確率統計 A

第 2 章 確率変数と分布 (Part 4)

(2012 年 5 月 22 日)

講師:若木宏文

E-mail: wakaki@math.sci.hiroshima-u.ac.jp

Web: http://home.hiroshima-u.ac.jp/~wakaki/lecture/probstatA12/

2.5. 離散型・連続型分布
2.5.1 1次元の場合

連続型確率変数と確率密度関数

身長, 温度, バスの待ち時間などのように連続する値をとる場合を考える. すなわち, X のとりうる値は実数全体のある部分集合であるとする. とくに, X の分布関数が

$$\mathrm{P}(X \le x) = F(x) = \int_{-\infty}^{x} f(t)dt, \quad {}^{\forall}x \in \mathbb{R}$$

と表せるとき, **連続型確率変数** (**Continuous Type Random Variable**) といい, その確率分布を**連続型分布**という. また, $\mathbb{R}$ 上の関数 $f(x)$ を**確率密度関数** (**Probability Density Function**, **pdf**) という.

確率密度関数の性質

先の定義において,

$$\mathrm{P}(a < X \le b) = \int_{a}^{b} f(t)dt,$$

であり, $f(x)$ の連続点で

$$\frac{d}{dx}F(x) = f(x),$$

が成立する. 定義から確率密度関数は次の性質をもつ.

$$[1].\ f(x) \ge 0, \quad {}^{\forall}x \in \mathbb{R}$$

$$[2].\ \int_{-\infty}^{\infty} f(t)dt = 1$$

一様分布

一様分布 (Uniform Distribution)

確率密度関数が

$$f(x) = \begin{cases} \dfrac{1}{b-a} & (a < x \le b) \\ 0 & (\text{その他}) \end{cases}$$

で与えられる分布を区間 $(a,b]$ 上の**一様分布**といい, $U(a,b]$ で表す.

定理 2.9

連続型確率変数 X の分布関数 $F(x)$ は, $a<x<b$ において, 狭義単調増加で $F(a)=0$, $F(b)=1$ であるとする ($a=-\infty$ あるいは $b=\infty$ でもよい). このとき, 確率変数 $Y=F(X)$ は一様分布 $(0, 1]$ に従う. 逆に Yが一様分布 $(0,1]$ に従うとする. このとき $X=F^{-1}(Y)$の分布関数はFになる.

定理 2.9 の証明

乱数の発生法

定理2.9より, 一様分布に従う乱数

$y_1, \ldots, y_n$

から, 変換 $x_j=F^{-1}(y_j)$ により, 分布関数 F をもつ乱数

$x_1, \ldots, x_n$

を発生させることができる.

正規分布

正規分布 (Normal Distribution)

確率密度関数が

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

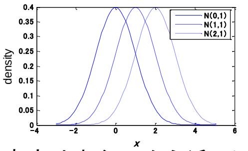
で与えられる分布を平均 μ, 分散 σ^2 の**正規分布**といい, $N(\mu,\sigma^2)$ で表す. ここに $-\infty<\mu<\infty$, $\sigma^2>0$. 特に $\mu=0$, $\sigma^2=1$ のとき, $N(0,1)$ を**標準正規分布** (**Standard Normal Distribution**) という. $N(0,1)$ の確率密度関数は

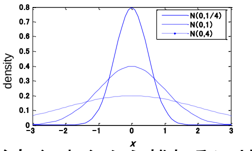
$$\phi(x) = f(x;0,1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$$

と表される.

正規分布の特性

1. μ は中心の位置をσ^2 は分布の裾の広がりを表している.
   - もし μ だけ変われば, 中心の位置だけずれる.
   - もし σ^2 だけ大きくなれば, 分布の裾が広くなる.





2. 密度は中心 μ から近いほど高く, 中心から離れるほど低くなる.
   - $\mu-\sigma$ から $\mu+\sigma$ までの範囲の累積確率は約 68%.
   - $\mu-2\sigma$ から $\mu+2\sigma$ までの範囲の累積確率は約 95%.
   - $\mu-3\sigma$ から $\mu+3\sigma$ までの範囲の累積確率は約 99.7%.

