平成27年度 数学科リレー講座「統計入門」

４日目「いろいろな分布」

柴山太郎

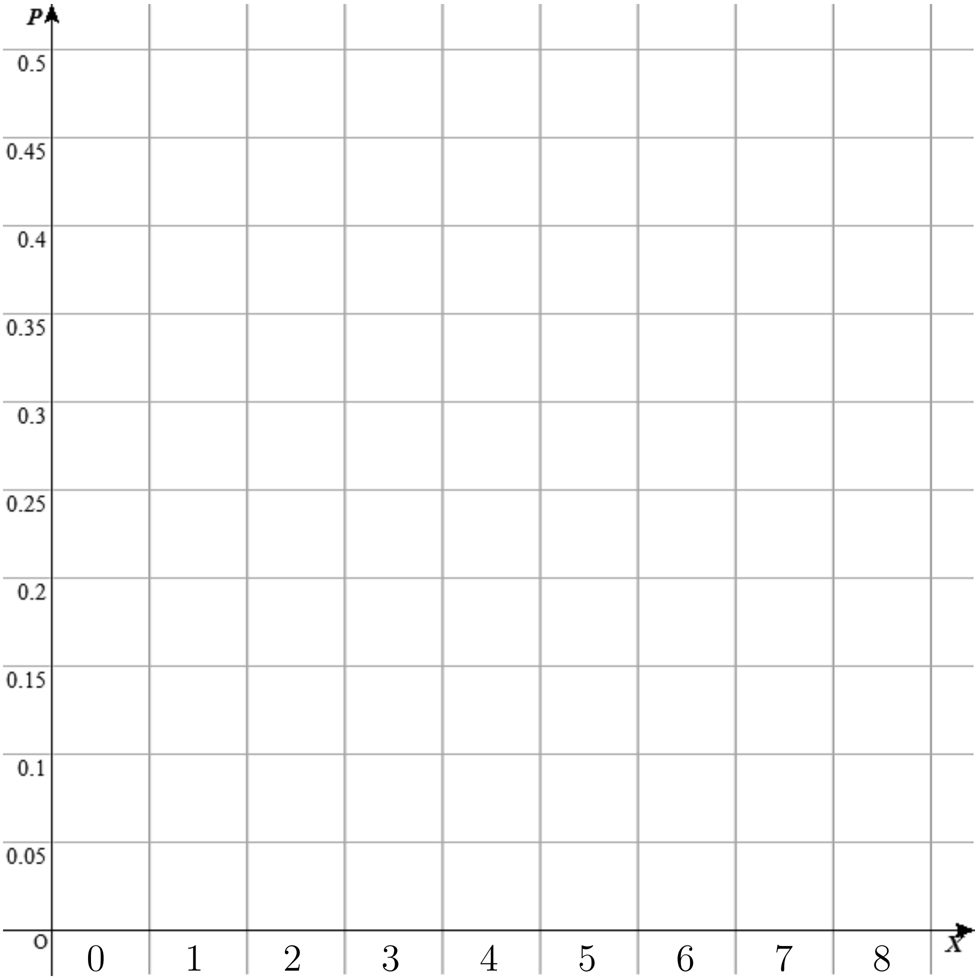
目次

「中高生にも雰囲気だけは伝わるように」レジュメを作成したつもりである．そのため，厳密さに欠ける表現が多く含まれるが，ご容赦いただきたい．

§ 1　確率分布の概念

問 1　硬貨を 2 回投げたとき，表が出た回数 X とそれに対応する確率 P の関係を下の表と棒グラフで表しなさい．

X	0	1	2	3	4	5	6	7	8	計
P										



平均 $E(X)$ と分散 $V(X)$ を求めなさい．

$$E(X) =$$

$$V(X) =$$

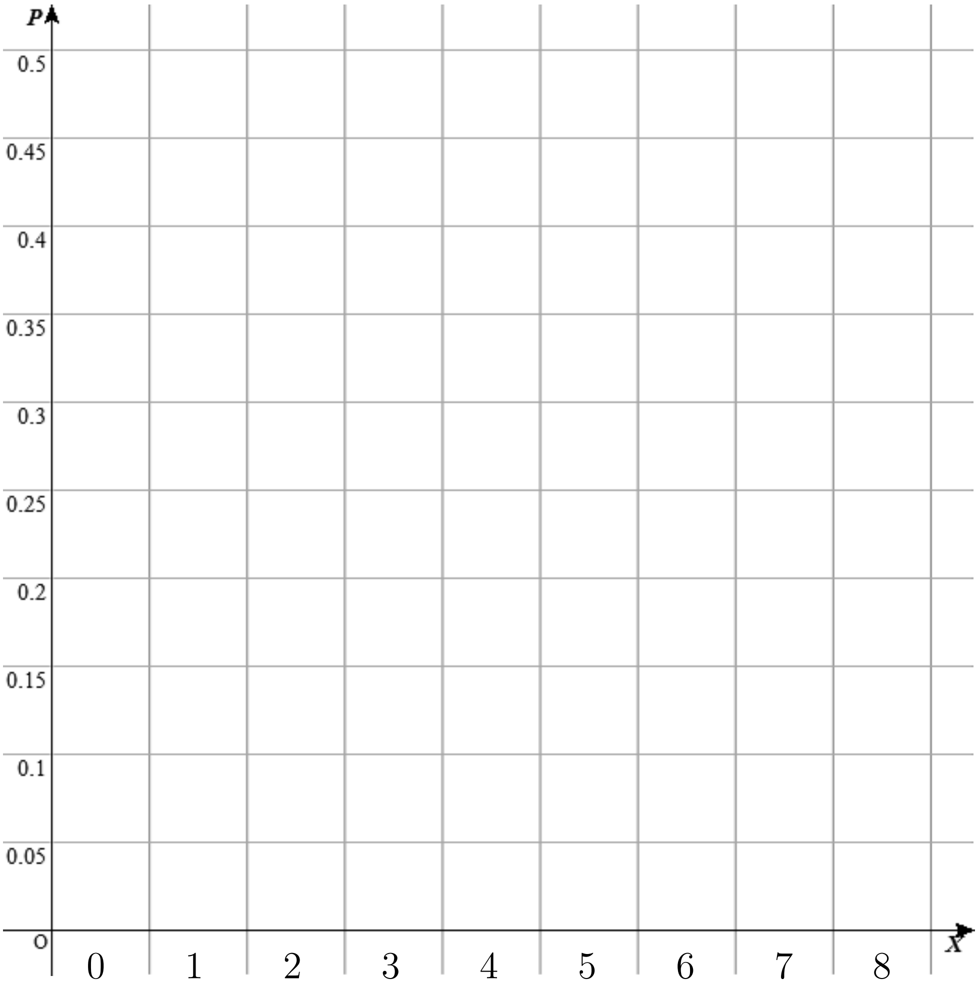
$X = a$ となる確率を $P(X = a)$，$a \leqq X \leqq b$ となる確率を $P(a \leqq X \leqq b)$ のように表します．$P(1 \leqq X \leqq 2)$ を求めなさい．また，この値は棒グラフのどの面積と一致するか．

$$P(1 \leqq X \leqq 2) =$$

【計算・メモ欄】

問 2　硬貨を 7 回投げたとき，表が出た回数 X とそれに対応する確率 P の関係を下の表と棒グラフで表しなさい．

X	0	1	2	3	4	5	6	7	8	計
P										



※棒グラフに表す際，計算が大変なので電卓を用いてよい！

平均 $E(X)$ と分散 $V(X)$ を求めなさい.

$$E(X) =$$

$$V(X) =$$

$X = a$ となる確率を $P(X = a)$，$a \leqq X \leqq b$ となる確率を $P(a \leqq X \leqq b)$ のように表します．$P(4 \leqq X \leqq 6)$ を求めなさい．また，この値は棒グラフのどの面積と一致するか.

$$P(4 \leqq X \leqq 6) =$$

【計算・メモ欄】

今までの例のように，確率変数のとり得る値と，その値をとる確率との対応を示したものを，その確率変数の**確率分布**または単に**分布**といい，確率変数 X はこの分布に従うという.

