

確率と統計

横田 壽

目次

第 1 章	データの整理	3
1.1	度数分布表 (Frequency Tabulations)	4
1.2	標本の散布度, 相関関係	8
1.3	相関表, 回帰直線	12
第 2 章	確率	17
2.1	順列・組み合わせ	18
2.2	確率の定義	20
2.3	確率の公理	22
第 3 章	確率変数	25
3.1	確率変数と確率分布	26
3.2	2次元確率分布	28
3.3	確率変数の平均値と分散	29
3.4	多次元確率分布	31
第 4 章	理論分布	33
4.1	2項分布	34
4.2	ポワソン分布	35
4.3	正規分布	36
4.4	幾何分布	38
4.5	超幾何分布	40
4.6	指数分布	41
4.7	ガンマ分布	42
第 5 章	統計的推定法	43
5.1	統計量と標本分布	44
5.2	信頼区間	47
5.3	母比率の区間推定 (大標本の場合)	50
5.4	母比率の区間推定 (小標本の場合)	51
5.5	重要な標本分布	52
5.6	χ^2 分布	53
5.7	t 分布	55
5.8	F 分布	56

第 6 章	統計的検定	57
6.1	統計的検定の考え方	58
6.2	母集団が正規分布で 2 標本の場合	62
6.3	比率の検定	66
6.4	適合度検定	69
第 7 章	演習問題解答	75

第1章 データの整理

1.1 度数分布表 (Frequency Tabulations)

サンプリング

データの収集過程やサイコロを投げるなどの試行を繰り返すことにより結果を得る過程をサンプリング (sampling) といいます。サンプリングの結果えられたものをサンプル (sample) または標本といいます。

例えば、次のようなデータを得たとしましょう。

例題 1.1

表 1.1: がん患者のヘモグロビン濃度

13.6	14.8	13.7	14.2	11.5
11.9	13.8	14.6	14.2	12.7
13.4	11.5	11.9	14.8	12.7
12.4	15.3	15.2	13.5	15.0
12.4	12.0	13.8	11.7	10.0
13.2	15.5	14.0	13.5	15.0
12.7	12.9	13.7	15.1	13.5
15.7	12.7	15.7	10.9	14.0
14.8	14.0	13.8	12.7	11.9
12.0	11.4	11.1	13.7	13.2

このデータを見ただけでは、どんな結果がでたのか分かりにくいので、これらのデータを整理して分かりやすい表にすることを考えます。データの整理の方法として度数分布表 (frequency table) を用いることがよくありますので、度数分布表の作り方を学びます。

データの値を x_i で表すとき、 x_i が現れる回数を度数 (frequency) といい、 f_i で表すと、

$$f_1 + f_2 + \cdots + f_k = n$$

ただし、 n はデータの数です。これより、度数を表にしたものを作成することができます。

可能な値	度数
x_1	f_1
x_2	f_2
\vdots	\vdots
x_k	f_k
合計	n

しかし、データの多くは小数点以下を切り捨てたり、四捨五入したりして得たものなので、 x_i という値で表を作成する代わりに、 a 以上 b 未満での度数という形で表を作成します。このとき、データを a 以上 b 未満と

というようないくつかの区間に分けて集計するときの各区間を階級 (class interval) といい、 $a \sim b$ で表します。そして、区間の幅つまり $b - a$ を階級幅 (class interval width) といいます。それぞれの区間の端点の相加平均 $\frac{a+b}{2}$ を階級値 (midpoint) といいます。また、全標本の個数 n に対する各階級の度数の割合 f_i/n を相対度数 (relative frequency) といいます。さらに、統計解析のために f_i 以下の度数の合計

$$F_i = f_1 + f_2 + \cdots + f_i$$

を考えます。これを累積度数といえます。これらを用いて表したものが度数分布表 (frequency distribution) です。では、上記のデータを用いて度数分布表を作成してみましょう。データの値が 10.0 ~ 15.7 なので、階級幅を 0.9 にとると 7 個の階級を用いることにより全てのデータを含むことが可能です。ただし、データの値が階級の境界値となるのはおかしいので、最初の階級を 9.95 から始めます。

表 1.2: 度数分布表

階級	階級値	度数	相対度数	累積度数	累積相対度数
9.95 ~ 10.85	10.4	1	0.02	1	0.02
10.85 ~ 11.75	11.3	6	0.12	7	0.14
11.75 ~ 12.65	12.2	7	0.14	14	0.28
12.65 ~ 13.55	13.1	12	0.24	26	0.52
13.55 ~ 14.45	14.0	12	0.24	38	0.76
14.45 ~ 15.35	14.9	9	0.18	47	0.94
15.35 ~ 16.25	15.8	3	0.06	50	1.00

度数分布表を図 (棒グラフ) で表したものをヒストグラム (histogram) といいます。また、変量の小さい階級から順に度数を加えていったものを累積度数 (cumulative distribution function) といいます。

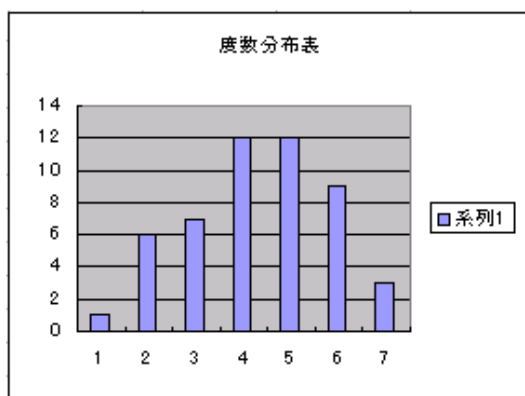


図 1.1: ヒストグラム

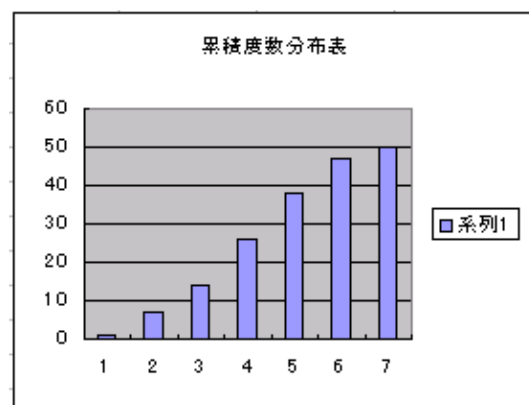


図 1.2: 累積度数分布図

Sturges の式

データ数 n に対して階級数を決める一つの目安にスタージスの式があります。

$$\text{階級数} = 1 + \frac{\log_{10} n}{\log_{10} 2}$$

この式を用いて例題 (1.1) の階級数を求めてみると、標本数 n が 50 より、階級数 k は

$$k = 1 + \frac{\log_{10} 50}{\log_{10} 2} = 1 + 3.32 \log_{10} 50 = 1 + 3.32(1.699) = 6.64 \approx 7$$

となります。また、階級幅は (最大値-最小値)/階級数で求まるので、階級幅は $(15.7-10.0)/6.64 = 0.86$ となります。したがって、階級幅を 0.9 とすると階級数は 7 となります。

標本の代表値

度数分布表が得られると、データ全体を視覚的に把握することができるようになります。しかしながら、それはあくまで直感的なことです。そこで、直感に頼るのではなく理論的にデータを処理するために、データの特徴を数値で表します。

代表値：分布の特徴を代表する数値

変数 x に関する n 個のデータ x_1, x_2, \dots, x_n が与えられたとき、

$$T = x_1 + x_2 + \dots + x_n = \sum_{i=1}^n x_i$$

を標本総計値 (total) といいます。また、

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{T}{n}$$

を標本平均値 (mean) といいます。また変数 x の値が x_1, x_2, \dots, x_k で、その度数が f_1, f_2, \dots, f_k で与えられているとき、標本総計値は

$$T = x_1 f_1 + x_2 f_2 + \dots + x_n f_n$$

となるので、標本平均値は

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + \dots + x_n f_n}{n} = \frac{1}{n} \sum_{i=1}^k x_i f_i$$

で与えられます。

変数の測定値を、大きさの順に並べたとき、中央の位置にくるものを、ミディアン (median) または中央値といえます。データの数 n が偶数のときは第 $\frac{n}{2}$ 番目と第 $\frac{n}{2} + 1$ 番目の変数の平均が中央値。また、データの

数 n が奇数のときは第 $\frac{n+1}{2}$ 番目の変数が中央値となります。

度数が最も大きい標本値 x_i 、または階級値 m_i をモード (mode) または最頻値といえます。

確認問題

- 例題 (1.1) の中央値を求めよ。
- 例題 (1.1) のモードを求めよ。

演習問題 1.1

1. 次のデータについて、スタージェスの式をもちいて度数分布表・ヒストグラム・累積度数分布図を作成しよう。また、平均値、最大値、最小値、中央値、最頻値を求めよう。

コンクリート円柱の引っ張りの強さ (g/cm^2)

320	380	340	410	380	340	360	350	320	370
350	340	350	360	370	350	380	370	300	420
370	390	390	440	330	390	330	360	400	370
320	350	360	340	340	350	350	390	380	340
400	360	350	390	400	350	360	340	370	420
420	400	350	370	330	320	390	380	400	370
390	330	360	380	350	330	360	300	360	360
360	390	350	370	370	350	390	370	370	340
370	400	360	350	380	380	360	340	330	370
340	360	390	400	370	410	360	400	340	360

1.2 標本の散布度，相関関係

散布度 (dispersion)

5点満点のテストを行なったところ次のような度数分布表を得ました．

階級	f_i	f_i/n	F_i	F_i/n
0	2	0.02	2	0.02
1	13	0.13	15	0.15
2	33	0.33	48	0.48
3	35	0.35	83	0.83
4	16	0.16	99	0.99
5	1	0.01	100	1.00

また，代表値として次の表を得ました．

標本数	T	\bar{x}	$\sum f_i x_i^2$	s^2	s
100	253	2.53	741	1.01	1.00

この表の s^2 と s について説明します．

散布度：データが平均のまわりに集中して分布するか，平均のまわりから散らばって分布するかの程度を表わすのが，散布度です．

度数分布表において，各階級数 x_i ($i = 1, 2, \dots, k$) に対する度数を f_i とするとき，変量 x の平均 \bar{x} からの偏差の平方の平均:

$$\begin{aligned} s^2 &= \frac{(x_1 - \bar{x})^2 f_1 + (x_2 - \bar{x})^2 f_2 + \cdots + (x_k - \bar{x})^2 f_k}{n} \\ &= \frac{1}{n} \sum_{i=1}^k (x_i - \bar{x})^2 f_i \end{aligned}$$

を標本分散 (variance) といいます．また，標本分散の正の平方根:

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n} \sum_{i=1}^k (x_i - \bar{x})^2 f_i}$$

を標本標準偏差 (sample standard deviation) といいます．テキストによっては，標本分散の定義が

$$\begin{aligned} s^2 &= \frac{(x_1 - \bar{x})^2 f_1 + (x_2 - \bar{x})^2 f_2 + \cdots + (x_k - \bar{x})^2 f_k}{n - 1} \\ &= \frac{1}{n - 1} \sum_{i=1}^k (x_i - \bar{x})^2 f_i \end{aligned}$$

となっています．この2つの式の違いは，前者は観測対象の全ての観測値が求められた場合に用います．後者はそれ以外のときに用います．

実際の問題では階級に分ける前にすべてのデータを打ち込むので, 変量 x に関する n 個のデータ x_1, x_2, \dots, x_n が与えられたとき, 標本分散は次の式で与えられます.

$$\begin{aligned} s^2 &= \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n} \\ &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \end{aligned}$$

ここで標本分散を簡単に計算する実用的な方法として次の簡便計算法があります.

$$s^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - (\bar{x})^2$$

例題 1.2

上の式を導きなさい.

解答

$$\begin{aligned} s^2 &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n (x_i^2 - 2x_i\bar{x} + \bar{x}^2) \\ &= \frac{1}{n} \left(\sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + \sum_{i=1}^n \bar{x}^2 \right) = \frac{1}{n} \left(\sum_{i=1}^n x_i^2 \right) - 2\bar{x}\bar{x} + \bar{x}^2 \\ &= \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 \end{aligned}$$

標本標準偏差は平均値のまわりのデータの散らばりの大きさを表す量ですが, 標本標準偏差が 10 点であるといっても平均点が 30 点のときと, 60 点のときでは違いがあることが分かります. この違いを表す量として, 変動係数とよばれるものがあります. 変動係数はデータの平均値 \bar{x} で標準偏差 s を割った割合 $\frac{s}{\bar{x}}$ で表します. したがって, 変動係数は平均値に対する相対的な散らばりの大きさを表します.

例題 1.3

あるクラスの英語の試験の平均点 \bar{x} は 67 で標準偏差 s_x は 8.5. また, 数学の試験の平均点 \bar{y} は 53 で標準偏差 s_y は 12.6 でした. このクラスの A 君の成績は英語が 75 点で数学が 68 点でした. A 君のクラスでの成績は, 英語と数学のどちらの順位が上でしょうか.

解答 2 つの異なるものを比較するには, 共に同じ土俵をもってこななくてはなりません. その方法として標準化とよばれるものがあります.

$$z_i = \frac{x_i - \bar{x}}{s}$$

とおくと, $\{z_i\}$ の平均は 0 に分散 s^2 は 1 になります. そこで, 英語の成績と数学の成績の標準化を行うと,

$$\begin{aligned} z_{\text{english}} &= \frac{75 - 67}{8.5} = 0.94 \\ z_{\text{math}} &= \frac{68 - 53}{12.6} = 1.19 \end{aligned}$$

となり，A君のクラスでの成績は数学の方が英語より上であるといえます。

相関関係

2次元データの分布の特徴は2つの変数の平均値と分散だけでは表わすことが困難です．そこで (x, y) の n 組のデータを

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$$

とすると，2つの変数の間の関係を調べるものに共分散 (covariance) と相関係数 (correlation coefficient) よばれるものがあり，次のように定義されます．

$$s_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x}\bar{y}$$

$$r = \frac{s_{xy}}{s_x s_y}$$

ここで s_x は変数 x の標準偏差， s_y は変数 y の標準偏差を表わします．

確認問題

1. 相対累積度数が $p/100$ であるような標本値 x を p パーセント点といいます．特に，25%点 Q_1 を第1四分位数，50%点 Q_2 を第2四分位数，75%点 Q_3 を第3四分位数といいます．次の20個のデータの第1四分位数を求めよ．

67	54	54	66	56	65	46	35	45	45
83	72	54	58	47	60	43	82	76	92

演習問題 1.2

1. 次のデータについて，共分散，相関係数を求めよう．

表 1.3: 二酸化硫黄と二酸化窒素の濃度

時刻	二酸化硫黄 x	二酸化窒素 y	時刻	二酸化硫黄 x	二酸化窒素 y
1	23	43	13	38	21
2	21	28	14	51	37
3	18	17	15	109	65
4	17	16	16	90	65
5	17	16	17	78	50
6	15	10	18	75	58
7	13	5	19	34	42
8	14	5	20	33	52
9	16	8	21	29	55
10	17	13	22	31	55
11	17	11	23	25	55
12	35	28	24	25	51

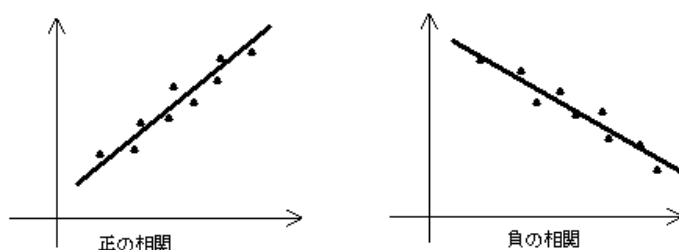
1.3 相関表，回帰直線

相関表

ある場所で1時間おきに二酸化硫黄と二酸化窒素の濃度を測定しました．このとき，二酸化硫黄と二酸化窒素の濃度の間にはどんな関係があるのか調べるために準備をします．

ある時刻での二酸化硫黄の濃度を x_i ，二酸化窒素の濃度を y_i とし， x 軸に二酸化硫黄の濃度を y 軸に二酸化窒素の濃度をとり，座標 (x_i, y_i) を持つ点を図示したものを相関図 (correlation diagram) といいます．

x_i が増加するとき， y_i も増加する傾向があるとき， x_i と y_i は正の相関 (positive correlation) があるといえます．これに反し， x_i が増加するとき， y_i が減少する傾向があるとき， x_i と y_i は負の相関 (negative correlation) があるといえます．



相関表

相関図ではデータの数が多い場合には，その図示が困難となる場合があります．そのような場合には，2つの変量を同時に考えた度数分布表として表わすと便利です．このような表を相関表といえます．

回帰直線

二酸化硫黄と二酸化窒素の関係のように，変量 x の値 x_1, x_2, \dots, x_n に，変量 y の値 y_1, y_2, \dots, y_n がそれぞれ対応していると仮定します．このとき，平面上の n 個の点：

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$$

に対して， y 軸方向の距離が，ある意味で最も近い直線 (回帰直線 (linear regression)):

$$y = ax + b$$

を求めるを考えます，

もし回帰直線が求められていれば， x_i に対する y の値 (予測値 \hat{y}_i) は

$$\hat{y}_i = ax_i + b$$

となります．ところが x_i に対する実際の値 (観測値) は y_i です．そこでこの2つの値の差 $y_i - \hat{y}_i$ を予測誤差といい d_i で表わします．

$$d_i = y_i - \hat{y}_i = y_i - ax_i - b$$

ここで，

$$\sum_{i=1}^n d_i^2 = \sum_{i=1}^n (y_i - ax_i - b)^2$$

を最小にするような a, b の値を定め, 最適直線を求める方法を最小 2 乗法 (method of least square) といいます.

では, どうすれば $\sum_{i=1}^n (y_i - ax_i - b)^2$ を最小にする a, b を求めることができるでしょうか. ここで, a, b の値が変化することにより, $\sum_{i=1}^n (y_i - ax_i - b)^2$ の値が変化するので,

$$F(a, b) = \sum_{i=1}^n (y_i - ax_i - b)^2$$

とおくことにします. ここで, 微分積分学の授業で学んだ微分可能な関数 $y = f(x)$ が極値をとる点では何が起きているかを思い出すと, $y' = f'(x) = 0$ となります. このことを, a と b という変数に対して行くと,

$$\begin{aligned} \frac{\partial F}{\partial a} &= \sum_{i=1}^n [2(y_i - ax_i - b)(-x_i)] = -2 \sum_{i=1}^n (y_i - ax_i - b)x_i \\ \frac{\partial F}{\partial b} &= \sum_{i=1}^n [2(y_i - ax_i - b)(-1)] = -2 \sum_{i=1}^n (y_i - ax_i - b) \end{aligned}$$

この式を書き直すと正規方程式

$$\begin{aligned} \sum_{i=1}^n x_i y_i - a \sum_{i=1}^n x_i^2 - b \sum_{i=1}^n x_i &= 0 \\ \sum_{i=1}^n y_i - a \sum_{i=1}^n x_i - bn &= 0 \end{aligned}$$

を得ます. ここで, $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$, $\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$ であることに注意すると, 上の 2 式は次のように書きなおせます.

$$\begin{aligned} \sum_{i=1}^n x_i y_i - a \sum_{i=1}^n x_i^2 - bn\bar{x} &= 0 \\ n\bar{y} - an\bar{x} - bn &= 0 \end{aligned}$$

第 2 式から, $\bar{y} - a\bar{x} - b = 0$ となるので, これを第 1 式に代入すると,

$$\begin{aligned} \sum_{i=1}^n x_i y_i - a \sum_{i=1}^n x_i^2 - (\bar{y} - a\bar{x})n\bar{x} &= 0 \\ \sum_{i=1}^n x_i y_i - n\bar{x}\bar{y} &= a \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right) \end{aligned}$$

となります. ここで, 両辺を n で割ると,

$$\frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x}\bar{y} = a \left(\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 \right)$$

となり, 左辺は共分散 s_{xy} , 右辺は a かける x の分散 as_x^2 です. したがって,

$$s_{xy} = as_x^2$$

となり，求める a は

$$a = \frac{s_{xy}}{as_x^2}$$

となります．最後に， $\bar{y} - a\bar{x} - b = 0$ より， b を求めると，

$$b = \bar{y} - a\bar{x} = \bar{y} - \frac{s_{xy}}{as_x^2}\bar{x}$$

これより， x 上の y の回帰直線

$$y - \bar{y} = \frac{s_{xy}}{s_x^2}(x - \bar{x})$$

が求まります．ここまでを整理すると，

n 個のデータ $\{(x_i, y_i) \ (i = 1, 2, \dots, n)\}$ について， y の x への回帰係数は

$$a_{yx} = \frac{s_{xy}}{s_x^2} = \frac{nT_{xy} - T_x T_y}{n \sum x_i^2 - T_x^2}$$

y の x への回帰直線 l の方程式は

$$y - \bar{y} = a_{yx}(x - \bar{x})$$

回帰直線を用いることができるのはデータが x, y の時だけではありません．例えば，自動車の部品メーカーでは，あるセラミック部品の寸法 y を精度よく予測する課題に取り組むことになった． y を予測するための変数として，次の3つの変数が考えられている．

- x_1 : 注入速度
- x_2 : 材料の粒度
- x_3 : 水分量

この問題では， y を予測するために3つの変数 x_1, x_2, x_3 を用いることになる．このとき， y を目的変数， x_1, x_2, x_3 を説明変数といいます．この場合，回帰直線は

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3$$

の形をとり，この式を求めることを重回帰分析といいます．

演習問題 1.3

1. 次のデータ から得られる相関表より回帰直線を求めよう．

表 1.4: 二酸化硫黄と二酸化窒素の濃度

時刻	二酸化硫黄 x	二酸化窒素 y	時刻	二酸化硫黄 x	二酸化窒素 y
1	23	43	13	38	21
2	21	28	14	51	37
3	18	17	15	109	65
4	17	16	16	90	65
5	17	16	17	78	50
6	15	10	18	75	58
7	13	5	19	34	42
8	14	5	20	33	52
9	16	8	21	29	55
10	17	13	22	31	55
11	17	11	23	25	55
12	35	28	24	25	51

第2章 確率

2.1 順列・組み合わせ

ある事柄が何通りの起こり方があるかを考えるとき、その起こり方の個数を場合の数という。

番号のついた n 個の異なったものをある規則のもとに順に並べたものを順列 (permutation) といい、順列の総数を順列の数という。

- 順列 異なる n 個のものから r 個とって 1 列に並べる順列の数は

$${}_n P_r = n(n-1)(n-2)\cdots(n-r+1) = \frac{n!}{(n-r)!}$$

である。特に、 ${}_n P_n = n!$ 、 $0! = 1$ とする。

- 重複順列 異なる n 個のものから重複を許して r 個とる順列の数は

$${}_n \Pi_r = n^r$$

- 円順列 異なる n 個のものを円形に並べる順列の数は

$$(n-1)!$$

- 同じものを含む順列 n 個のものうち、 p 個、 q 個、 \dots 、 r 個がそれぞれ同じであるとき、これらを 1 列に並べる順列の数は

$$\binom{n}{p, q, \dots, r} = \frac{n!}{p!q!\cdots r!}, p + q + \cdots + r = n$$

並べる順序は考えずに、複数個から幾つか選んだ組を組合せ (combination) という。

- 組み合わせ 異なる n 個のものから r 個とる組み合わせの数は

$${}_n C_r = \frac{n!}{r!(n-r)!}$$

である。特に、 ${}_n C_0 = 1$ 、 ${}_n C_n = 1$ とする。

- 幾つかの組に分ける n 個の物を、 p 個、 q 個、 \dots 、 r 個の組に分ける組合せの数は

$$\binom{n}{p, q, \dots, r} = \frac{n!}{p!q!\cdots r!}, p + q + \cdots + r = n$$

- 重複組合せ n 個の異なるものから，繰り返しを許して， r 個とるときの組合せの数は

$${}_nH_r = \binom{n+r-1}{r} = \frac{n(n+1)\cdots(n+r-1)}{r!},$$

演習問題 2.1

- 0 から 6 までの 7 個の数字を取り出して並べるとき，次のような 4 けたの整数はいくつあるか求めよう。
 - すべての数字が異なる場合
 - 5 の倍数
 - 同じ数字が重複してもよい場合
- 1 から 10 までの番号のついたカードから 6 枚を取り出すとき，次のような場合の数を求めよう。
 - すべての場合
 - 1 と 2 のカードを含む場合
 - 1 または 2 のカードを含む場合
- 1 枚の硬貨を 5 回投げるとき，次の場合は何通りあるか求めよう。
 - 表の出る回数が 0 回，1 回，2 回，3 回，4 回，5 回のそれぞれの場合
 - 起こりえるすべての場合
- a, b, c, d, e, f の文字を一列に並べるのに次の場合は何通りあるか求めよう。
 - a, b が隣あう場合
 - a, b が隣合わない場合
 - a, b が両端にくる場合
- a, a, a, a, b, b, c, d の 8 個の文字を並べる順列の総数は $\binom{8}{4,2,1,1}$ で与えられることを示せ。

2.2 確率の定義

さいころを投げてどの目が出るか，1枚の硬貨を投げて表がでるか裏が出るかは，実際に投げてみないと分からない．このように，ほぼ一定の条件のもとで，繰り返し起こる現象を観察したり実験することを試行 (experiment) という．試行によって起こる結果 (outcome) は一般に多数あるが，その起こる事柄を事象 (event) という．また，試行によって起こりえるすべての事象の集まりを標本空間 (sample space) といい，一般に Ω で表す．事象のうち，これ以上簡単なものに分解できないような事象を根元事象という．

例 2.1 さいころを投げるという試行を行う．このとき，根元事象は

$$1, 2, 3, 4, 5, 6$$

の6つであり，標本空間 Ω は $\Omega = \{1, 2, 3, 4, 5, 6\}$ である．ここで，出た目が偶数であるという事象 E を考えると， $E = \{x : x \text{ は偶数} \} = \{2, 4, 6\}$ となる．

- 理論的確率 (数学的確率) ある試行において起こりえる標本空間 Ω の根元事象が全部で n 個あり，それらのどれが起こることも同様に確からしいとする．このとき，ある事象 A の起こる根元事象が r 個であるとき，事象 A の起こる確率を

$$P(A) = \frac{r}{n}$$

で定義する．

- 統計的確率 (経験的確率) ある試行を繰り返し n 回おこなったとき，事象 A が r 回起こったとする．いま，試行の回数 n を増やしていくとき，相対度数 r/n が一定の値 p に近づけば，事象 A の起こる確率を p と定義する．これを統計的確率という．

演習問題 2.2

1. 1個のさいころを6回投げるとき，次の確率を求めよう．
 - (a) 1の目が1回出る．
 - (b) 1の目が4回出る．
 - (c) 1の目が出るのは4回以下である．
 - (d) 1の目が出るのは5回以上である．
2. 1つの袋に白玉が5個，赤玉が3個，黒玉が2個入っている．その中から4個の球を取り出すとき，次の確率を求めよう．
 - (a) 全部が白である場合．
 - (b) 白がちょうど2個である場合．
 - (c) 白が2個以内．

- (d) 白が 2 個，赤が 2 個．
 - (e) 白，赤，黒がともに含まれている場合
3. 1 から 10 までの番号のついたカードがある．これらのカードを勝手に 1 列に並べるとき，次の確率を求めよう．
- (a) 1 から 10 までがその順に並ぶ場合
 - (b) 4 のカードがちょうど 4 番目にある場合
 - (c) 1 が最初に，4 が 4 番目にある場合
4. 広い方眼紙に縦横に 8cm 間隔に線を引いて，8cm 四方の正方形が沢山かかれているとする．直径が 3cm の円板を投げるとき，次の確率を求めよう．
- (a) 円板が 1 つの正方形の中に入る．
 - (b) 円板が正方形の辺にかかる．
 - (c) 円板が 4 つの正方形にまたがる．
5. 4 個の白玉と 6 個の赤玉のはいつている袋がある．この袋から，同時に 2 個を取り出すとき次の確率を求めよう．
- (a) 2 個とも白玉である確率
 - (b) 1 個だけ白玉である確率
 - (c) 少なくとも 1 個は白玉である確率

2.3 確率の公理

事象 A と事象 B とが同じであるとき, $A = B$ と書く. 事象 A が事象 B に含まれているとき, $A \subset B$ と書く.

