第 7 章　標本分布と正規近似

1　統計量と標本分布

統計量（statistic）：実際の標本が抽出される前に，標本の関数として定義される平均や分散などの特性値（抽出される標本に依存して変わりうる確率変数）

統計値：標本が与えられたときに計算される統計量の実現値

標本分布：統計量の確率分布

（注）母集団分布は母集団の確率分布である．しかし，標本分布を標本の確率分布というのは誤り．

2　正規母集団からの標本分布

以下，正規母集団からのランダム標本を考える．

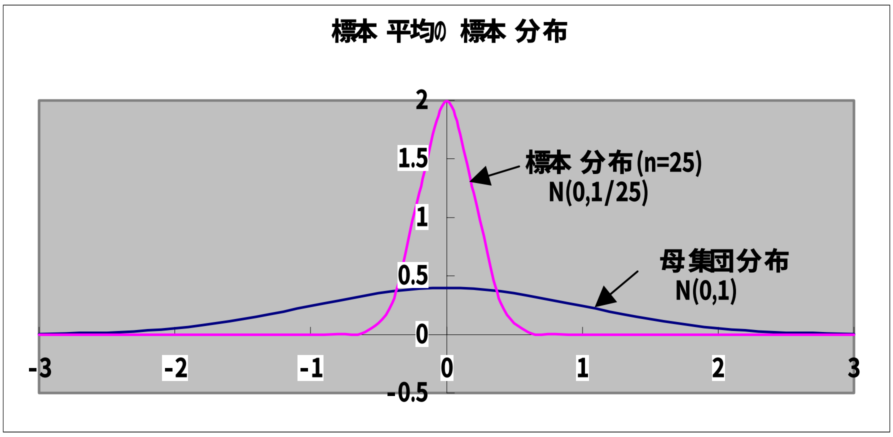
$N(\mu,\ \sigma^2)$ の母集団分布からのサイズ n のランダム標本

$$\Leftrightarrow \quad X_1, \cdots, X_n \ \sim \ i.i.d.N(\mu,\ \sigma^2)$$

$$X_i \ \sim \ i.i.d.N(\mu,\ \sigma^2) \ \Rightarrow \ X_1 + \cdots + X_n \ \sim \ N(n\mu, n\sigma^2)$$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \ \sim \ N(\mu, \frac{\sigma^2}{n})$$

3　和の標本分布



〔例題 7.1〕　正規母集団からのサイズ n のランダム標本から得られる平均と母集団平均の差が，絶対値において母集団標準偏差を超える確率を，n=1，4，9 の場合に求めよ．

（解）求める確率は,

$$P\left(\left|\overline{X}-\mu\right|>\sigma\right)=P\left(\frac{\sqrt{n}\left|\overline{X}-\mu\right|}{\sigma}>\sqrt{n}\right)=P\left(|Z|>\sqrt{n}\right),\qquad Z \sim N(0,1)$$

から計算することができる．

n	1	4	9
確率	0.317	0.0455	0.0027

4　2 乗和の標本分布

χ^2 分布： $Z_1, \cdots, Z_n \sim i.i.d.N(0,1)$ **のとき，**

$W = Z_1^2 + \cdots + Z_n^2$

の標本分布を自由度 n **の** χ^2 **分布といい，** $W \sim \chi^2(n)$ **と表す．**

$E(W) = n, \qquad V(W) = 2n$

標本分散 $\quad S^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(X_i-\overline{X})^2$ **の標本分布**

$X_1, \cdots, X \sim i.i.d.N(\mu, \sigma^2)$ **のとき，**

$$\sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{\sigma}\right)^2=\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1),\qquad Note:\ \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2(n)$$

〔例題 7.3〕 正規母集団からのサイズ n のランダム標本から得られる標本分散と母集団分散の比の分布の上側 5% 点を n=2，5，10，20 の場合に求めよ．