（離散型）確率分布の性質

確率変数 X のとる値が $x_1, x_2, \cdots, x_n$ であるとき，$P(X = x_i) = p_i$ とおけば，

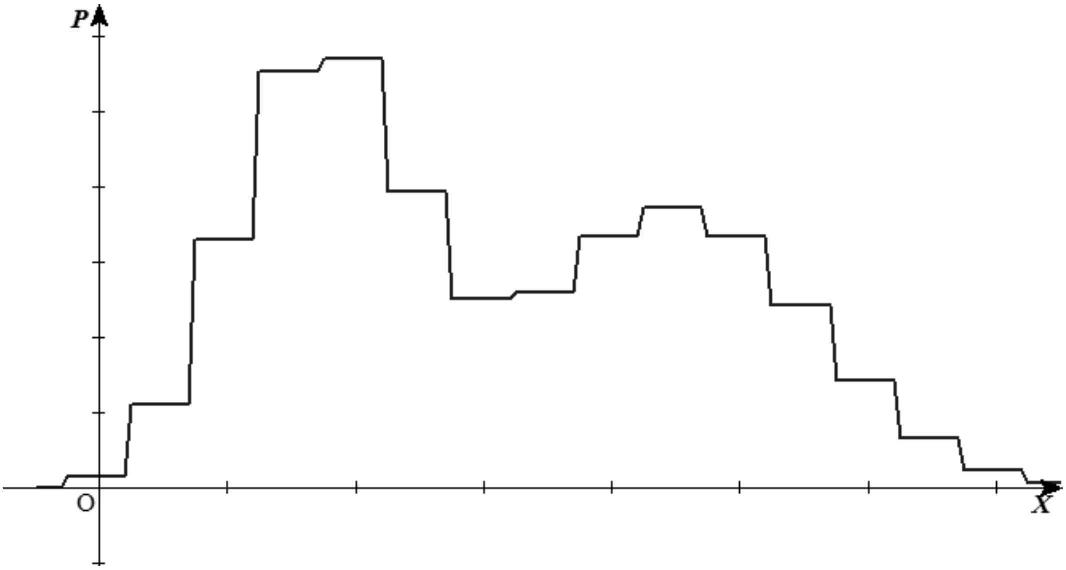
① $p_i \geqq 0$

② $p_1 + p_2 + \cdots + p_n = 1$

が成り立つ.

以下のような表を，**確率分布表**という.

X	x_1	x_2	x_3	$\cdots$	x_n	計
P	p_1	p_2	p_3	$\cdots$	p_n	1



§ 2　二項分布

一般に，異なる n 個のものから，順序を問題にしないで，r 個を取り出して 1 組としたものを n **個から** r **個をとる組み合わせ**といい，その総数を ${}_n\mathrm{C}_r$ で表す．

$$ {}_n\mathrm{C}_r = \frac{n!}{r!(n-r)!} = \frac{n \times (n-1) \times \cdots \times (n-r+1)}{r \times (r-1) \times \cdots \times 1} $$

（ただし，$n! = n \times (n-1) \times \cdots \times 1$ を，n **の階乗**という．）

問 3　硬貨を 7 回投げたとき，表がちょうど 3 回出る確率をもう一度求めなさい．[*1]

問 4　サイコロを 6 回ふるとき，1 の目がちょうど 2 回出る確率を求めなさい．

[*1] 問 2 の $P(X=3)$ に等しい．

定義 （二項分布） n は正の整数，p は $0 < p < 1$ をみたす定数，$q = 1 - p$ とする．
確率変数 X のとる値が $0, 1, 2, \cdots, n$ で，

$$P(X = r) = {}_n\mathrm{C}_r p^r q^{n-r} \quad (r = 0, 1, 2, \cdots, n)$$

であるとき，この確率分布のことを，確率 p に対する次数 n の**二項分布**といい，$B(n, p)$ という記号で表す．

定理 1　二項分布 $B(n, p)$ にしたがう確率変数 X の平均, 分散は, 平均 $E(X) = np$, 分散 $V(X) = npq$ （ただし，$q = 1 - p$） で与えられる．

問 5　$B(2, 0.5)$，$B(7, 0.5)$ それぞれにおいて，定理 1 の公式を用いて平均，分散を求めなさい．

イメージしやすいようにグラフをいくつか載せておく．

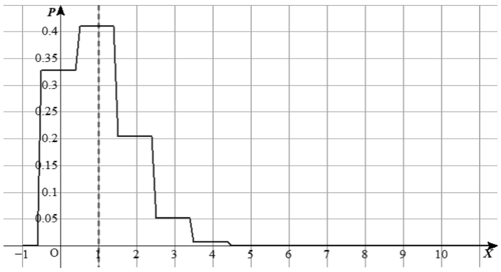


図 1　$B(5, 0.2)$

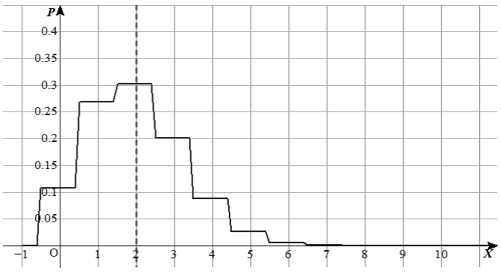


図 2　$B(10, 0.2)$

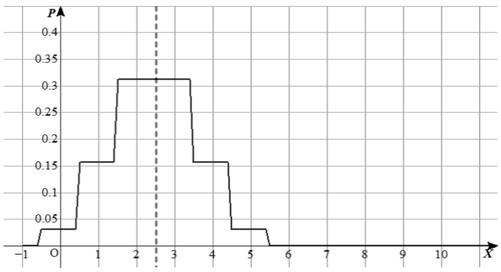


図 3　$B(5, 0.5)$

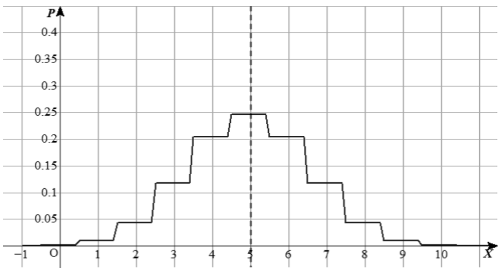


図 4　$B(10, 0.5)$

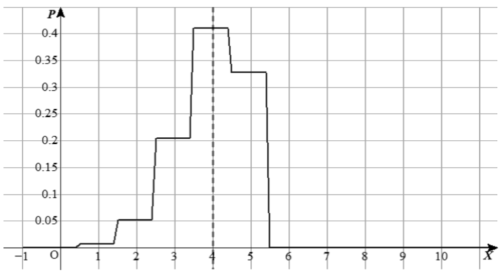


図 5　$B(5, 0.8)$

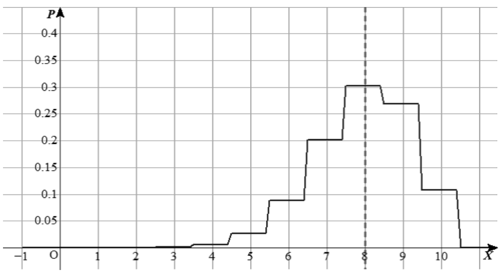
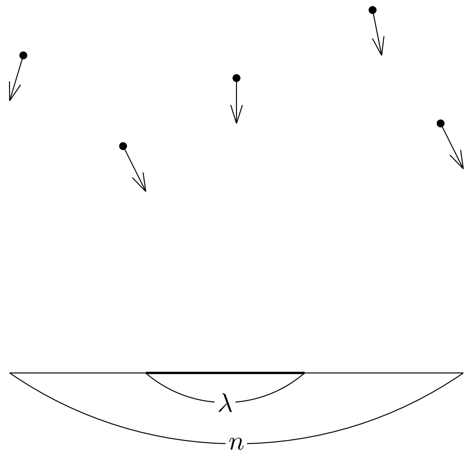


