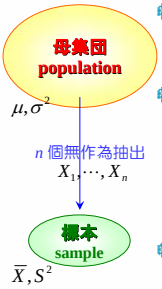
データ分布と予測

母集団と標本

堀田　敬介

2006/11/11,Sat.～

母集団と標本：統計的推論



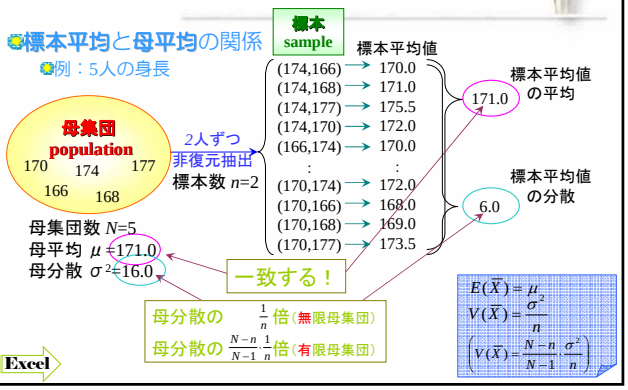
- 母集団の性質を表す数値
  - 母平均：μ
  - 母分散：σ^2　（母標準偏差：σ）
- 母集団からの標本
  - n個のデータを無作為に抽出　$X_1, \cdots, X_n$
- $X_1,\ldots,X_n$ は互いに独立
- 各確率変数 X は母集団と同じ分布に従う
- $X_1,\ldots,X_n$ から作られる確率変数
  - 標本平均：$\bar{X} = \frac{X_1 + \cdots + X_n}{n}$
  - 標本分散：$S^2 = \frac{1}{n}\{(X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2\}$

無作為抽出には乱数などを利用

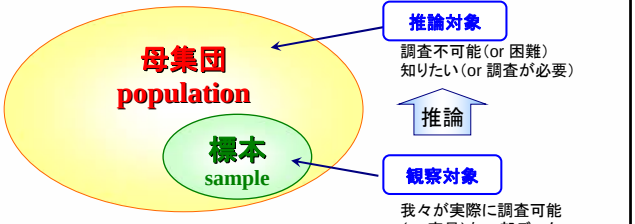
確率変数！
無作為抽出より，実際に取る値は偶然による
〔標本調査は試行である〕

Contents

母集団と標本：標本平均



母集団と標本：統計的推論



- 推測統計学 statistical estimate / statistical inference
  - ○母集団が大きすぎて調査不可能な場合
    - 全国大学生の身長
  - ○全数調査（悉皆調査）が不可能な場合
    - 品質検査
    - 料理の味見

注意：今後特に断りのない限り，無限母集団を考える．

補足：標本平均の平均と母平均・標本平均の分散と母分散の関係（証明）

$$E(\bar{X}) = E\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\cdot n\mu = \mu$$

$$V(\bar{X}) = E(\bar{X} - E(\bar{X}))^2$$
$$= E\left(\frac{X_1 + \cdots + X_n}{n} - E(\bar{X})\right)^2$$
$$= \frac{1}{n^2}E\{X_1 - E(X_1)\} + \cdots + \{X_n - E(X_n)\}^2$$
$$= \frac{1}{n^2}E\left(\{X_1 - E(X_1)\}^2 + \cdots + \{X_n - E(X_n)\}^2 + 2\sum_{i<j}(X_i - E(X_i))(X_j - E(X_j))\right)$$
$$= \frac{1}{n^2}\left\{\sum_{i=1}^{n} E(X_i - E(X_i))^2 + 2\sum_{i<j}(X_i - E(X_i))(X_j - E(X_j))\right\}$$
$$= \frac{1}{n^2}\left\{\sum_{i=1}^{n} V(X_i) + 2\sum_{i<j} Cov(X_i, X_j)\right\}$$
$$= \frac{1}{n^2}\left\{n\sigma^2 - 2\cdot\frac{n(n-1)}{2}\cdot\left(-\frac{1}{N-1}\sigma^2\right)\right\}$$
$$= \frac{1}{n}\cdot\frac{N-n}{N-1}\sigma^2$$

