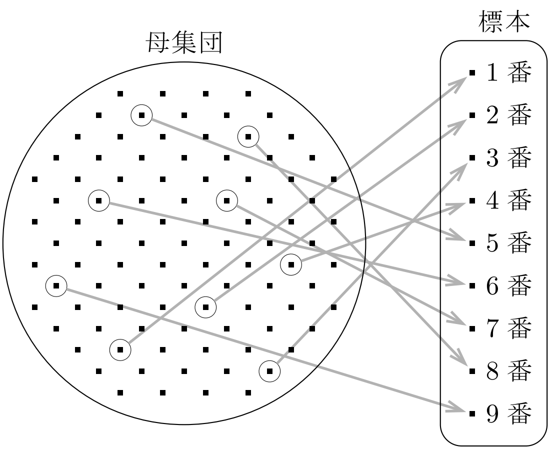
§5.1　標本調査

物とか事とかの属性を表す値を，その物事の属性値といいます．例えば，人を対象とするとき，各個人の身長とか体重とかは属性値です．また例えば，工業製品の一個一個について，重量とか耐久時間とか，性能を示す数量などは属性値です．通常，事物の属性値は，その事物を測定あるいは観測すると得られます．ある事物の集合があって，その集合に属す事物のある属性値の平均とか分散とか標準偏差とかを知りたいとします．これらの値はその集合に属す総ての事物の属性値を調べれば分かります．考慮の対象となる総ての事物の属性値を調べることを全数調査 (total inspection) または悉皆（しっかい）調査といいます．しかし，調査によっては，全数調査が事実上不可能であったり，不経済であったり，全数調査をすると調査の意味が無くなるなどということがあります．例えば，内閣の支持率を調べるには，選挙権を持つ人の全員に支持するかしないか尋ねるのが一番確実ですが，これは手間と時間とが掛かりすぎて非現実的です．また例えば，ある工程で製造された製品の強度の平均は，その工程で製造された製品の総ての強度を調べれば正確に分かりますが，強度の検査が破壊検査であれば，総ての製品が破壊された後で強度の平均が分かってもあまり意味がありません．



このようなとき，考慮の対象となる事物の中から一部分だけを選び出してそれらの属性値を調べて，その結果に基づいて全体の平均や分散などを推し測ります．このように，考慮の対象となる事物の中から選び出した一部分だけの事物の属性値を調べることを標本調査 (sample survey) といいます．このとき，考慮の対象となる物事の全体を**母集団** (population) といいます．母集団に属す総ての事物の属性値の，平均，分散，標準偏差を，各々，**母（集団）平均**，**母（集団）分散**，**母（集団）標準偏差**といいます．このような，母集団の性質を表す量を母数 (population parameter) といいます．また，母集団に属す事物の個数を母集団の大きさ (population size) といいます．

標本調査のために母集団から抽出した事物の列を**標本** (sample) といいます．標本に属す事物の属性値の，平均，分散，標準偏差を，各々，**標本平均**，**標本分散**，**標本標準偏差**といいます．母集団の中から事物を n 回抽出してできる標本を**大きさ** (size) n の標本といいます．つまり，大きさ n の標本とは，母集団から抽出してできる n 個の事物の列です．

標本調査において大きさ n の標本を抽出するとします．

$$
\begin{aligned}
&1\text{ 番の事物の属性値を表す変数を } X_1\,,\\
&2\text{ 番の事物の属性値を表す変数を } X_2\,,\\
&3\text{ 番の事物の属性値を表す変数を } X_3\,,\\
&\qquad\vdots\\
&n\text{ 番の事物の属性値を表す変数を } X_n\,,
\end{aligned}
$$

とおきます．母集団が一定で，かつ抽出の仕方が定められれば，通常，これらの変数 $X_1, X_2, X_3, \ldots, X_n$ は確率変数と考えられます．これらを**標本確率変数**といいます．更に，標本確率変数の実現値が分かれば値を計算できる確率変数を（**標本**）**統計量** (statistic) といいます．例えば，標本確率変数 $X_1, X_2, X_3, \ldots, X_n$ に対して以下のような確率変数は標本統計量です：

$$\text{標本の平均を表す確率変数}\ \ \overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k\ ,$$

$$\text{標本の偏差平方和を表す確率変数}\ \ S = \sum_{k=1}^{n}\left(X_k - \overline{X}\right)^2\ ,$$

$$\text{標本の分散を表す確率変数}\ \ V = \frac{1}{n}\sum_{k=1}^{n}\left(X_k - \overline{X}\right)^2 = \frac{S}{n}\ ,$$

$$\text{標本の不偏分散を表す確率変数}\ \ U = \frac{1}{n-1}\sum_{k=1}^{n}\left(X_k - \overline{X}\right)^2 = \frac{S}{n-1}\ .$$