図 6　$B(10, 0.8)$

§ 3　ポアソン分布

問 6　長さ n の線分に n 個の粒子を落とすことを考える．（つまり，単位長さあたりの平均は 1 個である．）どこに落ちるかは均等と仮定する．



このとき，長さ λ の特定の線分の上にちょうど r 個の粒子が落ちる確率を求めなさい．
[ヒント] 1 個の粒子がそこに落ちる確率は $\dfrac{\lambda}{n}$ である．[*2]

ここで，n を限りなく大きくして，無限に長い線分（直線！）に無限に多くの粒子を落とすことを考える．単位長さ当たりの平均は 1 個という割合は変えないものとすると，どんな分布になるだろうか？

⇢ 定理 3

[*2] この事実は感覚的に明らかだと思うが，後でもう少し詳しくみる．

定義 （ポアソン分布）λ を $\lambda > 0$ をみたす定数とする．
確率変数 X のとる値が $0, 1, 2, \cdots$ で，

$$P(X = r) = \frac{e^{-\lambda}\lambda^r}{r!} \quad (r = 0, 1, 2, \cdots)$$

であるとき，この確率分布のことを，**ポアソン分布**といい，$Po(\lambda)$ という記号で表す．

> **定理 2**　ポアソン分布 $Po(\lambda)$ にしたがう確率変数 X の平均, 分散は, 平均 $E(X) = \lambda$，分散 $V(X) = \lambda$ で与えられる．

イメージしやすいようにグラフをいくつか載せておく．

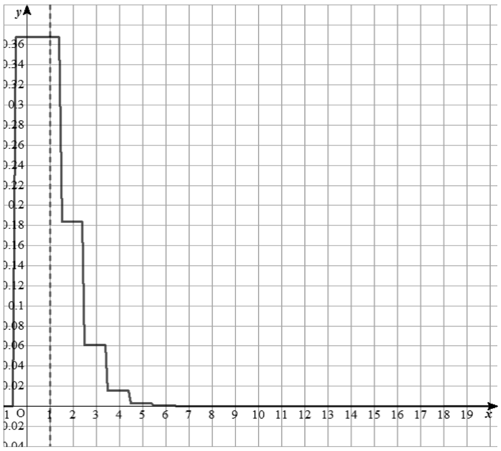


図 7　$Po(1)$

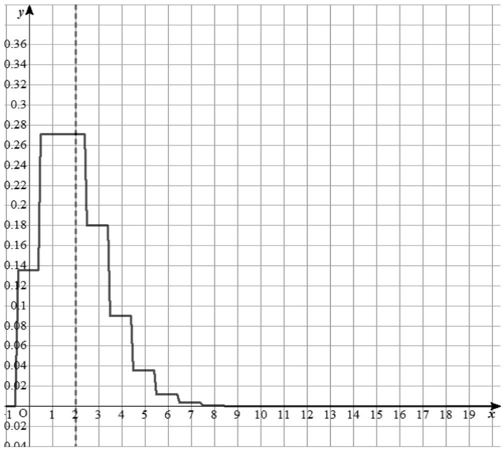


図 8　$Po(2)$

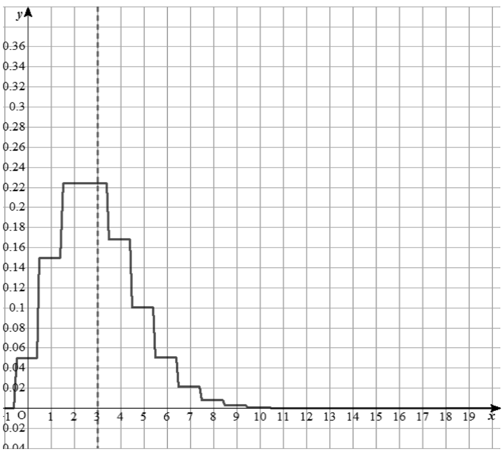


図 9　$Po(3)$

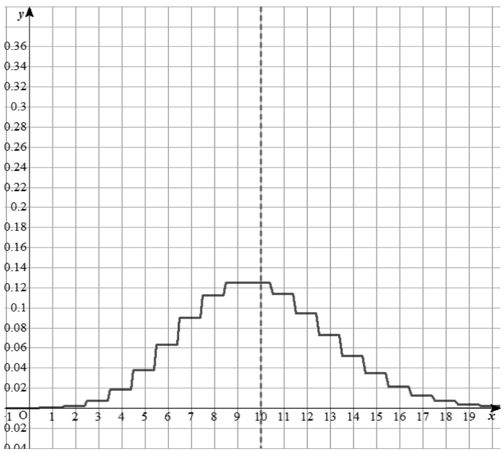


図 10　$Po(10)$

e はネイピア数といい，$e = 2.71828182845904523536028747135 2\cdots$ という無理数である．自然対数の底として知られる．

ポアソン分布の有用性は，次の定理に依る．$\boxed{\text{問 6}}$ を思い出そう．

定理 3　（二項分布のポアソン近似） 二項分布 $B\left(n, \dfrac{\lambda}{n}\right)$ において，n を限りなく大きくすると，ポアソン分布 $Po(\lambda)$ に近づく．

近づいている様子をプロジェクターにて確認します！

$\boxed{\text{問 6}}$，定理 3 から，ポアソン分布 $Po(\lambda)$ とは，「1 区画あたり平均 1 個の粒子を落としたとき，特定の λ 個の区画上に実際に r 個落ちる確率」とみることができる．つまり，