$$Cov(X_i, X_j) = E\{(X_i - E(X_i))(X_j - E(X_j))\}$$
$$= E\{(X_i - \mu)(X_j - \mu)\}$$
$$= \frac{1}{N(N-1)}(x_1 - \mu)(x_2 - \mu) + \cdots + \frac{1}{N(N-1)}(x_N - \mu)(x_{N-1} - \mu)$$
$$= \frac{1}{N(N-1)}\{(x_1 - \mu) + \cdots + (x_N - \mu)\}^2 - \{(x_1 - \mu)^2 + \cdots + (x_N - \mu)^2\}$$
$$= \frac{1}{N(N-1)}\left\{\left\{\frac{X_1 + \cdots + X_N}{N} - \mu\right\}^2 - \{(x_1 - \mu)^2 + \cdots + (x_N - \mu)^2\}\right\}$$
$$= \frac{1}{N(N-1)}(0^2 - N\sigma^2) = -\frac{1}{N-1}\sigma^2$$

補足：有限母集団修正

- 母集団が有限の場合
  - 標本平均の分散と母分散の関係は，

$$V(\bar{X}) = \frac{N-n}{N-1} \cdot \frac{\sigma^2}{n}$$

有限修正項

Nが余り大きくない場合や，n/Nが大きい場合

標本数nに比べて母集団の数Nが大きくないとき，有限修正項を考慮する．
無限母集団（Nが十分大きい）時は，有限修正項は1となるので無視して良い．

- 母集団が無限の場合
  - 標本平均の分散と母分散の関係は，

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

演習１：標本平均

1．世界に4匹しかいない貴重な昆虫がいる．その集団を母集団としよう．
- 神様はこの4匹の全長を全て知っており，それぞれ (2, 6, 7, 5) である．
- 神様は母平均の値を求めた．いくつか？ $\mu = ?$
- 神様は母分散の値を求めた．いくつか？ $\sigma^2 = ?$

2．探検家は2匹捕まえる．それが標本となる．
- 各探検家は重複なく2匹を捕まえた．
  （つまり，非復元抽出で2匹捕らえ，全長測定後放す）
- 各探検家は自分が捕まえた2匹の標本の平均値を求めた．
- それぞれ，いくつか？ 全ての組合せについて計算せよ．$\bar{X} = ?$

3．１と２の結果から，$E(\bar{X}) = \mu$ と $V(\bar{X}) = \frac{N-n}{N-1} \cdot \frac{\sigma^2}{n}$ が成立していることを確認しよう．
ただし，Nは母集団の大きさ，nは標本の大きさである．

母集団と標本：標本平均

- なぜ「標本平均の分散」の方が，「母分散」より小さくなるのか？
  - 例：5人の身長
    174,166,168,177,170

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

実際には 1/n 程度小さい

「**標本平均値**の**散らばり具合**」の方が，
「**母集団**の**散らばり具合**」より**小さい**！

母集団	標本平均値
166	
167	● ●
168	● ●
169	● ●
170	● ●
171	● ●
172	● ● ● ●
173	● ●
174	● ●
175	
176	● ●
177	

母集団と標本：大数の法則

- 「標本平均 $\bar{X}$ の期待値は母平均 μ に等しい」 $E(\bar{X}) = \mu$
- 「標本平均 $\bar{X}$ の分散は母分散 σ^2 の1/nに等しい」$V(\bar{X}) = \frac{\sigma^2}{n}$

有限母集団の場合 $\frac{N-n}{N-1} \cdot \frac{1}{n}$ 倍

大数の法則

標本数 n が大きくなるにつれて，標本平均

$$\bar{X} = \frac{1}{n}(X_1 + \cdots X_n)$$

が母平均 μ に近い値をとる確率は 1 に近づく．

標本数 n が十分大きければ，標本は母集団を正しく表すと考えてもよいでしょう．

母集団と標本：標本平均（まとめ）

- 標本平均 $\bar{X}$

母集団からn個
無作為抽出

$$\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$$

- $X_1, \ldots, X_n$はそれぞれ確率変数
- それから作られる標本平均も確率変数

- 注意：「標本平均」と「標本平均値」は意味が違う
  - 標本平均 … 上で定義される確率変数
  - 標本平均値 … 確率変数「標本平均」が標本ごとに実際に取る値
- 「標本平均 $\bar{X}$ の期待値は母平均 μ に等しい」 $E(\bar{X}) = \mu$
- 「標本平均 $\bar{X}$ の分散は母分散 σ^2 の1/nに等しい」$V(\bar{X}) = \frac{\sigma^2}{n}$