（解）求める分位点は,

$$0.05=P\left(S^2/\sigma^2>x\right)=P\left((n-1)S^2/\sigma^2>(n-1)x\right)$$

となる x であり,

$$(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$$

であることから,

$x=\chi^2(n-1)$ **分布の 上側** 5% **点を** $n-1$ **で割った値**

となる．したがって，付表 3（247 ページ）より求めることができる．

5　比の標本分布

t **分布：**　Z **と** W **が互いに独立で，** Z **〜** $N(0,1)$, W **〜** $\chi^2(n)$ **のとき，**

$$\mathrm{U} = \mathrm{Z}\big/\sqrt{W/n}$$

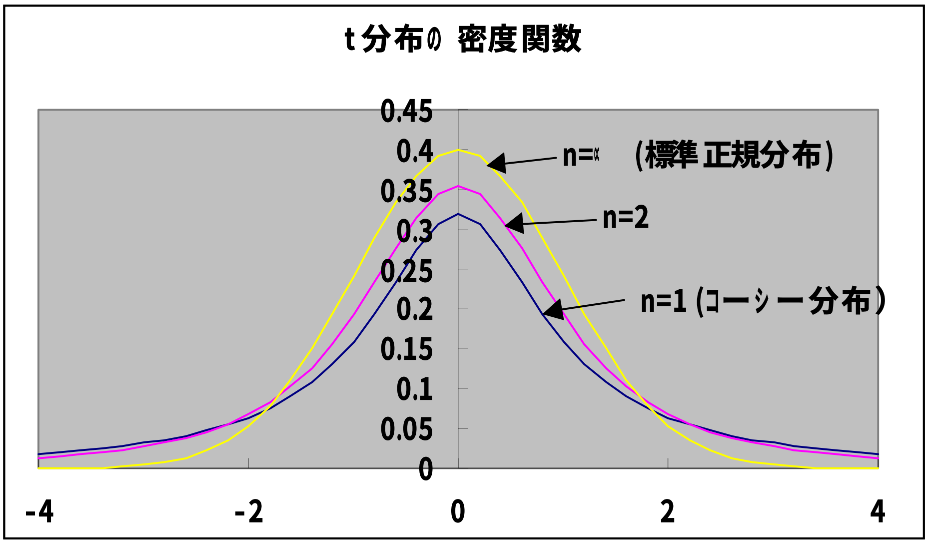
は自由度 n **の** t **分布に従うといい，** U **〜** $t(n)$ **と表す．**

$E(U)=0 \quad (n \geq 2) \qquad V(U)=n/(n-2) \quad (n \geq 3)$

例題 7.4　$t(n)$ **分布の** 95%**，**97.5%**，**99.5% **点を，**
$n=1, 2, 3, 10, 20$**，**30**，**∞ **の場合に求めよ．**

（解）248 ページの**付表** 4 から次の結果を得る．自由度が ∞ の分布は N(0, 1).

自由度	1	2	3	10	20	30	∞
95%	6.314	2.920	2.353	1.812	1.725	1.697	1.645
97.5%	12.706	4.303	3.182	2.228	2.086	2.042	1.960
99.5%	63.657	9.925	5.841	3.169	2.845	2.750	2.576



F 分布： W_1, W_2 **が互いに独立で，**$W_1 \sim \chi^2(m)$, $W_2 \sim \chi^2(n)$ **のとき，**

$$R = \frac{W_1/m}{W_2/n}$$

は，自由度(m, n) **の** F **分布に従うといい，**$R \sim F(m,n)$ **と表す．**

(**例**) $X \sim F(3,10)$ **のとき，上側**5% **点，**1% **点は，付表** 5 **より，それぞれ** 3.71, 6.55 **となる．**

(**問題**) $X \sim t(n)$ **のとき，**$X^2 \sim F(1,n)$ **となることを示せ．**

６　標本分布の正規近似

母集団が正規分布でない場合には，統計量の正確な標本分布を導出することは一般に困難である．このとき考えられる 1 つの方法はコンピュータを使うことである．

コンピュータ・シミュレーションに基づく標本分布：想定された母集団分布に従う確率変数をコンピュータから乱数として発生させることにより，数多くの標本を作り，得られる統計値のヒストグラムを標本分布とみなす方法