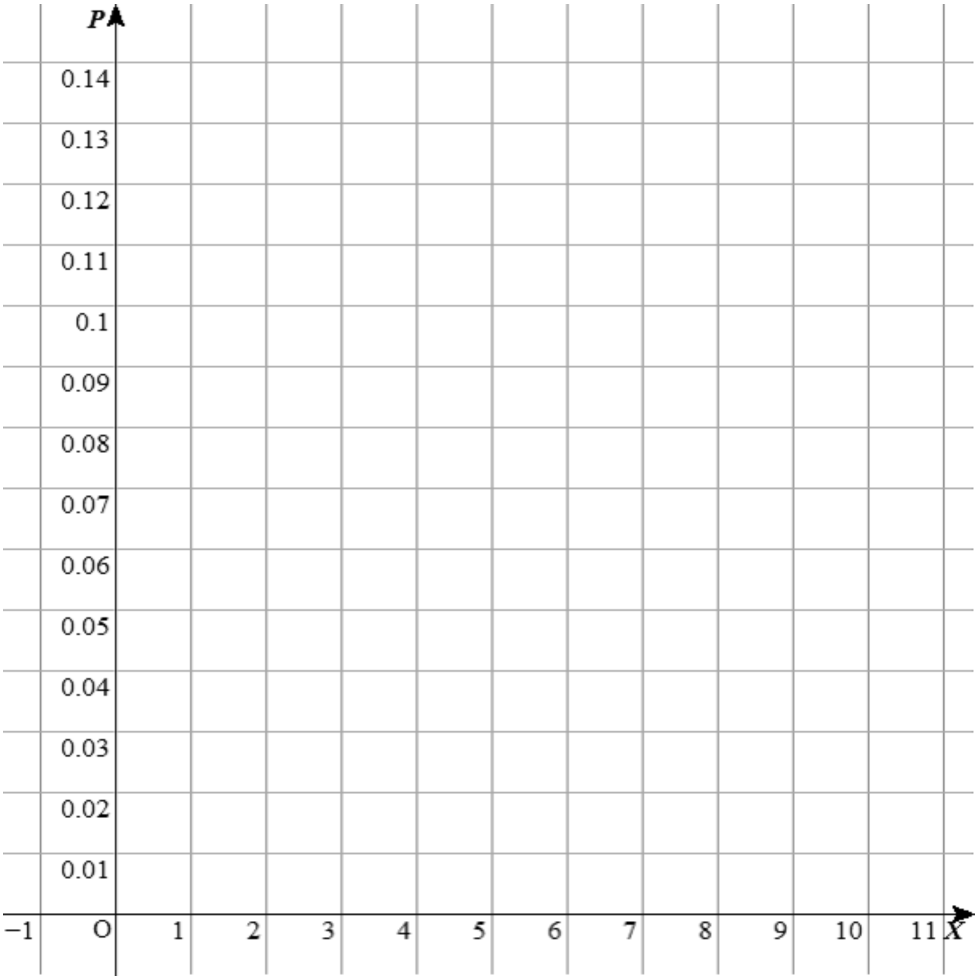
- ある素材の単位面積あたりの傷の数
- ぶあつい本の中にある 1 ページごとのミスプリント数

などはポアソン分布に従う．

§ 4　連続確率分布

問 7　数直線の $0 \leqq X \leqq 10$ の部分に 1 個の粒子を落とすことを考える．$0 \leqq X \leqq 10$ のどこに落ちるかは均等と仮定する．落ちた位置 X とそれに対応する確率 P の関係を下の表とグラフで表しなさい．*3

X	0	1	2	3	4	5	6	7	8	9	10	計
P												



上のグラフを見れば，縦軸が $P(X = 5)$ などを表しているわけではないが，面積を考えることで $P(2 \leqq X \leqq 5)$ などがわかる．

*3 **問 6** で，1 個の粒子がそこに落ちる確率は $\dfrac{\lambda}{n}$ だという [ヒント] を用いたがそれについて考えている．

一般に，ある区間に属するすべての実数値をとる確率変数を**連続型確率変数**といい，これに対して，とびとびの値だけをとる確率変数を**離散型確率変数**という．

定義（確率密度関数）連続型確率変数 X において，$\alpha \leqq X \leqq \beta$ となる確率が，

$$P(\alpha \leqq X \leqq \beta) = \int_{\alpha}^{\beta} f(x)dx$$

のように表される場合，関数 $f(x)$ を X の**確率密度関数**という．

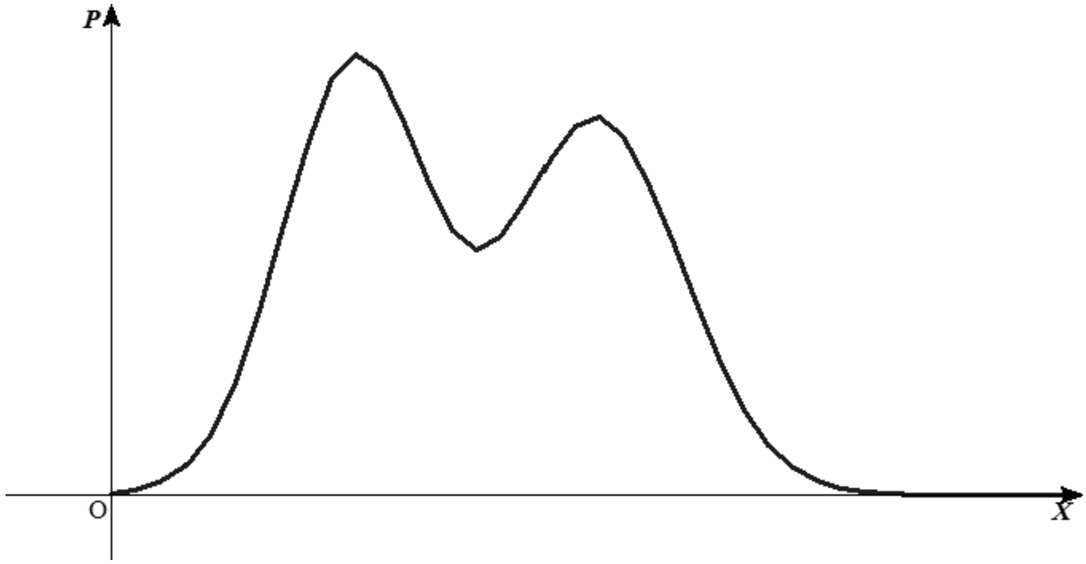
確率密度関数の性質

連続型確率変数 X の確率密度関数を $f(x)$ とおけば，

① $f(x) \geqq 0$

② $\int_{-\infty}^{\infty} f(x)dx = 1$

が成り立つ．



$\boxed{\text{問 7}}$ のように定数関数（一定の値をとる関数）の確率密度関数で定義される分布を**一様分布**という．一般に確率密度関数は，ぐにゃっとした曲線のグラフになることが多く，その例を次のページから紹介する．

§ 5　正規分布

定義（**正規分布**）確率密度関数 $f(x)$ が，

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

で与えられる確率分布を**正規分布**といい，$N(m, \sigma^2)$ で表す．ただし，$\sigma > 0$ とする．

> **定理 4**　正規分布 $N(m, \sigma^2)$ にしたがう確率変数 X の平均，分散は，平均 $E(X) = m$，分散 $V(X) = \sigma^2$ で与えられる．

