稀少な事象と確率（ポアソン分布を巡って）

SHIMURA Masato
jcd02773@nifty.ne.jp

2010 年 7 月 23 日

目次

概要

稀少確率に関するポアソン分布の種々の事例と script の適用を検討すると共に、ガンマ分布などポアソン分布の類縁分布の応用事例の script を作成する

指数ファミリー

指数ファミリーと呼ばれる正規分布を中心とした分布関数のファミリーの概要図。

超幾何級数 (H.) は非復元型の 2 項級数で、ガウス由来であり、多くの特殊関数を表すことができる。*K.I.Iverson* が *J* に *H.* として組み込み、*E.Show* が *erf* 関数や Γ 関数を用いて種々の分布関数のスクリプトを提供した。

$$
\begin{array}{ccccccc}
 & \text{指数分布} & \leftrightarrow & \text{積率母関数} & & & \\
 & & \nearrow\swarrow & & \nwarrow\searrow & & \\
\text{ポアソン分布} & \leftrightarrow & 2\text{ 項分布} & \leftrightarrow & \text{正規分布} & \rightarrow & \text{対数正規分布} \\
 & & & & \uparrow & & \\
 & & H.(\text{超幾何級数}) & \rightarrow & erf(\text{誤差})\text{ 関数} & & \\
 & & \downarrow & & & & \\
 & & \Gamma\text{関数} & \rightarrow & \chi^2\text{関数} & & \\
 & & \downarrow & & & & \\
 & & \beta\text{関数} & \rightarrow & F\text{ 分布} & & \\
 & & \downarrow & & & & \\
p\text{ 値} & \leftarrow & t\text{ 分布} & & & &
\end{array}
$$

1　ポアソン分布

稀にしか起こらない現象を記述する確率分布にポアソン分布がある。

$$
{}_nC_x p^x (1-p)^{n-x} \rightarrow e^{-\lambda}\frac{\lambda^x}{x!}
$$

ポアソン分布は *2* 項分布の極限であり、試行回数が大きく生起確率が小さい場合に、*2* 項分布の良い近似式となる

$$
\begin{gathered}
P(X=x) = e^{-\lambda}\frac{\lambda^x}{x!} \\
(x = 0, 1, 2, 3, \cdots) \\
\lambda = xp
\end{gathered}
$$

$$
\begin{array}{cc}
\text{ポアソン分布の平均値} & E[x] = \lambda \\
\text{ポアソン分布の分散} & V[x] = \lambda \\
\text{大凡の守備範囲} & n \geq 50 \\
 & p \leq 0.1
\end{array}
$$

■ポアソン　$Siméon\ Denis\ Poisson(1781-1840 \quad France)$

父は下級軍人で除隊後下級役人のポストを得た。ポアソンは最初の子ではなかったが、生きられた最初の子であり、父から教育を授けらた。*8* 才の時にフランス革命が起こり、父は中部フランスの *Pithivier* の行政長官になった。父はポアソンを医者にしたいと思ってフォンテンブローの外科医の叔父の元にを送ったが、医学と外科が嫌いなポアソンは逃げ帰った。*1796* 年に父は再びポア

ソンをフォンテンブローの *École Centrale* で学ばせた。ここで数学の才能を見いだされ、パリの *École Politecnique* に進んでラプラスとラグランジュに数学を学んだ。*1802* 年にいきなり母校の准教授の地位を得た。ラプラスやラグランジュの研究を引継ぎ、微分方程式や物理学（熱学、音響）天文学に業績を残す。
École Politecnique の申し子で、激動の時代に本人もポリテクニークもその波から護った。小柄で少し下腹が膨らんだと記されている。

ガロアを失望させ、死に追いやった *3* 人の学者の一人で、熱力学ではフーリエとの確執もあった。

1.1 ポアソン分布のスクリプト

J の *ADDON* *addons/stats/base/distribution.ijs*
SUZUKI と *E.Shaw* *SUZUKI* と *E.Show* のスクリプト（添付）

G.Suzuki	*E.Shaw*	*J Addon*
pmf	***pmf***	*pmf*
`pden 0.1`	`0.1 poissonpmf i.5`	`poissondist 0.1 4`
`0    0.904837`	`0    0.904837`	`0    0.904837`
`1   0.0904837`	`1   0.0904837`	`1   0.0904837`
`2  0.00452419`	`2  0.00452419`	`2  0.00452419`
`3 0.000154653`	`3 0.000150806`	`3 0.000150806`
微少領域は纏めている	`4  3.77016e_6`	`4  3.77016e_6`
pmf* と *cdf	***cdf***	
累積分布を求めて度数と併せて表示する *pden_cum* を作成する	`0.1 poissoncdf i.5`	
`pden_cum 0.1`	`0 0.904837`	
`0    0.904837 0.904837`	`1 0.995321`	
`1   0.0904837 0.995321`	`2 0.999845`	
`2  0.00452419 0.999845`	`3 0.999996`	
`3 0.000154653        1`	`4        1`	

pden=:3 : 0 k=.i.5+>.+:y p=.(^-y)*(y^k)%!k p,1-+/p=.(+/p>1e_3){.p)	poissonpmf ^ * ^@-@[% !@]	poissondist=:3 : 0 'm n'=. y (^-m)**/\1,m%}.i.>:n)
pden λ λ のみ	λ *poissonpmf n*	*poissondist* λ *n*

pmf *probability mass function* 離散的確率関数
cdf *cumulative distribution function* 累積分布関数