標本統計量が従う確率分布を標本分布 (sample distribution) といいます．標本統計量の実現値は標本調査によって実際に得られた属性値を標本確率変数に代入して計算した値です．

母集団から一定の方法で抽出した事物の属性値を表す変数は，普通，確率変数です．この確率変数を母（集団）確率変数といい，母集団確率変数が従う確率分布を母集団分布 (population distribution) といいます．母集団分布が正規分布である母集団を特に**正規母集団** (normal distribution) といいます．

母集団からの標本の抽出について，標本を選ぶ場合の一つ一つが全てが等確率であるとき，この抽出は**無作為** (random) であるといいます．つまり，標本の無作為抽出とは，標本を選ぶときの場合の数が全部で N 通りあるときどの場合になる確率も等しく $\dfrac{1}{N}$ になるような抽出のことです．

例として簡単な場合を考えます．大きさ 4 の母集団に属す 4 つの事物の各々を A, B, C, D とおきます．これら A, B, C, D の各々の属性値が $3, 1, 9, 3$ であるとします．この母集団から大きさ 2 の標本を抽出します．標本の 1 番目の事物と 2 番目の事物とは異なるものとして，無作為抽出します．このとき抽出の仕方の総数は 12 ですから，抽出の仕方の各々の確率は $\dfrac{1}{12}$ です．

1番目	1番目の属性値	2番目	2番目の属性値	標本平均	確率
A	3	B	1	$(3+1)/2=2$	1/12
A	3	C	9	$(3+9)/2=6$	1/12
A	3	D	3	$(3+3)/2=3$	1/12
B	1	A	3	$(1+3)/2=2$	1/12
B	1	C	9	$(1+9)/2=5$	1/12
B	1	D	3	$(1+3)/2=2$	1/12
C	9	A	3	$(9+3)/2=6$	1/12
C	9	B	1	$(9+1)/2=5$	1/12
C	9	D	3	$(9+3)/2=6$	1/12
D	3	A	3	$(3+3)/2=3$	1/12
D	3	B	1	$(3+1)/2=2$	1/12
D	3	C	9	$(3+9)/2=6$	1/12

この表より，標本平均を表す変数を $\overline{X}$ とおくと，$\overline{X}$ は次のような離散型確率変数です：

$$\mathrm{P}\left(\overline{X}=2\right) = \frac{1}{3}\,,\quad \mathrm{P}\left(\overline{X}=3\right) = \frac{1}{6}\,,\quad \mathrm{P}\left(\overline{X}=5\right) = \frac{1}{6}\,,\quad \mathrm{P}\left(\overline{X}=6\right) = \frac{1}{3}\,.$$

$2, 3, 5, 6$ 以外の実数 k について $\mathrm{P}\left(X=k\right)=0$．従って，標本平均を表す確率変数 $\overline{X}$ の平均値 $\mathrm{E}\left[\overline{X}\right]$ は

$$
\begin{aligned}
\mathrm{E}\left[\overline{X}\right] &= 2\times\frac{1}{3} + 3\times\frac{1}{6} + 5\times\frac{1}{6} + 6\times\frac{1}{3} = \frac{4+3+5+12}{6}\\
&= 4\,.
\end{aligned}
$$

標本平均を表す確率変数 $\overline{X}$ の分散 $\mathrm{V}\left[\overline{X}\right]$ は

$$
\begin{aligned}
\mathrm{V}\left[\overline{X}\right] &= (2-4)^2\times\frac{1}{3} + (3-4)^2\times\frac{1}{6} + (5-4)^2\times\frac{1}{6} + (6-4)^2\times\frac{1}{3} = \frac{8+1+1+8}{6}\\
&= 3\,.
\end{aligned}
$$

このようにして標本平均を表す確率変数の平均値や分散などが考えられます．