統計的な推測【標本比率と正規分布】 p.96～97

青チャートB: 例題 86～87

【態度目標】しゃべる、質問する、説明する、動く、協力する、貢献する

【内容目標】集団から抽出した標本の平均値が従う分布の性質について理解できるようになろう。

□標本比率と正規分布

たとえば，ある工場で製造された製品に含まれる不良品の割合を調べる場合のように，母集団においてある 1 つの特性をもつものの割合を調査の対象とすることがある。

このとき，一般に，母集団全体の中である特性 A をもつ要素の割合を，その特性 A の **母比率** という。また，これに対して抽出された標本の中で特性 A をもつ要素の割合を **標本比率** という。

特性 A の母比率が p である十分大きな母集団から，大きさが n の標本を無作為に抽出するとき，標本の中で特性 A をもつものの個数を T とすると，T は二項分布 $B(n,\ p)$ に従う。

よって，$q=1-p$ とすると，n が大きいとき，T は近似的に正規分布 $N(np,\ npq)$ に従う。

二項分布 $B(n,\ p)$ に従う確率変数 X は、n が大きいとき、近似的に正規分布 $N(np,\ npq)$ に従う

特性 A の標本比率を R とすると，$R=\dfrac{T}{n}$ である。R は標本平均 $\overline{X}$ と同様に確率変数で

$$E(R)=\frac{1}{n}E(T)=\frac{1}{n}\cdot np=p$$

$$V(R)=\frac{1}{n^2}V(T)=\frac{1}{n^2}\cdot npq=\frac{pq}{n}$$

したがって，

特性 A の母比率 p の母集団から抽出された大きさ n の無作為標本について，標本比率 R は，n が大きいとき，近似的に正規分布 $N\left(p,\ \dfrac{pq}{n}\right)$ に従うとみなすことができる。

標本比率 R は，次のように考えると，標本平均 $\overline{X}$ の特別な場合になる。

特性 A の母比率が p である母集団から，大きさが n の標本を無作為に抽出するとき，それらに対して，X_1，X_2，……，X_n の値を次のように定める。

特性 A をもつとき　　$X_k=1$，

特性 A をもたないとき　$X_k=0$

$(k=1,\ 2,\ \cdots\cdots,\ n)$

「もつか」、「もたないか」の二択で決まるので、二項分布に従うことになる

このとき，$T=X_1+X_2+\cdots\cdots+X_n$ を考えると，T は大きさ n の標本の中で特性 A をもつものの個数を表す確率変数であり，T は二項分布 $B(n,\ p)$ に従う。

和がそのまま特性 A をもつ要素の個数になる

$q=1-p$ とすると，n が大きいとき，

T は近似的に正規分布 $N(np,\ npq)$ に従う。

二項分布のときの説明と同じ

統計的な推測【標本比率と正規分布】 p.96〜97

青チャートB: 例題 86〜87

特性 A の標本比率 R は，これらのうち値が 1 であるものの割合であるから

$$R=\frac{X_1+X_2+\cdots\cdots+X_n}{n}$$

であり，標本平均 $\overline{X}=\frac{T}{n}=\frac{1}{n}(X_1+X_2+\cdots\cdots+X_n)$ であるから

これは，標本平均 $\overline{X}$ に他ならない。よって

$$\underline{R=\overline{X}}=\frac{T}{n}=\frac{1}{n}(X_1+X_2+\cdots\cdots+X_n)$$

であり，**R は標本平均 $\overline{X}$ と同様**に確率変数で

母集団から大きさ n の無作為標本を抽出し，それらの変量 x の値を X_1，X_2，……，X_n とするとき，$\overline{X}=\frac{X_1+X_2+\cdots\cdots+X_n}{n}$ を **標本平均** という

標本比率 $R=\frac{T}{n}$ は近似的に

正規分布 $N\left(\frac{np}{n},\ \frac{npq}{n^2}\right)$ すなわち $N\left(p,\ \frac{pq}{n}\right)$ に従う。

確率変数 X が正規分布 $N(m,\ \sigma^2)$ に従うとき$aX+b$ は，正規分布 $N(am+b,\ a^2\sigma^2)$ に従う

（$a=\frac{1}{n}$ とみる）

標本比率だけでなく，一般の標本平均 $\overline{X}$ の分布についても，次の法則が成り立つことが知られている。

> **標本平均の分布**
>
> 母平均 m，母標準偏差 σ の母集団から大きさ n の無作為標本を抽出するとき，
> 標本平均 $\overline{X}$ は，n が大きいとき，
>
> 近似的に正規分布 $N\left(m,\ \frac{\sigma^2}{n}\right)$ に従うとみなすことができる。