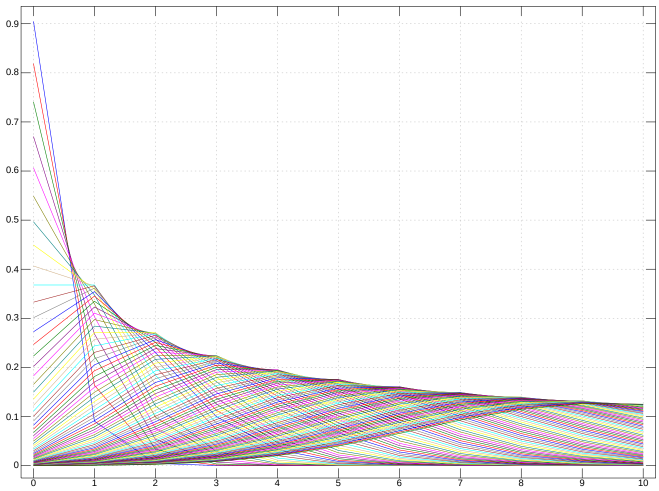


図 1 poisson distribution

```
poissondist L:0  {(0.1*i.100),.10
```

1.2 ポアソン分布の計算

1.2.1 ポアソン分布の計算

ある分野の高額販売物件と来客数

成約件数 $p = 0.002$ *(=0.2%)*

来客件数 $n = 1000$ 件 $\rightarrow np = 1000 * 0.02 = 2 = \lambda$

成立件数が $0, 1, 2, 3$ 件となる確率を求める

2 項確率 ${}_nC_x p^x(1-p)^{n-x}$ *3* 件の場合 ${}_{1000}C_3 0.002^3(0.998)^{997}$	`(3!1000),(0.002^3),0.998^997` `1.66167e8 8e_9 0.135878` `*/ (3!1000),(0.002^3),0.998^997` `0.180628`
ポアソン確率 $e^{-\lambda}\frac{\lambda^x}{x!}$ *3* 件の場合 $e^{-2}\frac{2^3}{3!}$	`(1x1^-2),(2^3)%!3` `0.135335 1.33333` `*/ (1x1^-2),(2^3)%!3` `0.180447`
pden *y* は λ *3* 件以下となる確率 `+/ 4{. pden 2` `0.857123`	`(i.9) ,. pden 2` `0   0.135335` `1   0.270671` `2   0.270671` `3   0.180447 NB.` `4  0.0902235` `5  0.0360894` `6  0.0120298` `7 0.00343709` `8 0.00109672`

1.2.2 計算事例集

ポアソン分布の計算に馴染むべく事例を集め、*Scipt* を用いて計算した。[*1]

一番使いやすいのは *pden* である。*poissonpmf* は右引数に打ち出す個数を指定する。

Example 1 或る県では交通事故による死亡事故が一日平均 *0.1* 件である。死亡事故 *0* の日の確率と、何件かの死亡事故の起こる確率を求める。

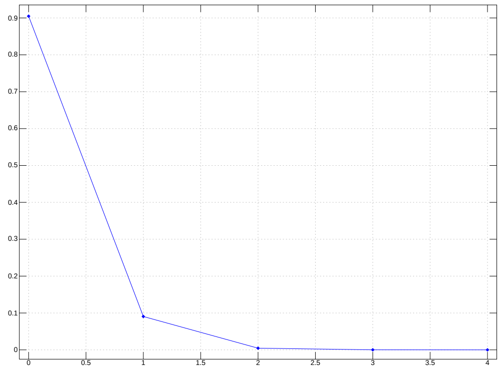
$$P_x(X=0)=\frac{0.1^0}{0!}e^{-0.1}=0.904837$$

[*1] ポアソン分布は発表から 60 年ほどは見向きもされなかったようだ。

$$P_x(X=1)=\frac{0.1^1}{1!}e^{-0.1}=0.0904837$$
$$P_x(X=2)=\frac{0.1^2}{2!}e^{-0.1}=0.00452419$$

```
 (i.5),. (0.1 poissonpmf i.5),. (pden 0.1),0
'marker,line'plot {|:  (i.5),. (0.1 poissonpmf i.5)
```

```
p  (poissonpmf)    (pden)
0     0.904837     0.904837
1    0.0904837    0.0904837
2   0.00452419   0.00452419
3  0.000150806  0.000154653
4   3.77016e_6            0
```



Example 2（出典 和達　十河 *p44*）

或る都府の一年間の死亡者数と発生する日の資料

死者数	0	1	2	3	4	5	6	*day*
該当日数	114	149	64	28	7	3	0	*fatal*

$$\lambda=\frac{1}{365}\sum x\times\text{度数}=\frac{404}{365}=1.1068$$

```
 +/ */"1 (i.7),.fatal
404

(i.7),.404r365 poissonpmf i.7
0     0.330599
1     0.365923
2     0.202511
3    0.0747164
4    0.0206749
5   0.00457681
6  0.000844306
```