有限母集団の場合：$V(\bar{X}) = \frac{N-n}{N-1} \cdot \frac{\sigma^2}{n}$

補足：大数の法則

大数の法則

$$P\left(\left|\bar{X} - \mu\right| < \varepsilon\right) \to 1 \quad (n \to \infty)$$

- 証明はチェビシェフの不等式 $P\left(\left|\bar{X} - \mu\right| > k\sigma\right) \le 1/k^2$ から

∵）$X_1, \ldots, X_n$ は独立で，同じ分布に従う

$\rightarrow E(X_i) = \mu, V(X_i) = \sigma^2 (i = 1, \cdots, n)$

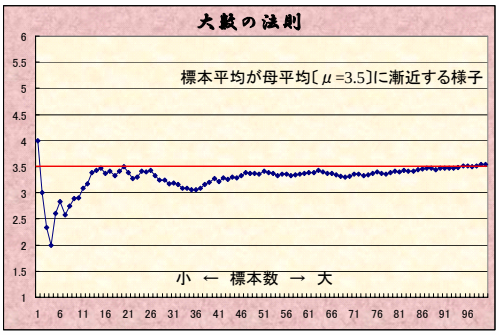
$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ とすると $E(\bar{X}) = \mu, V(\bar{X}) = \frac{\sigma^2}{n}$

ここで，チェビシェフの不等式から，$k\sigma := \varepsilon$ とおくと （$\sigma^2 := \sigma^2/n$）

$P\left(\left|\bar{X} - \mu\right| > \varepsilon\right) \le \sigma^2 / n\varepsilon^2 \to 0 \quad (n \to \infty)$ ∎

母集団と標本：大数の法則

- 大数の法則例：サイコロを振って出た目の平均〔μ=3.5〕



Excel

補足：中心極限定理

中心極限定理

$n \to \infty$のとき，

$$P\left(a \le (X_1 + \cdots + X_n - n\mu)/\sqrt{n}\sigma \le b\right) \to \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

が成りたつ．言い換えると，

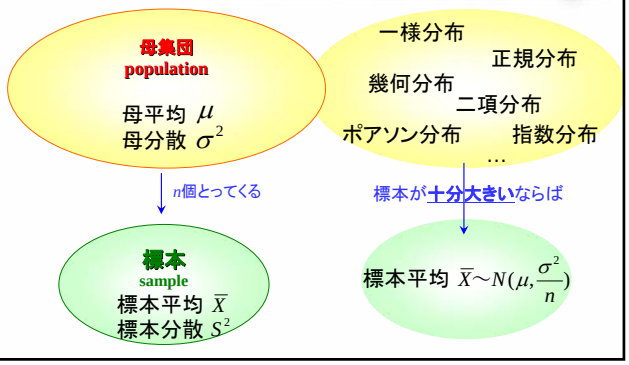
$$P\left(a \le \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \le b\right) \approx \phi(b) - \phi(a)$$

としてよいということ．
（右辺のϕは標準正規分布の累積分布関数）

標本分布：母集団が正規分布の時

- 標本平均 $\overline{X}$ はどんな確率分布に従うのか？
  - 母集団が，母平均μ，母分散σ^2の正規分布に従う
  - その母集団から無作為に抽出された大きさ n の標本（ n 個の互いに独立な確率変数 $X_1,\ldots,X_n$ ）もそれぞれ同じ正規分布 $N(\mu,\sigma^2)$ に従う
  - 標本平均 $\overline{X}$ は正規分布 $N(\mu,\sigma^2/\mathrm{n})$ に従う

