第 6 章　さまざまな母集団分布

1　母集団分布の例

母集団を描写するための有用な確率分布の例として，

離散的な場合・・・　ベルヌーイ分布，二項分布，ポアソン分布

連続的な場合・・・　一様分布，指数分布，正規分布

について説明する．

2　離散的な母集団分布

ベルヌーイ分布　母集団が無数の 0 と 1 からなり，1 の割合が p であるとする．例として，次のような場合を考えることができる．

(a) 母集団がすべての有権者からなり，現在の内閣を支持する者は 1，その他の者は 0 を割り振られている場合．このとき，p は母集団における内閣支持率となる．

(b) 母集団が労働力の対象となるすべての人々からなっていて，失業中の者は 1，就業中の者は 0 の場合．このとき，p は母集団失業率となる．

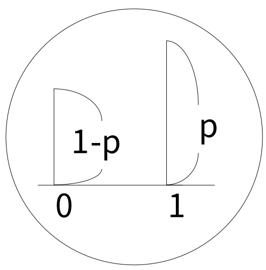
(c) 将棋の駒を投げたとき，ある特定の面が下側になるときは 1，その他のときは 0 とする場合．このとき，p は特定の面が下側になる確率となる．

以上のような場合の母集団分布は，確率変数 X を使って，次のように表すことができる．これを**ベルヌーイ分布**という．

x	0	1
P(X = x)	**1− p**	**p**

ベルヌーイ分布を図示すると次のようになる．

ベルヌーイ母集団



期待値： $E(X) = p$　　　**分散：** $V(X) = p(1-p)$

二項分布 成功の確率 p, 失敗の確率 1-p の n 回の独立試行において, 成功する回数 X の分布

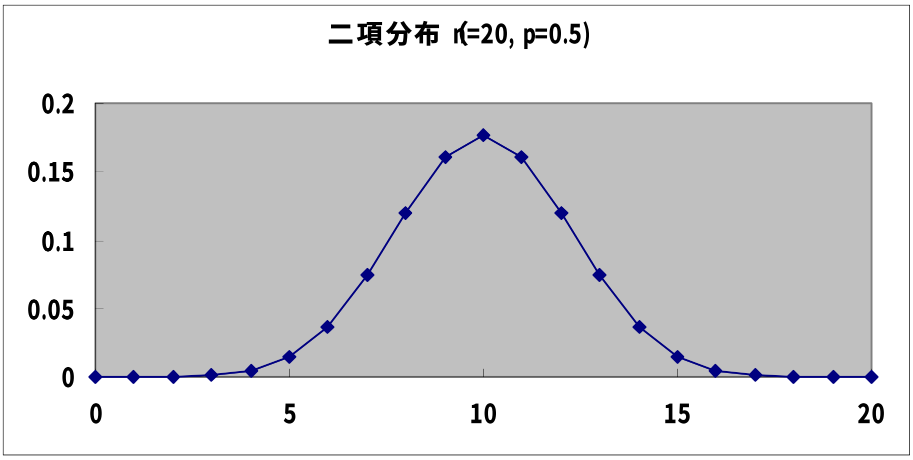
$$X \sim B(n,p) \quad \Leftrightarrow \quad P(X=k)={}_nC_k\, p^k (1-p)^{n-k} \qquad (k=0,\ 1,\ \cdots,\ n)$$

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

二項分布は, 独立なベルヌーイ確率変数の和の分布である.

このことから, 期待値と分散は次のようになる.

期待値: $E(X)=np$ **分散**: $V(X)=np(1-p)$



〔例題 6.2〕 プロ野球の 3 割打者が, 20 打席で 7 本以上ヒットを打つ確率と, 400 打席で 140 本以上ヒットを打つ確率を求めよ.

(解) どちらの確率も, 打席数は異なるものの, 3 割 5 分以上の成績を残す確率である. 打席数を n, ヒット数を X とするとき,

$$X \sim B(n,\ p) \qquad p=0.3$$

と考える. このとき, k 本以上ヒットを打つ確率は,

$$P(X\ge k)=\sum_{j=k}^{n} {}_nC_j\, p^j (1-p)^{n-j}$$

で与えられる. これより, 次の結果を得る.