Example 3
クラス（５０人）に *Morzwalt* と誕生日 *(1756/1/27)* が同じ人がいる確率

```
 (i.5),.  50r365 poissonpmf i.5
0    0.871982
1     0.11945
2  0.00818148
3 0.000373584
4   1.2794e_5
```

（参照）

```
 _1 x:  50r365
0.136986
```

Example 4
不良率が *0.1%* の *LED*
100 個入りの袋の中に不良品が少なくとも一個は行っている確率。
$\lambda = 0.001 \times 100 = 0.1$

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

不良品がない： $P_{0.1}(0) = \frac{0.1^0}{0!}e^{-0.1} = \frac{1}{1} \times 0.904837 = 0.904837$

不良品が 1 個含まれる： $P_{0.1}(1) = \frac{0.1^1}{1!}e^{-0.1} = \frac{0.1}{1} \times 0.904837 = 0.0904837$

不良品が 2 個含まれる： $P_{0.1}(2) = \frac{0.1^2}{2!}e^{-0.1} = \frac{0.01}{2} \times 0.904837 = 0.00452419$

*(*以上は *Example1* と同じ*)*

このバルク *50* 袋（*50* 袋 ×*100* 個）で *2* 個の不良品が含まれる確率

$\lambda = np = 5000 * 0.001 = 5$

poissondist を使ってみる。右引数に λ と打ち出す個数を指定する

```
 (i.13),.(+/\ a),. a=. poissondist 5 12
 0 0.00673795 0.00673795
 1  0.0404277  0.0336897
 2   0.124652  0.0842243   NB. 2 is 12.5%
 3   0.265026   0.140374
```

```
 4   0.440493   0.175467
 5   0.615961   0.175467
 6   0.762183   0.146223
 7   0.866628   0.104445
 8   0.931906   0.065278
 9   0.968172  0.0362656
10   0.986305  0.0181328
11   0.994547 0.00824218
12   0.997981 0.00343424
```

Example 5
故障率 *0.2* 件/年のシステムの *2* 年間の故障率

```
 (i.5),.  0.2 poissonpmf i.5
0   0.818731
1   0.163746
2  0.0163746
3 0.00109164
4 5.45821e_5
```

$$0.16 * 2 \quad = 0.32 \quad \text{故障率}$$
$$0.68 \quad \text{信頼度}$$

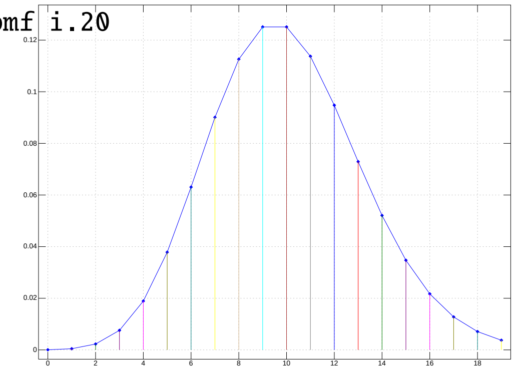
`1x1^-0.2` *NB.* $e^{-0.2}$
0.818731 NB. 参考

Example 6
花の種の袋（*1000* 粒入り。発芽率 *99%)*
発芽しない数が *5* 以内の確率を求める。

$\lambda = np = 1000 \times 0.01 = 10$

Example 7
工場の労働災害
1 年間に事故に遭う確率 *0.5%*
ここで働く或る労働者が *10* 年間事故に遭わない確率

```
(+/ a),˜ a=. 6{. (i.20),.  10 poissonpmf i.20
 0  4.53999e_5
 1 0.000453999
 2     0.00227
 3  0.00756665
 4   0.0189166
 5   0.0378333
sum    0.067086
```




```
(i.5),.  0.05 poissonpmf i.5
0   0.951229
1  0.0475615
2 0.00118904
3 1.98173e_5
4 2.47716e_7
```

$$\lambda = np = 10 \times 0.005 = 0.05$$
$$P(X = 0) = \frac{0.05^0 e^{-0.05}}{0!}$$

2 更に稀少な領域のテスト

2.1 ポアソン分布を用いた適合度検定

■Example **ポラリス潜水艦の部品交換確率** .