イメージしやすいようにグラフをいくつか載せておく．

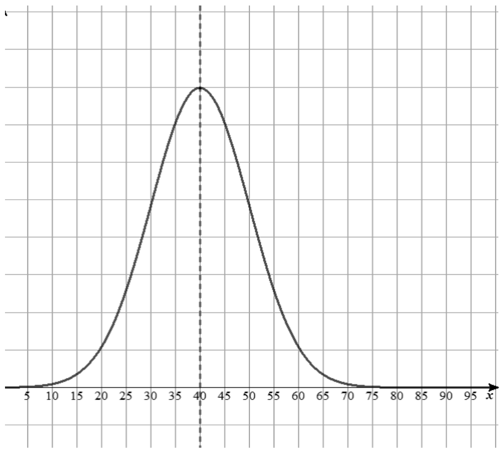


図 11　$N(40, 100)$

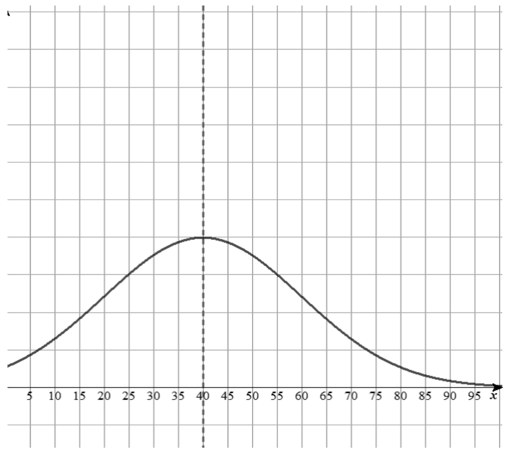


図 12　$N(40, 400)$

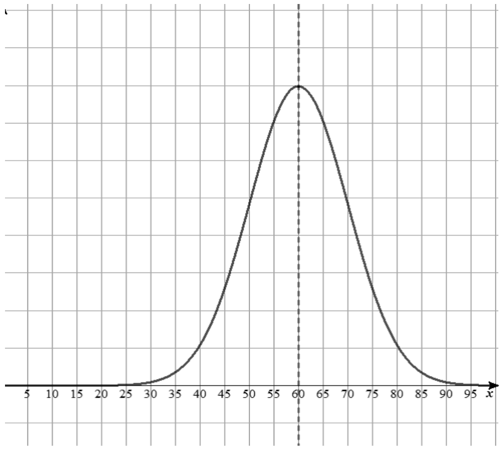


図 13　$N(60, 100)$

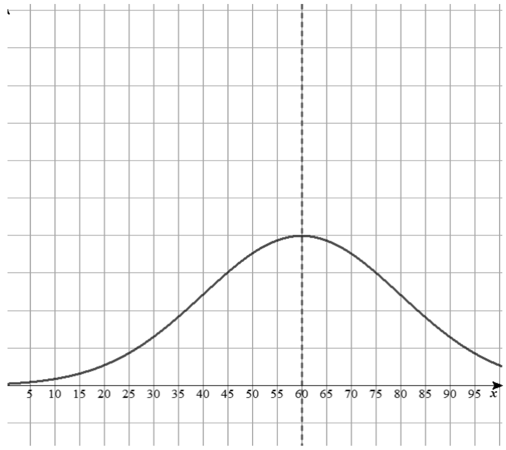
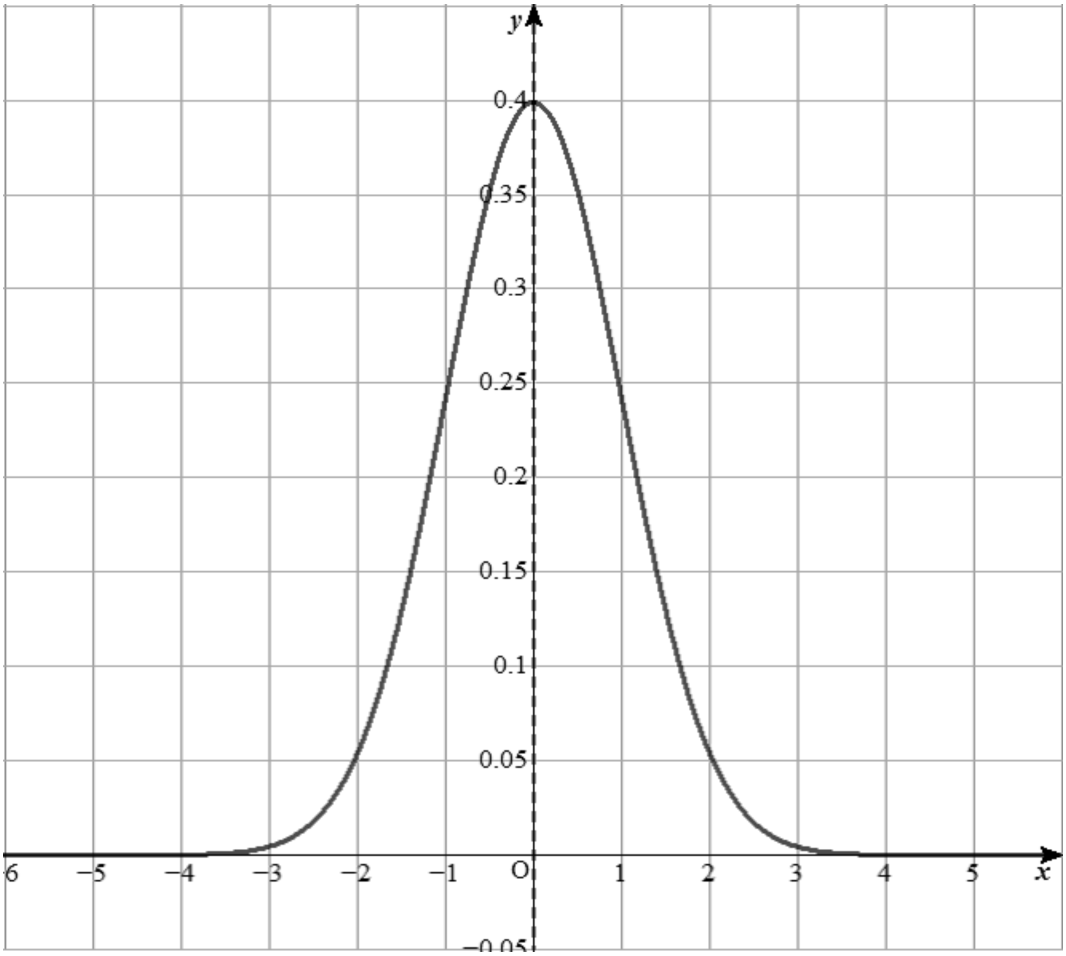


図 14　$N(60, 400)$

π は円周率で，$\pi = 3.14159265358979323846264338327 9\cdots$ という無理数である．

定義 （標準正規分布）

平均 0，分散 1 の正規分布 $N(0,1)$ のことを標準正規分布という．標準正規分布の確率密度関数は $f(x) = \dfrac{1}{\sqrt{2\pi}}\ e^{-\frac{x^2}{2}}$ である．

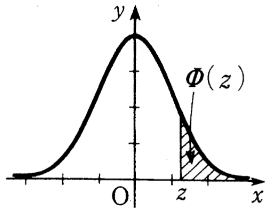


> **定理 5**　X が正規分布 $N(m,\sigma^2)$ をもつ確率変数であるとき，確率変数 $Z = \dfrac{X-m}{\sigma}$ の確率分布は標準正規分布 $N(0,1)$ である．

定理 5 により，標準正規分布の特徴が分かっていれば，すべての正規分布を調べることができることを意味している．

いわゆる「正規分布表」は正の数 z に対し，$Q(z) = \int_z^\infty \dfrac{1}{\sqrt{2\pi}}\ e^{-\frac{x^2}{2}} dx$ の値を表にしたものである．この表を用いることで，正規分布に従う事象の確率を簡単に計算することができる．