- 全事象 標本空間 Ω 自身で表される事象
- 和事象 事象 A, B の少なくとも一方が起こる事象. $A \cup B$ で表す.
- 積事象 事象 A, B がともに起こる事象. $A \cap B$ で表す.
- 余事象 事象 A に対して A が起こらない事象. \bar{A} で表す.
- 空事象 決して起こらない事象. ϕ で表す.

事象の演算について, 集合の場合と同様に次の関係式が成り立つ.

1. 全事象を Ω , 空事象を ϕ とするとき,

$$\Omega = \Omega \cup \phi, \phi = \Omega \cap \phi$$

2. 任意の事象 A に対して

$$A \cap \bar{A} = \phi, A \cup \bar{A} = \Omega$$

3. 任意の事象 A, B, C に対して

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

4. 任意の事象 A, B に対して (DeMorgan の法則)

$$\overline{(A \cup B)} = \bar{A} \cap \bar{B}, \overline{(A \cap B)} = \bar{A} \cup \bar{B}$$

5. 任意の事象 A, B, C に対して,

$$(A \cap B) \cap (A \cap \bar{B}) = \phi, A = (A \cap B) \cup (A \cap \bar{B})$$

定理 2.1 (確率の公理) 標本空間 Ω の任意の事象 A に対して, 次の公理を満足する実数 $P(A)$ が定まるとき, $P(A)$ を事象 A の確率といい, 確率が考えられる事象を確率事象という.

- 1.

$$0 \leq P(A) \leq 1$$

- 2.

$$P(\Omega) = 1, P(\phi) = 0$$

3. 事象の列 $A_1, A_2, \dots, A_i \dots$ のいずれの2つも互いに排反であれば,

$$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

が成り立つ. この性質を確率に関する完全加法性という.

条件確率 事象 A を固定して, 事象 B の関数として

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

と定義すると, この関数は確率の公理を満たす. これを事象 A が起こったときの事象 B の条件付き確率という.

事象 A_1, A_2, \dots, A_n が互いに排反であり,

$$A_1 \cup A_2 \cup \dots \cup A_n = \Omega$$

ならば, 任意の事象 B に対して

$$B = B \cap \Omega = (B \cap A_1) \cup (B \cap A_2) \cup \dots \cup (B \cap A_n)$$

と表せる. よって,

$$P(B) = P(B \cap A_1) + P(B \cap A_2) + \dots + P(B \cap A_n)$$

ここで, 条件付き確率より, $P(B \cap A_i) = P(A_i)P(B|A_i)$ となるので,

$$P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \dots + P(A_n)P(B|A_n)$$

で求めることができる. これを Bayes の定理という.

演習問題 2.3

1. $A =$ [さいころを4回投げて少なくとも1回6の目がでる]. $B =$ [2個のさいころを同時に24回投げて少なくとも1回2個とも6の目がでる] とするとき,

(a) $P(A)$ を求めよう.

(b) $P(B)$ を求めよう.

enshu:2-3-2 ある患者がある種の症状を訴えてきた. 医師の経験から, 同じ年齢層の人がその症状を訴えるとき, 約5%の人がガンであることを知っている. 一方, ある精密検査によって真のガン患者に対しては85%の陽性反応を示し, ガン患者でない人にも5%の陽性反応を示す. もしある患者がその精密検査の結果陽性反応を示した場合, その患者がガン患者である確率を求めよう.

3. 次の関係を示そう.

(a) $P(\overline{A \cup B}) = P(\overline{A} \cap \overline{B})$

(b) $P(\overline{A \cap B}) = P(\overline{A} \cup \overline{B})$

(c) B と C が互いに排反ならば, $P((B \cup C) | A) = P(B | A) + P(C | A)$

第3章 確率変数

3.1 確率変数と確率分布

- 確率変数 x_1, x_2, \dots, x_n なる n 個の値をとる変数 X に対して, $X = x_i$ なる確率 p_i が与えられているとき, X を確率変数という.
- 確率分布 確率変数 X とそれに対応する確率 $P(X = x_i)$ との対応関係を確率分布という.
- 分布関数 確率変数 X の値がある値 x までとる確率を $F(x)$ で表し, 確率変数 X の分布関数という. つまり, 分布関数は $F(x) = P(X \leq x)$ で与えられる.

確率変数 X のとる値が有限個または, 無限個であっても自然数で番号が付けられる場合, 確率変数 X は離散型であるという. また, 確率変数 X がある区間内の全ての実数を取り得る場合, 連続型であるという.

離散型の場合

確率変数 X のとる値を x_1, x_2, \dots, x_n とし, 各事象 ($X = x_i$) の確率を p_1, p_2, \dots, p_n とするとき,

$$P(X = x_i) = p_i \quad (i = 1, 2, \dots, n) \quad \sum p_i = 1, (p_i \geq 0)$$

で表される. これより, X の確率分布 f は

X の値 x_i	x_1	x_2	\dots	x_n
$P(X = x_i) = p_i = f(x_i)$	p_1	p_2	\dots	p_n

また, 確率変数 X のとる値を $x_1 < x_2 < \dots < x_n$ とするとき, その分布関数 $F(x_r)$ は次のように求められる.

$$F(x_r) = P(X \leq x_r) = p_1 + p_2 + \dots + p_r = \sum_{i=1}^r p_i$$

確率分布 f と分布関数 F は次の性質をもつ.

1. $0 \leq p_i = f(x_i) \leq 1 \quad (i = 1, 2, \dots, n)$
2. $F(x_n) = P(X \leq x_n) = p_1 + p_2 + \dots + p_n = 1$
3. $P(a < X \leq b) = F(b) - F(a)$
4. $a < b \implies F(a) < F(b)$

連続型の場合

確率変数 X が連続的な値をとるとき, 事象 $\{X \leq x\}$ の確率が連続関数 $F(x)$ によって,

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(x) dx$$

で与えられるとき, $F(x)$ を X の分布関数といい, $f(x)$ を確率密度関数という.

演習問題 3.1

- (1) 男児と女児の出生率が等しいと仮定して, 4 児を持つ家庭の確率変数 X の値と確率分布 f を求めよ .
 (2) 1 つの袋に赤玉 4 個と白玉 6 個が入っている . 同時に 3 個の球を取り出す場合, 赤玉の個数を表わす確率変数 X と確率分布 f を求め, そのグラフをかこう . また, $P(X = 1), P(1 \leq X \leq 3)$ を求めよう .

3. 与えられた a, b ($a < b$) に対して, 関数 $f(x) = \begin{cases} k & (a < x \leq b) \\ 0 & (x \leq a, x > b) \end{cases}$
 が与えられている .

(a) $f(x)$ が確率密度関数であるためには, 定数 k はどのような値であるか .

(b) $a \leq c \leq b$ である c に対して $P(X \leq c)$ を求めよ .

4. 確率密度が

$$f(x) = \begin{cases} 0, & x \leq 0 \\ 6x(1-x) & 0 < x \leq 1 \\ 0 & x > 1 \end{cases}$$

で与えられている .

(a) 分布関数 $F(x)$ を求めよ .

(b) $P(X \leq 0.7), P(0.2 < X \leq 0.8)$ を求めよ .

5. 関数

$$f(x) = \begin{cases} e^{-x} & , x \geq 0 \\ 0 & , x \leq 0 \end{cases}$$

が与えられている .

(a) $f(x)$ は確率密度関数を与えることを示せ .

(b) $P(X \leq a) = 0.1, P(X > b) = 0.05$ であるような a, b を求めよ .

3.2 2次元確率分布

離散型の場合，2つの確率変数の取り得る値をそれぞれ

$$x_1, x_2, \dots, x_k$$

$$y_1, y_2, \dots, y_l$$

とし，

$$P_r(X = x_i) = p_i$$

$$P_r(Y = y_j) = q_j$$

とする。ここで、「 $X = x_i$ かつ $Y = y_j$ 」という事象の確率を

$$P_r(X = x_i, Y = y_j) = p_{ij}, \quad i = 1, \dots, k, \quad j = 1, \dots, l$$

で表す。このとき，

$$\begin{cases} \sum_{j=1}^l p_{ij} = p_i, \\ \sum_{i=1}^k p_{ij} = q_j \\ \sum_{i=1}^k \sum_{j=1}^l p_{ij} = \sum_{i=1}^k p_i = \sum_{j=1}^l q_j = 1 \end{cases}$$

となる。このような分布を2次元確率分布という。

演習問題 3.1

1. 2個のサイコロの出る目の確率変数を X, Y とするとき，積 XY の確率分布と分布関数を求めよ。
2. 一枚の銅貨を投げて表が出れば1，裏が出れば0と表すことにする。3枚の銅貨を投げるとき，それぞれの銅貨の表が出ることの確率変数を X, Y, Z とする。

(a) 和 $X + Y + Z$ の確率分布を求めよ。

(b) 和 $X + Y + Z$ の分布関数を求めよ。

3.3 確率変数の平均値と分散

確率変数 X に対して，次の式で定義される値 $E(X)$ を X の平均値または期待値 (Expectation) といい， $V(X)$ を X の分散という。

離散型の場合 $p_i = P(X = i)$

$$1. E(X) = x_1p_1 + x_2p_2 + \cdots + x_kp_k = \sum_{i=1}^k x_i p_i$$

$$2. V(X) = (x_1 - E(X))^2 p_1 + (x_2 - E(X))^2 p_2 + \cdots + (x_k - E(X))^2 p_k = \sum_{i=1}^k (x_i - E(X))^2 p_i$$

連続型の場合 $P(X \leq x) = F(x) = \int_{-\infty}^x f(x)dx$

$$1. E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

$$2. V(X) = \int_{-\infty}^{\infty} (x - E(x))^2 f(x) dx$$

$aX + b$ の期待値と分散

$$\begin{aligned} E(aX + b) &= \int_{-\infty}^{\infty} (ax + b)f(x)dx = a \int_{-\infty}^{\infty} x f(x) dx + b \int_{-\infty}^{\infty} f(x) dx \\ &= aE(X) + b \quad \left(\int_{-\infty}^{\infty} f(x) dx = 1 \text{ に注意} \right) \end{aligned}$$

$$\begin{aligned} V(aX + b) &= \int_{-\infty}^{\infty} [(ax + b) - (aE(x) + b)]^2 f(x) dx = \int_{-\infty}^{\infty} a^2 (x - E(x))^2 f(x) dx \\ &= a^2 V(X) \end{aligned}$$

演習問題 3.2

$$1. f(x) = \begin{cases} 1 - |x| & (-1 \leq x \leq 1) \\ 0 & (\text{その他}) \end{cases}$$

のとき次の問いに答えよう。

- (a) $f(x)$ は確率密度関数であることを示そう。
 - (b) 分布関数 $F(x)$ を求めよう。
 - (c) $P_r(-\frac{1}{2} \leq X \leq \frac{1}{2})$ を求めよう。
 - (d) $E(X), V(X)$ を求めよう。
2. 1つのサイコロを，3回投げるとき，1の目が出る回数を X とする。
- (a) X の確率分布を求めよう。
 - (b) X の平均と標準偏差を求めよう。
 - (c) X を標準化した確率変数 Z を求め，更に， Z の確率分布を求めよう。

3. Bernoulli の定理を利用して、次の確率を求めよう。

(a) 1 枚の硬貨を 1000 回投げたとき、表が現われる回数 r が

$$\left| \frac{r}{1000} - \frac{1}{2} \right| < \frac{1}{10}$$

を満たす確率。

(b) 2 枚の硬貨を 1000 回投げたとき、2 枚とも表が現われる回数 r が、

$$\left| \frac{r}{1000} - \frac{1}{4} \right| < \frac{1}{10}$$

をみたす確率。

(c) 1 枚の硬貨を 2000 回投げて、表の出る回数が 1000 回より 50 回以内の偏りである確率を求めよ。

(d) 1 枚の硬貨を 2000 回投げて、表の出る割合と理論的確率 0.5 との差が 5% 以内である確率が 99% 以上であるようにするには、何回以上投げればよいか。

3.4 多次元確率分布

演習問題 3.3

1. 1枚の銅貨を投げて表が出れば1,裏が出れば0と表わすことにする.3枚の銅貨を投げるとき,それぞれの銅貨の表がでることの確率変数を X, Y, Z として,次の問に答えよ.

(a) その和 $X + Y + Z$ の確率分布と分布関数を求めよ.

(b) 表の出る枚数 $X + Y + Z$ の平均値と分散を求めよ.

2. 確率変数 X の確率密度が次の式

$$f(x) = \begin{cases} 2x & 0 \leq x < 1 \\ 0 & \text{その他} \end{cases}$$

で与えられるとき,次の確率変数 Y, Z の確率密度,平均値,分散を求めよ.

(a) $Y = 2X + 3$

(b) $Z = X^2$

3. 2つのさいころを投げて出た目を確率変数 X, Y とする.

(a) 積 XY の期待値を求めよ.

(b) 積 XY の分散を求めよ.

第4章 理論分布

4.1 2項分布

次の1~3を満たす試行をベルヌーイ試行という。

1. 各試行において，その事象が発生するか否かのみを問題にする。
2. 各試行は統計的に独立。
3. 対象とする事象が発生する確率は，各試行を通じて一定。

1回の試行において，ある事象 X が発生する確率を p とする。 n 回のベルヌーイ試行列において，ちょうど i 回事象 X が発生する確率は

$$P_r(X = i) = \binom{n}{i} p^i (1-p)^{n-i}$$

で表され，このとき X の確率分布を2項分布といい， $X \sim B(n, p)$ と表す。

演習問題 4.1

1. 2項分布 $B(8, 0.4)$, $B(8, 0.2)$ の確率分布を求めよ。
2. ある会社で製造したプラモデルは25セット中2セットの割合で部品がかけられているとする。客が任意に3セット選ぶとき，いずれも完全なセットである確率を求めよ。
3. ある試薬をネズミに注射すると一定期間のうちに40%が死亡するという。8匹のネズミにその試薬を注射した場合，
 - (a) 8匹全部が一定期間以上生存する確率を求めよ。
 - (b) 95%の確率で，8匹のうち何匹以上が生存するということができるか。

4.2 ポワソン分布

次の 1 ~ 5 の条件を満たすものをポワソン過程という。

1. 事象はいかなる時点でもランダムに発生しうる。
2. 与えられた時間区間での事象の発生は、それと重複しない他の区間に対して独立である。
3. 微小時間 Δt における事象の発生確率は Δt に比例して小さくなっている。
4. 微小時間 Δt の間に事象が 2 回以上発生する確率は無視できる。
5. 時間 t の間に当該事象が発生する平均発生回数 λ がおおむね 5 以下である。

X をポワソン過程における事象の発生回数とすると、

$$P_r(X = r) = \frac{\lambda^r}{r!} e^{-\lambda}$$

となり、 $X \sim P_o(\lambda)$ と表す。ただし、 λ はポワソン過程における事象の平均発生回数。

ポワソン過程には、テープの傷、交換台にかかってくる電話、電球の破損、タクシー待ちなどがある。

演習問題 4.2

1. 交通事故による死亡者が 1 日平均 0.8 人であるとき、次の確率はいくらか。
 - (a) 死亡者 0 の日。
 - (b) 死亡者 6 名以上。
2. ある放射性物質から 1 秒間に放出される粒子の数は平均して 3 個である。1 秒間に 0, 1, 2, 3, 4, 5, 6 個の粒子が放出される確率を求めよ。1 秒間に少なくとも 1 個の粒子が放出される確率はいくらか。
3. ある磁気テープには、平均して 100m あたりに 2 個の傷があることが分かっている。このとき、300m の長さのテープ一巻中に傷が全くない確率を求めよ。

4.3 正規分布

確率変数 X の確率密度関数が

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma} \text{EXP} \left[-\frac{(x-\mu)^2}{2\sigma^2} \right], \quad -\infty < x < \infty$$

で与えられるとき，確率変数 X は正規分布に従うといい， $X \sim N(\mu, \sigma^2)$ と表わします．
また，

$$Pr(0 \leq Z \leq z) = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-\frac{x^2}{2}} dx$$

の値は標準正規分布表として与えられている。

標準化

確率変数 X の平均 $E(X)$ を 0 に，分散 $V(X)$ を 1 に直すことを標準化といいます．

標準化の方法

$$Z = \frac{X - E(X)}{\sqrt{V(X)}}$$

とおくと

$$E(Z) = 0, \quad V(Z) = 1$$

になります．

一様分布

確率変数 X の確率密度関数が

$$f(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{otherwise} \end{cases}$$

で与えられるとき，確率変数 X は一様分布に従うといい， $X \sim U(a, b)$ と表わします．

例題 4.1 電車が 20 分間隔で走っているとする．ランダムにホームに着いたとき，15 分以上待つ確率を求めよ．

解 X を電車の待ち時間とすると， $X \sim U(a, b)$ である．このとき，

$$f(x) = \begin{cases} \frac{1}{20}, & 0 < x < 20 \\ 0, & \text{otherwise} \end{cases}$$

よって，15 分以上待つ確率は，

$$Pr(15 \leq X \leq 20) = \int_{15}^{20} f(x) dx = \int_{15}^{20} \frac{1}{20} dx = \frac{1}{4}$$

正規分布の応用

定理 4.1 X_1, X_2, \dots, X_n が互いに独立で正規分布 $N(\mu, \sigma^2)$ に従っているとすると,

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

定理 4.2 (中心極限定理) X_1, X_2, \dots, X_n が互いに独立で同じ分布に従っているとすると、このとき、 n が十分大きければ

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} \text{ は近似的に } \sim N\left(\mu, \frac{\sigma^2}{n}\right) \text{ に従う。}$$

定理 4.3 (ラプラスの定理) $X \sim B(n, p)$ のとき、十分大きな n に対して

$$X \text{ は近似的に } N(np, np(1-p)) \text{ に従う。}$$

演習問題 4.3

1. $X \sim N(170, 10^2)$ のとき、次の確率を求めよう。

(a) $P_r(X \leq 160)$

(b) $P_r(160 \leq X \leq 175)$

2. $Z \sim N(0, 1)$ のとき、次の式を満たす λ を求めよ。

(a) $P_r(Z > \lambda) = 0.05$

(b) $P_r(|Z| > \lambda) = 0.05$

3. 全国の 20 才に男子の身長は正規分布 $N(168.9, 5.6^2)$ に従うものとする。

(a) 身長の高さに順に総数を 10 等分するためには、境界値をいくらにすればよいか。

(b) 20 才の男子 120 名を抽出して、身長の平均値が 168.9cm より 1.3cm 以上かたよる確率を求めよ。

4. 2 項分布、ポワソン分布、正規分布について、次のことがいえます。

$X \sim B(n, p)$ のとき、

$$X \sim \begin{cases} P_o(\mu) & , np \leq 5 \\ N(\mu, \sigma^2) & , np > 5 \end{cases} \text{ で近似される}$$

このことを用いて次の質問に答えよう。

(a) $X \sim B(100, 0.02)$ のとき、 $P_r(X \geq 2)$ を求めよ。

(b) $X \sim B(100, 0.2)$ のとき、 $P_r(X \geq 25)$ を求めよ。

(c) 1 個のさいころを 600 回投げて、1 の目が出る回数 S が 90 回以上 100 回以下である確率を近似せよ。

(d) 100 枚の偏りのない硬貨を同時に投げる。表の出る回数が 40 以上 60 以下の確率を近似せよ。

4.4 幾何分布

繰返し独立に試行を行うとき、注目している事象が初めて起こる直前までの試行回数を X とし、 p を注目している事象が起こる確率とすると、注目している事象が $i + 1$ 回目に起こる確率は、

$$P_r(X = i) = (1 - p)^i p$$

となる。このとき、確率変数 X は幾何分布 $G_e(p)$ に従うといい、

$$X \sim G_e(p)$$

と表す。また、

$$E(X) = \frac{q}{p}, \quad V(X) = \frac{q}{p^2}$$

で与えられる。

期待値 $E(X)$ を求める一つの方法に、母関数 (generating function) を用いる方法がある。

離散型の場合

$\eta(t) = E(t^X) = \sum_{k=0}^{\infty} t^k P(X = k)$ とおくと、 $\eta'(t) = E(Xt^{X-1})$ より、 $t = 1$ のとき、 $\eta(1) = E(X)$ となる。よって、期待値は $\eta(t)$ が求まれば微分することにより求めることができる。

連続型の場合

$\phi(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} dF(x)$ とおき、 e^{tx} を整級数で置き換えると、

$$\begin{aligned} \phi(t) &= \int_{-\infty}^{\infty} e^{tx} dF(x) \\ &= \int_{-\infty}^{\infty} \sum_{k=0}^{\infty} \frac{(tx)^k}{k!} dF(x) \\ &= \sum_0^k \frac{t^k}{k!} \int_{-\infty}^{\infty} x^k dF(x) \\ &= \sum_0^k \frac{t^k}{k!} E(X^k) \end{aligned}$$

これより、

$$\phi'(t) = \sum_1^k \frac{k t^{k-1}}{k!} E(X^k)$$

ここで、 $t = 0$ とおくと、

$$\phi'(0) = E(X)$$

となる。

演習問題 4.4

1.

4.4. 幾何分布

(a) $X \sim G_e(p)$ のとき,

$$E(X) = \frac{q}{p}, \quad V(X) = \frac{q}{p^2}$$

を示そう. ただし, $q = 1 - p$.

(b) 勝率 3 割のチームは平均して何試合目にはじめて勝つか求めよ.

