確率過程ゼミ ♯2
確率分布の基礎

2012.05.26(Sun)

担当:M1 若林由弥

発表内容

- 二項分布
- ポアソン分布
- 幾何分布
- 超幾何分布

離散型の分布

- 一様分布
- 指数分布
- 正規分布
- ガンマ分布
- ワイブル分布

連続型の分布

- 大数の法則
- 特性関数
- 中心極限定理

0 用語の確認

▪確率密度関数 $\boldsymbol{p(x)}$

確率変数がある値xをとる確率の関数

▪確率分布関数 $\boldsymbol{F(x)}$

確率変数がx以下の値をとる確率．つまり，

$$F(x) = \int_{-\infty}^{x} p(\omega)d\omega$$

▪分布の平均 $\boldsymbol{E[X]}$

[離散型]

$$E[X] = \sum_{i=1}^{N} x_i p(x_i)$$

[連続型]

$$E[X] = \int_{-\infty}^{\infty} xp(x)dx$$

▪分布の分散 $\boldsymbol{V[X]}$

xの分布が期待値からどれだけ散らばっているかを示す値

[離散型]

$$V(X) = \sum_{i=1}^{N} (x_i - E[X])^2 p(x_i)$$

[連続型]

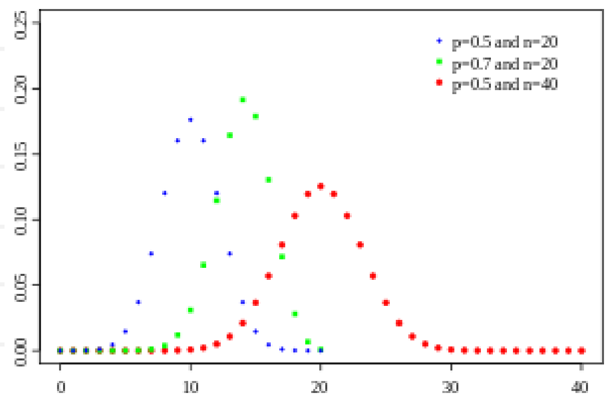
$$V[X] = \int_{-\infty}^{\infty} (x - E[X])^2 p(x)dx$$

1 二項分布

一定の確率で2つの結果だけが得られる独立な試行(ベルヌーイ試行)を繰り返す場合を考える

確率pで成功するベルヌーイ試行をn回行った結果，k回成功する確率密度$B_{n,p}(k)$を求めると

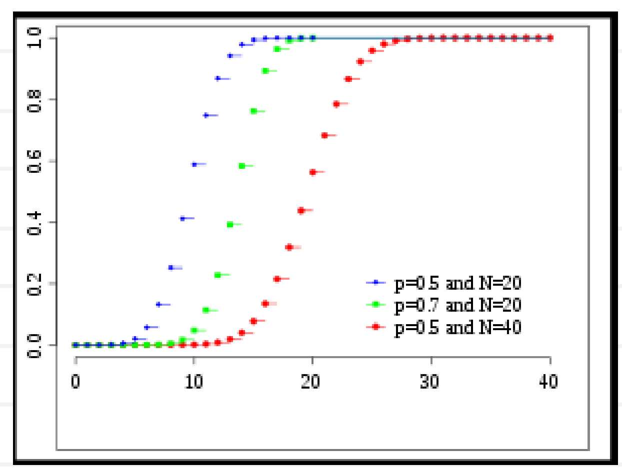
$$B_{n,p}(k) = {}_nC_k\, p^k (1-p)^{n-k}$$



また，確率分布関数は

$$F(k) = \sum_{i=0}^{k} B_{n,p}(i)$$

$$= \sum_{i=0}^{k} {}_nC_k\, p^i (1-p)^{n-i}$$



平均：

$$E[K] = np$$

分散：

$$V[K] = np(1-p)$$

2 ポアソン分布

ある交差点で1ヶ月に起こる交通事故の件数など，少ないがある程度起こりうる事象の確率を表すときにはポアソン分布を用いる

$$np \to \lambda \qquad n \to \infty \qquad p \to 0$$

となる極限においては

$$ {}_nC_k \, p^k (1-p)^{n-k} \to \frac{e^{-\lambda}\lambda^k}{k!}$$

が成り立つ(ポアソンの小数の法則)