正規分布表



z	0	1	2	3	4	5	6	7	8	9
0.0	.50000	.49601	.49202	.48803	.48405	.48006	.47608	.47210	.46812	.46414
0.1	.46017	.45620	.45224	.44828	.44433	.44038	.43644	.43251	.42858	.42465
0.2	.42074	.41683	.41294	.40905	.40517	.40129	.39743	.39358	.38974	.38591
0.3	.38209	.37828	.37448	.37070	.36693	.36317	.35942	.35569	.35197	.34827
0.4	.34458	.34090	.33724	.33360	.32997	.32636	.32276	.31918	.31561	.31207
0.5	.30854	.30503	.30153	.29806	.29460	.29116	.28774	.28434	.28096	.27760
0.6	.27425	.27093	.26763	.26435	.26109	.25785	.25463	.25143	.24825	.24510
0.7	.24196	.23885	.23576	.23270	.22965	.22663	.22363	.22065	.21770	.21476
0.8	.21186	.20897	.20611	.20327	.20045	.19766	.19489	.19215	.18943	.18673
0.9	.18406	.18141	.17879	.17619	.17361	.17106	.16853	.16602	.16354	.16109
1.0	.15866	.15625	.15386	.15151	.14917	.14686	.14457	.14231	.14007	.13786
1.1	.13567	.13350	.13136	.12924	.12714	.12507	.12302	.12100	.11900	.11702
1.2	.11507	.11314	.11123	.10935	.10749	.10565	.10383	.10204	.10027	.09853
1.3	.09680	.09510	.09342	.09176	.09012	.08851	.08692	.08534	.08379	.08226
1.4	.08076	.07927	.07780	.07636	.07493	.07353	.07215	.07078	.06944	.06811
1.5	.06681	.06552	.06426	.06301	.06178	.06057	.05938	.05821	.05705	.05592
1.6	.05480	.05370	.05262	.05155	.05050	.04947	.04846	.04746	.04648	.04551
1.7	.04457	.04363	.04272	.04182	.04093	.04006	.03920	.03836	.03754	.03673
1.8	.03593	.03515	.03438	.03363	.03288	.03216	.03144	.03074	.03005	.02938
1.9	.02872	.02807	.02743	.02680	.02619	.02559	.02500	.02442	.02385	.02330
2.0	.02275	.02222	.02169	.02118	.02068	.02018	.01970	.01923	.01876	.01831
2.1	.01786	.01743	.01700	.01659	.01618	.01578	.01539	.01500	.01463	.01426
2.2	.01390	.01355	.01321	.01287	.01255	.01222	.01191	.01160	.01130	.01101
2.3	.01072	.01044	.01017	.00990	.00964	.00939	.00914	.00889	.00866	.00842
2.4	.00820	.00798	.00776	.00755	.00734	.00714	.00695	.00676	.00657	.00639
2.5	.00621	.00604	.00587	.00570	.00554	.00539	.00523	.00508	.00494	.00480
2.6	.00466	.00453	.00440	.00427	.00415	.00402	.00391	.00379	.00368	.00357
2.7	.00347	.00336	.00326	.00317	.00307	.00298	.00289	.00280	.00272	.00264
2.8	.00256	.00248	.00240	.00233	.00226	.00219	.00212	.00205	.00199	.00193
2.9	.00187	.00181	.00175	.00169	.00164	.00159	.00154	.00149	.00144	.00139
3.0	.00135	.00131	.00126	.00122	.00118	.00114	.00111	.00107	.00104	.00100
3.1	.00097	.00094	.00090	.00087	.00084	.00082	.00079	.00076	.00074	.00071
3.2	.00069	.00066	.00064	.00062	.00060	.00058	.00056	.00054	.00052	.00050
3.3	.00048	.00047	.00045	.00043	.00042	.00040	.00039	.00038	.00036	.00035
3.4	.00034	.00032	.00031	.00030	.00029	.00028	.00027	.00026	.00025	.00024
3.5	.00023	.00022	.00022	.00021	.00020	.00019	.00019	.00018	.00017	.00017
3.6	.00016	.00015	.00015	.00014	.00014	.00013	.00013	.00012	.00012	.00011
3.7	.00011	.00010	.00010	.00010	.00009	.00009	.00008	.00008	.00008	.00008
3.8	.00007	.00007	.00007	.00006	.00006	.00006	.00006	.00005	.00005	.00005
3.9	.00005	.00005	.00004	.00004	.00004	.00004	.00004	.00004	.00003	.00003

定理 6　X が正規分布 $N(m, \sigma^2)$ をもつ確率変数なら，次が成り立つ．

$$P(m - \sigma \leqq X \leqq m + \sigma) = 0.6827$$

$$P(m - 2\sigma \leqq X \leqq m + 2\sigma) = 0.9545$$

$$P(m - 3\sigma \leqq X \leqq m + 3\sigma) = 0.9973$$

また，次が成り立つ．

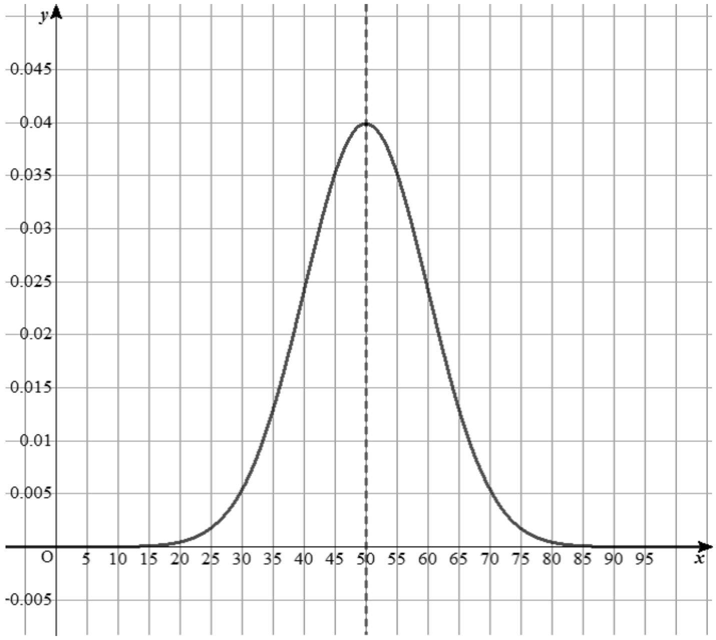
$$P(m - 1.645\sigma \leqq X \leqq m + 1.645\sigma) = 0.9000$$