4.5 超幾何分布

壺のなかに白玉が N_1 個，黒玉が N_2 個入っている。この中から 1 個ずつ，元に戻さない（非復元抽出）で n 個の玉を取り出すときの白玉の個数を X とする。白玉の数が i 個の確率は

$$P_r(X = i) = \frac{\binom{N_1}{i} \binom{N_2}{n-i}}{\binom{N}{n}}$$

となる。このとき，確率変数 X は超幾何分布に従うといい，

$$X \sim H_g(N, N_1, n)$$

と表す。また，

$$E(X) = \frac{nN_1}{N}, \quad V(X) = \frac{nN_1}{N} \cdot \frac{n - N_1}{N} \cdot \frac{N - n}{N - 1}$$

演習問題 4.5

- 10 本のくじがあり，そのうち 4 本が当たりで残り 6 本が外れである．いま 3 本くじを引いて，そのうち 2 本が当たる確率を求めよ．

4.6 指数分布

確率変数 X の確率密度関数が

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

で与えられるとき，確率変数は指数分布に従うといい， $X \sim E_x(\lambda)$ で表します．ここで，

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

で与えられる．

$$\begin{aligned} E(X) &= \int_0^{\infty} x(\lambda e^{-\lambda x}) dx = \lambda \int_0^{\infty} x e^{-\lambda x} dx \left(\begin{array}{l} u = x \quad dv = e^{-\lambda x} dx \\ du = dx \quad v = -\frac{1}{\lambda} e^{-\lambda x} \end{array} \right) \\ &= \lambda \left[-\frac{1}{\lambda} x e^{-\lambda x} + \frac{1}{\lambda^2} e^{-\lambda x} \right]_0^{\infty} \\ &= \frac{1}{\lambda} \end{aligned}$$

これより，期待値 $E(X) = \frac{1}{\lambda}$ となる．

分散は母関数を用いて求める．

$$\begin{aligned} \Phi(t) &= E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} \lambda e^{-\lambda x} dx \\ &= \int_{-\infty}^{\infty} \lambda e^{(t-\lambda)x} dx \\ &= \lambda \frac{1}{t-\lambda} e^{(t-\lambda)x} \Big|_{-\infty}^{\infty} \\ &= -\frac{\lambda}{t-\lambda} \end{aligned}$$

これより，

$$\Phi'(t) = \frac{\lambda}{(t-\lambda)^2}, \quad \Phi'(0) = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}$$

演習問題 4.6

1. ある人の通話時間 X は平均 3 分の指数分布に従うものとする．

(a) その人の通話が 4 分以内に終わる確率を求めよ．

(b) また，その人が通話を始めてから 2 分が経過しているとき，その後 4 分以内に終わる確率を求めよ．

4.7 ガンマ分布

確率変数 X の確率密度関数が

$$f(x) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}, & 0 < x < \infty \\ 0, & -\infty < x < 0 \end{cases}$$

で与えられるとき，確率変数はガンマ分布に従うといい， $X \sim \Gamma(\alpha, \beta)$ で表します．ここで，

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx \quad (4.1)$$

で与えられます．

ガンマ関数の特徴

$$\begin{aligned} \Gamma(1) &= \int_0^\infty e^{-x} dx = 1 \\ E(X) &= \alpha\beta, \quad V(X) = \alpha\beta^2 \end{aligned}$$

式 (4.1) の両辺を $\Gamma(\alpha)$ で割ると，

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

ここで， x を $\frac{x}{\beta}$ で置き換えると，

$$1 = \int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}} dx$$

ここで，被積分関数を $f(x)$ とおくと，ガンマ関数の定義が得られる．また，この式より，ガンマ関数は確率密度関数であることも分かる．

演習問題 4.7

1. あるガステーションで1週間に売れるガソリンの量 X キロリットルは，次の密度関数をもつ分布に従っているとする．

$$f(x) = \begin{cases} ce^{-1}\{-(x-2)^2 + 2\}, & 1 < x < 3 \\ ce^{-(x-2)}, & x \geq 3 \\ 0, & \text{その他} \end{cases}$$

(a) 定数 c の値を定めよ．

(b) $2.5 \leq X \leq 3.5$ となる確率を求めよ．

2. 日本人の血液型は A 型 35%，B 型 25%，AB 型 10%，O 型 30% であるといわれている．いま，4 人いたとき，4 人とも血液型が異なる確率を求めよ．

3. 電話での通話時間は平均 2 分の指数分布に従うとする．電話ボックスにすでに話し中の人も含めて 3 人いたとき，10 分以上待たなければならない確率を求めよ．

第5章 統計的推定法

5.1 統計量と標本分布

日本の小学6年生の身長を調査するとします。このとき、対象全体についての調査を全数調査といいます。しかし、全数調査は労力や経費の点から不可能なことがよくあります。そこで、全数調査に代わるものとして、対象全体から何らかの方法で一部の対象を選び出し調査を行い、それにより対象全体についての推測する方法を標本調査といいます。このとき、調査対象となる小学6年生の身長の集まりを母集団 (population) といいます。また、調査のために選び出された6年生の身長の集まりを標本 (sample) といいます。

標本抽出

日本の小学6年生を Π とし、小学6年生の各人の身長を X とすると、母集団は (Π, X) と表せます。この母集団から取り出した n 個の要素の組 (x_1, x_2, \dots, x_n) を大きさ n の標本といいます。このとき、個々の x_i は X と同じ分布をする確率変数 X_i が実現した数値でなければなりません。そこで、確率変数の組 (X_1, X_2, \dots, X_n) を大きさ n の確率標本変数といいます。確率標本変数 (X_1, X_2, \dots, X_n) に要求される数学的条件は、各 X_i が母集団 (Π, X) の X と同じ分布をする独立な確率変数であることです。では、実際に標本を選ぶときには、どのようにしたらよいのでしょうか。それには、個々の標本が全く偶然に、つまり同じ確率で現れるように選ばれる必要があります。例えば、6人から1人を選ぶには、正しいサイコロを振って決めるとか、52人から2人を選ぶとき、トランプのカードに各人を対応させて、よく切ったあと2枚を選ぶなどがあります。このようにして、標本を選ぶことを無作為抽出またはランダム抽出といいます。そして、このようにして選ばれた標本を確率標本といいます。

この母集団から無作為に抽出された標本を

$$X_1, X_2, \dots, X_n$$

とします。標本確率変数 $X_i (i = 1, 2, \dots, n)$ は互いに独立に母集団分布に従います。よって、

$$E(X_i) = \mu, \quad V(X_i) = \sigma^2$$

となります。ここで、標本 X_1, X_2, \dots, X_n を用いて母平均と母分散をすいてすることを考えます。まず、素朴に考えて、 X_1, X_2, \dots, X_n を n 個のデータの集まりとして、その平均と分散を求めます。すると、

$$\begin{aligned} \text{標本平均} \quad \bar{X} &= \frac{1}{n} \sum_{i=1}^n X_i \\ \text{標本分散} \quad S^2 &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \end{aligned}$$

を得ます。このとき、 \bar{X} は母平均 μ を、 S^2 は母分散 σ^2 を推測するのに、適当な統計量かという疑問がでます。

例題 5.1

標本平均の分散と標準偏差を求めよう。

解答

$$\begin{aligned}
V(\bar{X}) &= E(\bar{X}^2) - E(\bar{X})^2 \\
&= E\left(\frac{1}{n}\left(\sum_{i=1}^n X_i^2\right)\right) - \mu^2 \\
&= \frac{1}{n^2}E(X_1^2 + \cdots + X_n^2 + 2(X_1X_2 + \cdots + X_{n-1}X_n)) - \mu^2 \\
&= \frac{1}{n^2}\left(\sum_{i=1}^n E(X_i^2) + 2\sum_{1 \leq i, j \leq n} E(X_iX_j)\right) - \mu^2 \\
&= \frac{1}{n^2}\sum_{i=1}^n(\sigma^2 + \mu^2) + \frac{2}{n^2}\binom{n}{2}\mu^2 - \mu^2 = \frac{\sigma^2}{n}
\end{aligned}$$

したがって、標本平均の標準偏差は $\frac{\sigma}{\sqrt{n}}$

定理 5.1 (チェビシェフの定理)

確率変数 X の平均値を μ 、標準偏差を σ とすると、定数 $\lambda > 1$ に対して

$$P(|X - \mu| \geq \lambda\sigma) \leq \frac{1}{\lambda^2}$$

または

$$P(|X - \mu| < \lambda\sigma) \geq 1 - \frac{1}{\lambda^2}$$

例題 5.2

母平均 μ 、母分散 σ^2 の母集団 (Π, X) がある。ここから抽出した大きさ n の標本平均を \bar{X} とする。いま \bar{X} と μ との差が標準偏差 σ の $\frac{1}{5}$ より小さい確率を 0.9 以上にしたい。 n をいくらにとればよいか。

解答 題意を式で表すと

$$P(|\bar{X} - \mu| < \frac{\sigma}{5}) \geq 0.9$$

一方、 \bar{X} の標準偏差は $\frac{\sigma}{\sqrt{n}}$ であるから、チェビシェフの定理を用いると

$$P(|\bar{X} - \mu| < \frac{\lambda\sigma}{\sqrt{n}}) \geq 1 - \frac{1}{\lambda^2}$$

よって、 $\frac{\lambda}{\sqrt{n}} = \frac{1}{5}$ で

$$P(|\bar{X} - \mu| < \frac{\sigma}{5}) \geq 1 - \frac{25}{n} \geq 0.9$$

とすればよい。これから、 $n \geq 250$ とすればよいことが分かる。

統計的推定

母集団から無作為に抽出された標本

$$X_1, X_2, X_3, \dots, X_n$$

から、標本平均

$$\bar{X} = \frac{1}{n}[X_1 + X_2 + X_3 + \cdots + X_n]$$

標本分散

$$S^2 = \frac{1}{n}[(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2]$$

といった標本の統計量の値 (統計値) を用いて, 母集団の分布に含まれる母数 (母平均, 母分散) の値を推定することを統計的推定といいます.

点推定

点推定は母数を1個の数値で定めようとする方法のことです. 全数調査ができれば, 母集団の母数は簡単に求めることができます. しかし, 大事なことは, 全数調査ができないときに, 標本を通して母数の情報を得ることです.

母数を θ とし, これに対し大きさ n の標本変数 $\{x_1, x_2, \dots, x_n\}$ の統計量 $T(x_1, x_2, \dots, x_n)$ を考えます. この関数に抽出された標本値 (X_1, X_2, \dots, X_n) を代入した値 $\hat{\theta} = T(X_1, X_2, \dots, X_n)$ でもって, θ の値であると推定することを, θ の点推定という.

不変推定量

ある推定値 $\hat{\theta} = T(X_1, X_2, \dots, X_n)$ について,

$$E(\hat{\theta}) = E(T(X_1, X_2, \dots, X_n)) = \theta$$

のとき, $\hat{\theta}$ を θ の不偏推定量という.

母集団 $N(\mu, \sigma^2)$ において, 次の統計量是不偏推定量である.

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad U^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

この意味で, U^2 を不偏分散という.

例題 5.3

\bar{X} は不偏推定量であることを示そう.

解 $E(X + Y) = E(X) + E(Y)$ より,

$$nE(\bar{X}) = E(X_1) + E(X_2) + \cdots + E(X_n) = n\mu$$

したがって, $E(\bar{X}) = \mu$.

しかし, 母分散 σ^2 を推定するには, 不偏性とは異なる立場をとると, U^2 よりも標本分散

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{n-1}{n} U^2$$

の方が良いこともあります. このようになる理由は, n を増やしても U^2 や S^2 が σ^2 を中心とする狭い区間に入る確率がなかなか大きくなりません.

演習問題 5.1

1. 次のデータの不偏分散を求めよう. また, 標準偏差を求めよう.

110, 121, 133, 124, 126, 118, 112, 125, 131, 120(cm)

5.2 信頼区間

区間推定

母数 θ がある区間 $[\theta_1, \theta_2]$ に入るだろうと推定するのが区間推定です。詳しくいうと、母数 θ を推定するために、母集団から無作為に抽出された標本から 2 つの統計値 θ_1, θ_2 を定める。このとき、あらかじめ指定された小さな確率 α ($0 < \alpha < 1$) に対して、常に

$$P_r(\theta_1 < \theta < \theta_2) = 1 - \alpha$$

が満たされるとき、区間 (θ_1, θ_2) を θ の信頼区間、 θ_1, θ_2 を信頼限界、 $100(1 - \alpha)\%$ を信頼係数または信頼度といます。信頼区間 $[\theta_1, \theta_2]$ を求めることを区間推定といます。

θ は一定値ですが、区間 $[\theta_1, \theta_2]$ は標本によっていろいろ変わり、この区間に θ が入る確率が $1 - \alpha$ です。

区間推定法

母集団が正規分布 $N(\mu, \sigma^2)$ に従い、母分散 σ^2 が分かっているとします。このとき、母平均 μ はどの範囲にあるかを、どのくらい信頼できるかを考えて表わしてみましょう。

準備

標本 X_1, X_2, \dots, X_n が、 $X_i \sim N(\mu, \sigma^2)$ のとき、

$$E(\bar{X}) = \mu, V(\bar{X}) = \frac{\sigma^2}{n}$$

より

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

と表せます。また、

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \text{ の期待値 } E(S^2) = \frac{n-1}{n} \sigma^2$$

より

$$S'^2 = \frac{n}{n-1} S^2 \text{ の期待値 } E(S'^2) = \sigma^2$$

と表せます。

母平均 μ の区間推定 (σ^2 既知)

ここでは $\alpha = 0.05$ つまり、95%信頼区間を推定します。まず、

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

より標準化を行なうと、

$$Z = \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \sim N(0, 1)$$

これより、

$$P_r(|Z| \leq z_{\frac{\alpha}{2}}) = 1 - \alpha = 0.95$$

ここで、 $z_{\frac{\alpha}{2}}$ は、

$$P_r(Z \geq z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$$

を満たす点です．このとき， $z_{\frac{\alpha}{2}}$ を標準正規分布表の両側確率で求めると， $\alpha = 0.05$ のとき， $z_{\frac{\alpha}{2}}$ は

$$z_{\frac{\alpha}{2}} = 1.96$$

となります．よって求める信頼区間は次の不等式を満たします．

$$|Z| = \left| \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \right| \leq z_{\frac{\alpha}{2}}$$

この不等式を μ について解くと

$$\bar{X} - z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{n}} \leq \mu \leq \bar{X} + z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{n}}$$

を得ます．これが母平均 μ の信頼区間となります．

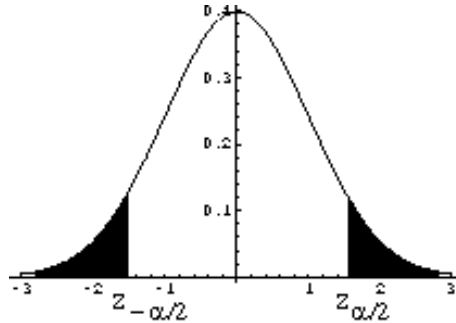


図 5.1: 正規分布

例題 5.4 標本 28, 24, 31, 27, 22 が与えられたとして，標準偏差が 2.2 である正規母集団の平均に対する 95% 信頼区間を求めよう．

解答 標準偏差が 2.2 より，母分散 $\sigma^2 = 6.25$ は既知である．この母集団から無作為に選んだ標本 X_i は $X_i \sim N(\mu, 6.25)$ の正規分布に従っていると考えられる．したがって，

$$\bar{X} \sim N(\mu, \sigma^2/5)$$

となる．ここで， \bar{X} を求めると，

$$\bar{X} = \frac{1}{5}[28 + 24 + 31 + 27 + 22] = \frac{132}{5} = 26.4$$

標準化を行なうと，

$$Z = \frac{\bar{X} - \mu}{\sqrt{\sigma^2/5}} \sim N(0, 1)$$

となる．95%信頼区間より， $P_r(|Z| \leq z_{\frac{\alpha}{2}}) = 0.95$ ．また， $z_{\frac{0.05}{2}} = 1.96$ ．したがって，

$$\bar{X} - z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{5}} \leq \mu \leq \bar{X} + z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{5}}$$

$$26.4 - 1.96\sqrt{6.25/5} \leq \mu \leq 26.4 + 1.96\sqrt{6.25/5}$$

$$24.21 \leq \mu \leq 28.59$$

次に，母集団が正規分布に従うことは分かっているが母分散 σ^2 が不明である場合を考えます．
平均値の区間推定 (σ^2 未知)

ここでは $\alpha = 0.05$ つまり，95%信頼区間を推定します．この場合，2つの母数 μ, σ^2 が必要となりますが， σ^2 が未知なので， σ^2 を推定する不偏分散 S'^2 を σ^2 の代わりに用います．すると，母分散に無関係に

$$T = \frac{\bar{X} - \mu}{\sqrt{S'^2/n}}$$

は，自由度 $n - 1$ の t 分布に従うことが知られています．これより，

$$P_r(|T| \leq t_{n-1, \alpha/2}) = 1 - \alpha$$

となります．ここで， $t_{n-1, \alpha/2}$ は，

$$P_r(T \geq t_{n-1, \alpha/2}) = \frac{\alpha}{2}$$

を満たす点である．このとき， $t_{n-1, \alpha/2}$ を t 分布表の両側確率で求めると， $\alpha = 0.05$ ， $n = 10$ のとき， $t_{9, 0.05/2}$ は

$$t_{9, 0.05/2} = 2.26$$

よって求める信頼区間は次の不等式を満たします．

$$\left| \frac{\bar{X} - \mu}{\sqrt{S'^2/n}} \right| \leq t_{n-1, \alpha/2}$$

この不等式を μ について解くと

$$\bar{X} - t_{n-1, \alpha/2} \sqrt{\frac{S'^2}{n}} \leq \mu \leq \bar{X} + t_{n-1, \alpha/2} \sqrt{\frac{S'^2}{n}}$$

を得ます．

演習問題 5.2

1 ある水域の一定区間における水質 BOD(ppm) はほぼ正規分布に従い，その母分散は $\sigma^2 = 6.25(\text{ppm})^2$ であることがわかっている．いま $n = 15$ 個の標本をとり，標本平均 $\bar{X} = 7.2\text{ppm}$ を得た．このとき信頼度 95% で，この水質の母平均の区間推定をせよ．

2 標本 145.3, 145.1, 145.4, 146.2 が与えられたとして，母平均が 146 である正規母集団の分散に対する 95% 信頼区間を求めよう．

5.3 母比率の区間推定 (大標本の場合)

母集団の中で、ある属性に対して事象 A の起こる割合 p を事象 A の母比率といいます。母比率が p の二項母集団から抽出された大きさ n の標本を (X_1, \dots, X_n) とします。ここで、

$$X_i = \begin{cases} 1 & A \text{ のとき} \\ 0 & \bar{A} \text{ のとき} \end{cases}$$

とします。このとき、 $X = X_1 + \dots + X_n$ とすると、 X は標本中 A であるものの個数を表す統計量で、 $\frac{X}{n}$ は事象 A の標本比率といいます。

$\frac{X}{n}$ は母比率 p の不偏推定量である

母比率 p の二項母集団から大きさ n の標本 (X_1, \dots, X_n) をとり、 $X = X_1 + \dots + X_n$ とすると X は二項分布 $B(n, p)$ に従います。ここで n が十分大きいときにはラプラスの定理によって、 X は近似的に正規分布 $N(np, np(1-p))$ に従い、標本比率 $\frac{X}{n} = \hat{p}$ は近似的に正規分布 $N\left(p, \frac{p(1-p)}{n}\right)$ に従います。よって、標準化を行うと

$$\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0, 1)$$

したがって、

$$P\left(\left|\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}\right| \leq z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

が成り立ちます。この式を書き直すと

$$\hat{p} - z_{\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}} \leq p \leq \hat{p} + z_{\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}$$

この両辺は母数 p を含んでいるが、 n が非常に大きいときには \hat{p} で近似できるので、母比率 p の信頼度 $100(1-\alpha)\%$ の信頼区間は

$$\left(\hat{p} - z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

となります。

定理 5.2 (ラプラスの定理) $X \sim B(n, p)$ のとき、十分大きな n に対して

X は近似的に $N(np, np(1-p))$ に従う。

演習問題 5.3

1 サイコロを 600 回投げたところ、1 の目が 108 回出たという。1 の目が出る母比率 p を信頼度 95% で区間推定せよ。

5.4 母比率の区間推定 (小標本の場合)

標本の大きさが大きければ大きいほど狭い区間で母比率の推定が可能となる。したがって、多くの標本があればよいのですが、薬品のような人体実験の場合、標本を多くとることが難しい場合があります。標本の大きさ n が小さいとラプラスの定理を用いることができません。そこで、標本の大きさ n が小さい場合には、次のような方法をとって区間推定を行います。母比率 $p = P(A)$ の二項母集団から大きさ n の標本を抽出したとき、 A であるものの個数 (標本和) が x であったとします。 n があまり大きくないとき、母比率 p の信頼度 $100(1 - \alpha)\%$ の信頼区間 (p_1, p_2) は次のようにして得ることができます。

$$\left(\frac{n_2}{n_1 f_1 + n_2}, \frac{m_1 f_2}{m_1 f_2 + m_2} \right)$$

ただし、 $n_1 = 2(n - x + 1)$, $n_2 = 2x$ とし、自由度 (n_1, n_2) の F 分布に従う確率変数 F が

$$P(F > f_1) = \frac{\alpha}{2}$$

となる f_1 を F 分布表から求める。

演習問題 5.4

1 あるテレビ番組が、無作為標本 900 台のテレビのうち 180 台で見られていることが分かった。この番組の視聴率を 95%の信頼度で区間推定せよ。

2 ある大学の学生から無作為に 300 人を選んで、アルバイトをしているかどうかの調査をしたところ、187 人が何かのアルバイトをしていた。その大学の全学生のうちアルバイトをしている学生の比率の信頼度 95%の信頼区間を求めよ。

5.5 重要な標本分布

第2章で、確率分布の基礎となる2項分布と正規分布の話をしました。ここでは、まず、正規分布の性質について考えます。

正規分布の加法性

確率変数 X, Y が独立で、それぞれ正規分布 $N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2)$ に従うとき、和 $aX + bY$ は正規分布

$$N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$$

に従う。

定理 5.3 X_1, X_2, \dots, X_n は互いに独立で正規分布 $N(\mu, \sigma^2)$ に従えば、

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

証明 正規分布の加法性より、

$$X_1 + X_2 + \dots + X_n \sim N(n\mu, n\sigma^2)$$

したがって、

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \dots + X_n) \sim N\left(\mu, \frac{n\sigma^2}{n^2}\right) = N\left(\mu, \frac{\sigma^2}{n}\right)$$

例題 5.5

ある年齢の生徒の身長は 120cm を平均値として、標準偏差が 4.5cm の正規分布に従っているとすると、このとき、50 人の身長の平均が 120.6cm よりも大きくなる確率を求めよ。解 X_i をある年齢の生徒の身長とすると、 $X_i \sim N(120, 4.5^2)$ 。50 人の相加平均 \bar{X} は定理 5.3 より $\bar{X} \sim N(120, \frac{4.5^2}{50})$ 。よって

$$\begin{aligned} P(\bar{X} > 120.5) &= P\left(Z > \frac{120.5 - 120}{\frac{4.5}{\sqrt{50}}}\right) = P(Z > 0.9428) \\ &= 0.5 - P(0 < Z < 0.9528) = 0.5 - 0.32710 \approx 0.173 \end{aligned}$$

次に、 X_i が正規分布に従わない場合を考える。

[中心極限定理] X_1, X_2, \dots, X_n が互いに独立で、平均値 μ 、分散 σ^2 の同じ確率分布に従うとする。このとき、 n が十分大きいならば

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} \text{ 近似的に } N\left(\mu, \frac{\sigma^2}{n}\right) \text{ に従う}$$

$n > 100$ ならば近似度はよくなる。

5.6 χ^2 分布

χ^2 分布は1つの自然数 n を含む連続型分布で, $\chi^2(n)$ と表し n をその自由度という。 χ^2 分布の密度関数 $f_n(x)$ は次の式で与えられる。

$$f_n(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{1}{2}x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

ここで, ガンマ関数 $\Gamma(x)$ は

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt \quad (x > 0)$$

で定義される。

χ^2 分布の名前は次の性質から来ている。

定理 5.4 確率変数 X_1, X_2, \dots, X_n が同一の標準正規分布 $N(0, 1)$ に従い, 互いに独立ならば, その統計量

$$\chi_n^2 = X_1^2 + X_2^2 + \dots + X_n^2$$

は自由度 n の χ^2 分布に従う。その期待値と分散は

$$E(\chi_n^2) = n, \quad V(\chi_n^2) = 2n$$

定理 5.5 (χ^2 分布の加法性) χ_n^2, χ_m^2 がそれぞれ自由度 n, m の χ^2 分布に従い, 互いに独立ならば, $\chi^2 = \chi_n^2 + \chi_m^2$ は自由度 $n + m$ の χ^2 分布に従う。

標本分散 S^2 に関して, 次の定理がある。

定理 5.6 $N(\mu, \sigma^2)$ の正規分布に従う母集団から無作為で得た標本を $\{X_1, X_2, \dots, X_n\}$ とすると,

$$Y = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{nS^2}{\sigma^2}$$

は自由度が $n - 1$ の χ^2 分布に従って分布する。

母集団が正規分布であるとする。 n が 20 の標本から標本分散を求めたところ, その値は 1.5 であった。母分散が 1 のとき, 標本分散が 1.5 より大きい確率を求めよ。

解 $P(S^2 \geq 1.5)$ を求める。 $X_i \sim N(\mu, 1)$ より,

$$Y = \frac{20S^2}{1} = \sum_{i=1}^{20} (X_i - \bar{X})^2$$

は自由度 19 の χ^2 分布に従う。したがって,

$$\begin{aligned} P(S^2 \geq 1.5) &= P(20S^2 \geq 28.5) \\ &= P(\chi_{19}^2 \geq 28.5) \end{aligned}$$

ここで、 χ^2 分布表を用いると、 $P(\chi_{19}^2 > 27.20) = 0.10$ で $P(\chi_{19}^2 > 30.14) = 0.05$ より、

$$P(\chi_{19}^2 \geq 28.5) = 0.05 + \frac{28.5 - 27.20}{30.14 - 27.20}(0.10 - 0.05) \approx 0.07$$

5.7 t 分布

定義 5.1 確率密度関数 $f_n(x)$ が

$$f_n(x) = \frac{\gamma(\frac{n+1}{2})}{\sqrt{n\pi}\gamma(\frac{n}{2})} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} \quad (n \geq 1)$$

で与えられる分布 T_n を自由度 n の t 分布という. $n \geq 3$ のとき, その期待値と分散は

$$E(T_n) = 0, \quad V(T_n) = \frac{n}{n-2}$$

となる.