よって二項分布を変形して，

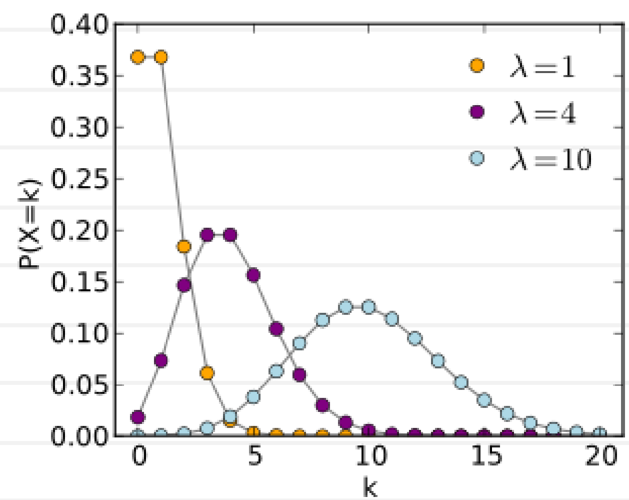
$$p_\lambda = \frac{\lambda^k e^{-\lambda}}{k!}$$

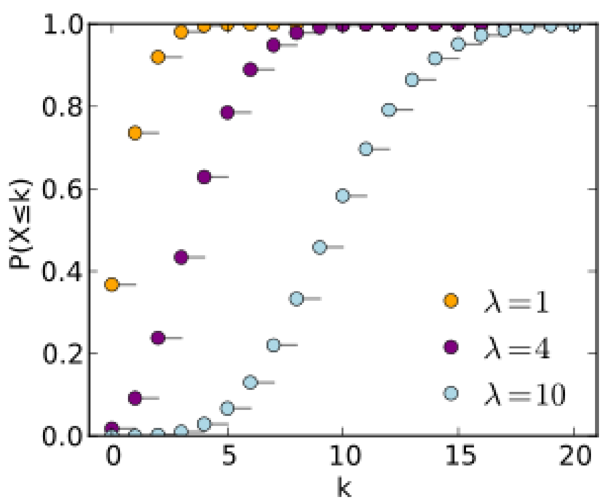
平均：

$$E[K] = \lambda$$

分散：

$$V[K] = \lambda$$





3 幾何分布

ある事象Aの起こる確率がpであるベルヌーイ試行を考える
このとき，k回目に初めてAが起こる確率は

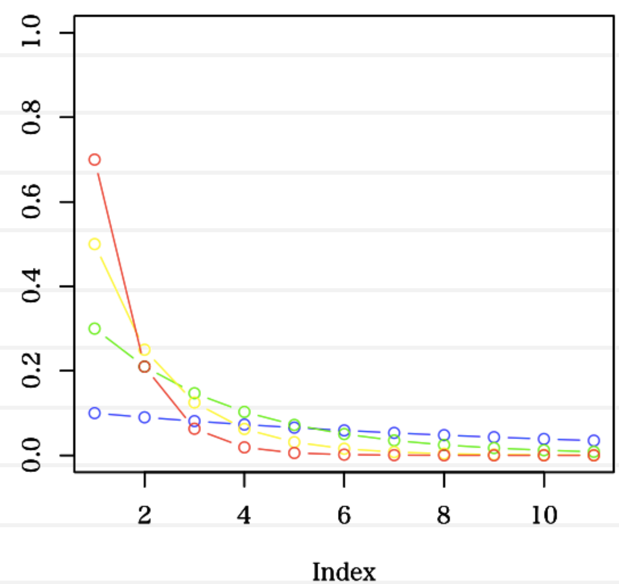
$$G_p(k) = (1-p)^{k-1}p$$

このような分布を**幾何分布**と呼ぶ

また，確率分布関数は

$$F(k) = \sum_{i=1}^{k} G_p(i)$$

$$= \sum_{i=1}^{k} (1-p)^{k-1} p$$



平均：

$$E[K] = \frac{1}{p}$$

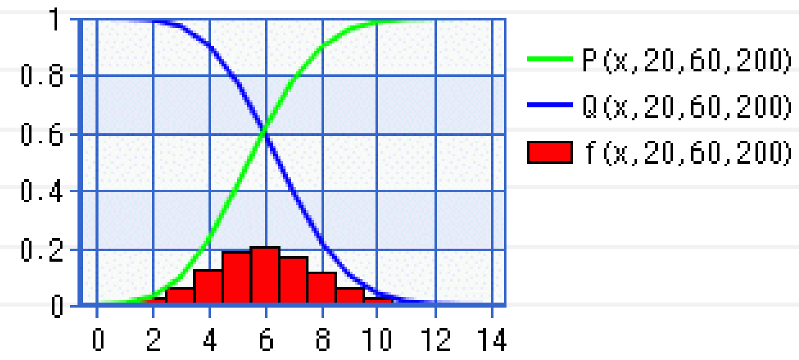
分散：

$$V[K] = \frac{q}{p^2}$$

4 超幾何分布

A, BからなるN個のものがあり，Aの個数がM個であるとする．この集団からn個取り出した時にAがx個である確率を考える

$$f(x) = \frac{{}_M C_x \cdot {}_{N-M} C_{n-x}}{{}_N C_n}$$