$n=20$ のとき, $P(X\ge 7)=0.392$

$n=400$ のとき, $P(X\ge 140)=0.0177$

二項分布は，各回で 0 か 1 をとる独立なベルヌーイ確率変数の和の分布として解釈されるが，とりうる値を 0，1 以外にしたときの和の分布を考えることもできる．例えば

$$P(Y_i = a) = p, \qquad P(Y_i = -b) = 1 - p \qquad (i = 1,\ 2,\ \cdots,\ n)$$

となるような n 個の独立な確率変数を導入して，

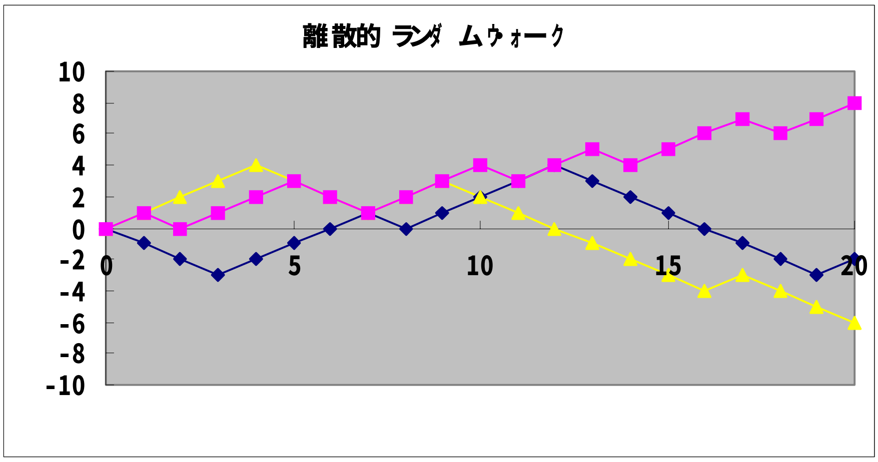
$$S_n = Y_1 + Y_2 + \cdots + Y_n$$
$$= S_{n-1} + Y_n \qquad (S_0 = 0)$$

を定義する．ここで，a，b は正の数である．この系列は，上昇あるいは下降をくり返して得られる時点 n における累積結果であり，**離散的ランダム・ウォーク**という．

確率分布： $P\left(S_n = ka - (n-k)b\right) = {}_nC_k\, p^k (1-p)^{n-k} \qquad (k = 0,\ 1,\ \cdots,\ n)$

期待値： $E(S_n) = n(ap - b(1-p))$ 　　　**分散：** $V(S_n) = n(a+b)^2 p(1-p)$

ここで考えたランダム・ウォークは，とりうる値が 2 通りしかない確率変数から生成されたものである．しかし，とりうる値が多いもっと一般の離散的確率変数あるいは連続的確率変数から生成することもできる．そのような一般化ランダム・ウォークは，株価や為替レートの動きを描写するモデルとして，**オプション理論**で使われている．特に，これらの対数値系列が一般化ランダム・ウォークに従うと仮定する場合が多い．

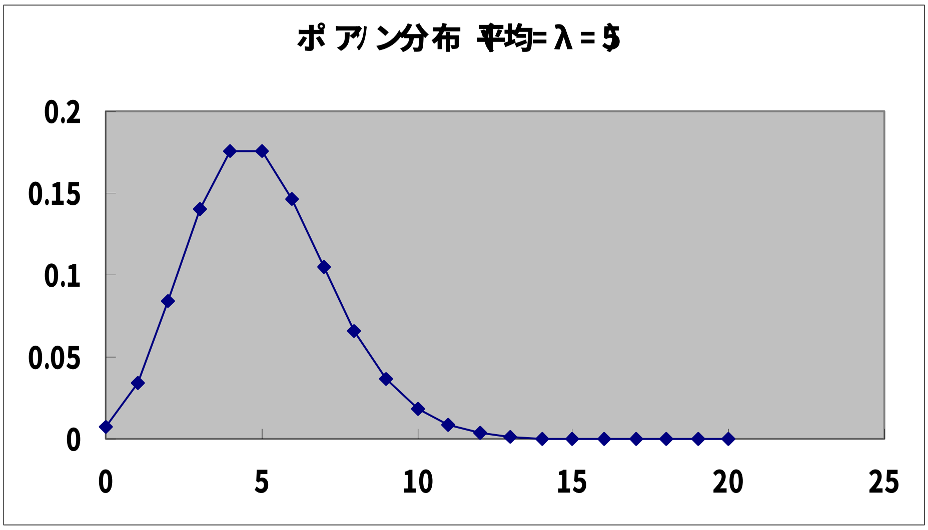


ポアソン分布

特定の時間内で発生する事象の回数，例えば，駐車場に入って来る車の台数，レストランに来る客の数，かかってくる電話の回数，飛行機事故の回数，放射性物質の微粒子放出の回数などの母集団分布を描写する確率分布