定理 5.7 X_1, X_2, \dots, X_n がいずれも正規分布 $N(\mu, \sigma^2)$ に従う互いに独立な確率変数とする. このとき,

$$U^2 = S'^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

とし, 正の平方根を U とすると,

$$T = \frac{\bar{X} - \mu}{U/\sqrt{n}}$$

は自由度 $n-1$ の t 分布に従う。

定理 5.8 Z を標準正規確率変数, χ_n^2 を自由度 n の χ^2 確率変数とする. さらに, Z と χ_n^2 が互いに独立ならば, 標本分布

$$T_n = \frac{Z}{\sqrt{\chi_n^2/n}}$$

は自由度 n の t 分布に従う。

自由度 n の t 分布を $t(n)$ と表す.

t 分布の正規近似

t 分布は自由度が大きければ標準正規分布で近似でき,

$$Pr(T_n \leq c) \approx Pr(Z \leq c)$$

となる. χ_n^2 は n 個の独立な確率変数の和であるから, n が大きければ χ_n^2/n は大数の法則により, 1 に収束する. T_n 確率変数の分母が 1 に近づくから, T_n 確率変数は分子の標準正規確率変数と変わらなくなる。

5.8 F 分布

定義 5.2 確率密度関数 $f_{m,n}(x)$ が

$$f_{m,n}(x) = \begin{cases} \frac{\gamma(\frac{m+n}{2})}{\gamma(\frac{m}{2})\gamma(\frac{n}{2})} m^{\frac{m}{2}} n^{\frac{n}{2}} \frac{x^{\frac{m}{2}-1}}{(mx+n)^{\frac{m+n}{2}}} & (x > 0) \\ 0, & \end{cases}$$

で与えられる分布 $F_{m,n}$ を自由度対 (m, n) の F 分布という。その期待値と分散は

$$\begin{aligned} E(F_{m,n}) &= \frac{m}{n-2} \quad (n > 2) \\ V(F_{m,n}) &= \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)} \quad (n > 4) \end{aligned}$$

となる。

定理 5.9 $X_{11}, X_{12}, \dots, X_{1n_1}$ がいずれも正規分布 $N(\mu_1, \sigma_1^2)$ に従う互いに独立な n_1 個の確率変数で、その相加平均を \bar{X}_1 、不偏分散を U_1^2 とする。これらと独立に、 X_{21}, \dots, X_{2n_2} は正規分布 $N(\mu_2, \sigma_2^2)$ に従う互いに独立な n_2 個の確率変数を考え、その相加平均を \bar{X}_2 、不偏分散を U_2^2 とする。

$$\begin{aligned} \bar{X}_1 &= \frac{1}{n_1 \sum_i X_{1i}}, \quad U_1^2 = \frac{1}{n_1 - 1} \sum_i (X_{1i} - \bar{X}_1)^2 \\ \bar{X}_2 &= \frac{1}{n_2 \sum_i X_{2i}}, \quad U_2^2 = \frac{1}{n_2 - 1} \sum_i (X_{2i} - \bar{X}_2)^2 \end{aligned}$$

このとき、

$$F = \frac{U_1^2/\sigma_1^2}{U_2^2/\sigma_2^2} = \frac{\sigma_2^2 U_1^2}{\sigma_1^2 U_2^2} \sim F(n_1 - 1, n_2 - 1)$$

例題 5.6

$F_{n_2}^{n_1}(0.05) = F_{10}^5(0.05)$ を求めよ。

解

$$F_{10}^5(0.05) = 3.3258$$

$F_{n_2}^{n_1}(1 - \alpha)$ を求めるには、次の公式を用いる

$$F_{n_2}^{n_1}(1 - \alpha) = \frac{1}{F_{n_1}^{n_2}(\alpha)}$$

例題 5.7

$F_{11}^5(1 - 0.05)$ を求めよ。

解

$$F_{11}^5(1 - 0.05) = \frac{1}{F_5^{11}(0.05)} = \frac{1}{3.10} = 0.32$$

第6章 統計的検定

6.1 統計的検定の考え方

超心理学では透視実験に ESP カードを用います。カードは 5 種類からなっています。そこで 1 枚のカードを引いて裏向きに置いて、このカードの種類をあてさせます。カードを元に戻し同じ実験を 5 回繰り返したところ、ある学生は 3 回の中しました。そこで問題です。この学生の透視能力についてどのような判断を下すべきか考えてみましょう。

2 つの結論が考えられます。

結論 1. 透視能力が無くても、5 回中 3 回ぐらいは偶然でも的中すると考えられるので、これだけのデータでは透視能力があるとはいえない。

結論 2. 5 回中 3 回の中することは滅多にないことだから透視能力があると考えの方がもっともらしい。

この 2 つの結論のどちらを選ぶべきかの基準に確率が用いられます。

まずこの学生がカードをあてる確率は毎回一定で、その確率は p とします。5 回の実験中的中する回数を X とすると、 $X \sim B(5, p)$ に従います。ここで透視能力がないということは $[p = 0.2]$ 、透視能力があるということとは $[p > 0.2]$ と表わせます。

そこで

$$H_0 : p = 0.2 \quad (\text{透視能力がない})$$

と仮定してみます。

ここで 5 回中 3 回以上の中する確率は

$$\begin{aligned} P_r(X \geq 3) &= P_r(X = 3) + P_r(X = 4) + P_r(X = 5) \\ &= \binom{5}{3}(0.2)^3(0.8)^2 + \binom{5}{4}(0.2)^4(0.8) + \binom{5}{5}(0.2)^5 = 0.057922 \end{aligned}$$

となり、この確率を有意確率といいます。

実験の結果がそれほど稀な現象ではない、つまり有意確率がそれほど小さくないと判断した場合は結論 1 を得ます。このことを仮説 H_0 を容認するといいます。

実験の結果がきわめて稀な現象である、つまり有意確率がきわめて小さいと判断した場合は結論 2 になります。このことを仮説 H_0 を棄却するといいます。

有意確率がどの程度小さければ、 H_0 を棄却したらよいかという基準を有意水準 (significance level) α とよび、 α として 0.05, 0.01 等が良く用いられます。この問題で有意水準を 0.05 とすると、仮説 H_0 は棄却されない (容認されます)。つまりこの学生は透視能力がないと判断されます。

仮説 H_0 が棄却される X の範囲は $P_r(X \geq 4) = 0.00672$ より $X \geq 4$ です。この範囲を棄却域 (critical region) といいます。また $H_0 : p = 0.2$ を帰無仮説 (null hypothesis)、 $H_1 : p > 0.2$ を対立仮説 (alternative hypothesis) といいます。

母数 θ に関する帰無仮説 $H_0 : \theta = \theta_0$ に対し対立仮説として次の 3 つがあります。

$$H_1 : \theta > \theta_0, \quad H_1 : \theta < \theta_0, \quad H_1 : \theta \neq \theta_0$$

母数 θ に関する検定の手順

1. 帰無仮説、対立仮説を立てる。
2. 有意水準 α を定める。

3. 帰無仮説のもとで検定に用いる統計量の分布を求める .
4. 棄却域を定める .
5. 検定統計量の実現値が棄却域に入るときは帰無仮説を棄却する .

(1) 母平均の検定

母集団が正規分布 $N(\mu, \sigma^2)$ に従う, 正規母集団から抽出された大きさ n の標本変量 (X_1, X_2, \dots, X_n) を考えます . このとき, これらの相加平均 \bar{X} は $N(\mu, \frac{\sigma^2}{n})$ に従っています . ここで, 母平均 μ についての検定を考えます .

(a) σ^2 が既知の場合の μ の検定 (有意水準 α)

この場合は次の標本分布を用います .

$$Z = \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \sim N(0, 1)$$

(b) σ^2 が未知の場合の μ の検定 (有意水準 α)

この場合は次の標本分布を用います .

$$T = \frac{\bar{X} - \mu}{\sqrt{S'^2/n}} \sim t(n-1)$$

(2) 母分散の検定

母集団が正規分布 $N(\mu, \sigma^2)$ に従う, 正規母集団から抽出された大きさ n の標本変量 (X_1, X_2, \dots, X_n) を考えます . ここで, 母分散 σ^2 についての検定を考えます .

(a) μ が既知の場合の σ^2 の検定 (有意水準 α)

この場合は次の標本分布を用います .

$$\chi^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 \sim \chi_{\alpha, n}^2$$

(b) μ が未知の場合の σ^2 の検定 (有意水準 α)

この場合は次の標本分布を用います .

$$\chi^2 = \frac{nS^2}{\sigma^2} \sim \chi_{\alpha, n-1}^2$$

棄却域

(a) σ^2 が既知の場合の μ の検定 (有意水準 α)

帰無仮説 $H_0: \mu = \mu_0$ に対し次の 3 通りの対立仮説を考えます .

$$(1) H_1: \mu > \mu_0 \quad (2) H_1: \mu < \mu_0 \quad (3) H_1: \mu \neq \mu_0$$

帰無仮説のもとで (つまり, $\mu = \mu_0$ のとき)

$$Z_0 = \frac{\bar{X} - \mu_0}{\sqrt{\sigma^2/n}} \sim N(0, 1)$$

となります。有意水準 α に対して、

$P_r(Z_0 > z_\alpha) = \alpha$ であるような z_α を正規分布表から読み取れば、有意水準 α の対立仮説 (1) に対する帰無仮説の棄却域は $Z_0 > z_\alpha$ となります。

これより各対立仮説に対する帰無仮説の棄却域は次の通りです。

(1) の場合 $Z_0 > z_\alpha$
 (2) の場合 $Z_0 < z_\alpha$
 (3) の場合 $|Z_0| > z_{\frac{\alpha}{2}}$

例題 6.1

ある大学では一年生に対して毎年同じテストを行なっている。昨年度の一年生の成績は平均 64.5, 分散 20 の正規分布に従っている。今年度の一年生にも同じテストを行ない、無作為に 8 人抽出したところ点数は次の通りであった。

66	73	55	69	70	67	62	71
----	----	----	----	----	----	----	----

今年度の一年生の平均点は昨年度より高いか有意水準 5% で検定せよ。ただし、今年度の分散は、昨年度と変わらないものとする。

解

今年度の 1 年生の平均点を μ とおき、昨年度より高いかの検定を行なうので、以下の帰無仮説と対立仮説を立てる。

$$H_0 : \mu = 64.5$$

$$H_1 : \mu > 64.5$$

今年度の標本 X_i は $X_i \sim N(\mu, 20)$ と考えることができる。したがって、

$$\bar{X} = \frac{1}{8}[66 + 73 + 55 + 69 + 70 + 67 + 62 + 71] = 533/8 = 66.625$$

また、 $\bar{X} \sim N(\mu, 20/8)$ 。帰無仮説のもとで、標準化を行なうと、

$$Z_0 = \frac{66.625 - 64.5}{\sqrt{20/8}} = \frac{2.125}{1.58} = 1.345$$

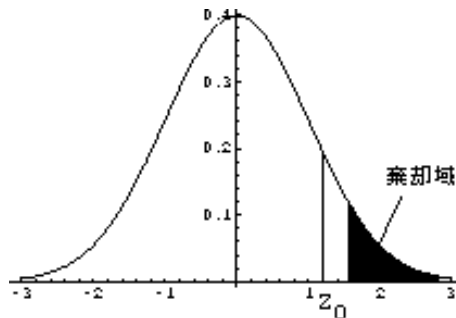


図 6.1: 正規分布

これより、 Z_0 は帰無仮説の棄却域に入っていない。したがって、帰無仮説を棄却できない。言い換えると、今年度の1年生は昨年度よりも優秀であるとは、結論付けることができない。

演習問題 6.1

1 ある工場での経験によると7mmボルトの規格の製品の寸法はほぼ正規分布をしており、その標準偏差は0.20mmであるという。ある日の製品から16個のボルトを無作為抽出したところ、その寸法の平均が7.09mmであった。この日の製品の寸法の平均は規格から外れているか。有意水準 $\alpha = 0.05$ で検定せよ。またこの日の製品の寸法の平均 μ の95%信頼区間を求めよう。

2 1台の機械が製造する鋼球の直径(単位 mm)は正規分布に従っており、その分散は従来からの経験から0.0016であるといわれている。ある日この機械が製造した鋼球から8個を抽出しその直径を測定したところ次のようになった。

11.97 12.02 12.06 12.03 11.99 11.98 12.12 12.05

この日の製品の直径の分散は0.0016より大きいといえるか。有意水準 $\alpha = 0.05$ で検定せよ。またこの日の製品の直径の分散の95%信頼区間を求めよう。

6.2 母集団が正規分布で2標本の場合

正規母集団 X は母平均 μ_1 , 母分散 σ_1^2 であるとし, 正規母集団から無作為抽出した標本を X_1, X_2, \dots, X_{n_1} , 標本平均を \bar{X} , 標本分散を S_1^2 とする. 同様に正規母集団 Y は母平均 μ_2 , 母分散 σ_2^2 であるとし, 正規母集団 Y から無作為抽出した標本を Y_1, Y_2, \dots, Y_{n_2} , 標本平均を \bar{Y} , 標本分散を S_2^2 とすると,

$$\bar{X} \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right), \bar{Y} \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$$

となります.

1. 母平均の差 $\mu_1 - \mu_2$ の検定

(a) σ_1^2, σ_2^2 既知の場合

$\bar{X} - \bar{Y}$ は正規分布の加法性より, 正規分布 $N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$ に従っています. よって,

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim N(0, 1)$$

例題 6.2

A校から30人, B校から50人の標本を抽出して身長を調べたところ, それぞれ平均値148.2cm, 146.4cmであった. この年齢の生徒の身長は, 標準偏差4.8cmの正規分布に従って分布するとする. 両校の身長に有意差があるか, 有意水準0.05で検定せよ.

解 A校の生徒は $N(\mu_1, 4.8^2)$, B校の生徒は $N(\mu_2, 4.8^2)$ で, 大きさ $n_1 = 30, n_2 = 50$ である. したがって,

$$\bar{X} \sim N\left(\mu_1, \frac{4.8^2}{30}\right), \bar{Y} \sim N\left(\mu_2, \frac{4.8^2}{50}\right)$$

検定するに当たっては, 題意より両校の身長に有意差があるかということより,

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

有意水準 $\alpha = 0.05$

統計量

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim N(0, 1)$$

H_0 のもとで,

$$Z_0 = \frac{148.2 - 146.4}{\sqrt{4.8^2/30 + 4.8^2/50}} = 1.6238$$

$z_{0.05/2} = 1.96$ より,

$$Z_0 = 1.62 < z_{0.05/2} = 1.96$$

したがって, H_0 は棄却されない.

(b) σ_1^2, σ_2^2 が未知だが $\sigma_1^2 = \sigma_2^2$ とみなせる場合.

$X_1, \dots, X_{n_1}, Y_1, \dots, Y_{n_2}$ に対して不偏分散をそれぞれ $S_1'^2, S_2'^2$ とします.

$$S_1'^2 = \frac{1}{n_1 - 1} \sum_i (X_i - \bar{X})^2, S_2'^2 = \frac{1}{n_2 - 1} \sum_i (Y_i - \bar{Y})^2$$

これを合併した不偏分散

$$\hat{\sigma}^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

を考えます。これは、両方の標本分散

$$S_1^2 = \frac{1}{n_1} \sum_i (X_i - \bar{X})^2, S_2^2 = \frac{1}{n_2} \sum_i (Y_i - \bar{Y})^2$$

を用いて

$$\hat{\sigma}^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}$$

としても同じです。この $\hat{\sigma}^2$ を (a) で用いた式に代入すると、

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\hat{\sigma}^2/n_1 + \hat{\sigma}^2/n_2}} \sim t(n_1 + n_2 - 2)$$

これは、自由度 $n_1 + n_2 - 2$ の t 分布に従うことが分かっています。これを利用して、母平均の差の検定を行うことができます。

例題 6.3

A, B 2 つの方法で化学物質を作ろうとした。それぞれ 5 回ずつ実験したらその純度の平均値および分散は次の通りであった。

平均値 $\bar{X}_A = 97.5\%$, $\bar{X}_B = 95.3\%$

分散 $S_A^2 = 1.23\%^2$, $S_B^2 = 1.56\%^2$

製法によって純度に差があるだろうか。有意水準 0.05 で検定せよ。

解 それぞれの母純度を $\mu_A\%$, $\mu_B\%$ とする。これらの分析値は、同じ分散の正規分布にしたがって分布するとする。

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

有意水準 $\alpha = 0.05$

統計量

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\hat{\sigma}^2/n_1 + \hat{\sigma}^2/n_2}} \sim t(n_1 + n_2 - 2), \hat{\sigma}^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}$$

$$H_0 \text{ のもとで, } \hat{\sigma}^2 = \frac{5(1.23) + 5(1.56)}{5+5-2} = 1.7428\%$$

$$T_0 = \frac{97.5 - 95.3}{\sqrt{1.7438/5 + 1.7438/5}} = 2.6342$$

$t_{0.05/2,8} = 2.3060$ より、

$$T_0 = 1.7428 < t_{0.05/2,8} = 2.3060$$

したがって、 H_0 は棄却されない。

(c) σ_1^2, σ_2^2 が未知

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{S_1^2/(n_1 - 1) + S_2^2/(n_2 - 1)}} \sim t(\phi),$$

$$\frac{1}{\phi} = \frac{c^2}{n_1 - 1} + \frac{(1-c)^2}{n_2 - 1}, \quad \frac{1}{c} = 1 + \frac{(n_1 - 1)S_2^2}{(n_2 - 1)S_1^2}$$

2. 母分散の比 σ_1^2/σ_2^2 の検定

$$\frac{n_1 S_1^2}{\sigma_1^2} \sim \chi^2(n_1 - 1), \quad \frac{n_2 S_2^2}{\sigma_2^2} \sim \chi^2(n_2 - 1)$$

より

$$F = \frac{\sigma_2^2 S_1'^2}{\sigma_1^2 S_2'^2} \sim F(n_1 - 1, n_2 - 1)$$

したがって,

1. 対立仮説が $H_1: \sigma_1^2 \neq \sigma_2^2$ のときは, 両側検定で, 棄却域は

$$W = \{F: F > F_{n_2-1}^{n_1-1}(\frac{\alpha}{2})\} \cup \{F: F < F_{n_2-1}^{n_1-1}(1 - \frac{\alpha}{2})\}$$

2. 対立仮説が $H_1: \sigma_1^2 > \sigma_2^2$ のとき棄却域は

$$W = \{F: F > F_{n_2-1}^{n_1-1}(\alpha)\}$$

3. 対立仮説が $H_1: \sigma_1^2 < \sigma_2^2$ のとき棄却域は

$$W = \{F: F < F_{n_2-1}^{n_1-1}(1 - \alpha)\}$$

しかし, $F_{n_2-1}^{n_1-1}(1 - \alpha)$ は数表にありません. そこで, この場合は,

$$F_{n_2-1}^{n_1-1}(1 - \alpha) = \frac{1}{F_{n_1-1}^{n_2-1}(\alpha)}$$

を用いて計算します.

例題 6.4

A, B 2つの機械から製造された製品から, それぞれの大きさ 10, 16 の標本を抽出して重量を調べたところ, 分散がそれぞれ $5.23g^2, 2.48g^2$ であった. 母分散に有意差があるか. 有意水準 5% で検定せよ.

解

$$\begin{aligned} n_A &= 10, S_A^2 = 5.23, S'^2 = \frac{10}{9} S_A^2 \\ n_B &= 16, S_B^2 = 2.24, S'^2 = \frac{16}{15} S_B^2 \end{aligned}$$

$$H_0: \sigma_1^2 = \sigma_2^2, H_1: \sigma_1^2 \neq \sigma_2^2$$

有意水準 $\alpha = 0.05$

統計量

$$F = \frac{\sigma_2^2 S_1'^2}{\sigma_1^2 S_2'^2} \sim F(n_1 - 1, n_2 - 1)$$

H_0 のもとで,

$$F_0 = \frac{\frac{10(5.23)}{9}}{\frac{16(2.24)}{15}} = 2.1967$$

$F_{15}^9(0.025) = 3.1227$ より,

$$F_0 = 2.1967 < F_{15}^9(0.05/2) = 3.1227$$

したがって, H_0 は棄却できない. すなわち両方の母分散に有意差はない.

演習問題 6.2

1 環境学部の A, B で数学の試験をした. A クラスから 10 名, B クラスから 12 名の成績を無作為に選んだら次の表を得た.

A	71	79	92	91	87	79	77	89	71	84		
B	63	84	71	81	80	84	71	84	64	84	69	77

A, B の成績は, それぞれ正規分布 $N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2)$ に従っており, 分散は同じであるとする. このとき, 仮説

$$H_0 : \mu_1 = \mu_2$$

を有意水準 5% で両側検定をしよう.

2 環境学部の A, B で数学の試験をした. A クラスから 10 名, B クラスから 12 名の成績を無作為に選んだら次の表を得た.

A	71	79	92	91	87	79	77	89	71	84		
B	63	84	71	81	80	84	71	84	64	84	69	77

A, B の成績は, それぞれ正規分布 $N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2)$ に従っているとする. このとき, 仮説

$$H_0 : \sigma_1 = \sigma_2$$

を有意水準 5% で左側検定をしよう.

6.3 比率の検定

母比率の検定 (大標本の場合) 母集団の中で, ある属性に対して事象 A の起こる割合 p を事象 A の母比率とします. この母比率に関する仮説を, 標本値から検定することを考えます.

母比率が p の二項母集団から抽出された大きさ n の標本を (X_1, \dots, X_n) とします. ここで,

$$X_i = \begin{cases} 1 & A \text{ のとき} \\ 0 & \bar{A} \text{ のとき} \end{cases}$$

とします. このとき, $X = X_1 + \dots + X_n$ とすると, X は標本中 A であるものの個数を表す統計量で, $\frac{X}{n}$ は事象 A の標本比率とします. そのとき, 母比率 p について, $p_0 (0 \leq p_0 \leq 1)$ を既知の値として, 帰無仮説

$$H_0: \text{「} p = p_0 \text{」}, \text{ 対立仮説 } H_1: \text{「} p \neq p_0 \text{」}$$

を検定することが問題となります.

母比率 p の二項母集団から大きさ n の標本 (X_1, \dots, X_n) をとり, $X = X_1 + \dots + X_n$ とすると X は二項分布 $B(n, p)$ に従います. ここで n が十分大きいときにはラプラスの定理によって, X は近似的に正規分布 $N(np, np(1-p))$ に従い, 標本比率 $\frac{X}{n} = \hat{p}$ は近似的に正規分布 $N\left(p, \frac{p(1-p)}{n}\right)$ に従います. よって, 標準化を行うと

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0, 1)$$

例題 6.5

サイコロを 600 回投げたところ, 1 の目が 108 回出たという. 1 の目が出る母比率 p は $\frac{1}{6}$ が有意水準 5% で検定せよ.

解答

$$H_0: \text{「} 1 \text{ の目の出る確率 } p = \frac{1}{6} \text{」}$$

$$H_1: \text{「} p \neq \frac{1}{6} \text{」}$$

有意水準 $\alpha = 0.05$

統計量

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0, 1)$$

H_0 のもとで, $\hat{p} = \frac{108}{600} = 0.18$ より,

$$\begin{aligned} Z_0 &= \frac{0.18 - \frac{1}{6}}{\sqrt{\frac{\frac{1}{6}(1-\frac{1}{6})}{600}}} \\ &= 0.088 \end{aligned}$$

対立仮説より, 標準正規分布の両側確率を用いる.

$$Z_0 < Z_{\frac{0.05}{2}} = 1.96$$

したがって, H_0 を容認.

母比率の差の検定

2つの母集団 A, B の中で1つの特性 C を持つものの母比率を p_1, p_2 とする. この母集団からそれぞれ大きさ n_1, n_2 個の標本を抽出し, その特性を持つものの個数を X_1, X_2 とする. このとき, 母比率について

帰無仮説 H_0 : 「 $p_1 = p_2$ 」と対立仮説 H_1 : 「 $p_1 \neq p_2$ 」

を検定することを考えます. 帰無仮説のもとで, 母比率の値 p_1, p_2 は未知ですが,

$$p_1 = p_2 = p, 1 - p = q$$

とおくと, n_1, n_2 がある程度大きければ X_1, X_2 の分布はそれぞれ正規分布 $(N(n_1p, n_1pq), N(n_2p, n_2pq))$ によって近似されるので, 統計量

$$\bar{X}_1 = \frac{X_1}{n_1} \sim (N(p, \frac{pq}{n_1}))$$

$$\bar{X}_2 = \frac{X_2}{n_2} \sim (N(p, \frac{pq}{n_2}))$$

$$\frac{X_1}{n_1} - \frac{X_2}{n_2} \sim N(0, (\frac{1}{n_1} + \frac{1}{n_2})pq)$$

これより,

$$Z = \frac{X_1/n_1 - X_2/n_2}{\sqrt{(\frac{1}{n_1} + \frac{1}{n_2})p(1-p)}} \sim N(0, 1)$$

ただし, 共通の母比率 p は未知数ですので, 観測比率

$$p = \frac{X_1 + X_2}{n_1 + n_2}$$

を用います.

例題 6.6

テレビの視聴率調査で, ある番組について男性は400人中の無作為標本中120人が, 女性は500人の無作為標本中180人が好きと答えた. 実際に男女の好みに差があるといえるか, 有意水準5%で検定せよ.