この確率分布を超幾何分布という

確率分布関数：

$$F(x) = \sum_{i=1}^{x} f(x)$$

$$= \sum_{i=1}^{x} \frac{{}_M C_i \cdot {}_{N-M} C_{n-i}}{{}_N C_n}$$

平均：

$$E[X] = \frac{nM}{N}$$

分散：

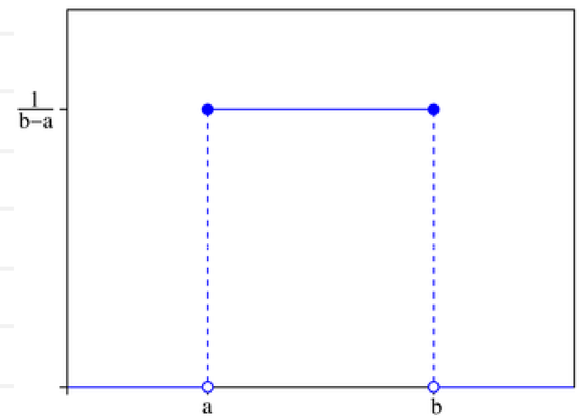
$$V[X] = \frac{nM(N-M)(N-n)}{N-1}$$

一次元一様分布

確率変数xが$[a,b]$上に一様に存在するような分布

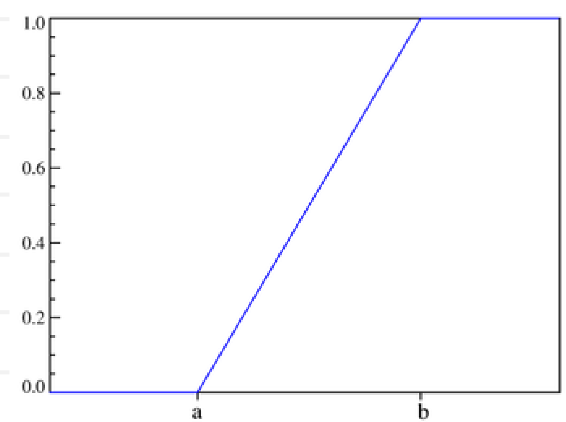
確率密度関数：

$$p(x)=\begin{cases}\dfrac{1}{b-a} & (a\leq x\leq b)\\ 0 & (x<a, b<x)\end{cases}$$



確率分布関数：

$$F(x)=\begin{cases}0 & (x\leq a)\\ \dfrac{x-a}{b-a} & (a\leq x\leq b))\\ 1 & 1\end{cases}$$



平均：

$$E[X]=\frac{b+a}{2}$$

分散：

$$V[X]=\frac{(b-a)^2}{12}$$

6 一様分布

球面上の一様分布

球座標(r,θ,ϕ)を用いて，半径$r=1$の単位球面上に一様に分布する点の確率密度関数$p(\theta,\phi)$を考える.