□標本平均の分布と正規分布

> **標本平均の分布**
>
> 母平均 m， 母標準偏差 σ の母集団から抽出された大きさ n の無作為標本について，
> 標本平均 $\overline{X}$ は，n が大きいとき，近似的に正規分布 $N\left(m,\ \frac{\sigma^2}{n}\right)$ に従うとみなすことができる。

この考え方を**「中心極限定理」**という 。簡単に言うと「母集団がどんな分布であっても、標本の数が大きくなると標本平均は正規分布に近似できる」ということである。

<u>母集団分布が正規分布のときは</u>，n が大きくなくても，<u>常に $\overline{X}$ は正規分布 $N\left(m,\ \frac{\sigma^2}{n}\right)$ に従う。</u>

正規分布 $N\left(m,\ \frac{\sigma^2}{n}\right)$ の標準偏差は $\frac{\sigma}{\sqrt{n}}$ であり，

「標本平均の分布」で述べた標本平均 $\overline{X}$ に対して，

母集団から大きさ n の無作為標本を抽出すると、この標本平均 $\overline{X}$ は n が大きいとき近似的に正規分布 $N\left(m,\frac{\sigma^2}{n}\right)$ に従うと見なすことができる

$Z=\dfrac{\overline{X}-m}{\frac{\sigma}{\sqrt{n}}}$ は，<u>n が大きいとき，近似的に標準正規分布 $N(0,\ 1)$</u> に従う。

$\overline{X}$ を標準化：標本の算術平均と母平均との誤差の確率分布が中心極限定理により近似的に０になる

例題９） 母平均 50，母標準偏差 20 をもつ母集団から，大きさ 100 の無作為標本を抽出するとき，その標本平均 $\overline{X}$ が 54 より大きい値をとる確率を求めよ。

考え方　$\overline{X}$ は近似的に正規分布 $N\left(50,\ \dfrac{20^2}{100}\right)$ に従う。

標準正規分布 $N(0,\ 1)$ に従う確率変数 Z に直して考える。

標本平均と正規分布

中心極限定理

解答　標本の大きさは $n=100$，母平均 $m=50$，母標準偏差は $\sigma=20$，

標本平均 $\overline{X}$ の分散は $\dfrac{\sigma^2}{n}=\dfrac{20^2}{100}=4$　標準偏差は $\dfrac{\sigma}{\sqrt{n}}=\dfrac{20}{10}=2$ であるから

この標本平均 $\overline{X}$ は近似的に正規分布 $N\left(50,\ \dfrac{20^2}{100}\right)$，すなわち $N(50,\ 4)$ に従う。

また，$Z=\dfrac{\overline{X}-50}{2}$ は近似的に標準正規分布 $N(0,\ 1)$ に従う。

$\overline{X}=54$ のとき　$Z=\dfrac{54-50}{2}=2$

よって

$$\begin{aligned} P(\overline{X}>54) &= P(Z>2) \\ &= 0.5-p(2) \\ &= 0.5-0.4772 \\ &= 0.0228 \end{aligned}$$

中心極限定理により、母集団の分布がどのような分布であっても、母平均 m と母分散 σ^2 が存在すれば、十分な大きさの標本をとるとき、標本平均 $\overline{X}$ はほぼ正規分布に従う。

つまり母平均 m と母分散 σ^2 がわかれば、標本平均の実際の値がどの範囲に何%の確率で現れるかがわかる

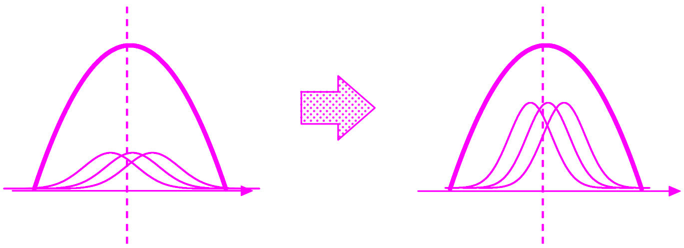
深める　標本の大きさを 100 より大きくするとき，標本平均 $\overline{X}$ が 54 より大きい値をとる確率は，0.0228 と比べてどのようになるだろうか（大きく・変わらず・小さく）