解答 ある番組を男性が好きな比率を p_1 , 女性が好きな比率を p_2 とすると, $p_1 = \frac{120}{400}, p_2 = \frac{180}{500}$

H_0 : 「男女で好みに差がない」 $p_1 = p_2$

H_1 : 「男女で好みに差がある」 $p_1 \neq p_2$

有意水準 $\alpha = 0.05$

統計量

$$Z = \frac{X_1/n_1 - X_2/n_2}{\sqrt{(\frac{1}{n_1} + \frac{1}{n_2})p(1-p)}} \sim N(0, 1)$$

H_0 のもとで, $p = \frac{120+180}{400+500} = \frac{1}{3}$ より,

$$\begin{aligned} Z_0 &= \frac{120/400 - 180/500}{\sqrt{(\frac{1}{400} + \frac{1}{500})\frac{1}{3}(1 - \frac{1}{3})}} \\ &= -1.897 \end{aligned}$$

対立仮説より，標準正規分布の両側確率を用いる．

$$|Z_0| < Z_{\frac{0.05}{2}} = 1.96$$

したがって， H_0 を容認．

演習問題 6.3

1 ある政党の支持率は従来 28%であったが，最近の世論調査で無作為に抽出された 3,000 人有権者のうち支持率は 25%であった．支持率が低下したと判断すべきか，有意水準 5%で検定せよ．

2 あるテレビ番組の視聴率調査を男女別に行った．その結果，男性の無作為標本 200 人のうち 25 人が，女性の無作為標本 300 人のうち 20 人が見ていると答えた．このとき，男女の視聴率に差があるといえるか，有意水準 5%で検定せよ．

6.4 適合度検定

データにある確率分布をあてはめ、あてはまりのよさを検定するのが適合度検定 (goodness of fit test) です。この検定の問題に対して、標本は元のデータに対応します。また、想定した確率分布には、ある確率変数 X が対応しています。

(1) 多項分布に対する適合度の検定

ある試行の結果、 k 個の事象 A_1, A_2, \dots, A_k のいずれかが現われるとします。ここで、 A_i が起こる確率を $P(A_i)$ とすると、

$$\begin{aligned} P(A_i) &= p_i \\ p_1 + p_2 + \dots + p_k &= 1 \end{aligned}$$

となります。この試行を n 回独立に行なうとき、 A_1, A_2, \dots, A_k がそれぞれ n_1, n_2, \dots, n_k 回現われる確率は

$$\binom{n}{n_1, n_2, \dots, n_k} p_1^{n_1} p_2^{n_2} \dots p_k^{n_k} = \frac{n!}{n_1! n_2! \dots n_k!} p_1^{n_1} p_2^{n_2} \dots p_k^{n_k}$$

ここで、 $k = 2$ のときは、2 項分布に他なりません。この 2 項分布の一般化を多項分布 (multinomial distribution) といい、 n 回の独立試行で事象 A_i が起こる回数を確率変数 X_i で表すと、

$$P(X_1 = n_1, X_2 = n_2, \dots, X_k = n_k) = \frac{n!}{n_1! n_2! \dots n_k!} p_1^{n_1} p_2^{n_2} \dots p_k^{n_k}$$

となります。

例題 6.7

それぞれの目が出る確率が等しいサイコロがある。これを 6 回投げたとき、1 から 6 ままで 1 回ずつ現れる確率を求めよ。

解 各数字が現れる確率は $\frac{1}{6}$ で、1 から 6 ままで 1 回ずつ現れる組み合わせは $\binom{6}{1,1,1,1,1,1}$ 通り。したがって、その確率は

$$\binom{6}{1,1,1,1,1,1} \left(\frac{1}{6}\right)^6 = \frac{6!}{6^6} = \frac{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{6 \cdot 6 \cdot 6 \cdot 6 \cdot 6 \cdot 6} = \frac{5}{324}$$

次に、 A_1, A_2, \dots, A_k の互いに排反な事象のいずれかが現われる多項分布を考えます。 $P(A_i) = p_i$ とすると、大きさ n の標本のうち A_i に入る期待値は $np_i = m_i$ となります。一方、大きさ n の標本のうち A_i の部分に入る個数を確率変数 X_i で表すと、次のことが知られています。 $m \geq 5$ のとき、

$$\chi^2 = \frac{(X_1 - m_1)^2}{m_1} + \frac{(X_2 - m_2)^2}{m_2} + \dots + \frac{(X_k - m_k)^2}{m_k}$$

は $\chi^2(k-1)$ に従う。

理論度数 m_i と実測度数 X_i がすべての i について近い値であれば、 χ^2 は全体として小さな値となります。したがって、 χ^2 が大きな値となったとき、その理論値 m_i に疑問が持たれます。このことから、次のような適合度の検定が得られます。

帰無仮説 $H_0: p_1 = p_{10}, p_2 = p_{20}, \dots, p_k = p_{k0}$

(p_{i0} は正数で $p_{10} + p_{20} + \dots + p_{k0} = 1$ となる数)

対立仮説 $H_1: p_1 = p_{11}, p_2 = p_{21}, \dots, p_k = p_{k1}$

ただし, $(p_{11}, p_{21}, \dots, p_{k1}) \neq (p_{10}, p_{20}, \dots, p_{k0})$ である.

ここでは問題の性質上, 片側検定にあたるものは考えられません. H_0 のもとで A_i に入る理論度数 m_i は,

$$m_i = np_{i0}$$

で与えられます. ここでは, n は十分大きな値で, すべての i に対して $m_i \geq 5$ となるとします. A_i に入る標本値が x_i であるとき

$$\chi^2 = \sum_{i=1}^k \frac{(x_i - np_{i0})^2}{np_{i0}} > \chi_{\alpha, k-1}^2 \text{ ならば } H_0 \text{ を棄却する.}$$

これによって適合度が検定できます.

例題 6.8

あるサイコロを 600 回投げたところ, 次のような表が得られた. 各目の現れる確率が等しいと考えられるか, 有意水準 0.05 で検定しよう.

目の数	1	2	3	4	5	6	計
回数	102	89	87	106	115	101	600

解

H_0 : 「各目の現れる確率は等しい」 ($p_1, p_2, p_3, p_4, p_5, p_6 = \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}$)

H_1 : 「各目の現れる確率は等しくない」 ($p_1, p_2, p_3, p_4, p_5, p_6 \neq \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}$)

有意水準 $\alpha = 0.05$

統計量

$$\chi^2 = \sum_{i=1}^6 \frac{(X_i - np_i)^2}{np_i} = \sum_{i=1}^6 \frac{X_i^2}{np_i} - n$$

H_0 のもとで,

$$\begin{aligned} \chi_0^2 &= \frac{102^2}{100} + \frac{89^2}{100} + \frac{87^2}{100} + \frac{106^2}{100} + \frac{115^2}{100} + \frac{101^2}{100} - 600 \\ &= 104.04 + 79.21 + 75.69 + 112.36 + 132.25 + 102.01 - 600 = 5.56 \end{aligned}$$

$\chi_{0.05, 6-1}^2 = 12.83$ より,

$$\chi_0^2 = 5.56 < \chi_{0.05, 5}^2 = 11.07$$

したがって, H_0 を容認.

演習問題 6.4

(1) ある遺伝形質は $A : B : C : D = 9 : 3 : 3 : 1$ のメンデル比に従って現れるとされているが, 実験の結果次の表を得た. メンデル比に従っているといえるか, 有意水準 5% で検定しよう.

遺伝形質	A	B	C	D	計
観測度数	243	72	78	15	408

(2) 確率分布に対する適合度の検定

ここでは、ある分布が正規分布に従う、あるいはポワソン分布に従う、ということ自体が帰無仮説となる適合度検定を考えます。つまり、

帰無仮説 H_0 : 「ある分布 D に従う」

を設定します。 D の分布は既知であって、母数 $\theta_1, \theta_2, \dots, \theta_l$ を含んでいるとします。例えば、正規分布では μ, σ^2 の 2 個の母数を含み、これらの値は不明であるとしてします。

次に排反な各階級 A_1, A_2, \dots, A_k に入る個数 (X_1, X_2, \dots, X_k) の実現値を (x_1, x_2, \dots, x_k) とし、母数 θ_i をこの値を用いて推定します。つまり、

$$\theta_i = \hat{\theta}_i(x_1, x_2, \dots, x_k) \quad (i = 1, 2, \dots, l)$$

この θ_i を用いて各階級 A_1, A_2, \dots, A_k に入るべき期待度数 m_1, m_2, \dots, m_k を求めます。ここで、 $m_i = np_{i0}$ 。つまり、

$$\chi^2 = \sum_{i=1}^k \frac{(X_i - m_i)^2}{m_i}$$

を求めます。このとき、 $\chi^2 \sim \chi_{k-l-1}^2$ であることが分かっています。そして、これを用いて H_0 の検定を行います。

例題 6.9

ある軍隊の 10 個の部隊において、1 年間に馬に蹴られて死亡した兵士の数とその部隊数を 10 年間調べた結果次のような表になった。

死亡者数	0	1	2	3	4	計
部隊数	109	65	22	3	1	200

この表はポワソン分布に従うか、有意水準 5% で検定しよう。

H_0 : 「ポワソン分布 $P(\lambda)$ に従っている」

有意水準 $\alpha = 0.05$

統計量

この表をポワソン分布とみて、死亡数の理論値を求める。これがポワソン分布 $P(\lambda)$ によるものと考えて、 λ の値を推定する。死亡者数 k のときの確率を p_k とすると、

$$\sum_{k=0}^{\infty} kp_k = E(X) = \lambda$$

死亡者数	k	0	1	2	3	4	計
部隊数	f_k	109	65	22	3	1	200
	kf_k	0	65	44	9	4	122
	p_k	0.5435	0.3313	0.1011	0.0206	0.0031	
理論度数	m_k	108.7	66.3	20.2	4.1	0.6	

ここで, $np_k \approx f_k$ より $\sum_k k f_k \approx \lambda n$. これより平均値 λ は

$$\lambda \approx \frac{1}{n} \sum_k k f_k = \frac{122}{200} = 0.61$$

死亡者数	k	0	1	2	3	4	計
部隊数	x_k	109	65	22	3	1	200
理論度数	m_k	108.7	66.3	20.2	4.1	0.6	

この表で, $k \geq 3$ の所の m_k は単独で 5 よりも小さいので, χ^2 検定ができない. そこで, 右から順に m_i を加えて 5 を越すまで合併すると, $k \geq 2$ の階級を 1 つにしなければならない. したがって,

$$\chi^2 = \sum_{i=0}^2 \frac{(x_i - m_i)^2}{m_i}$$

H_0 のもとで,

$$\begin{aligned} \chi_0^2 &= \frac{(109 - 108.7)^2}{108.7} + \frac{(65 - 66.3)^2}{66.3} + \frac{(26 - 25)^2}{25} \\ &= 0.066 \end{aligned}$$

$$\chi_{0.05, 3-1-1}^2 = 3.84 \text{ より,}$$

$$\chi_0^2 = 0.066 < \chi_{0.05, 1}^2 = 3.84$$

したがって, H_0 を容認.

母数 λ が標本から 1 個推定されたので, 自由度は $3 - 1 - 1 = 1$ となる.

(3) 独立性の検定

母集団の要素は, すべて A, B の 2 種類の属性をもち, A, B はそれぞれ排反な A_1, \dots, A_k および B_1, \dots, B_l に分かれています. 母集団から大きさ n の標本を抽出して, $A_i \cap B_j$ に入る観測度数を x_{ij} とすると, 次の表のように行列の形に整理できる.

	B_1	B_2	\cdots	B_l	和
A_1	x_{11}	x_{12}	\cdots	x_{1l}	$x_{1\cdot}$
A_2	x_{21}	x_{22}	\cdots	x_{2l}	$x_{2\cdot}$
A_3	\vdots	\vdots		\vdots	\vdots
A_k	x_{k1}	x_{k2}	\cdots	x_{kl}	$x_{k\cdot}$

ここで, x_i, x_j は周辺度数である. このような表を $k \times l$ 分割表 (contingency table) という.

これを用いて, 母集団の属性 A と B が無関係であるかを調べることを独立性の検定という. 独立性の検定には適合度の検定を応用することができる.

対 A_i, B_j の出現度数の確率変数を X_{ij} , A_i, B_j の実現する確率を p_i, q_j . また, A_i, B_j が同時に起こる確率を P_{ij} とする.

ここで、次のような適合度の検定を考える。

帰無仮説 : 「属性 A, B は独立である」

対立仮説 : 「属性 A, B は従属である」

帰無仮説 H_0 のもとで

$$P_{ij} = P_r(A_i \cap B_j) = P_r(A_i)P_r(B_j) = p_i q_j$$

が成り立つ。ここで、 p_i, q_j は母数なのでこれを最尤法によって推定すると、それらの推定値は

$$\hat{p}_i = \frac{x_{i.}}{n}, \quad \hat{q}_j = \frac{x_{.j}}{n}$$

で与えられる。このとき、 n が十分大きければ、帰無仮説 H_0 のもとで統計量

$$\chi^2 = \sum_{i=1}^k \sum_{j=1}^l \frac{(X_{ij} - nP_{ij})^2}{nP_{ij}}$$

が自由度 $(k-1)(l-1)$ のカイ 2 乗分布に従うことが知られている。観測度数 x_{ij} を用いると、統計量 χ^2 の実現値は

$$\begin{aligned} \chi_0^2 &= \sum_{i=1}^k \sum_{j=1}^l \frac{(x_{ij} - n\hat{p}_i\hat{q}_j)^2}{n\hat{p}_i\hat{q}_j} \\ &= \sum_{i=1}^k \sum_{j=1}^l \left\{ \frac{x_{ij}^2}{n\hat{p}_i\hat{q}_j} - 2x_{ij} + n\hat{p}_i\hat{q}_j \right\} = n \left\{ \sum_{i=1}^k \sum_{j=1}^l \frac{x_{ij}^2}{x_{i.}x_{.j}} - 1 \right\} \end{aligned}$$

となる。

第7章 演習問題解答

問題解答 1.1

1 スタージスの式から

階級数 $= 1 + \frac{\log 100}{\log 2} = 1 + 6.64 = 7.64$ また最大値 440 最小値 300 より, 階級幅は

階級幅 $= \frac{440-300}{7.64} = 18.4$ となるので, 階級幅を 18 とすることにします. これより度数分布表を作成します.

表 7.1: 度数分布表

階級	階級値	度数	相対度数	累積度数	累積相対度数
300 ~ 318	309	2	0.02	2	0.02
318 ~ 336	327	10	0.1	12	0.12
336 ~ 354	345	25	0.25	37	0.37
359 ~ 372	363	31	0.31	68	0.68
372 ~ 390	381	8	0.08	76	0.76
390 ~ 408	399	18	0.18	94	0.94
408 ~ 426	417	5	0.05	99	0.99
426 ~ 444	435	1	0.01	100	1.00

$$\text{平均値 } \bar{x} = \frac{1}{100} [318 \cdot 2 + 336 \cdot 10 + 354 \cdot 25 + 363 \cdot 31 + 381 \cdot 8 + 399 \cdot 18 + 417 \cdot 5 + 435 \cdot 1] = 365.16$$

最大値 440

最小値 300

$$\text{中央値 } \frac{360 + 360}{2} = 360$$

最頻値 363

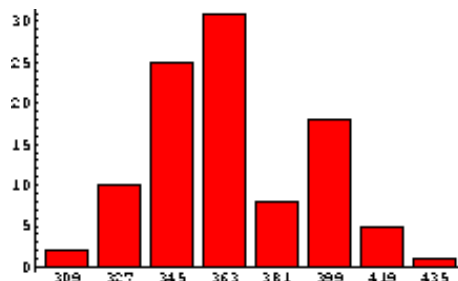


図 7.1: ヒストグラム

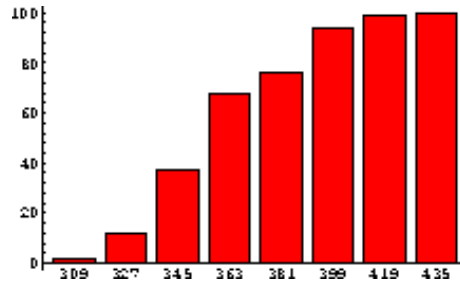


図 7.2: 累積度数分布表

問題解答 1.2

1

電卓を使う場合は必ず途中の値を書く必要があります。また、計算は小数点以下2桁までで表わすことにします。

$$T_x = 841 \quad T_y = 806$$

$$\bar{x} = 35.04 \quad \bar{y} = 33.58$$

$$T_{xx} = \sum_{i=1}^{24} x_i^2 = 45553 \quad T_{yy} = \sum_{i=1}^{24} y_i^2 = 36990$$

$$s_x = \sqrt{\frac{1}{24}T_{xx} - (\bar{x})^2} = 25.88 \quad s_y = \sqrt{\frac{1}{24}T_{yy} - (\bar{y})^2} = 20.34$$

$$T_{xy} = \sum_{i=1}^{24} x_i y_i = 37192$$

これより

$$\begin{aligned} s_{xy} &= \frac{1}{n}T_{xy} - \frac{T_x}{n} \frac{T_y}{n} \\ &= \frac{1}{24} \cdot 37192 - \frac{841}{24} \cdot \frac{806}{24} = 372.85 \end{aligned}$$

$$r = \frac{s_{xy}}{s_x s_y} = \frac{372.85}{25.88 \cdot 20.34} = 0.71$$

これより正の相関でかなり強い相関があるといえる。

問題解答 1.3

1

$$\text{階級数} = 1 + \frac{\log n}{\log 2} = 1 + \frac{\log 24}{\log 2} = 1 + 4.58 = 5.58$$

また x の最大値 109 最小値 13 より

$$\text{階級幅} = \frac{109-13}{5.58} = 17.20$$

これより x の階級幅を 17 と取ります。また y の最大値 65 最小値 5 より

$$\text{階級幅} = \frac{65-5}{5.58} = 10.75$$

これより y の階級幅を 10 と取ります。

表 7.2: 相関表

	x	10 ~ 27	27 ~ 44	44 ~ 61	61 ~ 78	78 ~ 95	95 ~ 112	
y	階級値	18.5	35.5	52.5	69.5	86.5	103.5	計
0 ~ 10	5	3						3
10 ~ 20	15	6						6
20 ~ 30	25	1	2					3
30 ~ 40	35			1				1
40 ~ 50	45	1	1					2
50 ~ 60	55	2	3			2		7
60 ~ 70	65					1	1	2
		13	7	1		3	1	24

$$T_x = 841 \quad T_y = 806$$

$$\bar{x} = 35.04 \quad \bar{y} = 33.58$$

$$T_{xx} = \sum_{i=1}^{24} x_i^2 = 45553 \quad T_{yy} = \sum_{i=1}^{24} y_i^2 = 36990$$

$$s_x = \sqrt{\frac{1}{24}T_{xx} - (\bar{x})^2} = 25.88 \quad s_y = \sqrt{\frac{1}{24}T_{yy} - (\bar{y})^2} = 20.34$$

$$T_{xy} = \sum_{i=1}^{24} x_i y_i = 37192$$

$$\begin{aligned} s_{xy} &= \frac{1}{n}T_{xy} - \frac{T_x}{n} \frac{T_y}{n} \\ &= \frac{1}{24} \cdot 37192 - \frac{841}{24} \cdot \frac{806}{24} = 372.85 \end{aligned}$$

$$r = \frac{s_{xy}}{s_x s_y} = \frac{372.85}{25.88 \cdot 20.34} = 0.71$$

これより, x 上の y の回帰直線は

$$y - 33.58 = \frac{372.85}{670.24}(x - 35.04) = 0.56(x - 35.04)$$

したがって

$$y = 0.56x + 13.96$$

問題解答 2.1

1.
 - (a) ここでは□□□□の□の中に異なる数字を入れて4桁の数字を何個作れるか考えてみます。まず, 4桁の数字なので, 千の位の□には0を使うことができません。そこで千の位は1から6までの6個の中から1個選ぶことになるので, 6通り。百の位から一の位までは0から6までのどの数字も使うことができます。しかし, 数字を取り出して並べるため, 同じ数字は2度使えないことに注意して下さい。では並べてみましょう。

まず、百の位には千の位で用いられた数字以外どれでも使えるので、6通りあります。十の位は千の位と百の位で用いられた数字以外どれでも使えるので、5通りあります。最後に一の位は千の位、百の位、十の位で用いられたもの以外すべて使えるので4通り。よって全部で

$$6 \cdot 6 \cdot 5 \cdot 4 = 720$$

個の4桁の数字を作ることができます。

ここで百の位、十の位、一の位は千の位で用いられた数字以外どれでも1回ずつ使うことができることに注意すると、6個の中から3個を取り出し順序をつけて並べる順列の数になるので ${}_6P_3$ と表わせます。よって

$$6 \cdot {}_6P_3 = 6 \cdot 6 \cdot 5 \cdot 4 = 720$$

(b) まず、5の倍数は一の位が必ず、0か5になることに注意します。次に、一の位が0の場合と5の場合を別々に考えます。

一の位が0の場合。

千の位、百の位、十の位は1から6までの数字を1回ずつ使うことができるので、6個の中から3個を取り出し並べる順列の数になるので ${}_6P_3$ と表わせます。よって120通り

一の位が5の場合。

千の位に0が使えないので、千の位は0、5以外の5通り、百と十の位は一と千の位で用いた数字以外どれでも1回ずつ使えるので ${}_5P_2$ 通り。よって $5 \cdot {}_5P_2 = 5 \cdot 5 \cdot 4 = 100$ 通り

ここで、一の位が0の場合と5の場合は同時におきないので、全部で

$$120 + 100 = 220 \text{ 通り}$$

(c) 同じ数字が重複してもよいので、千の位は1から6までの6通り、百の位は0から6までの7通り、十の位も0から6までの7通り、最後に一の位も0から6までの7通り。したがって、

$$6 \cdot {}_7\Pi_3 = 6 \cdot 7 \cdot 7 \cdot 7 = 2058 \text{ 通り}$$

2.

(a) 1から10までの番号のついたカードから6枚を取り出すとき、何通りの取り出し方があるかを考えます。このとき、一枚ずつ順に取り出し並べるわけではないので、取り出す順序を考える必要はありません。よって、何通りの組み合わせがあるかを考えればよいでしょう。

10個の中から6個を取り出す組み合わせは ${}_{10}C_6$ なので、

$${}_{10}C_6 = \frac{10!}{4!6!} = \frac{10 \cdot 9 \cdot 8 \cdot 7}{4 \cdot 3 \cdot 2 \cdot 1} = 210 \text{ 通り}$$

(b) 1と2のカードを含むとは取り出した6枚の中に必ず1と2のカードが入っていないなければならないということです。

これは1と2を先に引いておいて残りの4枚を3から10までの8枚から選ぶ選び方と考えることができます。よって ${}_8C_4 = 70$ 通り。

(c) 1のカードを含む場合を C_1 、2のカードを含む場合を C_2 とします。すると C_1 は ${}_9C_5$ 通り、 C_2 も ${}_9C_5$ 通りとなるので、1または2を含む場合は

$${}_9C_5 + {}_9C_5 - 70 = 252 - 70 = 182 \text{ 通り}$$

別解 1 または 2 のカードを含むの否定は 1 と 2 のカードを含まないとなるので, 1 と 2 のカードを含まない場合を考えます.

1 と 2 を含まないので残りの 8 枚のカードから 6 枚を取り出すことになるので, その取り出し方は ${}_8C_6$ 通り. よって, 1 または 2 のカードを含む場合は

$$210 - {}_8C_6 = 210 - \frac{56}{2} = 210 - 28 = 182 \text{ 通り}$$

3.
(a) $A_i =$ [表の出る回数が i 回] とおきます. すると表は 5 回中 i 回どこで出てよく出る順序は関係ないので, その組み合わせは ${}_5C_i$ となります.

A_0 は ${}_5C_0 = 1$ 通り, A_1 は ${}_5C_1 = 5$ 通り, A_2 は ${}_5C_2 = 10$ 通り, A_3 は ${}_5C_3 = 10$ 通り, A_4 は ${}_5C_4 = 5$ 通り, A_5 は ${}_5C_5 = 1$ 通り

(b) 起こりえるすべての場合は ${}_2\Pi_5 = 2^5 = 32$ 通り. または,

$$1 + 5 + 10 + 10 + 5 + 1 = 32 \text{ 通り}$$

4.
(a) a, b が隣り合うので, a, b の順に並んでいるのを A とおくと, この問題は 5 文字を一行に並べる並べ方は何通りあるかという問題と同じになります. したがって,

$${}_5P_5 = 5! = 120$$

また, b, a の順に並んでいるのを B とおくと,

$${}_5P_5 = 5! = 120$$

したがって, 全部で 240 通り.

(b) a, b が隣り合わないのはすべての並び方から a, b が隣り合う場合を引いたもの. よって

$$6! - 240 = 720 - 240 = 480 \text{ 通り}$$

(c) a, b が両端にくるのは a, b と b, a の場合があることに注意すると, 中の 4 文字の並び方は ${}_4P_4$ 通りとなるので, 全部で

$$2 \cdot {}_4P_4 = 2 \cdot 4! = 48 \text{ 通り}$$

5. $\binom{8}{4,2,1,1} = \frac{8!}{4!2!1!1!}$. 全てに番号をつけて並べると, ${}_8P_8 = 8!$ 通りの並べかたがある. しかし, 実際は 4 個の a の重複 ${}_4P_4 = 4!$, 2 個の b の重複 ${}_2P_2 = 2!$ は区別できないので, その組み合わせは $\frac{8!}{4!2!1!1!}$ である.

問題解答 2.2

1.
(a) まず, 場合の数を使って確率を求めてみましょう. $X_i =$ [6 回中 i 回 1 の目が出る] とおきます. さいころを 6 回投げると, 目のでかたは全部で ${}_6\Pi_6 = 6^6$ 通りあります. 次に 1 の目が 6 回中 1 回出る場合の数は何通りあるか数えてみましょう.