球面上の点$(1,\theta,\phi)$における微小面積は

$$dS = sin\theta d\theta d\phi$$

θとϕは独立なので，

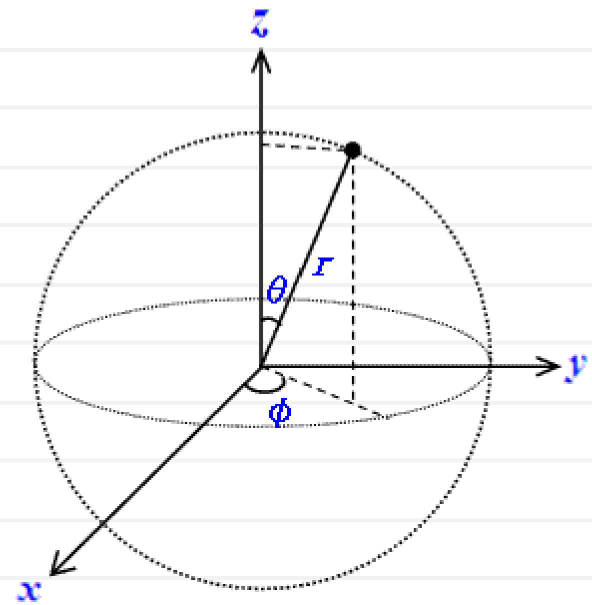
$$p(\theta,\phi) = p_\theta(\theta)p_\phi(\phi)$$

$$\left(\int_0^\pi p_\theta(\theta)d\theta = 1,\ \int_0^{2\pi} p_\phi(\phi)d\phi = 1\right)$$

この条件で$p(\theta,\phi)$を求めると，

$$p(\theta,\phi) = p_\theta(\theta)p_\varphi(\phi)$$

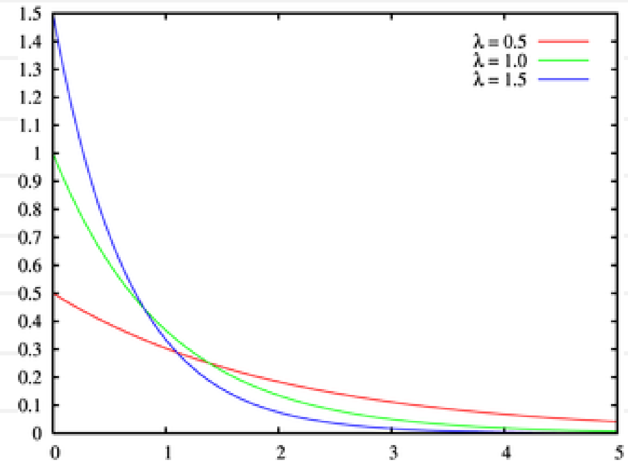
$$= \frac{1}{2}\sin\theta * \frac{1}{2\pi} = \frac{1}{4\pi}\sin\theta$$



7 指数分布

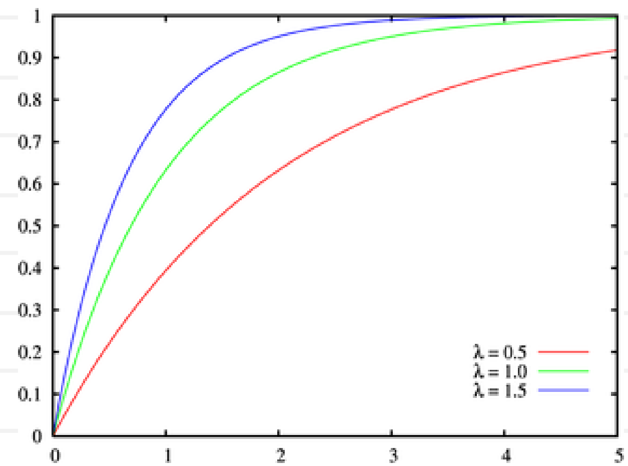
確率密度関数:

$$p(x) = \begin{cases} \lambda e^{-\lambda x} & (x \geq 0) \\ 0 & (x < 0) \end{cases}$$



分布関数:

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & (x \geq 0) \\ 0 & (x < 0) \end{cases}$$



平均:

$$E[X] = \frac{1}{\lambda}$$

分散:

$$V[X] = \frac{1}{\lambda}$$

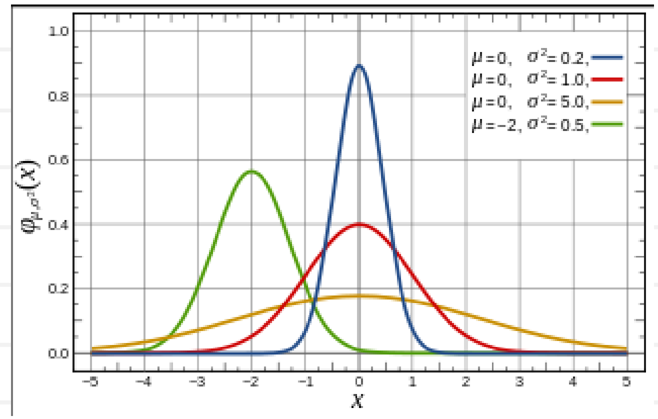
単位時間ごとに発生する事象の数がパラメータλのポアソン分布に従う時，その事象の発生間隔はパラメータλの指数分布に従う