$$X \sim Po(\lambda) \quad \Leftrightarrow \quad P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad (x = 0,\ 1,\ \cdots)$$

e は自然対数の底，λ（ラムダ）は正の値をとるパラメータである．



自然対数の底 e の性質

(1) **指数関数** $y = a^x$ **において，** $x = 0$ **における接線の傾きが** 1 **となる** a

(2) $e = \lim_{n\to\infty}\left(1+\dfrac{1}{n}\right)^n = 2.718282$

(3) e^x **の任意階の導関数は** e^x

(4) $e^x = 1 + x + \dfrac{x^2}{2} + \dfrac{x^3}{3!} + \cdots = \sum_{j=0}^{\infty}\dfrac{x^j}{j!}$ **（テーラー展開）**

〔**例題 6.6**〕 あるデパートの女性用注文服売り場には，1 時間当り平均 2 人の客が来るという．来客数がポアソン分布に従うとして，次の確率を計算せよ．
(a) 1 時間に 4 人以上の客が来る確率
(b) 1 時間に少なくとも 1 人の客が来る確率
(c) 2 時間の間まったく客が来ない確率
(d) 次の客が来るまでの間隔が 30 分を越える確率

（解）1 時間の来客数を X とすると，X は平均 λ= 2 のポアソン分布に従うと考えられる．このことから，次の結果を得る．

(a) $P(X \ge 4) = \sum_{x=4}^{\infty}\dfrac{e^{-2}2^x}{x!} = 1 - e^{-2}\sum_{x=0}^{3}\dfrac{2^x}{x!} = 0.143$

(b) $1 - P(X = 0) = 1 - e^{-2} = 0.865$

(c) 2 **時間の来客数を** Y **とすれば，** Y **〜** $Po(4)$ **であるから，求める確率は**

$P(Y = 0) = e^{-4} = 0.018$ $\left(P(X = 0)\times P(X = 0) = e^{-2}\times e^{-2}\right.$ **としてもよい**$\left.\right)$

(d) 30 **分間の来客数を** Z **とするとき，** Z **〜** $Po(1)$ **となるから，求める確率は**

$P(Z = 0) = e^{-1} = 0.368$

ポアソン分布と二項分布の関係

$X \sim B(n,p)$ **において，**$n \to \infty$**，**$p \to 0$**，ただし，**$np = \lambda =$**一定とするとき，**

$$P(X=x) = {}_nC_x p^x (1-p)^{n-x} \quad \to \quad \frac{e^{-\lambda}\lambda^x}{x!}$$ **となる．**

〔**例題** 6.7〕 ある病気は，感染しても発病する可能性が非常に低く，感染者が発病する確率は 0.03 であるという．感染者が 100 人の場合，発病する人数の正確な確率分布と近似分布を求めよ．

（解）

正確な 確率 $= P(X=x) = {}_nC_x p^x (1-p)^{n-x} \qquad n=100, \quad p=0.03, \quad x=0,1,\cdots,100$

近似確率 $= \dfrac{e^{-\lambda}\lambda^x}{x!} \qquad \lambda = np = 3, \qquad x = 0,1,\cdots$

x	0	1	2	3	4	5	6	7	8
正確	0.048	0.147	0.225	0.227	0.171	0.101	0.050	0.021	0.007
近似	0.050	0.149	0.224	0.224	0.168	0.101	0.050	0.022	0.008

3　連続的な母集団分布

一様分布

確率変数 X が，区間 [a, b] 上の値をランダムにとるような分布

$$X \sim U[a,\ b] \quad \Leftrightarrow \quad f(x) = \frac{1}{b-a} \qquad a \le x \le b$$