一投目で 1 の目が出ると, 残りの 5 回は 1 以外なので, ${}_5\Pi_5$ 通り. 同じことが二投目, 三投目, ... でも言えるので, 全部で

$${}_6C_1 \cdot {}_5\Pi_5 = 6 \cdot 5^5$$

したがって、1の目が6回中1回出る確率は

$$P_r(X_1) = \frac{6 \cdot 5^5}{6^6} = \left(\frac{5}{6}\right)^5$$

別解

まず、1の目が1回出るのは最初の1投目でも2投目でも6投目でもいいので全部で6通りあることに注意します。この6通りのそれぞれの確率を調べてみましょう。まず、1投目で1の目が出たとすると

$$\{1 \square \square \square \square \square\}$$

となります。このとき、 \square の中には1の目以外なんでも入ります。よってこのときの確率を求めると、

$$\frac{1}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} = \frac{1}{6} \cdot \left(\frac{5}{6}\right)^5$$

となります。これが全部で6通りあるので、1の目が1回出る確率は

$$6 \cdot \frac{1}{6} \cdot \left(\frac{5}{6}\right)^5 = \left(\frac{5}{6}\right)^5 = \frac{3125}{7776}$$

となります。

(b) 1の目が4回出る組み合わせは全部で ${}_6C_4$ 通り。そのとき、残りの2回は1以外なので 5^2 通り。よって全部で、

$${}_6C_4 \cdot 5^2 \text{ 通り}$$

したがって、1の目が6回中4回出る確率は

$$P_r(X_4) = \frac{{}_6C_4 \cdot 5^2}{6^6} = \frac{375}{46656}$$

別解

6回中4回1の目が出る組み合わせは ${}_6C_4$ 通り。また、それぞれの確率は

$$\left(\frac{1}{6}\right)^4 \cdot \left(\frac{5}{6}\right)^2$$

よって、

$$P_r(X_4) = {}_6C_4 \left(\frac{1}{6}\right)^4 \cdot \left(\frac{5}{6}\right)^2 = \frac{375}{46656} = \frac{125}{15552}$$

(c) 1の目が出るのは4回以下とは

$$X_0 \cup X_1 \cup X_2 \cup X_3 \cup X_4$$

のことです。ここで、それぞれの事象は同時に起こりえないことに注意すると、

$$\begin{aligned} P_r(X_0 \cup X_1 \cup X_2 \cup X_3 \cup X_4) &= P_r(X_0) + P_r(X_1) + P_r(X_2) + P_r(X_3) + P_r(X_4) \\ &= 1 - [P_r(X_5) + P_r(X_6)] \\ &= 1 - \left[\binom{6}{5} \cdot \left(\frac{1}{6}\right)^5 \left(\frac{5}{6}\right) + \binom{6}{6} \cdot \left(\frac{1}{6}\right)^6 \right] = \frac{46625}{46656} \end{aligned}$$

(d) 1の目が出るのは5回以上とは

$$X_5 \cup X_6$$

のことで、ここで、1の目が出るのが4回以下の確率は

$$P_r(X_0 \cup X_1 \cup X_2 \cup X_3 \cup X_4) = \frac{46625}{46656}$$

に注意すると、

$$P_r(X_5 \cup X_6) = 1 - P_r(X_0 \cup X_1 \cup X_2 \cup X_3 \cup X_4) = \frac{31}{46656}$$

2.

(a) 白玉5個、赤玉3個、黒玉が2個合わせて10個の中から4個を取り出す組み合わせは ${}_{10}C_4$ 通り。次に白玉4個を袋の中から取り出す組み合わせを考えてみましょう。

袋の中の5個の白玉から4個を取り出すしかないので、その組み合わせは ${}_5C_4$ 通り。したがって、取り出した4個が全て白玉の確率は

$$\frac{{}_5C_4}{{}_{10}C_4} = \frac{\frac{5!}{4!1!}}{\frac{10!}{6!4!}} = \frac{5}{10 \cdot 9 \cdot 8 \cdot 7 / 4 \cdot 3 \cdot 2} = \frac{1}{42}$$

別解 取り出した4個に注目。

1. 白玉4個取り出す組み合わせは ${}_5C_4 = 5$ 通り

2. 白玉1個ずつ取り出す確率は、 $\frac{4}{10} \cdot \frac{3}{9} \cdot \frac{2}{8} \cdot \frac{1}{7}$

したがって、取り出した4個が全て白玉の確率は

$$5 \cdot \frac{4}{10} \cdot \frac{3}{9} \cdot \frac{2}{8} \cdot \frac{1}{7} = \frac{1}{42}$$

(b) 白玉5個、赤玉3個、黒玉が2個合わせて10個の中から4個を取り出す組み合わせは ${}_{10}C_4$ 通り。次に白玉2個を袋の中から取り出す組み合わせを考えてみましょう。

2個だけ白玉ということは5個の白玉から2個取り出し、残りの赤玉と黒玉から2個取り出す場合の数なので、 ${}_5C_2 \times {}_5C_2$ 。したがって、取り出した4個のうち白がちょうど2個である確率は

$$\frac{{}_5C_2 \times {}_5C_2}{{}_{10}C_4} = \frac{10}{21}$$

(c) 白が2個以内とは白が0個、1個、2個の事象の場合である。よって、その確率は

$$\frac{{}_5C_0 \cdot {}_5C_4 + {}_5C_1 \cdot {}_5C_3 + {}_5C_2 \times {}_5C_2}{{}_{10}C_4} = \frac{31}{42}$$

(d) 白が2個赤が2個を取り出す組み合わせは、 ${}_5C_2 \times {}_3C_2$ である。したがって、取り出した4個のうち白が2個赤が2個である確率は

$$\frac{{}_5C_2 \cdot {}_3C_2}{{}_{10}C_4} = \frac{1}{7}$$

(e) 取り出した4個に注目。

1. 白、赤、黒がともに含まれるということは、白、赤、黒のどれかが2個になる組み合わせを考えればよい。よって $\binom{4}{2,1,1}$ 通り。

2. 白、赤、黒1つずつ取り出す確率は $\frac{5}{10} \cdot \frac{3}{9} \cdot \frac{2}{8}$ 。

3. したがって、白、赤、黒がともに含まれる確率は

$$\binom{4}{2,1,1} \frac{5}{10} \frac{3}{9} \frac{2}{8} = \frac{1}{2}$$

3.

(a) 1 から 10 までがその順に並ぶ場合を考えているので、まずは、1 から 10 を勝手に並べた並べ方は何通りあるか考えてみましょう。

先頭にくるのは 1 から 10 の内どれでもよいので、10 通り、次は 9 通り、... となるので、全部で ${}_{10}P_{10} = 10!$ 通りとなります。

次に 1 から 10 までがその順に並ぶ場合は一通り。したがって、その確率は

$$\frac{1}{10!}$$

別解 1 が先頭にくる確率は $\frac{1}{10}$ 、1 が先頭にきたことが分ったあと、2 が 2 番目にくる確率は $\frac{1}{9}$ 、1 が先頭、2 が 2 番目にきたことが分ったあと、3 が 3 番目にくる確率は $\frac{1}{8}$ 、... となるので、1 から 10 までがその順に並ぶ確率は

$$\frac{1}{10} \cdot \frac{1}{9} \cdots \frac{1}{1} = \frac{1}{10!}$$

(b) 4 のカードがちょうど 4 番目ということは、それ以外の 9 枚のカードはどこにあってもいいので、4 のカードがちょうど 4 番目にくるのは ${}_9P_9 = 9!$ 通りあります。よってその確率は

$$\frac{{}_9P_9}{{}_{10}P_{10}} = \frac{9!}{10!} = \frac{1}{10}$$

別解 4 のカードがちょうど 4 番目にくる確率は $\frac{1}{10}$

(c) 1 が最初に、4 が 4 番目にあるということは、それ以外の 8 枚のカードはどこにあってもいいので、全部で ${}_8P_8$ 通り。よって、1 が最初に、4 が 4 番目にくる確率は

$$\frac{{}_8P_8}{{}_{10}P_{10}} = \frac{8!}{10!} = \frac{1}{10 \cdot 9} = \frac{1}{90}$$

別解 1 が最初にくる確率は $\frac{1}{10}$ 。次に 1 が最初にきたことが分ったあと、4 が 4 番目にくる確率は $\frac{1}{9}$ 。よって、1 が最初に、4 が 4 番目にくる確率は

$$\frac{1}{90}$$

4.

(a) 円板の半径は 1.5cm よりちょうど正方形の中に入るには、円板の中心が 1 辺 5cm の正方形の中にあればよい。したがって、その確率は

$$\frac{1 \text{ 辺 } 5\text{cm の正方形の面積}}{1 \text{ 辺 } 8\text{cm の正方形の面積}} = \frac{25}{64}$$

(b) $A =$ 「円板が正方形の辺にかかる」の余事象は $\bar{A} =$ 「円板が正方形の中にある」となる。したがって、その確率は

$$1 - \frac{1 \text{ 辺 } 5\text{cm の正方形の面積}}{1 \text{ 辺 } 8\text{cm の正方形の面積}} = \frac{39}{64}$$

(c) 円板が4つの正方形にまたがるには、その中心が4つの正方形の境界から1.5cm以内になければならない。また、その面積は $\pi(1.5)^2$ 。したがって、円板が4つの正方形にまたがる確率は

$$\frac{\pi(1.5)^2}{64}$$

5.

(a) 白玉4個、赤玉6個、合わせて10個の中から2個を取り出す組み合わせは ${}_{10}C_2$ 通り。白玉2個を袋の中から取り出す組み合わせを考えてみましょう。

袋の中の4個の白玉から2個を取り出すしかないので、その組み合わせは ${}_4C_2$ 通り。したがって、取り出した2個が両方白玉の確率は

$$\frac{{}_4C_2}{{}_{10}C_2} = \frac{\frac{4!}{2!2!}}{\frac{10!}{8!2!}} = \frac{4 \cdot 3}{10 \cdot 9} = \frac{2}{15}$$

(b) 1個だけ白玉ということは4個の白玉から1個取り出し、6個の赤玉から1個取り出す場合の数なので、 ${}_4C_1 \times {}_6C_1$ 。よってその確率は

$$\frac{{}_4C_1 \times {}_6C_1}{{}_{10}C_2} = \frac{4 \times 6}{45} = \frac{8}{15}$$

(c) 少なくとも1個は白玉という事象は、2個とも白玉であるか、または1個だけ白玉であるかのどちらかです。ここで、これらの事象は排反事象(同時に起きない)であることに注意すると、全部で、 ${}_4C_2 + {}_4C_1 \times {}_6C_1$ 通り。したがって、求める確率は

$$\frac{{}_4C_2 + {}_4C_1 \times {}_6C_1}{{}_{10}C_2} = \frac{2}{15} + \frac{8}{15} = \frac{2}{3}$$

問題解答 2.3

1.

(a) 4回投げて少なくとも1回6の目がでるという事象 A の余事象 \bar{A} は、4回投げて一度も6の目がでないとなります。ここで、それぞれの回に6の目がでない確率は $\frac{5}{6}$ に注意すると、

$$P_r(\bar{A}) = \left(\frac{5}{6}\right)^4$$

となるので、これより、

$$P_r(A) = 1 - P_r(\bar{A}) = 1 - \left(\frac{5}{6}\right)^4 = \frac{671}{1296}$$

(b) 2個のさいころを同時に24回投げて少なくとも1回2個とも6の目がでるという事象 B を考えます。まず、2個のさいころを同時に投げたとき、2個とも6の目がでる確率は $\frac{1}{36}$ 。

ここで、 B の余事象 \bar{B} は、2個のさいころを同時に24回投げて一度も2個両方は6の目ではないとなります。したがって、

$$P_r(\bar{B}) = \left(1 - \frac{1}{36}\right)^{24} = \left(\frac{35}{36}\right)^{24}$$

これより

$$P_r(B) = 1 - P_r(\bar{B}) = 1 - \left(\frac{35}{36}\right)^{24} = 0.491$$

2 $A =$ 「真のガン患者」, $B =$ 「精密検査で陽性反応がでた患者」とおくと質問は, 患者がその精密検査の結果陽性反応を示した場合にガン患者である確率を求めることです. これは条件付き確率を用いると次のように表わせます.

$$P_r(A|B)$$

また, $P_r(A) = 0.05$, $P_r(B|A) = 0.85$, $P_r(B|\bar{A}) = 0.05$ が分っていることに注意します. ここで, Bayes の定理を用いると

$$B = B \cap \Omega = B \cap (A \cup \bar{A}) = (B \cap A) \cup (B \cap \bar{A})$$

より

$$\begin{aligned} (1) P_r(B) &= P_r(B \cap A) + P_r(B \cap \bar{A}) = P_r(A)P_r(B|A) + P_r(\bar{A})P_r(B|\bar{A}) \\ &= 0.05 \cdot 0.85 + 0.95 \cdot 0.05 = 0.9 \\ (2) P_r(A|B) &= \frac{P_r(A \cap B)}{P_r(B)} = \frac{P_r(A)P_r(B|A)}{P_r(B)} = \frac{P_r(A)P_r(B|A)}{P_r(B)} \\ &= \frac{0.05 \cdot 0.85}{0.05 \cdot 0.85 + 0.95 \cdot 0.05} = \frac{0.425}{0.9} = 0.47 \end{aligned}$$

別解

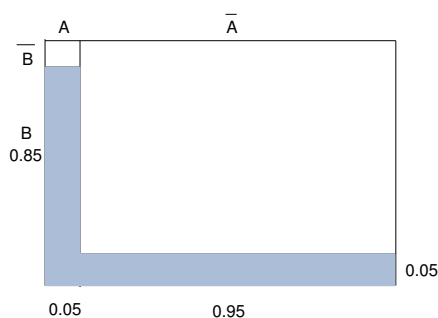


図 7.3: Bayes の定理

3.

(a)

$$\begin{aligned} x \in \overline{A \cup B} &\Leftrightarrow x \notin A \cup B \\ &\Leftrightarrow x \notin A \text{ and } x \notin B \\ &\Leftrightarrow x \in \bar{A} \text{ and } x \in \bar{B} \\ &\Leftrightarrow x \in \bar{A} \cap x \in \bar{B} \end{aligned}$$

これより,

$$\overline{A \cup B} = \bar{A} \cap \bar{B}$$

したがって,

$$P(\overline{A \cup B}) = P(\bar{A} \cap \bar{B})$$

(b)

$$\begin{aligned}
 x \in \overline{A \cap B} &\Leftrightarrow x \notin A \cap B \\
 &\Leftrightarrow x \notin A \text{ or } x \notin B \\
 &\Leftrightarrow x \in \overline{A} \text{ or } x \in \overline{B} \\
 &\Leftrightarrow x \in \overline{A} \cup x \in \overline{B}
 \end{aligned}$$

これより,

$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$

したがって,

$$P(\overline{A \cap B}) = P(\overline{A} \cup \overline{B})$$

(c) B と C が互いに排反ならば, $B | A$ と $C | A$ も互いに排反となるので,

$$P((B \cup C) | A) = P(B | A \cup C | A) = P(B | A) + P(C | A)$$

問題解答 3.11. X を 4 児をもつ家庭の男児の数とすると,

$$P(x=0) = \binom{4}{0} \left(\frac{1}{2}\right)^4 = \frac{1}{16}$$

$$P(x=1) = \binom{4}{1} \left(\frac{1}{2}\right)^4 = \frac{4}{16}$$

$$P(x=2) = \binom{4}{2} \left(\frac{1}{2}\right)^4 = \frac{6}{16}$$

$$P(x=3) = \binom{4}{3} \left(\frac{1}{2}\right)^4 = \frac{4}{16}$$

$$P(x=4) = \binom{4}{4} \left(\frac{1}{2}\right)^4 = \frac{1}{16}$$

よってその確率分布 f は

$$f(i) = P(X=i) = \binom{4}{i} \left(\frac{1}{2}\right)^4$$

で与えられる.

2. 10 個から 3 個取り出す組み合わせは ${}_{10}C_3 = 120$ 通り. また, 3 個中赤がゼロということは, 白が 3 個と同じことなので, 6 個の白から 3 個取り出すこととなり, その組み合わせは ${}_6C_3 = 20$ 通り. よって, X を赤玉の個数とおくと,

$$P_r(x=0) = \frac{{}_6C_3}{{}_{10}C_3} = \frac{6 \cdot 5 \cdot 4}{10 \cdot 9 \cdot 8} = \frac{1}{6}$$

3 個中 1 個赤ということは残りの 2 個は白なので

$$P_r(X=1) = \frac{{}_4C_1 \cdot {}_6C_2}{{}_{10}C_3} = \frac{4 \cdot 6 \cdot 5}{2 \cdot 1} \cdot \frac{3 \cdot 2 \cdot 1}{10 \cdot 9 \cdot 8} = \frac{1}{2}$$

同様にして,

X	0	1	2	3
$P(X = i)$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{6}{20}$	$\frac{1}{30}$

$$P_r(1 \leq X \leq 3) = 1 - P_r(X = 0) = 1 - \frac{1}{6} = \frac{5}{6}$$

3.

(a) $f(x)$ が確率密度関数であるためには,

1. $f(x) \geq 0$ ($-\infty < x < \infty$)

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

を満たしていることを示せばよい. そこで,

1. 定数 k を 0 以上とすれば $f(x) \geq 0$ は満たされる.

2. $\int_{-\infty}^{\infty} f(x)dx = \int a^b k dx = kx \Big|_a^b = k(b-a)$ より $k = \frac{1}{b-a}$ と定めればよい.

(b) $P(X \leq c) = F(c) = \int_{-\infty}^c f(x)dx$ より

$$P(X \leq c) = \int_a^c f(x)dx = \int_a^c \frac{1}{b-a} dx = \frac{c-a}{b-a}$$

となる.

(4)

(a) 分布関数 $F(x)$ は

$$F(x) = P_r(X \leq x) = \int_{-\infty}^x f(t)dt$$

で与えられることに注意します.

$x \leq 0$ のとき

$$F(x) = \int_{-\infty}^x f(t)dt = \int_{-\infty}^x 0dt = 0$$

$0 < x < 1$ のとき

$$\begin{aligned} F(x) &= \int_{-\infty}^x f(t)dt = \int_{-\infty}^0 0dt + \int_0^x 6t(1-t)dt \\ &= 0 + [3t^2 - 2t^3]_0^x = 3x^2 - 2x^3 \end{aligned}$$

$x > 1$ のとき

$$F(x) = \int_{-\infty}^x f(t)dt = \int_{-\infty}^0 0dt + \int_0^1 6t(1-t)dt + \int_1^x 0dt = 1$$

(b)

$$\begin{aligned} P_r(X \leq 0.7) &= P_r(0 < X \leq 0.7) = F(0.7) - F(0) = [F(x)]_0^{0.7} \\ &= [3x^2 - 2x^3]_0^{0.7} = 3(0.7)^2 - 2(0.7)^3 \end{aligned}$$

$$\begin{aligned}
 P_r(0.2 < X \leq 0.8) &= [F(x)]_{0.2}^{0.8} = [3x^2 - 2x^3]_{0.2}^{0.8} \\
 &= 3(0.8)^2 - 2(0.8)^3 - (3(0.2)^2 - 2(0.2)^3)
 \end{aligned}$$

4.

(a)

$f(x) \geq 0$ と $\int_{-\infty}^{\infty} f(x)dx = 1$ を示せばよいでしょう.

1. $f(x) = e^{-x}$ は指数関数より全ての x で $f(x) > 0$.

2.

$$\begin{aligned}
 \int_{-\infty}^{\infty} f(x)dx &= \int_{-\infty}^0 f(x)dx + \int_0^{\infty} f(x)dx \\
 &= \int_{-\infty}^0 0dx + \int_0^{\infty} e^{-x}dx \\
 &= [0 + -e^{-x}]_0^{\infty} = 1
 \end{aligned}$$

(b) $P_r(X \leq a) = 0.1$ を満たす a を求めます.

$$P_r(X \leq a) = P_r(X \leq 0) + P_r(0 \leq X \leq a) = \int_0^a e^{-x}dx = 1 - e^{-a} = 0.1$$

より $e^{-a} = 0.9$. したがって, $a = -\log 0.9$

$$P_r(X > b) = 1 - P_r(X \leq b) = 1 - (1 - e^{-b}) = e^{-b}$$

より $P_r(X > b) = 0.05$ を満たす b は

$$b = -\log 0.05$$

問題解答 3.2

1. $W = XY$ とおくと W の変域は $\{1, 2, 3, 4, 5, 6, 8, 9, 10, 12, 15, 16, 18, 20, 24, 25, 30, 36\}$ である. また, 確率分布 $h(i)$ は $P(W = i)$ で与えることに注意する. まず, $h(1)$ を考えてみよう. $h(1) = P(W = 1)$ より, 2個のサイコロを投げて両方1の目が出る確率を求めることと同じである. 場合の数を求めると, 両方1の目の組み合わせは36通り中の1通り. したがって $P(W = 1) = \frac{1}{36}$. 同様に, $W = 2, 3, 4, \dots, 36$ について求めると

i	1	2	3	4	5	6	8	9	10	12	15	16	18	20	24	25	30	36
$h(i)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{4}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{4}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
$H(i)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{8}{36}$	$\frac{10}{36}$	$\frac{14}{36}$	$\frac{16}{36}$	$\frac{17}{36}$	$\frac{19}{36}$	$\frac{23}{36}$	$\frac{25}{36}$	$\frac{26}{36}$	$\frac{28}{36}$	$\frac{30}{36}$	$\frac{32}{36}$	$\frac{33}{36}$	$\frac{35}{36}$	$\frac{36}{36}$

2.

(a) $W = X + Y + Z$ とおくと W の変域は $\{0, 1, 2, 3\}$ である. 確率分布 $h(i)$ は $P(W = i)$ で与えられる. まず, $h(0)$ を考えてみよう. $h(0) = P(W = 0)$ より, 3枚銅貨を投げて全て裏が出る確率を求めることと同じである. 場合の数を求めると, 全部裏の組み合わせは8通り中の1通り. したがって $P(W = 0) = \frac{1}{8}$. 同様に,

$W = 1, 2, 3$ について求めると

$$h(0) = P(W = 0) = \binom{3}{0} \left(\frac{1}{2}\right)^3 = \frac{1}{8}$$

$$h(1) = P(W = 1) = \binom{3}{1} \left(\frac{1}{2}\right)^3 = \frac{3}{8}$$

$$h(2) = P(W = 2) = \binom{3}{2} \left(\frac{1}{2}\right)^3 = \frac{3}{8}$$

$$h(3) = P(W = 3) = \binom{3}{3} \left(\frac{1}{2}\right)^3 = \frac{1}{8}$$

となる。(b) 分布関数 $H(i)$ を求める。 $H(i) = P(W \leq i)$ のことなので、

$$H(0) = P(W \leq 0) = \frac{1}{8}$$

$$H(1) = P(W \leq 1) = \frac{1}{8} + \frac{3}{8} = \frac{1}{2}$$

$$H(2) = P(W \leq 2) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} = \frac{7}{8}$$

$$H(3) = P(W \leq 3) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = 1$$

問題解答 3.3

(1)

(a)

$$1 - |x| = \begin{cases} 1 + x & -1 < x < 0 \\ 1 - x & 0 < x < 1 \end{cases}$$

より $f(x) \geq 0$. また

$$\begin{aligned} \int_{-\infty}^{\infty} f(x) dx &= \int_{-\infty}^{-1} 0 dt + \int_{-1}^0 (1+t) dt + \int_0^1 (1-t) dt + \int_1^{\infty} 0 dt \\ &= \left[t + \frac{t^2}{2} \right]_{-1}^0 + \left[t - \frac{t^2}{2} \right]_0^1 \\ &= -(-1 + \frac{1}{2}) + 1 - \frac{1}{2} = 1 \end{aligned}$$

別解 $f(x) \geq 0$ より, $\int_{-\infty}^{\infty} f(x) dx$ は $f(x)$ と x 軸の間の面積と考えられる。したがって, 求める面積は底辺 2 高さ 1 の三角形の面積より 1 .

(b)

$x < -1$ のとき

$$F(x) = P_r(X \leq x) = \int_{-\infty}^0 0 dt = 0$$

$-1 < x < 0$ のとき

$$\begin{aligned} F(x) = P_r(X \leq x) &= \int_{-\infty}^x f(t) dt = \int_{-\infty}^{-1} 0 dt + \int_{-1}^x (1+t) dt \\ &= \left[t + \frac{t^2}{2} \right]_{-1}^x \\ &= x + \frac{x^2}{2} - (-1 + \frac{1}{2}) = \frac{1}{2}(x+1)^2 \end{aligned}$$

$0 < x < 1$ のとき,

$$\begin{aligned} F(x) = P_r(X \leq x) &= \int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{-1} 0dt + \int_{-1}^0 (1+t)dt + \int_0^x (1-t)dt \\ &= \left[t + \frac{t^2}{2} \right]_{-1}^0 + \left[t - \frac{t^2}{2} \right]_0^x \\ &= -(-1 + \frac{1}{2}) + x - \frac{x^2}{2} = \frac{1}{2} + \frac{x}{2}(2-x) \end{aligned}$$

(c)

$$P_r(-\frac{1}{2} < X < \frac{1}{2}) = F(\frac{1}{2}) - F(-\frac{1}{2}) = \frac{1}{2} + \frac{1}{4}(2 - \frac{1}{2}) - \frac{1}{2}(\frac{1}{2})^2 = \frac{3}{4}$$

(d)

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} xf(x)dx \\ &= \int_{-\infty}^{-1} 0dx + \int_{-1}^0 x(1+x)dx + \int_0^1 x(1-x)dx + \int_1^{\infty} 0dx \\ &= \left[\frac{x^2}{2} + \frac{x^3}{3} \right]_{-1}^0 + \left[\frac{x^2}{2} - \frac{x^3}{3} \right]_0^1 = 0 \end{aligned}$$

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{\infty} x^2f(x)dx \\ &= \int_{-\infty}^{-1} 0dx + \int_{-1}^0 x^2(1+x)dx + \int_0^1 x^2(1-x)dx + \int_1^{\infty} 0dx \\ &= \left[\frac{x^3}{3} + \frac{x^4}{4} \right]_{-1}^0 + \left[\frac{x^3}{3} - \frac{x^4}{4} \right]_0^1 \\ &= -(-\frac{1}{3} + \frac{1}{4}) + \frac{1}{3} - \frac{1}{4} = \frac{2}{3} - \frac{1}{2} = \frac{1}{6} \end{aligned}$$

これより

$$V(X) = E(X^2) - E(X)^2 = \frac{1}{6}$$

2.