8 正規分布

確率密度関数が

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

で与えられるような分布を正規分布と呼ぶ(ガウス分布とも呼ばれる)



平均:

$$E[X] = m$$

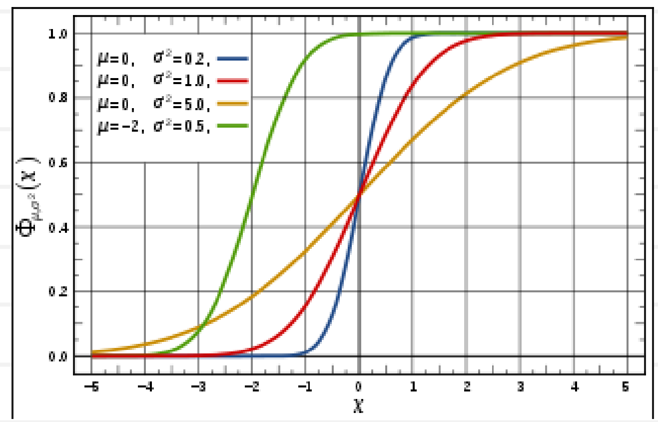
分散:

$$V[X] = \sigma^2$$

平均m, 分散σ^2の正規分布を

$$N(m, \sigma^2)$$

とかく



平均0, 分散1の正規分布$N(0,1)$を特に標準正規分布と呼び，この分布は二項分布を標準化し，$n \to \infty$としたときの近似系であることが一般的に知られている

9 ガンマ分布

確率密度関数:

$$p(x) = \begin{cases} \dfrac{\lambda}{\Gamma(a)}(\lambda x)^{a-1}e^{-\lambda x} & (x \geq 0) \\ 0 & (x < 0) \end{cases}$$

ここで, $\Gamma(a)$はガンマ関数と呼ばれ

$$\Gamma(a) = \int_0^\infty e^{-x}x^{a-1}\,dx \quad (a > 0)$$

$$\Gamma(a+1) = a\Gamma(a) \qquad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

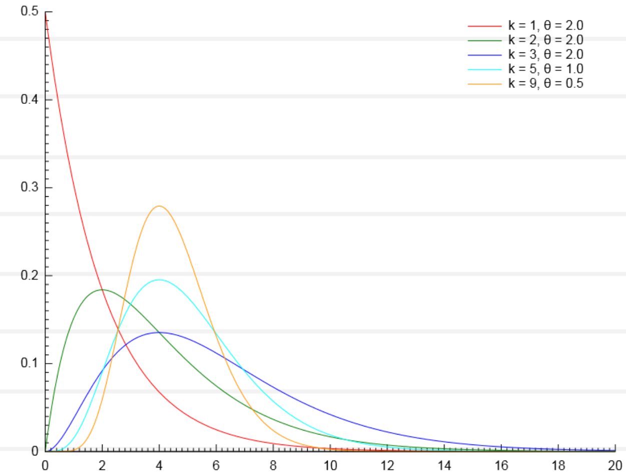
平均:

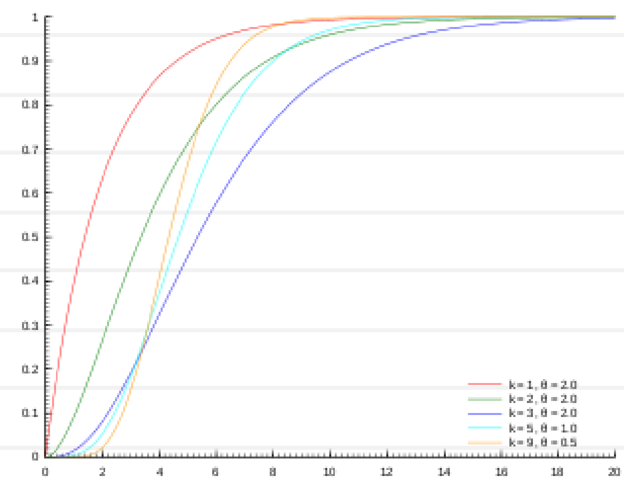
$$E[X] = \frac{a}{\lambda}$$

分散:

$$V[X] = \frac{a}{\lambda^2}$$

a: 形状パラメータ
λ: 尺度パラメータ
とも呼ばれる





a, λの値を適当に選ぶと他の分布が導出される

$a = 1$: →指数分布　$a = \dfrac{n}{2}$, $\lambda = \dfrac{1}{2}$: →指数分布

9 ガンマ分布

ガンマ関数とは

$n!$を実数で表現したい…
つまり,

$$\Gamma(n) = \boxed{(\mathrm{n}-1)!} \qquad \Gamma(x+1) = \mathrm{x}\Gamma(x) \qquad \Gamma(1) = 1$$

が成り立つような関数Γを考えればよい

$$\Gamma(x) = \int_0^\infty e^{-x} x^{a-1}\, dx$$

とおけば，部分積分により

$$\Gamma(x) = \left[-e^{-t}t^{x-1}\right]_0^\infty - \int_0^\infty (-e^{-t})(x-1)t^{x-2}dt$$

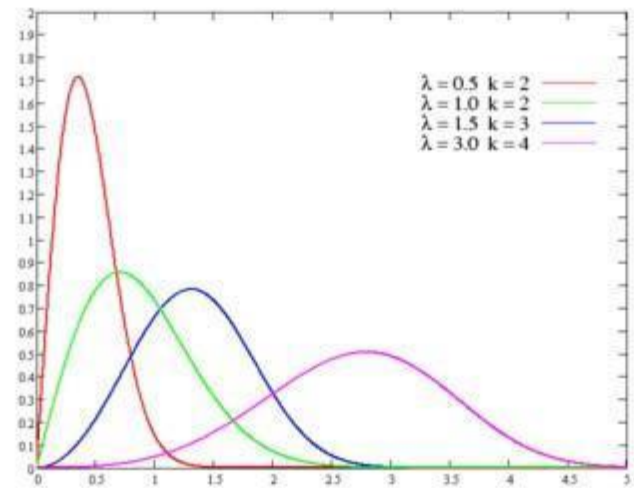
$$= 0 + (x-1)\int_0^\infty (-e^{-t})t^{x-2}dt$$

$$= (x-1)\Gamma(x-1)$$

10 ワイブル分布

確率密度関数:

$$p(x) = \begin{cases} \dfrac{ax^{a-1}}{\lambda} e^{-\frac{x^a}{\lambda}} & (x \geq 0) \\ 0 & (x < 0) \end{cases}$$



確率分布関数:

$$F(x) = 1 - e^{-\frac{t^a}{\lambda}}$$

平均:

$$E[X] = \lambda^{\frac{1}{a}} \Gamma\left(1 + \frac{1}{a}\right)$$

分散:

$$V[X] = \lambda^{\frac{2}{a}} \left[\Gamma\left(1 + \frac{2}{a}\right) - \left\{\Gamma\left(1 + \frac{1}{a}\right)\right\}^2\right]$$

ワイブル分布は部品や製品の寿命を表す分布として用いられる
故障率は

$$\lambda(t) = \frac{p(t)}{1 - F(t)} = \frac{at^{a-1}}{\lambda}$$

11 分布まとめ

離散型分布

分布名	確率密度関数	平均	分散
二項分布	$B_{n,p}(k) = {}_nC_k\, p^k (1-p)^{n-k}$	np	
ポアソン分布	$p_\lambda = \dfrac{\lambda^k e^{-\lambda}}{k!}$	λ	λ
幾何分布	$G_p(k) = (1-p)^{k-1}p$	$\dfrac{1}{p}$	$\dfrac{q}{p^2}$
超幾何分布	$f(x) = \dfrac{{}_MC_x \cdot {}_{N-M}C_{n-x}}{{}_NC_n}$	$\dfrac{nM}{N}$	$\dfrac{nM(N-M)(N-n)}{N-1}$