解答　標本の大きさ n が大きくなる

⇒　$\overline{X}$ の標準偏差が小さくなる

⇒　$\overline{X}$ は母平均に，より集中した分布となる

⇒　$\overline{X}$ が母平均から遠い値をとる確率は小さくなる



統計的な推測【標本比率と正規分布】p.96〜97

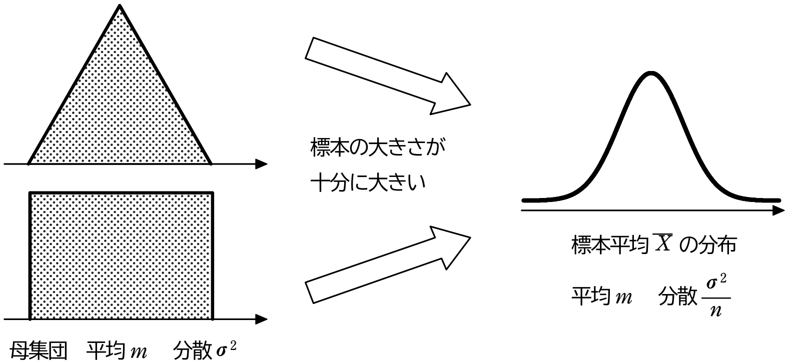
青チャートB: 例題 86〜87

補足 中心極限定理とは？

中心極限定理とは，平均 m ，分散 σ^2 の母集団から大きさ n の標本を無作為に取り出したとき，標本平均 $\overline{X}$ の分布は， n が十分大きいとき，近似的に平均 m ，分散 $\dfrac{\sigma^2}{n}$ の正規分布に従う。

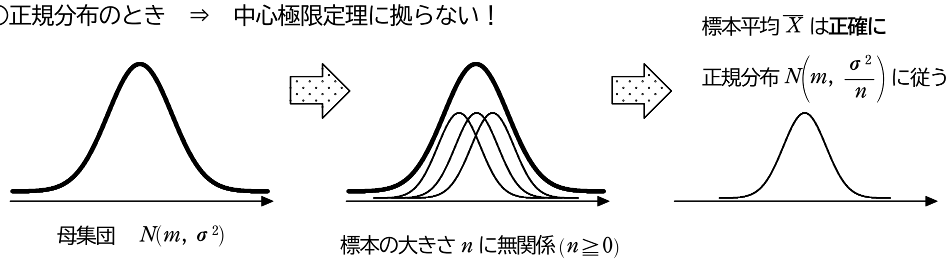
> 補足 確率変数 $X_1,\ X_2,\ \cdots\cdots,\ X_n$ が互いに独立で，平均が m ，分散が σ^2 の同じ分布に従うものとする。このとき， $X_1+X_2+\cdots\cdots+X_n$ を標準化した確率変数
>
> $$\frac{1}{\sqrt{n}\,\sigma}(X_1+X_2+\cdots\cdots+X_n-n\cdot m)=\frac{\sqrt{n}}{\sigma}(\overline{X}-m)$$ の分布は，
>
> $n\longrightarrow\infty$ のとき，標準正規分布 $N(0,\ 1)$ に収束する

注意 もとの母集団が正規分布でなくても，平均と分散をもつ分布であれば，標本平均の分布は標本の大きさが十分に大きければ正規分布に近似できる。

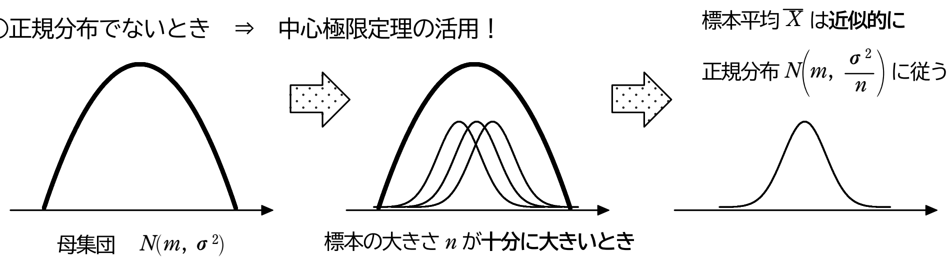


補足 中心極限定理は母集団が正規分布の場合と正規分布以外の場合で，どう違うのか？

○正規分布のとき　⇒　中心極限定理に拠らない！



○正規分布でないとき　⇒　中心極限定理の活用！



> 【まとめ】
>
> 中心極限定理は，母集団が平均と分散をもつ分布であれば，母集団が正規分布でなくても，標本の大きさを十分にとれば，標本平均が従う分布は近似的に正規分布になるという定理である。ちなみに対称な場合は $n\geqq 30$ ，非対称の場合は $n\geqq 50$ が適用できる目安となる。
>
> （母集団が正規分布のときは，中心極限定理によらなくても標本の大きさに関係なく標本平均は正規分布に従う。）