(a) 1 の目が出る回数を X とすると,

$$P_r(X=0) = \binom{3}{0} \left(\frac{5}{6}\right)^3 = \frac{125}{216}$$

$$P_r(X=1) = \binom{3}{1} \left(\frac{1}{6}\right) \left(\frac{5}{6}\right)^2 = \frac{25}{72}$$

$$P_r(X=2) = \binom{3}{2} \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right) = \frac{5}{72}$$

$$P_r(X=3) = \binom{3}{3} \left(\frac{1}{6}\right)^3 = \frac{1}{216}$$

別解 この試行はベルヌーイ試行である。つまり,

事象「1の目がでる」が発生するか否かを問う

この事象の発生はそれぞれの試行において独立である

この事象の発生は各試行において一定である

ベルヌーイ試行を n 回行なったとき、事象の発生回数 X を i 、事象の発生確率を p とすると

$$P_r(X = i) = \binom{n}{i} p^i (1-p)^{n-i}$$

となり、 X は2項分布に従うといい、 $X \sim B(n, p)$ と表わします。このとき、

$$E(X) = np, \quad V(X) = np(1-p)$$

が知られています。

(b)

$$E(X) = \sum_{i=0}^n i \cdot P_r(X = i) = 0 \cdot \frac{125}{216} + 1 \cdot \frac{25}{72} + 2 \cdot \frac{5}{72} + 3 \cdot \frac{1}{216} = \frac{1}{2}$$

$$E(X^2) = \sum_{i=0}^n i^2 \cdot P_r(X = i) = 0 \cdot \frac{125}{216} + 1 \cdot \frac{25}{72} + 2^2 \cdot \frac{5}{72} + 3^2 \cdot \frac{1}{216} = \frac{2}{3}$$

$$V(X) = E(X^2) - (E(X))^2 = \frac{2}{3} - \left(\frac{1}{2}\right)^2 = \frac{5}{12}$$

したがって、

$$D(X) = \sqrt{V(X)} = \sqrt{\frac{5}{12}}$$

別解

$$E(X) = 3 \cdot \frac{1}{6} = \frac{1}{2}, \quad V(X) = 3 \cdot \frac{1}{6} \cdot \frac{5}{6} = \frac{5}{12}$$

(c) X の標準化とは平均 μ を 0 に分散 σ^2 を 1 に変えることである。そこで X の標準化は

$$Z = \frac{X - \mu}{\sqrt{\sigma^2}} = \frac{X - \frac{1}{2}}{\sqrt{\frac{1}{6}}}$$

次に Z の確率分布を求めてみよう。

$$P_r(X = 0) = P_r\left(\frac{X - \frac{1}{2}}{\sqrt{\frac{1}{6}}} = \frac{0 - \frac{1}{2}}{\sqrt{\frac{1}{6}}}\right) = P_r\left(Z = -\sqrt{\frac{3}{5}}\right)$$

より

$$P_r\left(Z = -\sqrt{\frac{3}{5}}\right) = P_r(X = 0) = \frac{125}{216}$$

$$P_r(X = 1) = P_r\left(\frac{X - \frac{1}{2}}{\sqrt{\frac{1}{6}}} = \frac{1 - \frac{1}{2}}{\sqrt{\frac{1}{6}}}\right) = P_r\left(Z = \sqrt{\frac{3}{5}}\right)$$

より

$$P_r(Z = \sqrt{\frac{3}{5}}) = P_r(X = 0) = \frac{25}{72}$$

$$P_r(X = 2) = P_r\left(\frac{X - \frac{1}{2}}{\sqrt{\frac{5}{12}}} = \frac{2 - \frac{1}{2}}{\sqrt{\frac{5}{12}}}\right) = P_r(Z = \frac{3\sqrt{3}}{5})$$

より

$$P_r(Z = \frac{3\sqrt{3}}{5}) = P_r(X = 0) = \sqrt{\frac{25}{72}}$$

$$P_r(X = 3) = P_r\left(\frac{X - \frac{1}{2}}{\sqrt{\frac{5}{12}}} = \frac{3 - \frac{1}{2}}{\sqrt{\frac{5}{12}}}\right) = P_r(Z = \frac{5\sqrt{3}}{5})$$

より

$$P_r(Z = \frac{5\sqrt{3}}{5}) = P_r(X = 0) = \sqrt{\frac{1}{216}}$$

3

(a) Bernoulli の定理は試行回数が n , 事象発生回数が r , 事象発生確率が p のとき

$$P_r\left(\left|\frac{r}{n} - p\right| \leq \varepsilon\right) \geq 1 - \frac{p(1-p)}{n\varepsilon^2}$$

が成り立つといっているので ,

$$P_r\left(\left|\frac{r}{1000} - \frac{1}{2}\right| \leq \frac{1}{10}\right) \geq 1 - \frac{\frac{1}{2}(1 - \frac{1}{2})}{1000(\frac{1}{10})^2} = \frac{39}{40}$$

(b)

$$P_r\left(\left|\frac{r}{1000} - \frac{1}{4}\right| \leq \frac{1}{10}\right) \geq 1 - \frac{\frac{1}{4}(1 - \frac{1}{4})}{1000(\frac{1}{10})^2} = \frac{3}{160}$$

(c) この問題は試行回数が 2000 のとき , 事象発生回数 r の偏りが 50 回以内である確率を求めよということである . 言い換えると , 2000 回中の表が現われる割合 $\frac{r}{2000}$ と理論的確率 $p = \frac{1}{2}$ との誤差が $\frac{50}{1000}$ である確率を求めよということになる . そこで Bernoulli の定理を用いると

$$P\left(\left|\frac{r}{n} - p\right| \leq \varepsilon\right) \geq 1 - \frac{p(1-p)}{n\varepsilon^2}$$

より

$$P\left(\left|\frac{r}{2000} - \frac{1}{2}\right| < \frac{50}{1000}\right) \geq 1 - \frac{\frac{1}{2}(1 - \frac{1}{2})}{2000(\frac{1}{20})^2}$$

$$= 1 - \frac{\frac{1}{4}}{5} = 1 - 0.05 = 0.95$$

(d) この問題は n 回中の表が現われる割合 $\frac{r}{n}$ と理論的確率 $p = \frac{1}{2}$ との誤差が 0.05% である確率が 99% 以上になるには何回以上投げればよいかということを知っている . そこで Bernoulli の定理を用いると

$$P\left(\left|\frac{r}{n} - p\right| \leq \varepsilon\right) \geq 1 - \frac{p(1-p)}{n\varepsilon^2}$$

より $1 - \frac{p(1-p)}{n\varepsilon^2} \geq 0.99$ である n を求めればよい.

$$1 - \frac{\frac{1}{2}(1 - \frac{1}{2})}{n(0.05)^2} \geq 0.99 \Rightarrow \frac{\frac{1}{2}(1 - \frac{1}{2})}{n(0.05)^2} \leq 0.01 \Rightarrow$$

$$n \geq \frac{1}{4(0.05)^2 \cdot 0.01} = 10000$$

問題解答 3.4

1.

(a) $W = X + Y + Z$ とおくと W の変域は $\{0, 1, 2, 3\}$ である. 確率分布 $h(i)$ は $P(W = i)$ で与えられる. まず, $h(0)$ を考えてみよう. $h(0) = P(W = 0)$ より, 3枚銅貨を投げて全て裏が出る確率を求めることと同じである. 場合の数を求めると, 全部裏の組み合わせは 8通り中の 1通り. したがって $P(W = 0) = \frac{1}{8}$. 同様に, $W = 1, 2, 3$ について求めると

$$h(0) = P(W = 0) = \binom{3}{0} \left(\frac{1}{2}\right)^3 = \frac{1}{8}$$

$$h(1) = P(W = 1) = \binom{3}{1} \left(\frac{1}{2}\right)^3 = \frac{3}{8}$$

$$h(2) = P(W = 2) = \binom{3}{2} \left(\frac{1}{2}\right)^3 = \frac{3}{8}$$

$$h(3) = P(W = 3) = \binom{3}{3} \left(\frac{1}{2}\right)^3 = \frac{1}{8}$$

となる. 次に分布関数 $H(i)$ を求める. $H(i) = P(W \leq i)$ のことなので,

$$H(0) = P(W \leq 0) = \frac{1}{8}$$

$$H(1) = P(W \leq 1) = \frac{1}{8} + \frac{3}{8} = \frac{1}{2}$$

$$H(2) = P(W \leq 2) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} = \frac{7}{8}$$

$$H(3) = P(W \leq 3) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = 1$$

(b) W の平均値は $E(W)$ は

$$E(W) = E(X + Y + Z) = E(X) + E(Y) + E(Z)$$

で与えられることに注意すればすぐに求まる.

$$E(X) = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2}$$

$$E(Y) = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2}$$

$$E(Z) = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2}$$

より

$$E(W) = \frac{3}{2}$$

別解 W の確率分布が分かっているので, 直接求めることも可能である.

$$E(W) = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{3}{2}$$

次に W の分散 $V(W)$ を求める. 一般に $V(W) = E(W^2) - (E(W))^2$ を用いて求める方が簡単であるが, この問題では W の確率分布が分かっているので直接求める方が簡単である. $E(X) = \frac{3}{2}$ より

$$V(W) = \left\{ \left(0 - \frac{3}{2}\right)^2 \frac{1}{8} + \left(1 - \frac{3}{2}\right)^2 \frac{3}{8} + \left(2 - \frac{3}{2}\right)^2 \frac{3}{8} + \left(3 - \frac{3}{2}\right)^2 \frac{1}{8} \right\} = \frac{3}{4}$$

2.

(a) Y の確率密度関数 $g(y)$ を求めるには Y の分布関数 $G(y)$ について調べ, $g(y) = G'(y)$ の関係を用いる. X の分布関数 $F(x)$ は $F(x) = P(X \leq x)$, Y の分布関数 $G(y)$ は $G(y) = P(Y \leq y)$ で与えられるので,

$$G(y) = P(Y \leq y) = P(2X + 3 \leq y) = P(X \leq \frac{y-3}{2}) = F(\frac{y-3}{2})$$

を得る. 次に, X の確率密度関数は $f(x) = 2x$ ($0 \leq x < 1$) で与えられていることに注意すると, Y の確率密度関数 $g(y)$ は,

$$g(y) = G'(y) = F'(\frac{y-3}{2}) \frac{1}{2} = \frac{1}{2} f(\frac{y-3}{2}) = \frac{y-3}{2}$$

となる.

Y の期待値 $E(Y) = E(2X + 3) = 2E(X) + E(3) = 2E(X) + 3$ より $E(Y)$ を求めるには, $E(X)$ を求めればよいことが分かる. $E(X) = \int_{-\infty}^{\infty} xf(x) dx$ で与えられるので,

$$E(X) = \int_0^1 x(2x)dx = \int_0^1 2x^2 dx = \frac{2}{3}x^3 \Big|_0^1 = \frac{2}{3}$$

したがって, $E(Y) = E(2X + 3) = 2E(X) + 3 = \frac{4}{3} + 3 = \frac{13}{3}$

注: Y の確率密度関数を求めたので, 直接 $E(Y)$ を $E(Y) = \int_0^1 \frac{y-3}{2} dy$ で求めることもできる.

最後に Y の分散を求める. ここで $V(2X + 3) = V(2X)$, $V(2X) = 2^2V(X)$, $V(X) = E(X^2) - (E(X))^2$ に注意すると $V(Y)$ を求めるには $E(X^2)$ を求めればよいことが分かる.

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx = \int_0^1 2x^3 dx = \frac{1}{2}x^4 \Big|_0^1 = \frac{1}{2}$$

より

$$V(Y) = V(2X + 3) = 4V(X) = 4[E(X^2) - (E(X))^2] = 4[\frac{1}{2} - (\frac{2}{3})^2] = \frac{2}{9}$$

(b) Z の確率密度関数 $h(z)$ を求めるには Z の分布関数 $H(z)$ について調べ, $h(z) = H'(z)$ の関係を用いる.

X の分布関数 $F(x)$ は $F(x) = P(X \leq x)$, Z の分布関数 $H(z)$ は $H(z) = P(Z \leq z)$ で与えられるので,

$$H(z) = P(Z \leq z) = P(X^2 \leq z) = P(|X| \leq \sqrt{z}) = P(X \leq \sqrt{z}) - P(X \leq -\sqrt{z}) = F(\sqrt{z}) - F(-\sqrt{z})$$

を得る. これより Z の確率密度関数 $h(z)$ は,

$$h(z) = H'(z) = F'(\sqrt{z})(\sqrt{z})' - F'(-\sqrt{z})(-\sqrt{z})' = \frac{1}{2\sqrt{z}}(f(\sqrt{z}) + f(-\sqrt{z}))$$

となる. ここで

$$f(x) = \begin{cases} 2x & (0 \leq x < 1) \\ 0 & (\text{その他}) \end{cases}$$

より $f(\sqrt{z}) = 2\sqrt{z}$, $f(-\sqrt{z}) = 0$. したがって, $h(z) = \frac{1}{2\sqrt{z}} 2\sqrt{z} = 1$.

次に期待値を求める.

$E(Z) = \int_{-\infty}^{\infty} zf(z) dz$ で与えられるので,

$$E(Z) = E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx = \int_0^1 2x^3 dx = \frac{1}{2}x^4 \Big|_0^1 = \frac{1}{2}$$

最後に Z の分散を求める .

$$V(Z) = V(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx = \int_0^1 2x^3 dx = \frac{1}{2} x^4 \Big|_0^1 = \frac{1}{2}$$

より

$$V(Y) = V(2X + 3) = 4V(X) = 4[E(X^2) - (E(X))^2] = 4\left[\frac{1}{2} - \left(\frac{2}{3}\right)^2\right] = \frac{2}{9}$$

3.

(a) $W = XY$ とおくと, X, Y は独立である . したがって, その期待値は

$$E(W) = E(XY) = E(X)E(Y)$$

で求まる . X は 1 つのさいころを投げたとき出る目の値であるから, その期待値 $E(X)$ は

$$E(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{21}{6} = \frac{7}{2}$$

同様に, $E(Y) = \frac{7}{2}$ となるので, 求める期待値 $E(W)$ は $E(W) = \left(\frac{7}{2}\right)^2$ となる .

(b) $W = XY$ とおくと, その分散は

$$V(W) = E(W^2) - (E(W))^2 = E((XY)^2) - \left(\frac{7}{2}\right)^4$$

となるので, $E((XY)^2)$ を求めればよい . ここで X, Y は独立であることに注意すると

$$E((XY)^2) = E(X^2 Y^2) = E(X^2)E(Y^2)$$

さらに

$$E(X^2) = 1 \cdot \frac{1}{6} + 2^2 \frac{1}{6} + 3^2 \frac{1}{6} + 4^2 \frac{1}{6} + 5^2 \frac{1}{6} + 6^2 \frac{1}{6}$$

ここで

$$1^2 + 2^2 + 3^2 + \cdots + n^2 = \frac{1}{6}(n(n+1)(2n+1))$$

を用いると

$$E(X^2) = \frac{1}{6} \frac{1}{6} (6(7)(13)) = \frac{91}{6}$$

同様に $E(Y^2) = \frac{91}{6}$ となるので, 求める $V(W)$ は

$$V(W) = E((XY)^2) - \left(\frac{7}{2}\right)^4 = \left(\frac{91}{6}\right)^2 - \left(\frac{7}{2}\right)^4 = \frac{11515}{144}$$

となる .

問題解答 4.1

1. $p_r = P(X = r) = \binom{8}{r} (0.4)^r (0.6)^{1-r}$ より

r	0	1	2	3	4	5	6	7	8
P_r	0.0168	0.0896	0.209	0.2787	0.2322	0.1239	0.0413	0.0079	0.0007

$p_r = P(X = r) = \binom{8}{r} (0.2)^r (0.8)^{1-r}$ より

r	0	1	2	3	4	5	6	7	8
P_r	0.1678	0.3355	0.2936	0.1468	0.0459	0.0092	0.0011	0.0001	2.56×10^{-6}

2. プラモデルを取り出す試行はベルヌーイ試行. X を部品がかけているプラモデルのセットの数とすると, $X \sim B(3, \frac{2}{25})$. よって, 選んだ3セットが全て完全なセットである確率は

$$P(X=0) = \binom{3}{0} \left(\frac{2}{25}\right)^0 \left(1 - \frac{2}{25}\right)^3 = 0.7789$$

3. X を注射のあと一定期間生存できずに死亡するねずみの数とすると, $X \sim B(8, 0.4)$ となり, I で求めた確率分布が使える.

(a) 8匹全部が一定期間生存する確率は

$$P(X=0) = \binom{8}{0} (0.4)^0 (0.6)^8 = 0.0168$$

(b) r を 95%以上の確率で生存できないねずみの数となるとすると, X の分布関数 $F(x)$ は

$$F(r) = P(X \leq r) = \sum_{k \leq r} \binom{8}{k} (0.4)^k (0.6)^{8-k} \geq 0.95$$

を満たす. ここで, I の確率分布を用いると $P(X=0) = 0.168, P(X=1) = 0.0896$ より, $r=1$ となり, 95%で生存できるねずみの数は7匹以上となる.

問題解答 4.2

1. X を1日における交通事故の数とすると, 交通事故という事象は, ポワソン過程である. よって, 一日での事象の平均発生回数 λ は 0.8 となる.

(a) 死亡者0の日とは, $P_r(X=0)$ を求めることになる. よって,

$$P_r(X=0) = \frac{0.8^0}{0!} e^{-0.8} = 0.4493$$

死亡者6名以上の日とは, $P_r(X \geq 6)$ を求めることになる. よって,

$$P_r(X \geq 6) = 1 - P_r(X \leq 5)$$

$$P_r(X=0) = \frac{0.8^0}{0!} e^{-0.8} = 0.4493$$

$$P_r(X=1) = \frac{0.8^1}{1!} e^{-0.8} = 0.3595$$

$$P_r(X=2) = \frac{0.8^2}{2!} e^{-0.8} = 0.1438$$

$$P_r(X=3) = \frac{0.8^3}{3!} e^{-0.8} = 0.0383$$

$$P_r(X=4) = \frac{0.8^4}{4!} e^{-0.8} = 0.0077$$

$$P_r(X=5) = \frac{0.8^5}{5!} e^{-0.8} = 0.0012$$

これより,

$$P_r(X \geq 6) = 1 - 0.9998 = 0.0002$$

2. X を放射性物質から 1 秒間に放出される粒子の数とすると、粒子の放出はポワソン過程である。よって、一秒間での事象の平均発生回数 λ は 3 となる。これより、一秒間に 0 個の粒子が放出される確率は

$$P_r(X = 0) = \frac{3^0}{0!} e^{-3} = 0.0498$$

同様に、

$$P_r(X = 1) = \frac{3^1}{1!} e^{-3} = 0.1494$$

$$P_r(X = 2) = \frac{3^2}{2!} e^{-3} = 0.2240$$

$$P_r(X = 3) = \frac{3^3}{3!} e^{-3} = 0.2240$$

$$P_r(X = 4) = \frac{3^4}{4!} e^{-3} = 0.1680$$

$$P_r(X = 5) = \frac{3^5}{5!} e^{-3} = 0.1008$$

$$P_r(X = 6) = \frac{3^6}{6!} e^{-3} = 0.0504$$

最後に、1 秒間に少なくとも 1 個の粒子の放出は、 $P_r(X \geq 1)$ で与えられ、

$$P_r(X \geq 1) = 1 - P_r(X < 1) = 1 - P_r(X = 0) = 1 - 0.0498 = 0.9502$$

3. X をテープの傷の数とおくと、300m 中の傷の数の平均は 6 となり $X \sim P_6$ 。よって 300m の長さのテープ一巻中に傷が全くない確率は

$$P_r(X = 0) = e^{-6} \frac{6^0}{0!} = e^{-6} = 0.0025.$$

問題解答 4.3

1. $X \sim N(170, 10^2)$ より、 $Z = \frac{X-170}{10}$ とおくと、 $Z \sim N(0, 1)$ となる。

(a)

$$\begin{aligned} P_r(X \leq 160) &= P_r\left(\frac{X-170}{10} \leq \frac{160-170}{10}\right) = P_r(Z \leq -1) \\ &= 0.5 - P_r(0 \leq Z \leq 1) = 0.5 - 0.3413 = 0.1587 \end{aligned}$$

(b)

$$\begin{aligned} P_r(160 \leq X \leq 175) &= P_r\left(\frac{160-170}{10} \leq \frac{X-170}{10} \leq \frac{175-170}{10}\right) = P_r(-1 \leq Z \leq 0.5) \\ &= P_r(-1 \leq Z \leq 0) + P_r(0 \leq Z \leq 0.5) = 0.3413 + 0.1915 = 0.5328 \end{aligned}$$

2.

(a) $P_r(|Z| > \lambda) = 0.05$ を満たす λ を求める。標準正規分布表により、 $P_r(0 \leq Z \leq z)$ の値が与えられている。そこで、

$$P_r(|Z| > \lambda) = 1 - P_r(|Z| \leq \lambda)$$

と書き直すと,

$$P_r(0 \leq Z \leq \lambda) = \frac{1}{2}(1 - P_r(|Z| > \lambda)) = 0.225$$

これより, $\lambda = 1.645$ を得る。

(b) $P_r(|Z| > \lambda) = 0.05$ を満たす λ を求める。標準正規分布表により, $P_r(0 \leq Z \leq z)$ の値が与えられている。そこで,

$$P_r(Z > \lambda) = 0.5 - P_r(0 \leq Z \leq \lambda)$$

と書き直すと,

$$P_r(0 \leq Z \leq \lambda) = 0.5 - P_r(Z > \lambda) = 0.45$$

これより, $\lambda = 0.595$ を得る。

3.

(a) 平均 168.9, 分散 5.6^2 より正規化すると,

$$Z = \frac{X - 168.9}{5.6}$$

次に, 標準正規分布表より,

$$P_r(0 \leq Z \leq z) = 0.1, P_r(0 \leq Z \leq z) = 0.2, \dots, P_r(0 \leq Z \leq z) = 0.5$$

を満たす z を求めると,

$$P_r(0 \leq Z \leq 0.255) = 0.1$$

$$P_r(0 \leq Z \leq 0.525) = 0.2$$

$$P_r(0 \leq Z \leq 0.845) = 0.3$$

$$P_r(0 \leq Z \leq 1.281) = 0.4$$

$$P_r(0 \leq Z \leq 3.4) = 0.5$$

となる。これより, 総数を 10 等分するには, Z の値を次のように分ければよい。

$$-3.4 \sim -1.281 \sim -0.845 \sim -0.525 \sim -0.255 \sim 0 \sim 0.255 \sim 0.525 \sim 0.845 \sim 1.281 \sim 3.4$$

(b) 120 名の平均身長を \bar{X} とすると, 中心極限定理より

$$\bar{X} \sim N(168.9, \frac{5.6^2}{120})$$

ここで, 身長が平均が 168.9cm より 1.3cm 以上かたよる確率は

$$P_r(|\bar{X} - 168.9| \geq 1.3) = 1 - 2P_r(0 \leq \bar{X} - 168.9 < 1.3)$$

で求まる。ここで,

$$P_r(0 \leq \bar{X} - 168.9 < 1.3) = P_r(0 \leq \frac{\bar{X} - 168.9}{5.6/\sqrt{120}} \leq \frac{1.3}{5.6/\sqrt{120}}) = P_r(0 \leq Z \leq 2.543) = 0.4945$$

より,

$$P_r(|\bar{X} - 168.9| \geq 1.3) = 1 - 2(0.4945) = 0.011$$

4.

(a) $X \sim B(100, 0.02)$ より, 期待値 $E(X) = np = 2$ となり, ポワソン分布で近似できる。 $X \sim P_o(\lambda)$, $\lambda = E(X) = 2$ より,

$$\begin{aligned} P_r(X \geq 2) &= 1 - P_r(X < 2) = 1 - P_r(X = 0) - P_r(X = 1) \\ &= 1 - P_o(0) - P_o(1) = 1 - \frac{2^0}{0!}e^{-2} - \frac{2^1}{1}e^{-2} = 0.594 \end{aligned}$$

(b) $X \sim B(100, 0.2)$ より, 期待値 $E(X) = np = 20$ となり, 正規分布で近似できる。 $\mu = E(X) = np = 20$, $\sigma^2 = V(X) = np(1-p) = 16$ より, $X \sim N(20, 4^2)$ 。これより

$$\begin{aligned} P_r(X \geq 25) &= 1 - P_r(0 \leq X < 25) = 1 - P_r\left(0 \leq \frac{X - 20}{4} < \frac{25 - 20}{4}\right) \\ &= 1 - P_r(0 \leq Z < 1.25) = 1 - 0.3944 = 0.6056 \end{aligned}$$

(c) X を 1 の目の出る回数とすると, $X \sim B(600, 1/6)$ の 2 項分布に従う。そこで, 1 の目の出る回数 S が 90 回以上 100 回以下である確率を求めると

$$P_r(90 \leq X \leq 100) = P_r(X = 90) + P_r(X = 91) + \cdots + P_r(X = 100)$$

で求まるが, $\mu = E(X) = np = 100$, $\sigma^2 = V(X) = npq = 500/6$ より, $X \sim N(100, 500/6)$ の正規分布で近似できることが分かる。正規分布を用いると

$$\begin{aligned} P_r(90 \leq X \leq 100) &= P_r\left(\frac{90 - 100}{\sqrt{500/6}} \leq Z \leq 0\right) = P_r(0 \leq Z \leq \sqrt{1.2}) \\ &= P_r(0 \leq Z \leq 1.095) = 0.3621 \end{aligned}$$

(d) X を表の目の出る回数とすると, $X \sim B(100, 1/2)$ の 2 項分布に従う。そこで, 表の出る回数が 40 以上 60 以下の確率を求めると

$$P_r(40 \leq X \leq 60) = P_r(X = 40) + P_r(X = 41) + \cdots + P_r(X = 60)$$

で求まるが, $\mu = E(X) = np = 50$, $\sigma^2 = V(X) = npq = 25$ より, $X \sim N(50, 5^2)$ の正規分布で近似できることが分かる。正規分布を用いると

$$\begin{aligned} P_r(40 \leq X \leq 60) &= P_r\left(\frac{40 - 50}{5} \leq Z \leq \frac{60 - 50}{5}\right) = 2P_r(0 \leq Z \leq 2) \\ &= 2(.4772) = 0.9544 \end{aligned}$$

問題解答 4.4

1.