標準正規分布の上側確率の求め方

表D.1 (p318) はxをパーセント点 (閾値) としたときの標準正規分布の上側確率を表している. ($\mathrm{P}(Z>x)=?$)

横軸はxの小数点第2位を表している!

縦軸はxの小数点第1位までを表している!

x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010

もし $p=\mathrm{P}(Z>1.36)$となるpを求めたいなら, $1.36=1.3+0.06$

$\mathrm{P}(Z>1.36)=0.0869$

標準正規分布の上側パーセント点の求め方

表D.2 (p319) は上側確率を p としたときのパーセント点 (閾値) を表している. ($p=\mathrm{P}(Z>?)$)

$\mathrm{P}(Z>x)=0.082$
$x=1.392.$

横軸はpの小数第3位を表している!

縦軸はpの小数第2位までを表している!

p	.000	.001	.002	.003	.004	.005	.006	.007	.008	.009
.00	∞	3.090	2.878	2.748	2.652	2.576	2.512	2.457	2.409	2.366
.01	2.326	2.290	2.257	2.226	2.197	2.170	2.144	2.120	2.097	2.075
.02	2.054	2.034	2.014	1.995	1.977	1.960	1.943	1.927	1.911	1.896
.03	1.881	1.866	1.852	1.838	1.825	1.812	1.799	1.787	1.774	1.762
.04	1.751	1.739	1.728	1.717	1.706	1.695	1.685	1.675	1.665	1.655
.05	1.645	1.635	1.626	1.616	1.607	1.598	1.589	1.580	1.572	1.563
.06	1.555	1.546	1.538	1.530	1.522	1.514	1.506	1.499	1.491	1.483
.07	1.476	1.468	1.461	1.454	1.447	1.440	1.433	1.426	1.419	1.412
.08	1.405	1.398	1.392	1.385	1.379	1.372	1.366	1.359	1.353	1.347
.09	1.341	1.335	1.329	1.323	1.317	1.311	1.305	1.299	1.293	1.287
.10	1.282	1.276	1.270	1.265	1.259	1.254	1.248	1.243	1.237	1.232
.20	0.842	0.838	0.834	0.831	0.827	0.824	0.820	0.817	0.813	0.810
.25	0.674	0.671	0.668	0.665	0.662	0.659	0.656	0.653	0.650	0.646
.30	0.524	0.522	0.519	0.516	0.513	0.510	0.507	0.504	0.502	0.499
.40	0.253	0.251	0.248	0.246	0.243	0.240	0.238	0.235	0.233	0.230

もし$\mathrm{P}(Z>x)=0.082$となるxを求めたいなら, $0.082=0.08+0.002.$

その他の連続型確率分布

1. 指数分布 (Exponential Distribution): 確率密度関数が

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

で与えられる分布をパラメータ $\lambda\ (>0)$ を持つ**指数分布**といい, $Ex(\lambda)$ で表す.

2. ガンマ分布 (Gamma Distribution): 確率密度関数が

$$f(x)=\begin{cases}\frac{1}{\Gamma(r)}\lambda^r e^{-\lambda x}x^{r-1} & (x\ge 0)\\ 0 & (x<0)\end{cases}$$

で与えられる分布をパラメータ $r\ (>0)$, $\lambda\ (>0)$ を持つ**ガンマ分布**といい, $Ga(r,\lambda)$ で表す. ただし$\Gamma(x)$ はガンマ関数である. $Ga(1,\lambda)=Ex(\lambda)$ という関係式が成り立つ.

2.5. 離散型・連続型分布 2.5.2 2次元の場合

2 次元確率関数と確率密度関数の性質

- 確率関数 (離散型)

(1) $f(x_i,y_j)\ge 0,\ i,j=1,2,\ldots.$ (2) $\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}f(x_i,y_j)=1.$

同時分布関数は, $F(x,y)=\sum_{x_i\le x}\sum_{y_i\le y}f(x_i,y_j)$

- 確率密度関数 (連続型)

(1) $f(x,y)\ge 0,\ {}^{\forall}(x,y)\in\mathbb{R}^2$ (2) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)dxdy=1.$

同時分布関数は, $F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f(u,v)dudv$

$$\left(\frac{\partial^2}{\partial x\partial y}F(x,y)=f(x,y)\right)$$

2 項分布の拡張

- 1回の試行で事象 A, $B(=A^c)$ のいずれかが起こりうるものとし, A が起こる確率を p とすると, B が起こる確率は $1-p$ である. このとき n 回の独立な試行で, A が起こる回数を X とすれば, X は 2 項分布に従う.
- 1回の試行で事象 A, B, C のいずれかが起こりうるものとし, A が起こる確率を p_1, B が起こる確率を p_2 とすると, C が起こる確率は $1-p_1-p_2$ である. このとき n 回の独立な試行で, A が起こる回数を X, B が起こる回数を Y とすれば (X,Y) は 3 項分布に従う.

