2000
3	m	10		2	=RAND()		0.05	=G3	=G3+0.05	=COUNTIFS($E$2:$E$2001,">"&H3,$E$2:$E$2001,"<="&I3)/2000
4	v	10		3	=RAND()		0.1	=G4	=G4+0.05	=COUNTIFS($E$2:$E$2001,">"&H4,$E$2:$E$2001,"<="&I4)/2000
5	sd	=SQRT(B4)		4	=RAND()		0.15	=G5	=G5+0.05	=COUNTIFS($E$2:$E$2001,">"&H5,$E$2:$E$2001,"<="&I5)/2000
6				5	=RAND()		0.2	=G6	=G6+0.05	=COUNTIFS($E$2:$E$2001,">"&H6,$E$2:$E$2001,"<="&I6)/2000
7	Trial			6	=RAND()		0.25	=G7	=G7+0.05	=COUNTIFS($E$2:$E$2001,">"&H7,$E$2:$E$2001,"<="&I7)/2000
8	M	=AVERAGE(E:E)		7	=RAND()		0.3	=G8	=G8+0.05	=COUNTIFS($E$2:$E$2001,">"&H8,$E$2:$E$2001,"<="&I8)/2000
9	V	=VARP(E:E)		8	=RAND()		0.35	=G9	=G9+0.05	=COUNTIFS($E$2:$E$2001,">"&H9,$E$2:$E$2001,"<="&I9)/2000
10	SD	=SQRT(B9)		9	=RAND()		0.4	=G10	=G10+0.05	=COUNTIFS($E$2:$E$2001,">"&H10,$E$2:$E$2001,"<="&I10)/2000
11				10	=RAND()		0.45	=G11	=G11+0.05	=COUNTIFS($E$2:$E$2001,">"&H11,$E$2:$E$2001,"<="&I11)/2000
12	Renew: [F9]			11	=RAND()		0.5	=G12	=G12+0.05	=COUNTIFS($E$2:$E$2001,">"&H12,$E$2:$E$2001,"<="&I12)/2000



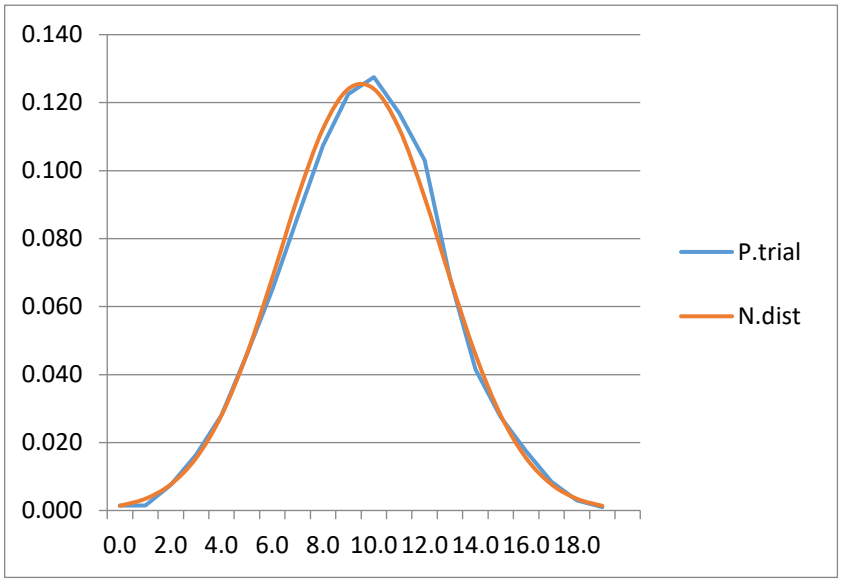
●正規分布の実験

正規分布の乱数 N(10, $10^{1/2}$)を発生させ，20 個の階級(class)のそれぞれの確率が Excel 関数で計算した確率になることを実験して確認します。

	A	B	C	D	E	F	G	H	I	J	K	L
1	N-dist			Trial	Z		Z	from	to	P.trial	N.dist	
2	Set			1	14.734		0.0	0.0	1.0	0.002	0.001	
3	m	10		2	9.152		1.0	1.0	2.0	0.002	0.003	
4	v	10		3	9.361		2.0	2.0	3.0	0.008	0.008	
5	sd	3.162		4	9.719		3.0	3.0	4.0	0.017	0.015	
6				5	9.114		4.0	4.0	5.0	0.028	0.028	
7	Trial			6	9.891		5.0	5.0	6.0	0.046	0.046	
8	M	10.069		7	15.173		6.0	6.0	7.0	0.065	0.068	
9	V	9.979		8	10.057		7.0	7.0	8.0	0.087	0.092	
10	SD	3.159		9	8.304		8.0	8.0	9.0	0.108	0.112	
11				10	0.094		9.0	9.0	10.0	0.123	0.124	
12	Renew: [F9]			11	11.962		10.0	10.0	11.0	0.128	0.124	

	A	B	C	D	E	F	G	H	I
1	N-dist			Trial	Z		Z	from	to
2	Set			1	=NORMINV(RAND(), $B$3, $B$5)		0	=G2	=G2+1
3	m	10		2	=NORMINV(RAND(), $B$3, $B$5)		1	=G3	=G3+1
4	v	10		3	=NORMINV(RAND(), $B$3, $B$5)		2	=G4	=G4+1
5	sd	=SQRT(B4)		4	=NORMINV(RAND(), $B$3, $B$5)		3	=G5	=G5+1

	J	K
1	P.trial	N.dist
2	=COUNTIFS($E$2:$E$2001,">"&H2,$E$2:$E$2001,"<="&I2)/2000	=NORMDIST(I2,$B$3,$B$5,1)-NORMDIST(H2,$B$3,$B$5,1)
3	=COUNTIFS($E$2:$E$2001,">"&H3,$E$2:$E$2001,"<="&I3)/2000	=NORMDIST(I3,$B$3,$B$5,1)-NORMDIST(H3,$B$3,$B$5,1)
4	=COUNTIFS($E$2:$E$2001,">"&H4,$E$2:$E$2001,"<="&I4)/2000	=NORMDIST(I4,$B$3,$B$5,1)-NORMDIST(H4,$B$3,$B$5,1)
5	=COUNTIFS($E$2:$E$2001,">"&H5,$E$2:$E$2001,"<="&I5)/2000	=NORMDIST(I5,$B$3,$B$5,1)-NORMDIST(H5,$B$3,$B$5,1)



[FIN]