$$E(X) = \frac{a+b}{2}, \quad V(X) = \frac{(b-a)^2}{12}$$

一様分布は水平な密度関数をもつが，和の分布の密度関数は水平のままでとどまることはできない．

〔**例題** 6.8〕 確率変数 X と Y が互いに独立に U[0, 1] に従うとき，X+Y の分布の密度関数を求めよ．

（解）まず，Z=X+Y の分布関数 G(z) を求めることにする．X と Y の同時密度関数を f(x, y) とすると，

$$f(x,\ y) = 1 \qquad 0 \le x \le 1, \quad 0 \le y \le 1$$

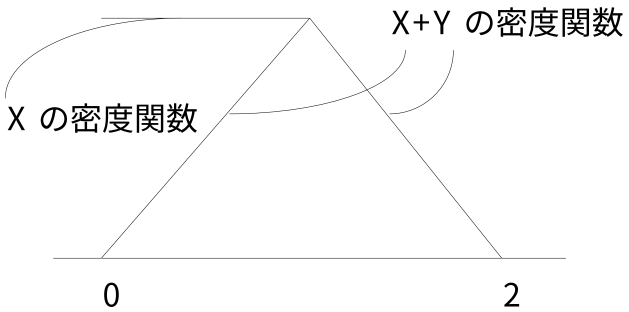
である．このことから，

$$G(z) = P(X+Y \le z) = \iint_{x+y\le z} f(x,y)\,dxdy = \iint_{x+y\le z} dxdy$$

$$= \begin{cases} z^2/2 & 0 \le z \le 1 \\ 1-(z-2)^2/2 & 1 \le z \le 2 \\ 0 & \text{その 他} \end{cases}$$

となる．したがって，求める密度関数 g(z) は次のようになる．

$$g(z)=\frac{dG(z)}{dz}=\begin{cases} z & 0\le z\le 1 \\ 2-z & 1\le z\le 2 \\ 0 & \text{その他} \end{cases}$$

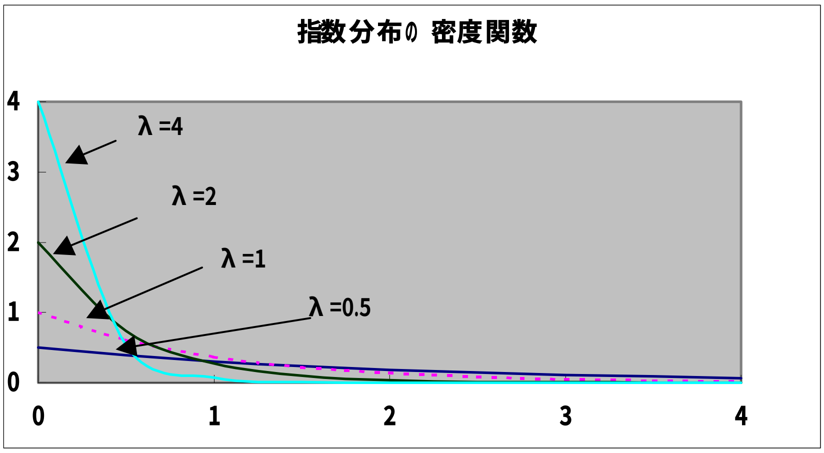


$\boxed{\text{指数分布}}$ 待ち時間や寿命の分布

$$X \sim Ex(\lambda) \quad \Leftrightarrow \quad f(x)=\lambda e^{-\lambda x} \qquad 0\le x$$

$$E(X)=1/\lambda,\quad V(X)=1/\lambda^2$$

λ は正の定数である．



ポアソン分布と指数分布との関係

ある事象の発生回数がパラメータ λ のポアソン分布に従うならば，その事象の発生時間間隔はパラメータ λ の指数分布に従う．

指数分布の「記憶喪失性」　　P(X>s+t|X>s)=P(X>t)

生存関数（survival function）　　P(X>t)

故障関数（hazard function）　　f(x)/P(X>t)

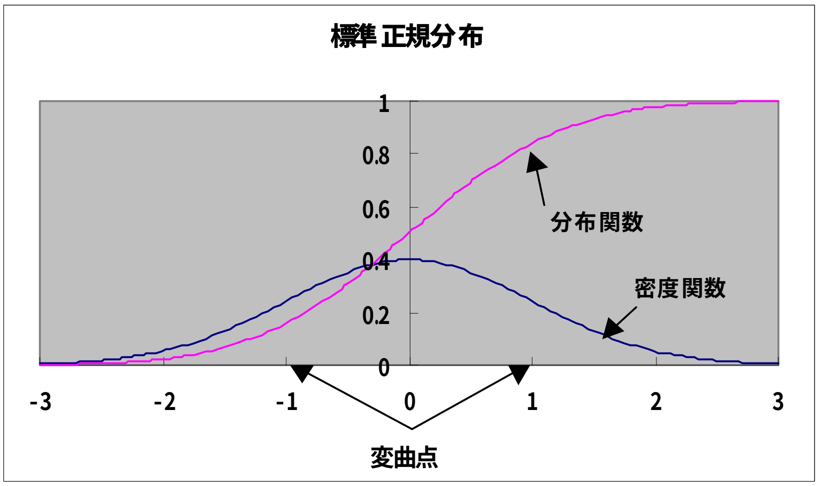
$\boxed{\text{正規分布}}$=Normal distribution (Normal とは，″ありふれた″ の意)