国防省の想定する故障率。

任務当たりの故障数	理論確率
0	0.368
1	0.368
2	0.184
3	0.061
4 以上	0.019

理論確率はポアソン確率を用いている。

```
(i.8),.1 poissonpmf i.8
0    0.367879
1    0.367879
2     0.18394
3   0.0613132
4   0.0153283
5  0.00306566
6 0.000510944
7   7.2992e_5
```

任務 *500* 回あたりの故障回数

任務当たりの故障数	任務回数
0	190
1	180
2	90
3	30
4 以上	10

出典 *Mansfield Ex10.2*

理論度数。

```
(i.5),.500*1 poissonpmf i.5
0  183.94
1  183.94
2 91.9699
3 30.6566
4 7.66416
```

$$\sum \frac{(f-e)^2}{e} = \frac{(190-183.9)^2}{183.9} + \frac{(180-183.9)^2}{183.9} + \frac{(90-91.9)^2}{91.9} + \frac{(30-30.6)^2}{30.6} + \frac{(10-9.4)^2}{9.46} = 0.367$$

```
  goodness_test DM0
0.36783
```

$\chi^2_{0.05}$ の値は数表から *9.48* と大きく理論分布は信頼できる。

2.1.1 Script

```
 goodness_test=: 3 : 0
NB. goodness fit test
NB. u DM0
TMP0=:;+/L:0<\.tmp=.1 poissonpmf i. 3+ # DM0
PX=:((<: # y) {. tmp), {: ( # y) {. TMP0
EX=: (+/ y) * PX
```

```
+/ EX %~  *: y - EX
)
```

DM1=: 190 180 90 30 10 NB. DATA

2.1.2 経過と解説

- 次のテーブルを作成する。*PX* はポアソン確率で最後の微少部分 *(4* 件以上）を累計している。

```
   DM1,.EX,.PX
190  183.94  0.367879
180  183.94  0.367879
 90 91.9699   0.18394
 30 30.6566 0.0613132
 10 9.48895 0.0189779
```

- $\dfrac{(f-e)^2}{e}$ を計算する
```
+/ EX %~  *: y - EX
```

2.2 p 値

2.2.1 p-value

最近の経済学などでは、*t* 値と併せて、*p* 値が記載されることが多い。

p value(i.e. probability value) は *t*-値から求められる。

p-value は帰無仮説のもとで、検定統計量の値を超える確率を示し、帰無仮説を棄却する最少の有意水準を示す。

D.Gjarati **[?]** *p value is defined as the lowest significance level of at which a null hypothesis can rejected*

例えば、自由度 *(df)* ８の *t-value* が *5.86* であった場合、*t-table* の α の値は *0.0010* では *5.041* である。 $tcdf$ で求めた値は *5.041* よりも大きく α は *0.001* より小さくなる。ｔ値から計算すると、ｐ値は *0.000189233* でエラーの確率は概ね *0.02* %、$\dfrac{2}{10000}$ であるといえる。

```
   8 tcdf 5.86
0.999811


  1- 8 tcdf 5.86
0.000189233   NB. p-value
```

3　ポアソン過程と待ち行列

　待ち行列には色々なタイプがあるが、ケンドールの分類 *M/M/n* は入力/出力共にポアソン/指数分布である。

- 解放型で思い起こされるのは、かって（*30* 年ほど前に）梅田阪急の正面階段の横にあったカレー屋「ピッコロ」である。ここが発祥の様で、「店内は椅子が *4* 席のみ、待つ客はトイレの通路に並ぶ。行列がトイレの通路から表の通路にかかると時間と見栄えからかそれ程行列は増えない。」
（店は今でも少し形を変えてあるようだ。）
- 開放型は到着時間と処理時間のやりくりで、計算も簡単な場合が多くシュミレーションにも適している。
- 閉鎖型は小さなモデルでも順列や小林/益川理論でも使われた分割数などを持ち出して少々複雑である。
- 順列や組み合わせはモデルを拡大すると計算量が飛躍的に増え、「計算問題」の課題も体験できる。

3.1　解法型の待ち行列 M/M/s

状態空間

$$S = \{0, 1, 2, 3, 4, \cdots, N\}$$

出生死亡連鎖の推移率　到着 ポアソン分布

$$q(i, i-1) = \lambda, \qquad (0 \le n \le N)$$

サービス 指数分布

$$q(i, i+1) = \mu, \qquad (0n \le N)$$

詳細釣り合い条件

$$\pi(0) = \frac{\lambda}{\mu}\pi(1)$$

$$\pi(1) = \frac{\lambda}{\mu}\pi(2)$$

$$\pi(2) = \frac{\lambda}{\mu}\pi(3)$$

$$\cdots$$

Example 床屋の待ち行列 *(M/M/s)Durret Sec4,Ex3.3*

- 床屋はパラメーター *3*（平均 *20* 分）でお客の髪を切る。$\lambda = 3$ の指数分布
- お客は平均 $\lambda = 2$（*30* 分）のポアソン過程に従って来店する
- 店には椅子が *2* つある。*2* つのイスが塞がっていると客は帰ってしまう
- お客数を状態空間 {*0,1,2,3*} で定義する

M/M/s 型はマルコフの推移行列が定常分布である場合にはパラメータ λ, μ を用いて確率が求められる。

状態空間

$$S = \{0, 1, 2, 3, 4\}$$

状態の推移

$$q(i, i-1) = 3, \qquad i = 1, 2, 3$$

$$q(i, i+1) = 2, \qquad i = 1, 2, 3$$

詳細釣り合い条件

$$2\pi(0) = 3\pi(1)$$

$$2\pi(1) = 3\pi(2)$$

$$2\pi(2) = 3\pi(3)$$

<table>
<tr>
<td>

in-out をマルコフの推移行列に仕立てる。
$\lambda, -\mu$

```
2 _3  0  0
0  2 _3  0
0  0  2 _3
```