12 分布まとめ

連続型分布

分布名	確率密度関数	平均	分散
一様分布	$\begin{cases} \dfrac{1}{b-a} & (a \leq x \leq b) \\ 0 & (x < a, b < x) \end{cases}$	$\dfrac{b+a}{2}$	$\dfrac{(b-a)^2}{12}$
指数分布	$\begin{cases} \lambda e^{-\lambda x} & (x \geq 0) \\ 0 & (x < 0) \end{cases}$	$\dfrac{1}{\lambda}$	$\dfrac{1}{\lambda}$
正規分布	$\dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}$	m	σ^2
ガンマ分布	$\begin{cases} \dfrac{\lambda}{\Gamma(a)} (\lambda x)^{a-1} e^{-\lambda x} & (x \geq 0) \\ 0 & (x < 0) \end{cases}$	$\dfrac{a}{\lambda}$	$\dfrac{a}{\lambda^2}$
ワイブル分布	$\begin{cases} \dfrac{a x^{a-1}}{\lambda} e^{-\frac{x^a}{\lambda}} & (x \geq 0) \\ 0 & (x < 0) \end{cases}$	$\lambda^{\frac{1}{a}} \Gamma\left(1 + \dfrac{1}{a}\right)$	$\lambda^{\frac{2}{a}} \left[\Gamma\left(1 + \dfrac{2}{a}\right) - \left\{ \Gamma\left(1 + \dfrac{1}{a}\right) \right\}^2 \right]$

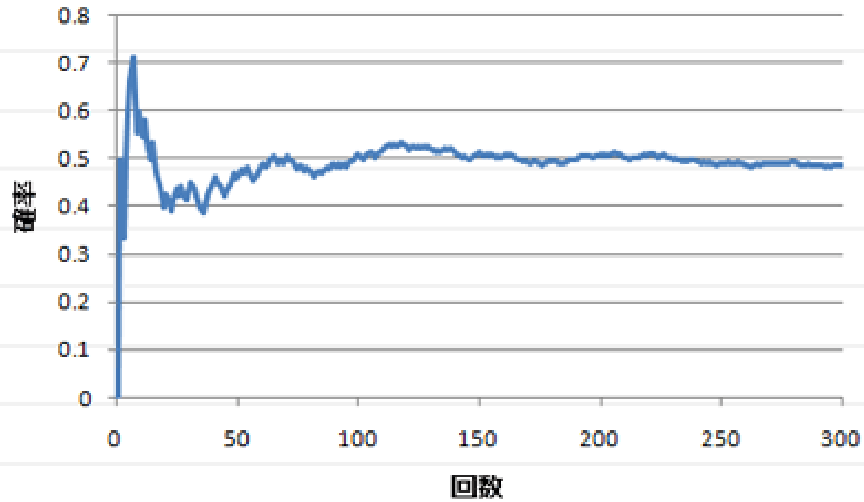
13 大数の法則

平均m，分散σ^2の分布に従う独立な確率変数$x_1, x_2, \dots, x_n$があるとする．この時，

$$y_n = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

で表される確率変数y_nは，
任意の$\epsilon > 0$に対して

$$\lim_{n \to \infty} P\{|y_n - m| < \epsilon\} = 1$$



が成り立つ．これを大数の弱法則という．

つまり，y_nが平均mに確率収束する

||

サンプル数が多いほど，観測値の平均は理論値に近づいていく

14 特性関数

特性関数という概念を用いると，確率分布の平均，分散を簡単に求めることができるようになる

確率密度分布$p(x)$に従う確率変数xの特性関数$\phi(t)$を以下のように求める

$$\phi(t) = E\left[e^{itx}\right] = \int_{-\infty}^{\infty} e^{itx} p(x)dx \longleftarrow p(x)\text{のフーリエ変換という}$$

この時，特性関数の対数をキュムラント母関数と呼び，

$$K(t) = \ln \phi$$

このキュムラント母関数の導関数から平均，分散を求めることができる

$$K'(t) = \frac{\phi'(t)}{\phi(t)} \qquad E[X] = \frac{K'(0)}{i}$$

$$K''(t) = \frac{\phi''(t)\phi(t) - \{\phi(t)\}^2}{\{\phi(t)\}^2} \qquad V[X] = -K''(0)$$

15 中心極限定理

中心極限定理

無限に多数の同じ分布をもった確率変数について適当な標準化を行うと，確率密度は正規分布$N(0,1)$に従う

独立な確率変数$x_1, x_2, \ldots, x_n$が平均m，分散σ^2の分布に従うとき，$x_i' = x_i - m$と変換すると，x_iはそれぞれ平均0，分散σ^2の分布に従う．この時，

$$y_n = \frac{x_1' + x_2' + \cdots + x_n'}{\sigma\sqrt{n}}$$

となる確率関数y_nを定義すると，特性関数の性質から

$$\phi_y(t) = \left\{1 - \frac{1}{2}\frac{t^2}{n} + O\left(\frac{1}{n}\right)\right\}^n$$

$$\lim_{n\to\infty} \phi_y(t) = e^{-\frac{t^2}{2}} \longleftarrow \text{標準正規分布の特性関数と同じ形}$$