対称，連続的な母集団を描写する分布として最も頻繁に使われる．

$$X \sim N(\mu,\ \sigma^2) \Leftrightarrow f(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

$$E(X) = \mu, \qquad V(X) = \sigma^2$$

N(0，1) の分布を標準正規分布という．



正規分布の性質－その 1

$$X \sim N(\mu,\ \sigma^2) \quad \Rightarrow \quad aX + b \sim N(a\mu + b,\ a^2\sigma^2)$$

$$Z = \frac{X-\mu}{\sigma} \sim N(0,1)$$ ＝**標準正規分布**

確率の計算

$$P(X < c) = P\left(\frac{X-\mu}{\sigma} < \frac{c-\mu}{\sigma}\right) = P\left(Z < \frac{c-\mu}{\sigma}\right) \rightarrow$$ **正規分布表（**245 **ページ）**

例 1) $X \sim N(50,100)$ **のとき，** $P(X > 70) = P\left(Z > \frac{70-50}{10}\right) = P(Z > 2) = 0.0228$

例 2) $X \sim N(\mu,\sigma^2)$ **のとき，**

$$P(|X-\mu| < 2\sigma) = P(|Z| < 2) = 1 - 2 \times P(Z > 2) = 0.9544$$

(例 3) **初婚の夫婦の年齢差** X **(＝夫の年齢－妻の年齢)～**$N(2.9,9)$ **のとき，妻の方が年上の確率** $= P(X < -1) = P(Z < (-1-2.9)/3) = P(Z < -1.3) = 0.097$

分位点の計算

$p = P(X < a)$ **をみたす** a **を求める．**

$p = P\left(Z < \dfrac{a-\mu}{\sigma}\right)$ $\rightarrow$ **正規分布表**(246 **ページ**)

例 1) X **〜** $N(50,100)$ **のときの上側** 5% **点**

$0.05 = P(X > x) = P\left(Z > \dfrac{x-50}{10}\right)$ **より，** $\dfrac{x-50}{10} = 1.645 \quad \therefore \quad x = 66.45$

（問題）ある試験の点数が平均 65，標準偏差 10 の正規分布に従うとする．5 段階評価 A，B，C，D，E の割合を 1:2:4:2:1 とするとき，各評価の境界点を求めよ．

（解）A と B の境界点は，

$0.1 = P(X > a) = P(Z > (a-65)/10)$ **より，** $(a-65)/10 = 1.282 \quad \therefore a = 77.82$

B と C の境界点は，

$0.3 = P(X > b) = P(Z > (b-65)/10)$ **より，** $(b-65)/10 = 0.524 \quad \therefore b = 70.24$

対称性から，C と D の境界点 = 65-5.24 = 59.76　D と E の境界点 = 65-12.82 = 52.18

正規分布の性質－その 2

$$X_i \ \sim\ N(\mu_i,\ \sigma_i^2) \quad \Rightarrow \quad \sum_{i=1}^{n} a_i X_i \ \sim\ N(\mu,\ \sigma^2)$$

ただし， $\mu = \sum_{i=1}^{n} a_i \mu_i,\quad \sigma^2 = V\left(\sum_{i=1}^{n} a_i X_i\right)$

特に， $X_1, \cdots, X_n$ **が互いに独立ならば，** $\sigma^2 = \sum_{i=1}^{n} a_i^2 \sigma_i^2$

(**例**) $X_1,\ X_2,\ X_3,\ X_4$ **が互いに独立に** $N(0,\sigma^2)$ **に従うとき，**

(a) **和が** $[-\sigma,\sigma]$ **の中にある確率**

$$= P(-\sigma < X_1 + X_2 + X_3 + X_4 < \sigma) = P(-1/2 < Z < 1/2)$$

(b) **少なくとも** 1 **つが** $[-\sigma,\sigma]$ **の外にある確率**

$$= 1 - P(|X_1| < \sigma, |X_2| < \sigma, |X_3| < \sigma, |X_4| < \sigma) = 1 - \{P(|Z| < 1)\}^4$$

(c) **最大値が** σ **以下となる確率**

$$= P(\max(X_1, X_2, X_3, X_4) < \sigma) = P(X_1 < \sigma, X_2 < \sigma, X_3 < \sigma, X_4 < \sigma) = \{P(Z < 1)\}^4$$

(d) **最小値が** σ **以下となる確率**

$$= P(\min(X_1, X_2, X_3, X_4) < \sigma) = 1 - P(\min(X_1, X_2, X_3, X_4) > \sigma) = 1 - \{P(Z > 1)\}^4$$