</td>
<td>

拡大係数行列に拡大し詳細釣り合い条件を入れる。

```
   D33
2 _3  0  0 0
0  2 _3  0 0
0  0  2 _3 0
1  1  1  1 1
```

</td>
</tr>
</table>

<table>
<tr>
<td>

クラメール法で解く

```
 1 x: cr  D33
1 0 0 0 27r65
0 1 0 0 18r65
0 0 1 0 12r65
0 0 0 1  8r65
```

</td>
<td>

$\pi(0)$ $\frac{27}{65}$
$\pi(1)$ $\frac{18}{65}$
$\pi(2)$ $\frac{12}{65}$
$\pi(3)$ $\frac{8}{65}$ *occupied* 2 *chair*

</td>
</tr>
</table>

クラメール法

```
cr=: %. }:"1
```

*2

---

*2

−32 を逆にする場合

```
a=.  _3 2 0 0 0,0 _3 2 0 0 ,0 0 _3 2 0,: 1 1 1 1 1
   cr=. %.}:"1
  1 x: cr a
1 0 0 0  8r65
0 1 0 0 12r65
0 0 1 0 18r65
0 0 0 1 27r65
```

4　ガンマ関数とガンマ分布

4.1　ガンマ関数

階乗を整数から分数や負の整数に拡大する試みはオイラーが成功し *(1729)*、ルジャンドルが装いを凝らし今日の表記にした。

オイラー	$\Gamma\frac{1}{2} = \dfrac{1}{2}\sqrt{\pi}$ π が現れたことから何か丸い曲線の下部の面積と関連していると考え積分と組み合わせ、次の式を得た。 $$[x] = \int_0^1 (-lnt)^x dt$$	`   ! 1r2` `0.886227` $\dfrac{1}{2}\sqrt{\pi}$ `-:  %: 1p1` `0.886227` `   1r2p0.5` `0.886227`
ルジャンドル	$y = -int$ と置き換える $[x] = -\int_{\infty}^{0} y^x e^{-y} dy$ $= \int_0^{\infty} y^x e^{-y} dy$ *1* ずらす $\Gamma(x) \equiv [x-1] = \int_0^{\infty} y^{x-1} e^{-y} dy$	

4.1.1　ガンマ関数のスクリプト

J のガンマ関数のスクリプトは `gamma=: <:@!`である。

竹内によるガンマ関数のスターリング近似 *(JAPLA Apr./2010)* も合わせて使ってみよう。

1 2 3 4 5	`gamma >:i.5` `1 1 2 6 24`	`Gamm_St >: i.5` `0.999499 0.999979 1.99999 6 24`
分数 $\frac{15}{2}$ 数式処理の解 *(Wolfram)* $\frac{135135\sqrt{\pi}}{128}$ `128%~135135*%:1p1` `1871.25`	`gamma 15r2` `1871.25`	`Gamm_St 15r2` `1871.25`
複素数	`gamma 3j4` `0.00522554j_0.172547`	`Gamm_St 3j4` `0.0052255j_0.172547`

4.2 ガンマ分布

J のパッケージやアドオンには *gammarand* はあるが分布関数は入っていないので作成する。

$$
\begin{array}{ccc}
pdf & x^{k-1}\dfrac{exp(-x/\theta)}{\Gamma(k)\theta^k} & \\
cdf & \dfrac{\gamma(k,x/\theta)}{\Gamma(k)} & \\
paramter & k>0 & shape \\
 & \theta>0 & scale
\end{array}
$$

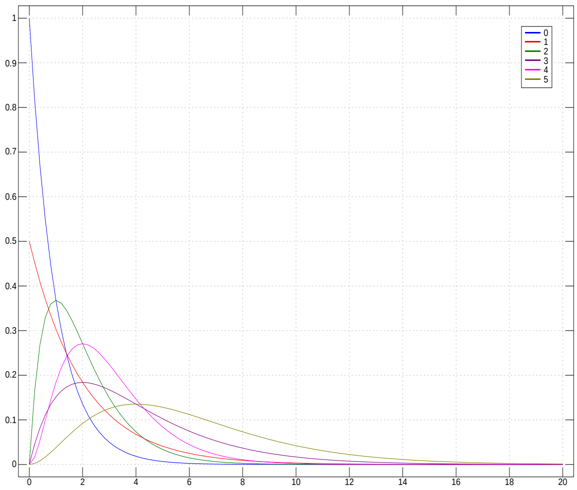
4.2.1 ガンマ分布のスクリプト

次の式に依った（出典 *wikipedia)* $f(x;k,\theta) = x^{k-1}\frac{e^{-x/\theta}}{\theta^k\Gamma(k)}$ $for\ x > 0\ and\ k,\theta > 0$	パラメータは次とした $k \to N0$ $\theta = \lambda \to MU$ *y= steps of y or i.y* `gammapmf0=: 4 : 0` `'N0 MU'=: x` `TMP0=.(1x1 ^ - y%MU)%(gamma N0)* MU^N0` `(y^ <: N0)*TMP0` `)`
これには $\alpha = k$,$\beta = \frac{1}{\theta}$ と置き換えた次の式もある。 $g(x;\alpha,\beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}$ $for\ x > 0$	`gammapmf1=: 4 : 0` `NB. beta=1/mu (= 1/lamda)` `'ALPHA MU'=. x` `BETA=. % MU` `((BETA^ALPHA)% !@<: ALPHA)*(y^<:ALPHA)*1x1^-BETA*y` `)`

