数学B　統計的な推測「正規分布」

~高校数学のまとめ~

連続型確率変数・確率密度関数・分布曲線

実数のある区間全体に値をとる確率変数 X に対して，

1 つの関数 $y=f(x)$ が対応して次の性質をもつとする．

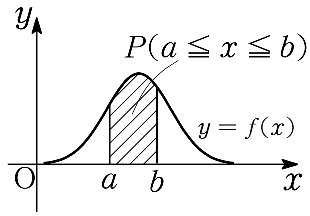
1. $f(x) \geqq 0$

2. X が $a \leqq x \leqq b$ の範囲の値をとる確率 $P(a \leqq X \leqq b)$ は

   曲線 $y=f(x)$ と x 軸および 2 直線 $x=a$, $x=b$ で

   囲まれた部分の面積に等しい．

   つまり $P(a \leqq X \leqq b)=\displaystyle\int_a^b f(x)\,dx$



3. 曲線 $y=f(x)$ と x 軸の間の面積は 1 である．

   つまり $\displaystyle\int_{-\infty}^{\infty} f(x)\,dx=1$

このとき

X を **連続型確率変数**（れんぞくがたかくりつへんすう）といい，関数 $f(x)$ を X の **確率密度関数**（かくりつみつどかんすう）という．

$y=f(x)$ のグラフをその **分布曲線**（ぶんぷきょくせん）という．

連続分布

確率密度関数によって確率分布が定められているとき，

その分布を **連続分布**（れんぞくぶんぷ）という．

離散型確率変数

連続型確率変数に対して，とびとびの値をとる確率変数を

離散型確率変数（りさんがたかくりつへんすう）という．

㊟補 大雑把に説明すると，連続型は $\int$，離散型は Σ

連続型確率変数の期待値（平均）と分散，標準偏差

連続型確率変数 X のとり得る値の範囲が $a \leqq X \leqq b$ であり，

その確率密度関数を $f(x)$ とするとき

1. 確率変数 X の期待値は $E(X)=\displaystyle\int_a^b xf(x)\,dx$
2. $E(X)=m$ とする．

   確率変数 X の分散は $$V(X)=\int_a^b (x-m)^2 f(x)\,dx = \int_a^b x^2 f(x)\,dx - m^2$$
3. 確率変数 X の標準偏差は $\sigma(X)=\sqrt{V(X)}$

補 離散型確率変数の Σ を $\int$ で置き換えている．

要

確率変数 X が離散型，連続型かで次のようになる．

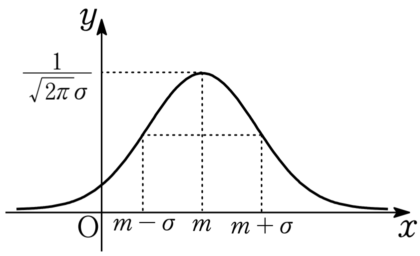
	離散型	連続型
期待値 $E(X)$	$\displaystyle\sum_{k=1}^{n} x_k P(X=k)$	$\displaystyle\int_a^b xf(x)\,dx$
分散 $V(X)$	$\displaystyle\sum_{k=1}^{n} (x_k-m)^2 P(X=k)$	$\displaystyle\int_a^b (x-m)^2 f(x)\,dx$
分散 $V(X)$	$\displaystyle\sum_{k=1}^{n} {x_k}^2 P(X=k)-\{E(X)\}^2$	$\displaystyle\int_a^b x^2 f(x)\,dx - m^2$

正規分布

連続型確率変数 X の確率密度関数 $f(x)$ が

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

で与えられているとき



X は 正規分布（せいきぶんぷ） $N(m,\,\sigma^2)$ に従う という.

ここで m は X の平均, σ^2 は X の分散である. なお, σ は X の標準偏差である.

$y=f(x)$ のグラフを正規分布曲線（せいきぶんぷきょくせん）といい, 次の性質がある.

1. 直線 $x=m$ に関して対称になる.
2. $f(x)$ の値は $x=m$ で最大値をとる.
3. x 軸を漸近線とする.
4. 曲線の山は, σ が大きくなるほど低くなって横に広がり, σ が小さくなるほど高くなって対称軸 $x=m$ のまわりに集まる.

(補) 正規分布を意味する英語は Normal distribution

(補) $f(x)$ が確率密度関数になることは知られているとする.

(考) 4 σ が大きいとばらつきも大きくなり, 小さいとばらつきが小さくなる.

正規分布の標準化

確率変数 X が正規分布 $N(m,\, \sigma^2)$ に従うとき

$$Z = \frac{X-m}{\sigma} \quad \text{すなわち} \quad Z = \frac{X-(X\text{の平均})}{(X\text{の標準偏差})}$$

とおくと 確率変数 Z は $N(0,\, 1)$ に従う.