中心極限定理



標本分布：母集団が正規分布でない時

- 標本平均 $\overline{X}$ はどんな確率分布に従うのか？
  - 標本数 n が十分大きければ…

中心極限定理

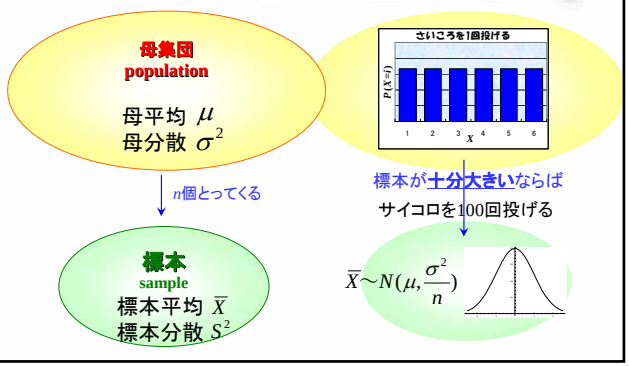
$X_1,\cdots,X_n$

母平均μ，母分散σ^2の母集団から大きさnの標本を無作為に抽出した時，nが十分大きければ，母集団の従う確率分布に関係なく，標本平均$\overline{X}$は期待値μ，分散σ^2/nの正規分布$N(\mu,\sigma^2/n)$に従うとみなすことができる

$$\begin{cases} X_1 + \cdots + X_n \sim N(n\mu, n\sigma^2) \\ \overline{X} = \frac{1}{n}(X_1 + \cdots + X_n) \sim N(\mu, \frac{\sigma^2}{n}) \end{cases}$$

nが十分大きければ，母集団分布が何であっても，和の確率分布$X_1+\ldots+X_n$の形は，大体正規分布と考えて良い！

中心極限定理



中心極限定理の応用

平均20,000回で，400回は±2%の誤差！ありふれたことだろう．．．

- 例題： 表裏が等確率で出るコインを40,000回投げるとき，表が20,400回より多いか，19,600回より少なく出る確率は？
  - i 回目：X_i=1,0　（1：表，0：裏）
  - 表の出る回数：$X=X_1+X_2+\cdots+X_n$

二項分布 $Bi(40000, 1/2)$ に従う

$$f(x)={}_nC_x p^x(1-p)^{n-x} \quad (x=0,1,\cdots,n)$$
$$E(X)=np, V(X)=np(1-p)$$

つまり $P(X>20400)+P(X<19600)$ はいくつか?

$$1-\sum_{x=19600}^{20400} {}_{40000}C_x (1/2)^x (1/2)^{40000-x}$$ を計算すればよい！

ところが ${}_{40000}C_x$ を計算するのは困難！

例えば，Excel2003で ${}_{40000}C_{19600}$ を計算すると，…

計算不能！

#NUM! =COMBIN(40000,19600)

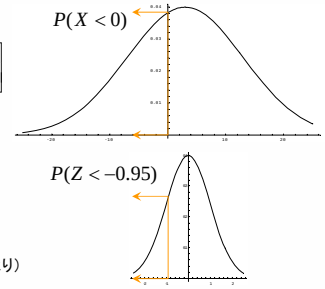
標本分布：標準化と標準正規分布

- 例題：確率変数X はある株式の利回り(%)で，正規分布 $N(3,10)$に従う．この株式への投資が損となる確率は？

$$P(X<0)$$
$$=P(\mu+\sigma Z<0)\left[\because Z=\frac{X-\mu}{\sigma}\right]$$
$$=P(Z<\frac{0-\mu}{\sigma})$$
$$=P(Z<\frac{0-3}{\sqrt{10}}=-0.94868)$$
$$\approx P(Z<-0.95)=0.17106$$

標準正規分布表から

=0.171391 (Excel関数 NORMDISTより)



中心極限定理の応用

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

- nが十分大きければ，二項分布は正規分布で近似できる！

$$P\left(a\le\frac{(X_1+\cdots+X_n-n\mu)}{\sqrt{n}\sigma}\le b\right)\rightarrow\int_a^b\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx$$

各X_i は二項分布 $Bi(1, 1/2)$ に従う

$$\begin{cases}\mu=E(X_i)=n_ip_i=1\times1/2=1/2,\\ \sigma^2=V(X_i)=n_ip_i(1-p_i)=1\times1/2\times1/2=1/4\end{cases}$$

$$\rightarrow\begin{cases}n\mu=40{,}000\times1/2=20{,}000\\ \sqrt{n}\sigma=\sqrt{40{,}000}\times1/4=100\end{cases}$$

$$\rightarrow P(19{,}600\le X_1+\cdots+X_{40{,}000}\le 20{,}400)$$
$$=P\left(-4\le\frac{X_1+\cdots+X_{40{,}000}-20{,}000}{100}\le4\right)$$
$$=\phi(4)-\phi(-4)$$
$$=0.9999\cdots$$