$$P(m - 1.96\sigma \leqq X \leqq m + 1.96\sigma) = 0.9500$$

$$P(m - 2.58\sigma \leqq X \leqq m + 2.58\sigma) = 0.9900$$

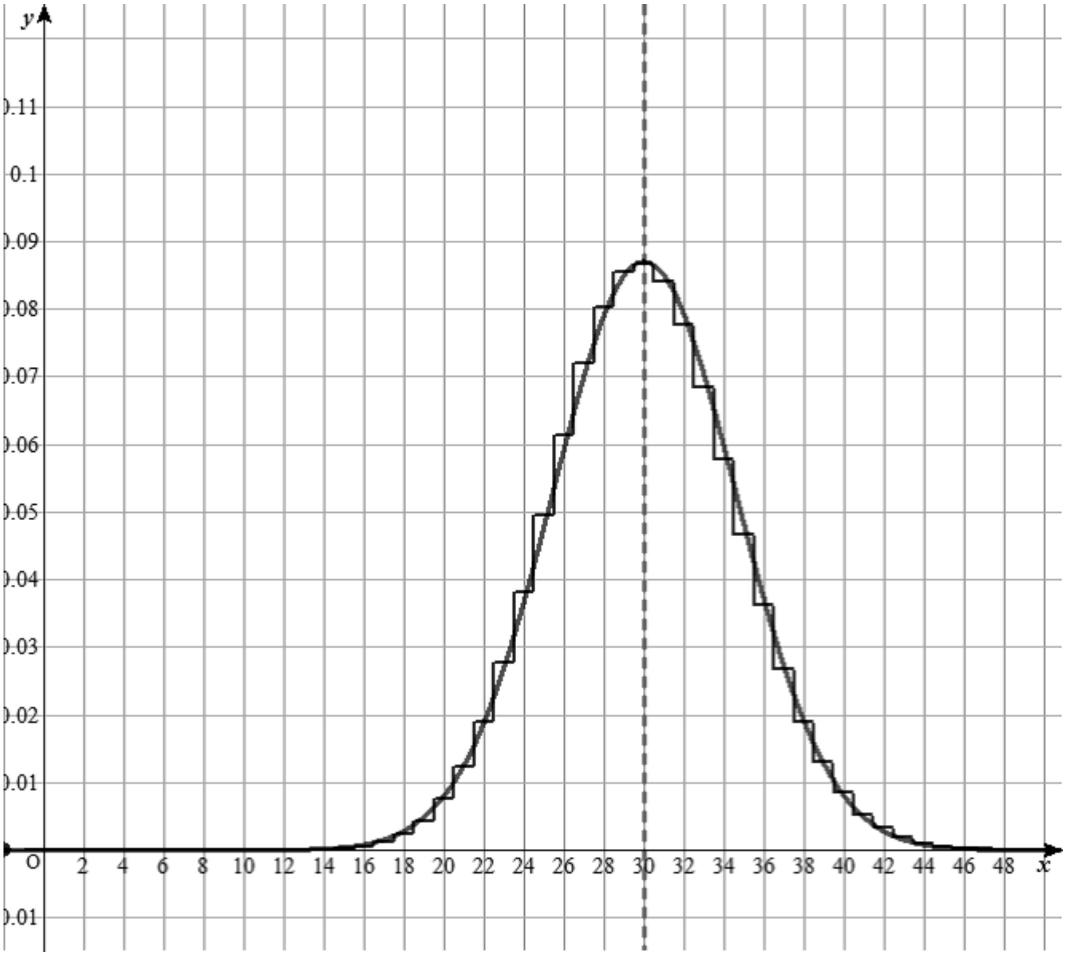
定理 6 は，定理 5 と正規分布表から得られる．

問 8　正規分布にしたがっていると考えれば，偏差値 70 以上の生徒は全体の何パーセントいるか．（ヒント：偏差値とは平均 50，標準偏差 10 に調整したものである．）



> 定理 7　（ド・モアブルーラプラスの定理）
> 平均と分散がそれぞれ等しい二項分布と正規分布は似ている．

次のグラフは $B(100, 0.3)$ の分布と，$N(30, 21)$ の確率密度関数のグラフである．



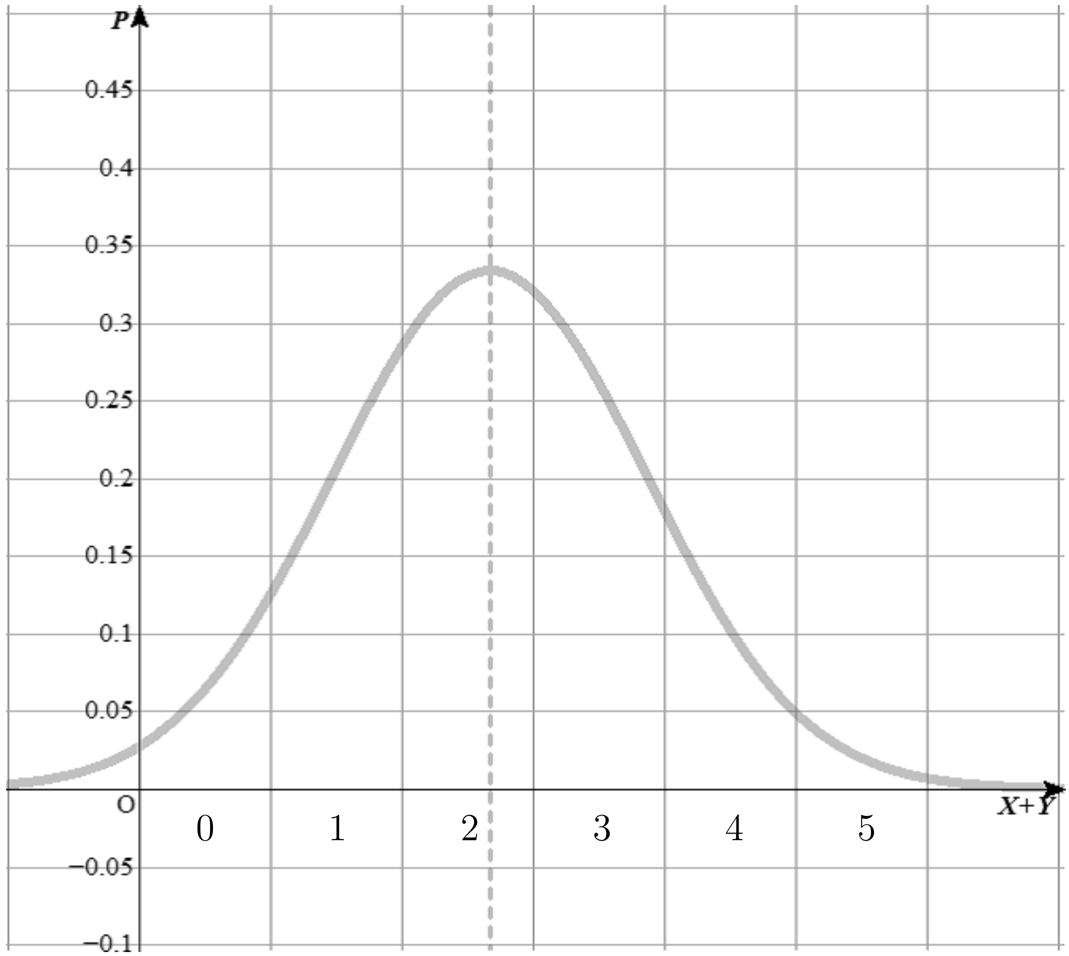
定理 7 のステイトメントはかなりざっくりと書いたが本当はもっと厳密であり，きちんと証明できる．

問 9　箱 A、箱 B にはどちらも数字が書かれたカードが入っています．箱 A には，0 が 4 枚，1 が 4 枚，2 が 1 枚入っています．また，箱 B には，0 が 1 枚，1 が 3 枚，2 が 3 枚，3 が 1 枚入っています．箱 A，箱 B からそれぞれ 1 枚ずつカードをひくとき，数字の和 $X+Y$ と，それに対応する確率 P の関係を下の表と棒グラフで表しなさい．

X	0	1	2	計
P	$\frac{4}{9}$	$\frac{4}{9}$	$\frac{1}{9}$	1

Y	0	1	2	3	計
P	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	1

$X+Y$	0	1	2	3	4	5	6	7	8	計
P										



平均 $E(X+Y)$ と分散 $V(X+Y)$ を求めなさい．

$$E(X+Y) =$$

$$V(X+Y) =$$

【計算・メモ欄】

なお，正規分布 $N\left(\frac{13}{6}, \frac{43}{36}\right)$ の密度関数のグラフがうすーくかいてありますが気にしないように．

問 9 では，ある 2 つの分布 X, Y に対し，$X+Y$ の分布を考え，それがたまたま正規分布に近くなっていた．2 つの場合では一般には成り立たないが，どんな分布でも，何度も何度も足し合わせることで正規分布に近くなるという大定理が知られている．