この Z を **標準化** した確率変数という.

㊟ X が $N(m,\, \sigma^2)$ に従うとき $E(X)=m,\ V(X)=\sigma^2$

確率変数 Z の平均は

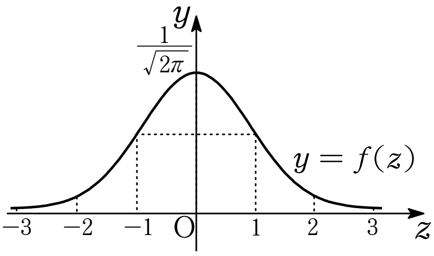
$$E(Z)=E\left(\frac{X-m}{\sigma}\right)=E\left(\frac{1}{\sigma}X-\frac{m}{\sigma}\right)=\frac{1}{\sigma}E(X)-\frac{m}{\sigma}=\frac{1}{\sigma}\cdot m-\frac{m}{\sigma}=0$$

確率変数 Z の分散は

$$V(Z)=V\left(\frac{X-m}{\sigma}\right)=V\left(\frac{1}{\sigma}X-\frac{m}{\sigma}\right)=\left(\frac{1}{\sigma}\right)^2V(X)=\frac{1}{\sigma^2}\cdot\sigma^2=1$$

標準正規分布

正規分布 $N(0,\ 1)$ を 標準正規分布（ひょうじゅんせいきぶんぷ） という.



標準正規分布の確率密度関数は

$$f(z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}$$

$y=f(z)$ のグラフを 標準正規分布曲線 といい，次の性質がある.

には次の性質がある.

$\boxed{1}$ $f(z)$ の値は $z=0$ で最大値をとる.

$\boxed{2}$ z 軸を漸近線とする.

$\boxed{3}$ y 軸に関して対称になるので $\displaystyle\int_0^{\infty}f(z)\,dz=\int_{-\infty}^{0}f(z)\,dz=0.5$

$\boxed{4}$ $P(0\leqq Z\leqq u)=\displaystyle\int_0^{u}f(z)\,dz$ は正規分布表から求まる.

要

確率変数 X が正規分布 $N(m,\ \sigma^2)$ に従うとき

$$Z=\frac{X-m}{\sigma}\quad すなわち\quad Z=\frac{X-(X\,の平均)}{(X\,の標準偏差)}$$

と標準化すると，確率変数 Z は標準正規分布 $N(0,\ 1)$ に従う.

標準正規分布曲線 $y=f(z)$ において z 座標 1 が σ に対応する.

例えば

$$m-\sigma\leqq x\leqq m+\sigma \iff -1\leqq z\leqq 1$$

すなわち

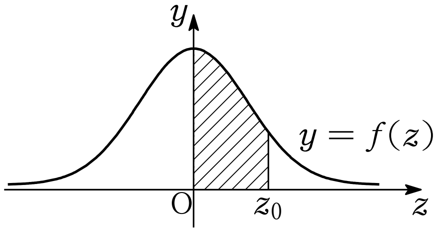
$$P(m-\sigma\leqq X\leqq m+\sigma)=P(-1\leqq Z\leqq 1)$$

正規分布表

標準正規分布曲線 $y = f(z)$ に対し

右図斜線部の面積

$$\int_0^{z_0} f(z)\,dz$$



の近似値を次のように表にしたものを **正規分布表** という.

z_0	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

上表は小数第５位を四捨五入して小数第４位までの値を表わしている.

確率変数 Z が標準正規分布 $N(0,\,1)$ に従うとすると，次の確率が求まる.

$$P(0 \leqq Z \leqq z_0) = \int_0^{z_0} f(z)\,dz$$

例 確率変数 Z が $N(0,\,1)$ に従うとして $P(0 \leqq Z \leqq 1.96) = \int_0^{1.96} f(z)\,dz = 0.4750$

二項分布の正規分布による近似

二項分布 $B(n,\ p)$ に従う確率変数 X は n が十分大きいとき

近似的に正規分布 $N\big(np,\ np(1-p)\big)$ に従う．

このことから

$$Z=\frac{X-np}{\sqrt{np(1-p)}}$$

とおくと 確率 Z は標準正規分布 $N(0,\ 1)$ に従う．

例 X が $B\left(180,\ \dfrac{1}{6}\right)$ に従うとき，180 が十分大きいので $N(30,\ 25)$ に従う．

$$Z=\frac{X-30}{\sqrt{25}}=\frac{X-30}{5}$$

とおくと Z は $N(0,\ 1)$ に従う．

$$P(30\leqq X\leqq 40)=P\left(0\leqq \frac{X-30}{5}\leqq 2\right)=P(0\leqq Z\leqq 2)=0.4772$$