3 2 plot_gammapmf 0 20 100

$x = k\ \theta$

$y = (steps)\,from\ 0\ to\ 20\ divide-by\ 100$



4.3 ガンマ分布の類縁

ガンマ分布の用途は大変広く色々な特殊関数の定義に用いられる。*Erlang* 分布はガンマ分布の *k* を整数に限定したものであり、ガンマ分布の関数をそのまま用いることができる。

ガンマ関数	$\Gamma(n)$	gamma=:<:@!
ベータ関数	$\beta(\alpha,\beta) = \dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$	
χ^2	$f_n(x) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{\frac{n}{2}} & (x > 0) \\ 0 & (x \leq 0) \end{cases}$	
t 分布	$f_n(x) = \dfrac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$	
	F 分布は β 関数を用いる。t 分布は β 関数を用いて作成することもできる	

```
NB. E Shaw
NB.==============
gamma=: ! & <: NB. Gamma function
ig0=: 4 : ' ( 1 H. (1+x ) % x  &((* ^) * (^ -)~)) y '
incgam=: ig0 % gamma@[  NB. incomplete gamma
beta=: *&gamma % gamma@+
ib0=: 4 : '(((,-.)/x ) H. (1+{. x ) * (^ % ])&({. x )) y '
incbet=: ib0 % [: beta / [ NB. incomplete beta
```

5　ベイズ統計の学習エンジンに指数分布やポアソン分布を用いる

ベイズ統計で用いる公式

$$\text{事後分布} \propto \text{尤度} \quad \times \text{事前分布}$$

ここで計算を簡略化するため、自然な共役分布として事前分布や事後分布に Γ 関数や β 関数がよく用いられる。この事前分布、尤度、事後分布の組み合わせの例。

事前分布	尤度	事後分布
β分布	二項分布	β分布
正規分布	正規分布	正規分布
逆Γ分布	正規分布	逆Γ分布
Γ分布	ポアソン分布	Γ分布

5.1　故障の確率

2 項分布

$$f(x|p) =_n C_x p^x (1-p)^{n-x} \qquad (x = 0, 1, 2, \cdots, n)$$

幾何分布　$_nC_x$ を除くと幾何分布になる

$$f(x|p) = p^x (1-p)^{n-x} \qquad (x = 0, 1, 2, \cdots)$$

指数分布　幾何分布の連続型

確率密度関数

$$f(x|\lambda) = \lambda e^{\lambda x} \qquad (x \geq 0), (0(x0)$$

5.2　指数分布と経験的ベイズ決定/信頼性のテスト

信頼性のテストでデータが全くないか少ない場合でもベイズ統計は推計ができる。信頼性の計測に指数分布を用いた例

Example　工業用エンジンの設計　（出典:松原 *10.2*）

- ある工業用エンジンの故障時間は平均 θ の指数分布 $E_x\left(\frac{1}{\theta}\right)$ に従う。

$$f(x|\theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}}, \qquad (x \geq 0)$$

- 経験豊かな設計者 A と経験の少ない設計者 B の共同作業
- 仕様上の信頼性 $\theta \geq 3000(h)$ が要求される

- 事前確率と予測
  1. *A,B* の事前確率

θ 範	(h) 囲	事 A	前 累計	確 B	率 累計
0	1000	0.01	0.01	0.15	0.15
1000	2000	0.04	0.05	0.15	0.3
2000	3000	0.2	0.25	0.2	0.5
3000	4000	0.5	0.75	0.2	0.7
4000	5000	0.15	0.9	0.15	0.85
5000	*over*	0.1	1	0.15	1

  2. $\theta \geq 3000(h)$ を超える確率は *A* は *0.75,B* は *0.7* と考えた
  3. $\theta \geq 2000$ は *A* は *0.05,B* は *0.3* と各々予測
- 試作エンジンを *2* 機作成しテストしたところ故障時間は $x_1 = 2000, x_2 = 2500$ であった
- この事実から事後確率 $w'_A(\theta|2000, 2500), w'_B(\theta|2000, 2500)$ を求める

1. 事後確率分布の式

$$w^{'}(\theta|x_1, x_2) \propto \left(\frac{1}{\theta}\right) e^{-\frac{x_1}{\theta}} \left(\frac{1}{\theta}\right) e^{-\frac{x_2}{\theta}} w(\theta) = \left(\frac{1}{\theta}\right)^2 e^{\left(-\frac{x_1+x_2}{\theta}\right)} w(\theta)$$

2. x_1, x_2 の関数の積を付け加える

$$w^{'}(\theta|x_1, x_2) \propto \left(\frac{x_1 + x_2}{\theta}\right) e^{\left(-\frac{x_1+x_2}{\theta}\right)} w(\theta)$$