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

故に，求める確率は 1%未満．殆ど起こりえないこと！

標本分布：標本平均の標準化

- 平均 μ，分散 σ^2の確率変数 X　の標準化

$$X\rightarrow Z=\frac{X-\mu}{\sigma}$$

- 平均 μ，分散 σ^2/nの標本平均 $\bar{X}$　の標準化

$$\bar{X}\rightarrow Z=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$$

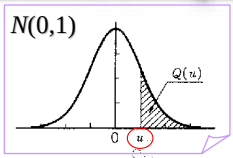
標本から母平均 μ を推定
「Z推定」「Z検定」

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

中心極限定理の応用

- 標準正規分布表の読み方

$Q(u)=1-\phi(u)=\int_u^\infty\phi(u)du$



$P(X\ge u)$

小数第2位

小数第1位

u	.00	.01	.02	.03	.04	.05	.06
.0	.50000	.49601	.49202	.48803	.48405	.48006	.4760
.1	.46017	.45620	.45224	.44828	.44433	.44038	.4364
.2	.42074	.41683	.41294	.40905	.40517	.40129	.3974
.3	.38209	.37828	.37448	.37070	.36693	.36317	.3594
.4	.34458	.34090	.33724	.33360	.32997	.32636	.3227
.5	.30854	.30503	.30153	.29806	.29460	.29116	.2877
.6	.27425	.27093	.26763	.26435	.26109	.25785	.2546
.7	.24196	.23885	.23576	.23270	.22965	.22663	.2236

例題：出展 技術評論社「確率・統計の仕組みがわかる本」例7.2

【問題】小学生の1ヶ月の小遣いが，平均2250円，標準偏差360円です．このとき，ランダムに選んだ36人の小学生の小遣い平均が2400円を超える確率は？

- 解答：母集団分布不明だが，n=36人は十分大きいので，中心極限定理から正規分布と仮定．標本平均 $\bar{X}$ の分布は
  平均：2250円（母集団と同じ），標準偏差：$\sqrt{\frac{\sigma^2}{n}}=\frac{360}{\sqrt{36}}=60$
  の正規分布に従う．これより標準化して，

$$Z=\frac{\bar{X}-2250}{60}$$

したがって $P(\bar{X}>2400)$

$$=P(60Z+2250>2400)$$
$$=P(Z>2.5)\cong0.0062$$

∴ 答え 0.62%

Coffee Break!

10^{100}と100^{10}はどっちが大きい?

- どちらが大きい? 計算して教えてよ!
  - 10^{100} = ?
  - 100^{10} = ?
- どちらが大きい?
  - 10^{100} = ?
  - 100! = ?
- スターリングの公式

$$N! \approx (N/e)^N \sqrt{2\pi N}$$

$$\left(\lim_{N\to+\infty} \frac{N!}{(N/e)^N\sqrt{2\pi N}} = 1\right)$$

充分大きなNについて，Nの階乗の近似値を与える

累乗の計算も大変だけど，階乗の計算はとんでもなく大変ね!

補足：標本分散の平均と母分散の関係（証明）

$$\begin{aligned}
E(S^2) &= E\left(\frac{1}{n}\left\{(X_1-\bar{X})^2+\cdots+(X_n-\bar{X})^2\right\}\right)\\
&= \frac{1}{n}E\left(\left\{(X_1-\mu)-(\bar{X}-\mu)\right\}^2+\cdots+\left\{(X_n-\mu)-(\bar{X}-\mu)\right\}^2\right)\\
&= \frac{1}{n}E\left(\sum_{i=1}^{n}\left\{(X_i-\mu)^2-2(X_i-\mu)(\bar{X}-\mu)+(\bar{X}-\mu)^2\right\}\right)\\
&= \frac{1}{n}\left(\sum_{i=1}^{n}E(X_i-\mu)^2-2E\left(\sum_{i=1}^{n}(X_i-\mu)(\bar{X}-\mu)\right)+\sum_{i=1}^{n}E(\bar{X}-\mu)^2\right)\\
&= \frac{1}{n}\left(\sum_{i=1}^{n}V(X_i)-2E\left(n(\frac{X_1+\cdots+X_n}{n}-\mu)(\bar{X}-\mu)\right)+nE(\bar{X}-\mu)^2\right)\\
&= \frac{1}{n}\left(n\sigma^2-2nE(\bar{X}-\mu)^2+nE(\bar{X}-\mu)^2\right)\\
&= \sigma^2-V(\bar{X})\\
&= \sigma^2-\frac{N-n}{N-1}\cdot\frac{1}{n}\sigma^2\\
&= \frac{N}{N-1}\cdot\frac{n-1}{n}\sigma^2
\end{aligned}$$