三項分布

3 項分布 (Trinomial Distribution):

確率関数が

$$f(x,y) = \mathrm{P}(X=x, Y=y)$$
$$= \frac{n!}{x!y!(n-x-y)!} p_1^x p_2^y (1-p_1-p_2)^{n-x-y}$$

で与えられる分布をパラメータ n, (p_1, p_2) $(0 \le p_1, 0 \le p_2, p_1+p_2 \le 1)$ の**3項分布**といい, $M_3(n,(p_1,p_2))$ で表す. ただし, $(x,y) \in D = \{(x,y);$ xとyは非負整数, $x+y \le n\}$

2 次元正規分布 (1/2)

2 次元正規分布 (Bivariate Normal Distribution)

確率密度関数が

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right\}\right]$$

で与えられる分布を**2 次元正規分布**という. ここに $-\infty < \mu_1 < \infty$, $\sigma_1 > 0$, $-\infty < \mu_2 < \infty$, $\sigma_2 > 0$, $|\rho| < 1$.

2 次元正規分布 (2/2)

$\boldsymbol{x} = (x,y)'$, $\boldsymbol{\mu} = (\mu_1, \mu_2)'$,

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$ (Σは対称な正則行列)

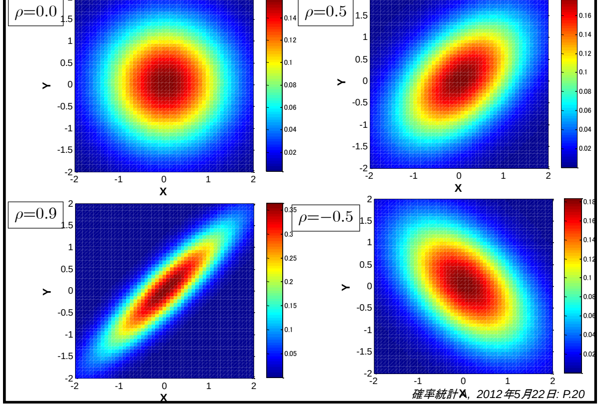
とおくと,

$$f(x,y) = \left(\frac{1}{2\pi}\right)^{2/2} |\,\Sigma\,|^{-1/2} \exp\ -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$$
$$= \left(\frac{1}{2\pi}\right) |\,\Sigma\,|^{-1/2} \exp\ -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$$

と書き換えることができるので, (X,Y) の分布を $N_2(\boldsymbol{\mu}, \Sigma)$

と表す.

2 次元正規分布の形



定理 2.10

(周辺分布)

1. (X,Y) が3項分布 $M_3(n,(p_1,p_2))$ に従うとする. このとき, X, Y はそれぞれ2項分布 $B(n,p_1)$, $B(n,p_2)$ に従う.
2. (X,Y) が2次元正規分布 $N_2(\boldsymbol{\mu},\Sigma)$ に従うとする. このとき, X, Y はそれぞれ正規分布 $N(\mu_1,{\sigma_1}^2)$, $N(\mu_2,{\sigma_2}^2)$ に従う.

定理 2.10 の証明

定理 2.11

(X,Y) は離散型あるいは連続型で, 同時確率密度関数 (または確率関数) および周辺確率密度関数 (または確率関数) を $f_{X,Y}(x,y)$, $f_X(x)$, $f_Y(y)$ とする. X と Y が独立であることと,

$$f_{X,Y}(x,y)=f_X(x)f_Y(y),$$
$$(x,y)\in D=\{(x,\ y);\ f_{X,Y}(x,y)>0\}$$

が成り立つことは同値である.

定理 2.11 の証明