3. 比例定数を求める *(A)* 積分の数値計算[*3]

$$\int_0^\infty \frac{4500}{\theta^2} exp\left(-\frac{4500}{\theta}\right) w_A(\theta) d\theta$$

$$= 0.01 \int_0^{1000} \frac{4500}{\theta^2} exp\left(-\frac{4500}{\theta}\right) d\theta$$

$$+0.04 \int_{1000}^{2000} \frac{4500}{\theta^2} exp\left(-\frac{4500}{\theta}\right) d\theta + \cdots +$$

$$+0.10 \int_{5000}^{\infty} \frac{4500}{\theta^2} exp\left(-\frac{4500}{\theta}\right) d\theta$$

$$= 0.01 exp\left(-\frac{4500}{1000}\right) + 0.04\left(exp\left(-\frac{4500}{2000}\right) - (exp\left(-\frac{4500}{1000}\right)\right) + \cdots +$$

$$0.10\left(1 - exp\left(-\frac{4500}{5000}\right)\right)$$

$$= 0.15$$

[*3] Gini 係数の積分計算もよく似ている

4. 比例乗数から *A* の *0-1000* の確率を求める

$$\frac{exp\left(-\frac{2000+2500}{1000}\right)w(\theta)}{0.15} = \frac{e^{-4.5}w(\theta)}{0.15} = \frac{(e^{-4.5})(0.01)}{0.15} = 0.0007406$$

```
|: ;("1),. 4500  bayes_m101 L:0 DM101;DM102
```

計算結果

区分		*A* の事後分布	累計	*B* の事後分布	累計
0	−1000	0.000741486	0.000741486	0.0103524	0.0103524
1000	−2000	0.0251742	0.0259156	0.0878684	0.0982208
2000	−3000	0.157162	0.183078	0.146284	0.244504
3000	−4000	0.338813	0.521891	0.126144	0.370648
4000	−5000	0.0820153	0.603906	0.0763381	0.446986
5000	*over*	0.396094	1	0.553014	1

そこそこの試作結果から *A* の事後分布は幾分改善され、*B* の信頼性は悲観が払拭されて大幅に改善されている

5.2.1 Script

```
bayes_m101=: 4 : 0
NB. 4500 bayes_m101 DM101
'RANK RELY'=: { y
TMP0=:(}: RELY)* ; -/ L:0 ^ (L:0)(-x)%L:0 |.(L:0)2<\RANK
TMP2=: TMP0 , ({:RELY)* 1- ^ (- x) % {: RANK NB. upper 5000
(1-+/ TMP3),~ TMP3=: TMP0 % +/ TMP2
)
```

5.2.2 経過と解説

1. *x*（左引数）$x_1 + x_2 = 4500$

2. *input A* のデータ

```
   DM101
   0 1000 2000 3000 4000 5000 NB. RANK
0.01 0.04  0.2  0.5 0.15  0.1 NB. RELY
```

3. 先ずは時間区分（ランク）の階差を計算する

```
a=: |.(L:0) 2<\ RANK
+------+---------+---------+---------+---------+
|1000 0|2000 1000|3000 2000|4000 3000|5000 4000|
```

```
+------+--------+--------+--------+---------+
```

4. $-4500 \div \frac{4500}{1000}$

```
 _4500 % L:0 a
+-------+---------+---------+----------+----------+
|_4.5 __|_2.25 _4.5|_1.5 _2.25|_1.125 _1.5|_0.9 _1.125|
+-------+---------+---------+----------+----------+
```

5. $exp - 4500 \div (\frac{4500}{1000})$

```
 ;-/ L:0 ^ (L:0) _4500 % L:0 a
0.011109 0.0942902 0.117731 0.101522 0.0819172
```

6. 事前確率を掛ける

```
tmp=.(}:RELY)* ;-/ L:0 ^ (L:0) _4500 % L:0 a
0.00011109 0.00377161 0.0235462 0.0507612 0.0122876
```

7. A の最終項は別計算し連結する

```
0.1* 1-^  _4500%5000
0.059343

] tmp2=. tmp,  0.1* 1-^  _4500%5000
0.00011109 0.00377161 0.0235462 0.0507612 0.0122876 0.059343
```

8. A の比例乗数を求める（積分）

```
   +/ tmp2
0.149821
```

9. *tmp* を比例定数で割ると時間区分での信頼度の事後確率が求まる

```
] tmp3=. tmp%+/tmp2
0.000741486 0.0251742 0.157162 0.338813 0.0820153
```

5.3 ポアソン分布を用いた尤度の計算

事前分布と事後分布に Γ 分布、尤度にポアソン分布を用いるベイズ統計
事前分布

$$\pi(\lambda) \propto \lambda^{\alpha-1} e^{-\lambda x}$$

尤度

n 個のデータ $x_1, x_2 \cdots x_n$ の同時分布

$\bar{x}$ は平均値

$$f(D|\lambda) = \frac{e^{-\lambda}\lambda^{x_1}}{x_1!}\frac{e^{-\lambda}\lambda^{x_2}}{x_2!}\cdots\frac{e^{-\lambda}\lambda^{x_n}}{x_n!} \propto e^{-n\lambda}\lambda^{n\bar{x}}$$