標本分布：標本分散

- 母集団からのn個の標本 $X_1,\cdots,X_n$ について，以下の確率変数を標本分散S^2という

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

注意） 「標本分散値」は確率変数「標本分散」が標本毎に実際に取る値

補足：有限母集団修正

- 母集団が有限の場合
  - 標本分散の平均と母分散の関係は，

$$E(S^2)=\frac{N}{N-1}\cdot\frac{n-1}{n}\sigma^2$$

有限修正項

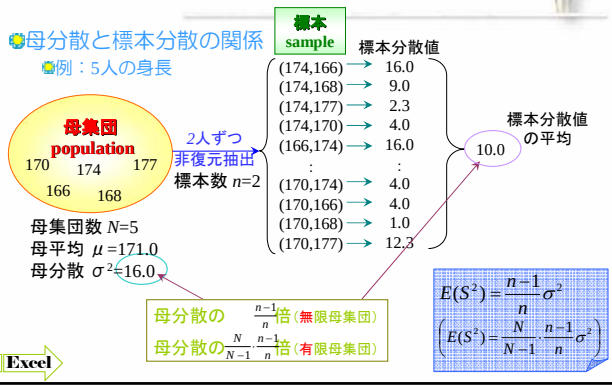
母集団の要素数Nが大きくないとき，有限修正項を考慮．
無限母集団（Nが十分大きい）時は，有限修正項は1となるので無視．

- 母集団が無限の場合
  - 標本分散の平均と母分散の関係は，

$$E(S^2)=\frac{n-1}{n}\sigma^2$$

母集団と標本：標本分散値の平均

- 母分散と標本分散の関係
  - 例：5人の身長



母集団と標本：標本分散（まとめ）

- 標本分散 S^2

母集団からn個
無作為抽出

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

- $X_1,\ldots,X_n$はそれぞれ確率変数
- それから作られる標本平均も確率変数
- よって，それから作られる標本分散も確率変数

- 注意：「**標本平均の分散** $V(\bar{X})$」と「**標本分散の平均** $E(S^2)$」を混同しないこと！

「標本分散値の平均」と「母分散」の関係

$$E(S^2)=\frac{n-1}{n}\sigma^2$$

有限母集団の場合：

$$E(S^2)=\frac{N}{N-1}\cdot\frac{n-1}{n}\sigma^2$$

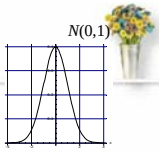
演習２：標本分散

1．世界に4匹しかいない貴重な昆虫がいる．その集団を母集団としよう．
- 神様はこの4匹の全長を全て知っており，それぞれ (2, 6, 7, 5) である．
- 神様は母分散の値を求めた．いくつか？ $\sigma^2 = ?$

2．探検家は2匹捕まえる．それが標本となる．
- 各探検家は重複なく2匹を捕まえた．
（つまり，非復元抽出で2匹捕らえ，全長測定後放す）
- 各探検家は自分が捕まえた2匹の標本の分散の値を求めた．
- それぞれ，いくつか？全ての組合せについて計算せよ．$S^2 = ?$

3．１と２の結果から，$E(S^2) = \frac{N}{N-1} \cdot \frac{n-1}{n} \sigma^2$ が成立することを確認しよう．
ただし，Nは母集団の大きさ，nは標本の大きさである．

χ^2分布とは？



- 標準正規分布 $N(0,1)$ に従う，互いに独立な n個の確率変数 $Z_1,\ldots,Z_n$ を考える

$$\chi^2 = Z_1^2 + \cdots + Z_n^2$$ ← 二乗和をとる

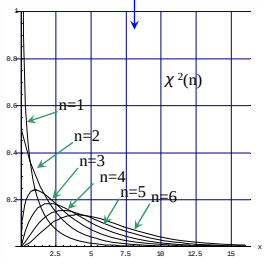
新たな確率変数

⇩

この確率変数 χ は，自由度n の χ^2分布に従う！

互いに自由に値をとることが出来る確率変数の個数

標本から母分散 σ^2を推定
「カイ二乗推定」「カイ二乗検定」



標本分布：標本分散と不偏分散

- 標本分散 S^2

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

- 不偏分散 s^2 ← この標本分散は，母分散 σ^2の不偏推定量

$$s^2 = \frac{1}{n-1}\left\{(X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2\right\}$$

$$E(S^2) = \frac{n-1}{n}\sigma^2 \qquad E(s^2) = \sigma^2$$

有限母集団の場合：
$$E(S^2) = \frac{N}{N-1} \cdot \frac{n-1}{n}\sigma^2 \qquad E(s^2) = \frac{N}{N-1}\sigma^2$$