16 多次元の確率分布

これまで1つの確率変数について取り扱ってきたが，現実には異なる2種の確率変数同士が関係を持つことも多い
例)平均歩行距離と健康状態，数学のテストと理科のテスト
これらの関係を考慮するために同時分布を考える

同時確率密度関数

ある確率変数X，Yについて$X = x$，$\mathrm{Y} = x$が同時に成り立つ確率を，$p(x, y)$と書く.これをX,Yの同時確率密度関数と呼ぶ

$$p(x, y) \geq 0 \qquad \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x, y) = 1$$

周辺確率分布

確率変数それぞれの分布は以下のようになる.この分布を周辺確率分布と呼ぶ.

$$g(x) = \int_{-\infty}^{\infty} f(x, y)dy \qquad h(y) = \int_{-\infty}^{\infty} f(x, y)dx$$

16 2変量正規分布

確率変数X, Yについて以下の値を考える

$$Cov(X, Y) = E\{(X - \mu_x)(Y - \mu_y)\}$$

この値をX, Yの共分散と呼び，それぞれの値が関連しあいながらばらつく度合いを表している

この時，XとYの関係の強さを表す相関係数は以下のよう

$$\rho_{XY} = \frac{Cov(X, Y)}{\sqrt{V(X)}\sqrt{V(Y)}}$$

2変量正規分布

平均(0,0)，分散(σ_1^2, σ_2^2)，共分散ρ_{12}を持つX_1, Y_1の分布が

$$f(x_1, x_2) = \frac{1}{2\pi\rho_1\rho_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{x_1^2}{\sigma_1^2} - \frac{2\rho_{12}x_1x_2}{\sigma_1\sigma_2} + \frac{x_2^2}{\sigma_2^2}\right)\right\}$$

で表されるとき，この分布を2変量正規分布という

17 分散共分散行列

平均(μ_1, μ_2)，分散(σ_1^2, σ_2^2)，共分散ρ_{12}の2変量正規分布は

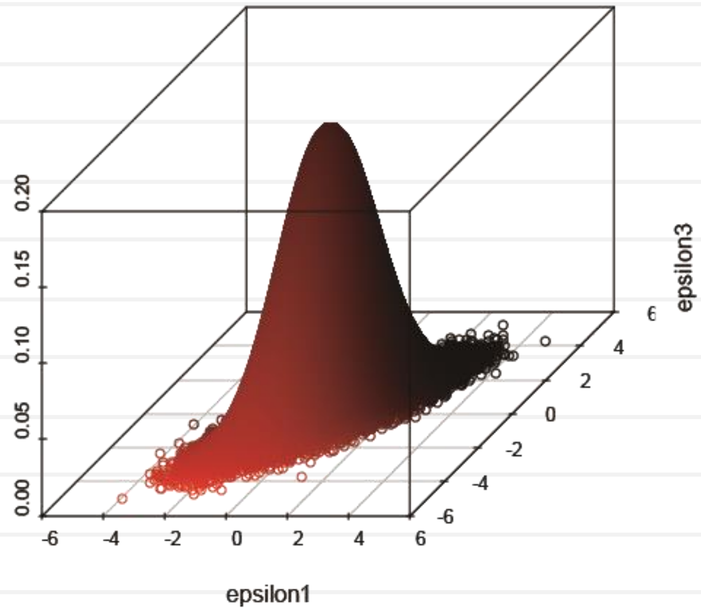
$$N((\mu_1\mu_2), (\sigma_1^2, \sigma_2^2, \rho_{12}))$$

あるいは

分散共分散行列

$$N\left((\mu_1\mu_2), \boxed{\begin{pmatrix} \sigma_1^2 & \rho_{12} \\ \rho_{12} & \sigma_2^2 \end{pmatrix}}\right)$$

のように表すことができる



多変量(3変数以上)にも適用可能で，平均と分散，共分散を与えることで同時正規分布を表現できる

18 コピュラ関数

金融分野では正規分布の代わりにコピュラ関数も使われることが多い