事後分布

$$\pi(\lambda|D) \propto 事前分布 \times 尤度$$

$$\propto e^{-n\lambda}\lambda^{n\bar{x}} \times e^{-\lambda x}\lambda^{\alpha-1} = e^{-(x+n)\lambda}\lambda^{\alpha+n\bar{x}-1}$$

5.3.1 Example

• ある都市の交通事故死者が *3* 日間で *0,1,2* 人であった • 昨年の交通事故の死者は 平均 *1* 人、標準偏差 *1* であった • 事後分布 *(θ)* を求める *(*出典　涌井 *4.5* の例*)*	ポアソン分布 $\frac{e^{-\lambda}\lambda^{x}}{x!}$ 尤度 $\frac{e^{-\lambda}\lambda^{0}}{0!}, \frac{e^{-\lambda}\lambda^{1}}{1!}, \frac{e^{-\lambda}\lambda^{2}}{2!}$

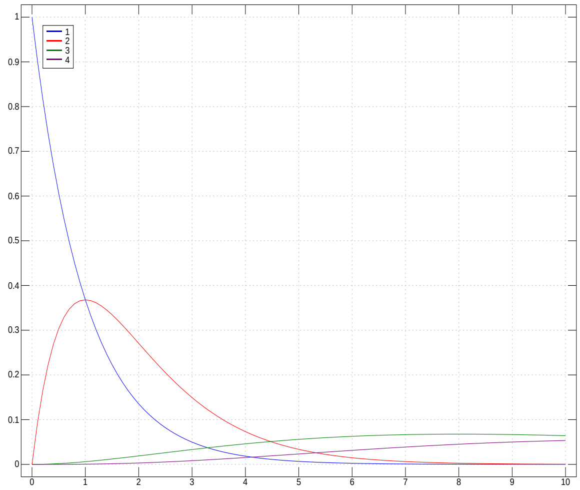
```
 (i.11),.|: ;("1),.  pden (L:0) {@> 0 1 2 3
  p=0      p=1         p=2        p=3
 0 1     0.367879   0.135335  0.0497871
 1 0     0.367879   0.270671   0.149361
 2 0      0.18394   0.270671   0.224042
 3 0    0.0613132   0.180447   0.224042
 4 0    0.0153283  0.0902235   0.168031
 5 0   0.00306566  0.0360894   0.100819
 6 0  0.000594185  0.0120298  0.0504094
 7 0            0 0.00343709   0.021604
 8 0            0 0.00109672 0.00810151
 9 0            0          0  0.0027005
10 0            0          0 0.00110249
```

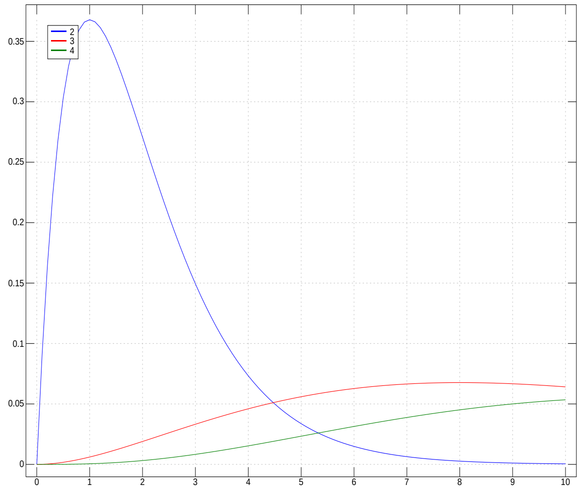
$$尤度 = \frac{e^{-\lambda}\lambda^0}{0!}\frac{e^{-\lambda}\lambda^1}{1!}\frac{e^{-\lambda}\lambda^2}{2!} \propto e^{-3\lambda}\lambda^3$$

$$事前分布 = e^{-\lambda}$$

$$事後分布 \propto 尤度 \times 事前分布 \propto e^{-3\lambda}\lambda^3 \times e^{-\lambda} = e^{-4\lambda}\lambda^3$$

事前分布は $Ga(1,1)$ で事後分布は $Ga(4,4)$ となる





References

和達三樹 · 十河清「キーポイント確率 · 統計」岩波書店　*1993*

Edwin Mansfield 蓑谷 · 高木 · 大津訳「統計学入門（下）」多賀出版　*1994*

William Dunham 一楽重雄　実川敏明訳「微積分名作ギャラリー」日本評論社 *2009*

R.Durret 今野/中村/曾雌/馬訳「確率過程の基礎」*Springer* 東京 *2005*

松原望　「入門　ベイズ統計」東京図書　*2008*

涌井良幸　「道具としてのベイズ統計」日本実業出版社　*2009*

reference.wolfram.com の *3.2.10* 特殊関数の項

Miscellance

J language DL available from

`http://www.jsoftware.com`

Script of this paper DL

`http://japla.sakura.ne.jp` *workshop june 2010*