Nが充分大きいならば，
$N/(N-1)$は1と考えて良い．

標本分布：標本分散

例題：ある正規母集団の母平均 μ=50,母分散 σ^2=25とする．ここから大きさ10の標本をとったとき，標本分散が50を超える確率は？

$$P(S^2 > 50)$$
$$= P(\frac{\chi^2\sigma^2}{n} > 50) \left[\because \chi^2 = \frac{nS^2}{\sigma^2}\right]$$
$$= P(\chi^2 > 50\frac{n}{\sigma^2})$$
$$= P(\chi^2 > 50\frac{10}{25} = 20) \in (0.025, 0.010)$$

=0.017912
(Excel関数 CHIDISTより)

自由度9の χ^2分布表から
$P(\chi^2(9)>19.0228) = 0.025$
$P(\chi^2(9)>21.6660) = 0.010$

標本分布：標本分散の従う確率分布

標本分散S^2 はどんな確率分布に従うのか？

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

$$\rightarrow \frac{n}{\sigma^2} \cdot S^2 = \frac{n}{\sigma^2} \cdot \frac{1}{n}\left\{(X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2\right\}$$

$$= \left(\frac{X_1 - \bar{X}}{\sigma}\right)^2 + \cdots + \left(\frac{X_n - \bar{X}}{\sigma}\right)^2$$

χ^2分布に従う ← n個の$N(0,1)$に従う確率変数の二乗和

$\sum(X_i - \bar{X}) = 0$
という制限のため，
自由に動ける変数
の個数はn-1となる．

- 母集団が正規分布 $N(\mu, \sigma^2)$ に従うとみなせる時，確率変数 $\frac{nS^2}{\sigma^2}$ は自由度n-1のχ^2(n-1)分布に従う．

t分布とは？

ギネスビールとは？
1756年創業のビール醸造会社
〔ダブリン（アイルランド）〕
ギネスビール（黒スタウト）を製造

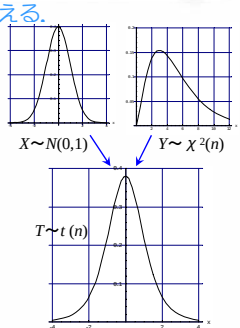
- 2個の互いに独立な確率変数 X, Y を考える．
  - X ： 標準正規分布 $N(0,1)$ に従う
  - Y ： 自由度 n のχ^2分布 $\chi^2(n)$ に従う

$$T := \frac{X}{\sqrt{Y/n}}$$

新たな確率変数

確率変数 T は，自由度 n の t 分布に従う！

Student の t分布
ゴセット (1876-1937)



ビール会社ギネスGuinessでビールの品質管理
標本が小さいとき，分散の値が(正規分布では上手くいかない…)
→ t 分布の発見（"Student"[W.S.Gossett] 'The probable error of a mean',Biometrika vol.6,1908）

標本分布：標本平均と標本分散

- 標本平均 $\bar{X}$ の標準化

$$\bar{X} \to Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$$

標準正規分布 $N(0,1)$ に従う

- 標本分散S^2にn/σ^2を掛けた確率変数

$$nS^2/\sigma^2$$

自由度 n-1 の χ^2分布 に従う

$$T = \left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)\cdot\frac{1}{\sqrt{\frac{1}{n-1}\cdot\frac{nS^2}{\sigma^2}}} = \frac{\bar{X}-\mu}{S/\sqrt{n-1}}$$

自由度 n-1 の t分布 に従う

標本から母平均μを推定
「t推定」「t検定」

補足：必要な標本の大きさ

- 例題：大きさ6000万の母集団の母比率pを，95%の確率で誤差が0.05以下になるようにしたい．必要な単純無作為抽出の大きさnはいくらか？ $|\bar{X}-\mu| \le 0.05$