> **定理 8　（中心極限定理）** $X_1, X_2, \cdots X_n$ が独立な確率変数で，それぞれの平均と分散が一定範囲にとどまっているならば，n が大きいとき，確率変数 $X = X_1 + X_2 + \cdots + X_n$ の分布は，$X_1, X_2, \cdots X_n$ の分布のいかんにかかわらず，正規分布に近い．

この定理から，自然界に正規分布に近い分布がしばしば現れることが説明される．

$X = X_1 + X_2 + \cdots + X_n$ だけでなく，$X = X_1^2 + X_2^2 + \cdots + X_n^2$ を知っておくことも統計学的には大事なことであり，以下の定理が知られている．$X_1, X_2, \cdots X_n$ が一般の場合ではないが，すべて標準正規分布 $N(0, 1)$ に従う場合の定理である．[*4]

> **定理 9** $X_1, X_2, \cdots, X_n$ がすべて独立で標準正規分布 $N(0, 1)$ に従うならば，$Y = X_1^2 + X_2^2 + \cdots + X_n^2$ は自由度 n の χ^2 分布 $\chi^2(n)$ に従う．

[*4] χ^2 分布の定義は次のセクションで扱う．

§ 6　χ^2 分布（カイ 2 乗分布）

χ^2 分布の有用性は定理 9 に依る．定義は難しいので，グラフの概形を知っておくだけでよい．

定義（χ^2 分布）確率密度関数 $f(x)$ が

$$f(x) = \frac{1}{\Gamma\left(\frac{n}{2}\right)}\left(\frac{1}{2}\right)^{\frac{n}{2}} x^{\frac{n}{2}-1}e^{-\frac{x}{2}} \quad (x > 0)$$

で与えられる確率分布を自由度 n の χ^2 **分布（カイ 2 乗分布）**といい，$\chi^2(n)$ という記号で表す．[*5]

定理 10　自由度 n の χ^2 分布 $\chi^2(n)$ にしたがう確率変数 X の平均，分散はそれぞれ平均　$E(X) = n$，分散　$V(X) = 2n$ で与えられる．

[*5] ここに現れる $\Gamma(s)$ はガンマ関数と呼ばれるもので，$\Gamma(s) = \int_0^\infty e^{-x}x^{s-1}dx$ として定義される．わからないと思うが，よく知られている大事な関数だと思っておけばよい．

イメージしやすいようにグラフをいくつか載せておく．

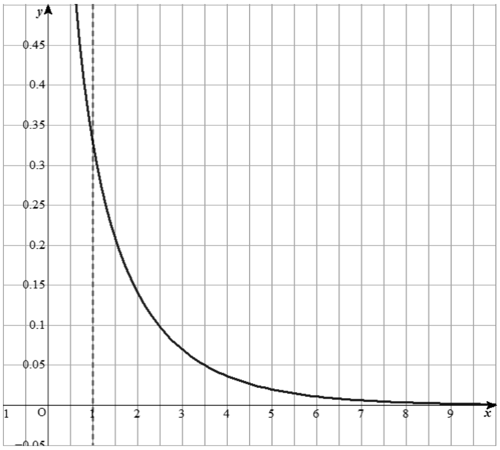


図 15　$\chi^2(1)$

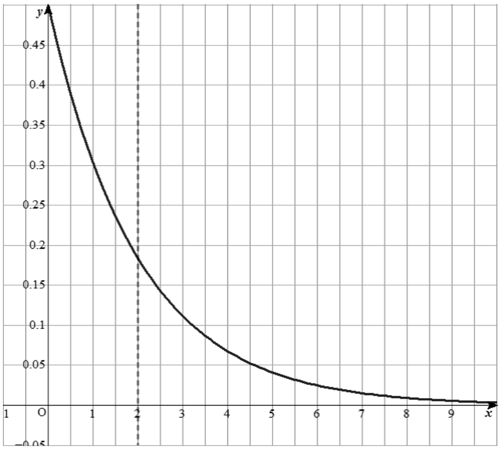


図 16　$\chi^2(2)$

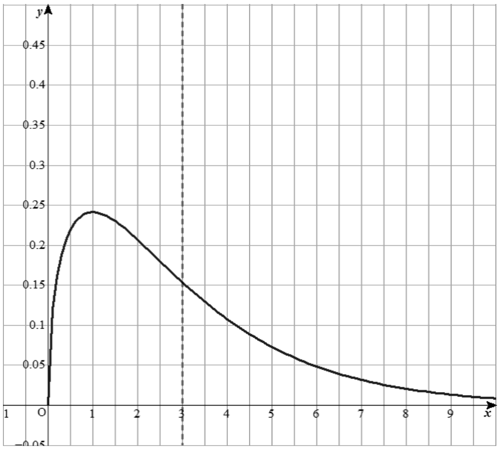


図 17　$\chi^2(3)$

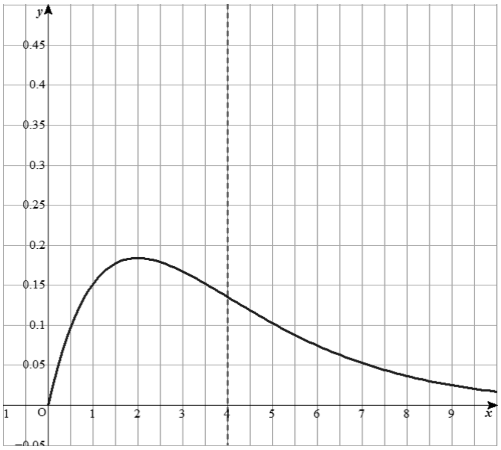


図 18　$\chi^2(4)$

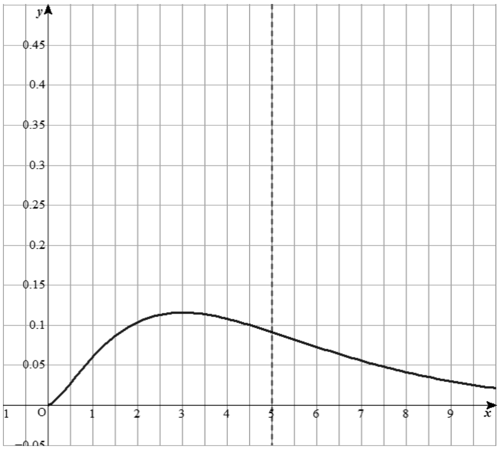


図 19　$\chi^2(5)$

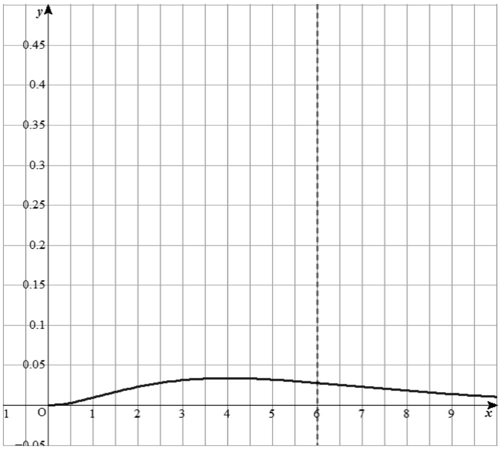


図 20　$\chi^2(6)$

参考文献

[1]　中田義元，根岸世雄，藤田宏：大学への確率・統計，研文書院，1985 年
[2]　小針あき宏：確率・統計入門，岩波書店，1973 年
[3]　福島正俊：確率論，裳華房，1998 年
[4]　俣野博，河野俊丈，ほか 27 名：数学 B，東京書籍，2015 年