(a)

$$\begin{aligned} \eta(t) &= E(t^X) = \sum_{k=0}^{\infty} t^k P(X = k) \\ &= \sum_{k=0}^{\infty} t^k (1-p)^k p \end{aligned}$$

より、両辺を t について微分すると、

$$\eta'(t) = \sum_{k=1}^{\infty} kt^{k-1}(1-p)^k p$$

ここで、 $t=0$ とおくと

$$\eta'(0) = E(X) = \sum_{k=1}^{\infty} k(1-p)^k p = p \sum_{k=1}^{\infty} kq^k$$

ちょっと、難しいが $S = \sum_{k=1}^{\infty} kq^k$ とおき、 $S - qS = S(1-q) = Sp$ を求めると、

$$\begin{aligned} Sp &= S - qS = \sum_{k=1}^{\infty} kq^k - \sum_{k=1}^{\infty} kq^{k+1} \\ &= \sum_{k=1}^{\infty} q^k = \frac{q}{1-q} = \frac{q}{p} \end{aligned}$$

となる。これより、 $S = \frac{q}{p^2}$ となり、

$$E(X) = p \frac{q}{p^2} = \frac{q}{p}$$

(b) X をはじめて勝つ直前までの試合回数とおくと、 $X \sim G_e(0.3)$ 。よってはじめて勝つまでの平均試合数は $E(X) + 1$ より

$$E(X) + 1 = \frac{q}{p} + 1 = \frac{0.7}{0.3} + 1 = 3.33.$$

問題解答 4.5

1. X を当たりくじの数とおくと、 $X \sim H_g(10, 4, 3)$ 。よって 3 本くじを引いて、そのうち 2 本が当たる確率は

$$Pr(X=2) = \frac{\binom{4}{2}\binom{6}{1}}{\binom{10}{3}} = \frac{3}{10}$$

問題解答 4.6

I. (a) X を通話時間とおくと、 $x \sim E_x(\lambda)$ 。また $E(X) = 1/\lambda$ より $\lambda = 1/3$ となるので $X \sim E_x(1/3)$ 。よって通話が 4 分以内に終わる確率は

$$Pr(0 \leq X \leq 4) = \int_0^4 \lambda e^{-\lambda x} dx = \int_0^4 \frac{1}{3} e^{-x/3} dx = 1 - e^{-4/3}.$$

(b) その人が通話をはじめてから 2 分が経過しているとき、その後 4 分以内に終わる確率は

$$\begin{aligned} Pr(2 \leq X \leq 6 | X \geq 2) &= \frac{Pr(2 \leq X \leq 6, X \geq 2)}{Pr(X \geq 2)} \\ &= \frac{\int_2^6 \frac{1}{3} e^{-x/3} dx}{\int_2^{\infty} \frac{1}{3} e^{-x/3} dx} = \frac{e^{-2/3} - e^{-2}}{e^{-2/3}} = 1 - e^{-4/3}. \end{aligned}$$

問題解答 4.7

I. $f(x)$ が密度関数になるには、

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

を満たす必要がある.

(a)

$$\begin{aligned}
 \int_{-\infty}^{\infty} f(x)dx &= \int_{-\infty}^1 f(x)dx + \int_1^3 f(x)dx + \int_3^{\infty} f(x)dx \\
 &= 0 + \int_1^3 ce^{-1}\{-(x-2)^2 + 2\}dx + \int_3^{\infty} ce^{-(x-2)}dx \\
 &= ce^{-1} \left[-\frac{1}{3}(x-2)^3 + 2x \right]_1^3 - \left[ce^{-(x-2)} \right]_3^{\infty} \\
 &= ce^{-1} \left(-\frac{1}{3} + 6 - \left(\frac{1}{3} + 2 \right) \right) - ce^{-1}(0 - (-1)) \\
 &= \frac{13}{3}ce^{-1}
 \end{aligned}$$

したがって, $c = \frac{3e}{13}$.

(b) $P(2.5 \leq X \leq 3.5) = \int_{2.5}^{3.5} f(x) dx$

2. X_1 を A 型の人的人数, X_2 を B 型の人的人数, X_3 を AB 型の人的人数, X_4 を O 型の人的人数とすると, $\vec{X} = (X_1, X_2, X_3, X_4) \sim M(4, 0.35, 0.25, 0.1, 0.3)$. よって

$$P_r(X_1 = 1, X_2 = 1, X_3 = 1, X_4 = 1) = \frac{4!}{1!1!1!1!} (0.35)(0.25)(0.1)(0.3) = 0.063.$$

3. X_1, X_2, X_3 を 3 人の通話時間とおくと, $X_1 \sim E_x(\lambda) = \Gamma(1, 1/\lambda)$, $X_2 \sim E_x(\lambda) = \Gamma(1, 1/\lambda)$, $X_3 \sim E_x(\lambda) = \Gamma(1, 1/\lambda)$. ここで $X = X_1 + X_2 + X_3$ とおくと X は 3 人の通話時間の合計を表わす. 言い換えると待ち時間を表わし, $X \sim \Gamma(3, 1/\lambda)$. また $E(X_1) = 1/\lambda = 2$ より $X \sim \Gamma(3, 2)$. よって 10 分以上待たなければならない確率は

$$\begin{aligned}
 P_r(X \geq 10) &= 1 - P_r(0 \leq X \leq 10) = \int_{10}^{\infty} \frac{1}{\Gamma(3)2^3} x^2 e^{-x/2} dx \\
 &= 1 - \frac{1}{16} \int_0^{10} x^2 e^{-x/2} dx = 1 - [-2x^2 e^{-x/2} \Big|_0^{10} + 4 \int_0^{10} x e^{-x/2} dx] \\
 &= 1 - \frac{1}{16} [-200e^{-5} + 4(-2xe^{-x/2} \Big|_0^{10} + 2 \int_0^{10} e^{-x/2} dx)] \\
 &= 1 - \frac{1}{16} [-200e^{-5} - 80e^{-5} - 16e^{-x/2} \Big|_0^{10}] \\
 &= 1 - \frac{1}{16} [-296e^{-5} + 16] = \frac{296e^{-5}}{16} = 0.1246
 \end{aligned}$$

問題解答 5.1

1.

$$\begin{aligned}
 \bar{X} &= \frac{1}{10}(110 + 121 + 133 + 124 + 126 + 118 + 112 + 125 + 131 + 120) = 122(\text{cm}) \\
 U^2 &= \frac{1}{9}((110 - 122)^2 + (121 - 122)^2 + \cdots + (120 - 122)^2) \\
 &= 55.111(\text{cm})
 \end{aligned}$$

また, $S^2 = \frac{9}{10}U^2 = 49.5999$ より, $S = 7.043$ となります.

問題解答 5.2

1 ある水域の一定区間における水質 BOD を X とおくと, $X \sim N(\mu, 6.25)$. 又, 標本数は 15 で, 標本平均 $\bar{X} = 7.2$ より,

$$\bar{X} \sim N\left(\mu, \frac{6.25}{15}\right)$$

ここで, \bar{X} を標準化すると,

$$Z = \frac{X - \mu}{\sqrt{\frac{6.25}{15}}} = \frac{7.25 - \mu}{\sqrt{\frac{6.25}{15}}}$$

これより,

$$P_r(|Z| \leq z_{\frac{\alpha}{2}}) = 0.95$$

を満たす, $z_{\frac{\alpha}{2}}$ を標準正規分布表を用いて求めると, $z_{\frac{\alpha}{2}} = 1.96$ したがって, 95%信頼区間は

$$|Z| = \left| \frac{7.25 - \mu}{\sqrt{\frac{6.25}{15}}} \right| \leq 1.96$$

つまり

$$7.25 - 1.96\sqrt{\frac{6.25}{15}} \leq \mu \leq 7.25 + \sqrt{\frac{6.25}{15}}$$

2 標準偏差が 2.2 より, 母分散 $\sigma^2 = 6.25$ は既知である. この母集団から無作為に選んだ標本 X_i は $X_i \sim N(\mu, 6.25)$ の正規分布に従っていると考えることができる. したがって,

$$\bar{X} \sim N(\mu, \sigma^2/5)$$

となる. ここで, \bar{X} を求めると,

$$\bar{X} = \frac{1}{5}[28 + 24 + 31 + 27 + 22] = \frac{132}{5} = 26.4$$

標準化を行なうと,

$$Z = \frac{\bar{X} - \mu}{\sqrt{\sigma^2/5}} \sim N(0, 1)$$

となる. 95%信頼区間より, $P_r(|Z| \leq z_{\frac{\alpha}{2}}) = 0.95$. また, $z_{0.05} = 1.96$. したがって,

$$\bar{X} - z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{5}} \leq \mu \leq \bar{X} + z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{5}}$$

$$26.4 - 1.96\sqrt{6.25/5} \leq \mu \leq 26.4 + 1.96\sqrt{6.25/5}$$

$$24.21 \leq \mu \leq 28.59$$

3 母平均 $\mu = 146$ であるが母分散 σ^2 は未知である. この母集団から無作為に選んだ標本 X_i は $X_i \sim N(146, \sigma^2)$ の正規分布に従っていると考えることができる. したがって,

$$\bar{X} \sim N(146, \sigma^2/4)$$

となる. ここで, \bar{X} を求めると,

$$\bar{X} = \frac{1}{4}[145.3 + 145.1 + 145.4 + 146.2] = \frac{582}{4} = 145.5$$

母分散の推定に S'^2 を用いると,

$$T = \frac{\bar{X} - \mu}{\sqrt{S'^2/4}} \sim t_{n-1, \alpha/2}$$

となる. そこで, S'^2 を求めると,

$$\begin{aligned} S'^2 &= \frac{1}{3}[(145.3 - 145.5)^2 + (145.1 - 145.5)^2 + (145.4 - 145.5)^2 + (146.2 - 145.5)^2] \\ &= \frac{1}{3}(0.04 + 0.16 + 0.01 + 0.49) = 0.23 \end{aligned}$$

となる.

95%信頼区間より, $P_r(|T| \leq t_{n-1, \alpha/2}) = 0.95$. また, $t_{3, 0.005/2} = 3.18$. したがって,

$$\begin{aligned} \bar{X} - t_{3, 0.05/2} \sqrt{\frac{S'^2}{4}} &\leq \mu \leq \bar{X} + t_{3, 0.05/2} \sqrt{\frac{S'^2}{4}} \\ 145.5 - 3.18 \sqrt{0.23/4} &\leq \mu \leq 145.5 + 3.18 \sqrt{0.23/4} \\ 144.73 &\leq \mu \leq 146.26 \end{aligned}$$

問題解答 5.3

1 標本比率は $\hat{p} = \frac{108}{600} = 0.18$. また, $z_{\frac{\alpha}{2}} = z_{\frac{0.05}{2}} = 1.96$ であるから, 上記の公式に代入すると

$$\left(0.18 - 1.96 \sqrt{\frac{(0.18)(0.82)}{600}}, 0.18 + 1.96 \sqrt{\frac{(0.18)(0.82)}{600}} \right)$$

より, (0.15, 0.21) となります.

問題解答 5.4

1 標本比率は $\hat{p} = \frac{180}{900} = 0.2$. また, $z_{\frac{\alpha}{2}} = z_{\frac{0.05}{2}} = 1.96$ であるから, 十分大きな n に対して, 統計量 \bar{X} の分布が正規分布 $N(p, \frac{pq}{n})$ で近似される. したがって, 与えられた α に対して

$$P\left(\frac{|\bar{X} - p|}{\sqrt{pqn}} \leq z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

が成り立つ. これより,

$$\bar{X} - z_{\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}} \leq p \leq \bar{X} + z_{\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}$$

が成り立つ. ここで, \bar{X} と p を \bar{p} で置き換えると,

$$\left(\bar{p} - z_{\frac{\alpha}{2}} \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}, \bar{p} + z_{\frac{\alpha}{2}} \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \right)$$

これより,

$$\left(0.2 - 1.96 \sqrt{\frac{(0.2)(1-0.2)}{900}}, 0.2 + 1.96 \sqrt{\frac{(0.2)(1-0.2)}{900}} \right)$$

より, (0.174, 0.226) となる.

1 標本比率は $\hat{p} = \frac{187}{300} = 0.623$. また, $z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$ であるから, 十分大きな n に対して, 統計量 \bar{X} の分布が正規分布 $N(p, \frac{pq}{n})$ で近似される. したがって, 与えられた α に対して

$$P\left(\frac{|\bar{X} - p|}{\sqrt{pq/n}} \leq z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

が成り立つ. これより,

$$\bar{X} - z_{\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}} \leq p \leq \bar{X} + z_{\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}$$

が成り立つ. ここで, \bar{X} と p を \bar{p} で置き換えると,

$$(\bar{p} - z_{\frac{\alpha}{2}} \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}, \bar{p} + z_{\frac{\alpha}{2}} \sqrt{\frac{\bar{p}(1-\bar{p})}{n}})$$

これより,

$$\left(0.623 - 1.96 \sqrt{\frac{(0.623)(1-0.623)}{300}}, 0.623 + 1.96 \sqrt{\frac{(0.623)(1-0.623)}{300}}\right)$$

より, (0.568, 0.678) となる.

問題解答 6.1

1 この工場の製品を X とすると, $X \sim N(\mu, 0.20^2)$ であることが分かる. この工場から 16 個の製品を取り出したとき, \bar{X} をそれらの標本平均とすると,

$$\bar{X} = 7.09, \bar{X} \sim N(\mu, 0.20^2/16)$$

となる. 次に, この日の製品が規格から外れているかの検定を行なう.

$$H_0 : \mu = 7$$

$$H_1 : \mu \neq 7$$

有意水準 $\alpha = 0.05$

統計量 σ^2 が既知より,

$$Z = \frac{\bar{X} - \mu}{\sigma^2/n} \sim N(0, 1)$$

H_0 のもとで,

$$Z_0 = \frac{7.09 - 7.00}{\sqrt{0.20^2/16}} = \frac{4(0.09)}{0.2} = 1.8$$

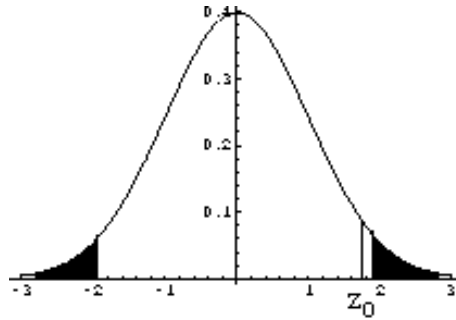


図 7.4: 正規分布

$z_{\frac{0.05}{2}} = 1.96$ より, H_0 を容認する.

95%信頼区間は

$$7.09 - 1.96\sqrt{\frac{0.20^2}{16}} \leq \mu \leq 7.09 + 1.96\sqrt{\frac{0.20^2}{16}}$$

より,

$$6.99 \leq \mu \leq 7.19$$

2 この工場の製品を X とすると, $X \sim N(\mu, 0.0016)$ であることが分かる. この工場から 8 個の製品を取り出したとき, \bar{X} をそれらの標本平均とすると,

$$\bar{X} = \frac{1}{8}(11.97 + 12.02 + \cdots + 12.05) = 12.028$$

となる. 次に, この日の製品の直径の分散は 0.0016 より大きいといえるかの検定を行なう.

$$H_0 : \sigma^2 = 0.0016$$

$$H_1 : \sigma^2 > 0.0016$$

有意水準 $\alpha = 0.05$

統計量 μ が既知で σ^2 の検定を行なうので

$$\chi^2 = \frac{nS^2}{\sigma^2} \sim \chi_{\alpha, n-1}^2$$

H_0 のもとで

$$S^2 = \frac{1}{8}[(11.97 - 12.02)^2 + \cdots + (12.05 - 12.028)^2] = 0.0021$$

$$\chi_0^2 = \frac{8(0.0021)}{0.0016} = 10.5$$

$$\chi_0^2 > \chi_{0.05, 7}^2 = 14.07$$

となり, H_0 は棄却される.

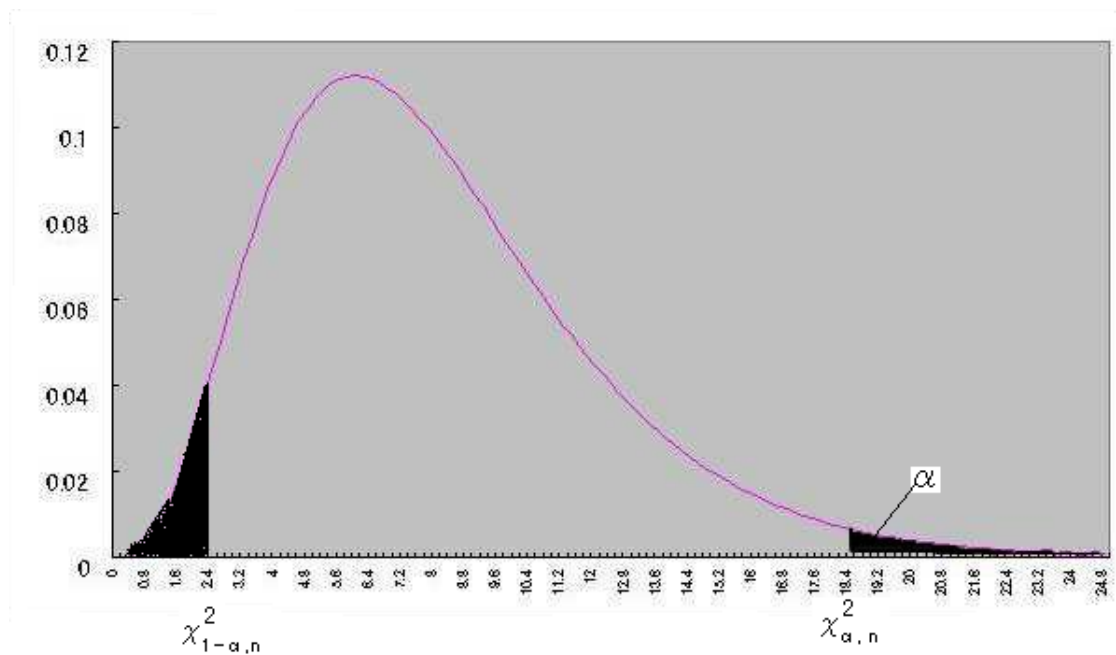


図 7.5: 2乗分布

95%信頼区間は

$$\chi^2_{1-0.05/2, 8-1} \leq \frac{0.0168}{\sigma^2} \leq \chi^2_{0.05/2, 8-1}$$

より,

$$1.690 \leq \frac{0.0168}{\sigma^2} \leq 16.01$$

$$0.0010 \leq \sigma^2 \leq 0.0099$$

問題解答 6.2

1

$$n_A = 10, \bar{X} = 82, S_A^2 = 54.41$$

$$n_B = 12, \bar{Y} = 76, S_B^2 = 59.17$$

母平均の差の検定である. $\sigma_1^2 = \sigma_2^2$ より,

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

有意水準 $\alpha = 0.05$

統計量

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\hat{\sigma}^2}{n_A} + \frac{\hat{\sigma}^2}{n_B}}} \sim t_{n_A + n_B - 2}$$

ただし,

$$\hat{\sigma}^2 = \frac{n_A S_A^2 + n_B S_B^2}{n_A + n_B - 2}$$

H_0 のもとで,

$$T_0 = \frac{82 - 76}{\sqrt{62.7/10 + 62.7/12}} = \frac{6}{\sqrt{6.27 + 5.23}} = 1.77$$

$t_{0.05/2,20} = 2.23$ より,

$$T_0 = 1.77 < t_{0.05/2,20} = 2.09$$

したがって, H_0 は棄却されない.

II.

$$n_A = 10, \bar{X} = 82, S_A^2 = 54.41, S'_A{}^2 = 60.44$$

$$n_B = 12, \bar{Y} = 76, S_B^2 = 59.17, S'_B{}^2 = 64.53$$

2 母分散の比の左側検定である. $\sigma_1^2 < \sigma_2^2$

$$H_0 : \sigma_A^2 = \sigma_B^2$$

$$H_1 : \sigma_A^2 < \sigma_B^2$$

有意水準 $\alpha = 0.05$

統計量

$$F = \frac{\sigma_B^2 S'_A{}^2}{\sigma_A^2 S'_B{}^2} \sim F_{n_A-1, n_B-1}$$

H_0 のもとで,

$$F_0 = \frac{60.44}{64.53} = 0.936$$

$$F_{1-0.05, 10-1, 12-1} = \frac{1}{F_{0.05, 12-1, 10-1}} = \frac{1}{3.105} = 0.322$$

より,

$$F_0 = 0.936 > F_{1-0.05, 10-1, 12-1} = 0.322$$

したがって, H_0 は棄却されない.

一般に,

$$F_{\alpha, n_1, n_2} = \frac{1}{F_{1-\alpha, n_2, n_1}}$$

が成り立つ.

問題解答 6.3

1 遺伝形質 $A : B : C : D = 9 : 3 : 3 : 1$

H_0 : 「メンデル比に従っている」 ($p_1, p_2, p_3, p_4 = \frac{9}{16}, \frac{3}{16}, \frac{3}{16}, \frac{1}{16}$)

H_1 : 「メンデル比に従っていない」 ($p_1, p_2, p_3, p_4 \neq \frac{9}{16}, \frac{3}{16}, \frac{3}{16}, \frac{1}{16}$)

有意水準 $\alpha = 0.05$

統計量

$$\chi^2 = \sum_{i=1}^4 \frac{(X_i - np_i)^2}{np_i} = \sum_{i=1}^4 \frac{X_i^2}{np_i} - n$$

H_0 のもとで,

$$\begin{aligned}\chi_0^2 &= \frac{243^2}{229.5} + \frac{72^2}{76.5} + \frac{78^2}{76.5} + \frac{15^2}{25.5} - 408 \\ &= 257.294 + 67.764 + 79.529 + 8.824 - 408 = 5.411\end{aligned}$$

$\chi_{0.05,4-1}^2 = 7.81$ より,

$$\chi_0^2 = 5.411 < \chi_{0.05,3}^2 = 7.81$$

したがって, H_0 を容認.

問題解答 6.4

1

H_0 : 「ポワソン分布 $P(\lambda)$ に従っている」

有意水準 $\alpha = 0.05$

統計量

この表をポワソン分布とみて, 死亡数の理論値を求める. これがポワソン分布 $P(\lambda)$ によるものと考えて, λ の値を推定する. 死亡者数 k のときの確率を p_k とすると,

$$\sum_{k=0}^{\infty} kp_k = E(X) = \lambda$$

死亡者数	k	0	1	2	3	4	計
部隊数	f_k	142	99	46	11	2	300
	kf_k	0	99	92	33	8	232
	p_k	0.473	0.33	0.153	0.036	0.0066	
理論度数	m_k	141.9	99	45.9	10.8	1.98	

ここで, $np_k \approx f_k$ より $\sum_k kf_k \approx \lambda n$. これより平均値 λ は

$$\lambda \approx \frac{1}{n} \sum_k kf_k = \frac{232}{300} = 0.77$$

死亡者数	k	0	1	2	3	4	計
部隊数	f_k	142	99	46	11	2	300
理論度数	m_k	141.9	99	45.9	10.8	1.98	

この表で, $k=4$ の所の m_k は単独で 5 よりも小さいので, χ^2 検定ができない. そこで, 右から順に m_i を加えて 5 を越すまで合併すると, $k \geq 3$ の階級を 1 つにしなければならない. したがって,

$$\chi^2 = \sum_{i=0}^3 \frac{(x_i - m_i)^2}{m_i}$$

H_0 のもとで,

$$\begin{aligned}\chi_0^2 &= \frac{(142 - 141.9)^2}{141.9} + \frac{(99 - 99)^2}{99} + \frac{(46 - 45.9)^2}{45.9} + \frac{(13 - 12.78)^2}{12.78} \\ &= 0.0040\end{aligned}$$

$\chi_{0.05, 3-1-1}^2 = 3.84$ より,

$$\chi_0^2 = 0.0040 < \chi_{0.05, 1}^2 = 3.84$$

したがって, H_0 を容認.

母数 λ が標本から 1 個推定されたので, 自由度は $3 - 1 - 1 = 1$ となる.

2

H_0 : 「飲酒と喫煙とは独立である。」

H_1 : 「飲酒と禁煙とは独立ではない。」

有意水準 $\alpha = 0.05$

統計量

$$\chi^2 = \sum_{i=1}^{350} \frac{(X_{ij} - nP_{ij})^2}{nP_{ij}}$$

H_0 のもとで,

$$\begin{aligned}\chi_0^2 &= 350 \left[\frac{1}{159} \left(\frac{39^2}{80} + \frac{54^2}{120} + \frac{49^2}{104} + \frac{17^2}{46} \right) \right. \\ &\quad \left. + \frac{1}{119} \left(\frac{27^2}{80} + \frac{43^2}{120} + \frac{40^2}{104} + \frac{9^2}{46} \right) + \frac{1}{72} \left(\frac{14^2}{80} + \frac{23^2}{120} + \frac{15^2}{104} + \frac{20^2}{46} \right) \right] - 1 \\ &= 367.55\end{aligned}$$

自由度 $(3 - 1) \times (4 - 1) = 6$ より, $\chi_{0.05, 6}^2 = 12.59$.

$\chi_0^2 = 367.55 > \chi_{0.05, 6}^2 = 12.59$ より, H_0 は棄却される. したがって, 飲酒と喫煙には関係がある.