Nが十分大きいので，

$$n \ge \frac{(1.96)^2\sigma^2}{\varepsilon^2} \ge \frac{(1.96)^2}{4\varepsilon^2} = \frac{(1.96)^2}{4(0.05)^2} \approx 384.16$$

$$\left(\sigma^2 = p(1-p) = -\left(p-\frac{1}{2}\right)^2 + \frac{1}{4} \le \frac{1}{4}\right)$$

σ^2の最大値は 0.25(p=0.5の時)

演習３：

- 2006年晩秋ゲーム機商戦たけなわ，ソニーのPlayState3と任天堂のWiiが発売された．ゲーム機を購入に来た客10人に聞いたところ，次のような結果を得た．（ただし，必ずどちらかを購入し，どちらも買わない客はいないとする）このとき，PS3を購入する比率（標本比率）を計算せよ．

PS3 PS3 Wii PS3 Wii Wii Wii PS3 Wii Wii

- 昨シーズン打率2割8分の打者が，今シーズンも同じ確率でヒットを打つものとし，450打数であるとすると，3割打者になれる確率はどれぐらいか？ また，この打者が，確率0.2以上で3割打者になろうとすると，打数はどのぐらいでなければならないか？

(出展：「統計学入門」東京大学出版会 p.173 練習問題8.3)

$X_1 + \cdots + X_n \sim Bi(n, 0.28)$ だが，計算は大変だし，nが未知

$X_i \sim Bi(1, 0.28)$ $(i = 1, \cdots, 450)$ のとき，

$P(X_1 + \cdots + X_{450} \ge 450 \times 3/10)$?

$P(X_1 + \cdots + X_n \ge 0.3n) \ge 0.2$?

$$\left(P\left(a \le \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le b\right) \approx \phi(b) - \phi(a) \quad \bar{X} \sim N(\mu, \sigma/\sqrt{n})\right)$$

参考文献

- 東京大学教養学部統計学教室編「統計学入門」東京大学出版会（1991）
- 東京大学教養学部統計学教室編「自然科学の統計学」東京大学出版会（1992）
- 鈴木達三・高橋宏一「標本抽出の計画と方法」放送大学（1991）
- 永田靖「サンプルサイズの決め方」朝倉書店（2003）
- 白石修二「例題で学ぶ Excel統計入門」森北出版（2001）
- 村上雅人「なるほど統計学」海鳴社（2002）
- 丹慶勝市「図解雑学 統計解析」ナツメ社（2003）
- 高橋信[著]・トレンドプロ[マンガ]「マンガでわかる統計学」オーム社（2004）

補足：必要な標本の大きさ

- 標本平均の実現値を母平均の推定値とする場合

$$|\bar{X}-\mu| \le \varepsilon \quad \left(\bar{X} \sim N(\mu, \sigma^2/n)\right)$$

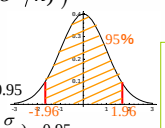
誤差　許容誤差

今，標本平均の従う正規分布から考えて

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1) \Rightarrow P(-1.96 \le \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le 1.96) = 0.95$$

$$\Leftrightarrow P(-1.96\frac{\sigma}{\sqrt{n}} \le \bar{X}-\mu \le 1.96\frac{\sigma}{\sqrt{n}}) = 0.95$$

$$\Leftrightarrow P(|\bar{X}-\mu| \le 1.96\frac{\sigma}{\sqrt{n}}) = 0.95$$



参考：
有限母集団の場合

$$n \ge \frac{1}{\frac{\varepsilon^2}{4\sigma^2}\left(1-\frac{1}{N}\right)+\frac{1}{N}} \quad \left(S^2 = \frac{N-n}{N-1}\cdot\frac{\sigma^2}{n}\right)$$

従って，許容誤差をεとしたとき

$$\Rightarrow 1.96\frac{\sigma}{\sqrt{n}} \le \varepsilon$$

$$\Leftrightarrow n \ge \frac{(1.96\sigma)^2}{\varepsilon^2}$$

定められた許容誤差$\varepsilon > 0$に対し，母集団の大きさNと母標準偏差σが既知の場合，単純無作為抽出の大きさnを，左不等式を満たすようにとれば，95%以上の確率で，誤差を許容誤差より小さくできる．