〔例題 7.6〕 [0，1] 上の一様母集団からのサイズ 10 のランダム標本の和の標本分布を，コンピュータ・シミュレーションにより求めよ．

（解）まず，コンピュータにより [0，1] 上の一様乱数を独立に 10 個生成して，その和を計算する．同様にして，別の 10 個の一様乱数を独立に生成して，その和を計算する．図 7-6（159 ページ）は，以上の手続きを 1,000 回繰り返して得られたヒストグラムである．これら 1,000 個のデータの平均は 4.99，標準偏差は 0.90 である．標本分布の正確な平均は 5，標準偏差は 0.91 である．

標本分布を導出する別の方法は，統計学の中で一番美しいといわれている次の定理に基づくものである．

中心極限定理

母集団平均 μ, 分散 σ^2 の母集団からのランダム標本の和の分布は，標本サイズが大きくなるとともに，母集団分布に関係なく，正規分布に近づく．すなわち，

$$\sum_{i=1}^{n} X_i \quad \to \quad N(n\mu,\ n\sigma^2)$$

$$\Leftrightarrow \quad \overline{X} \quad \to \quad N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$\Leftrightarrow \quad \frac{\sqrt{n}(\overline{X}-\mu)}{\sigma} \quad \to \quad N(0,1)$$

中心極限定理の実例

図 6-2（124 ページ）：非対称な分布をもつベルヌーイ母集団からの標本和の分布

図 6-6 および図 7-6（135，159 ページ）：水平な密度をもつ一様母集団からの標本和の分布

図 7-3（151 ページ）：カイ 2 乗分布が自由度の増加とともに正規分布に近づく様子

試験の点数の分布：各問ごとの得点の分布は非対称の場合も多いが，それらを加算した合計点は対称な分布になる．

（注意）：中心極限定理は，母集団分散の存在が必要．コーシー分布ではダメ．

(問題) $W \sim \chi^2(n)$ ならば，$(W-n)/\sqrt{2n}$ は，$n \to \infty$ のとき標準正規分布に収束することを示せ．

正規近似　　中心極限定理に基づいて，標本分布や確率の近似値を求めること

〔例題 7.7〕 次の確率を，正規近似を利用して求めよ．

(a) [10，20] 上の一様分布母集団からのサイズ 4 のランダム標本の和が 70 以上となる確率

(b) 成功の確率が 0.3 の独立試行において，20 回のうち 7 回以上成功する確率，および 400 回のうち 140 回以上成功する確率

（解）

(a) $Y = X_1 + X_2 + X_3 + X_4$ とおくと，

$$E(Y) = 4 \times E(X_1) = 60 \qquad V(Y) = 4 \times V(X_1) = 4 \times 100/12 = 100/3$$

したがって， $Y \approx N(60, 100/3)$であり，

$$P(Y \geq 70) = P\left(Z \geq \frac{70-60}{\sqrt{100/3}}\right) = P(Z \geq \sqrt{3}) \cong 0.042 \qquad \text{正確な値} = 0.042)$$

(b) X 〜 $B(20, 0.3)$ のとき，$E(X) = 6, V(X) = 4.2$ であるから，$X \approx N(6, 4.2)$.

$$P(X \geq 7) = P\left(Z \geq \frac{7-6}{\sqrt{4.2}}\right) = P(Z \geq 0.488) \cong 0.31 \qquad \text{正確な値} = 0.392)$$

X〜$B(400, 0.3)$のとき，$E(X) = 120, V(X) = 84$であるから，$X \approx N(120, 84)$.

$$P(X \geq 140) = P\left(Z \geq \frac{140-120}{\sqrt{84}}\right) = P(Z \geq 2.18) = 0.015 \qquad \text{正確な値} = 0.0177)$$

不連続補正　離散分布を連続な正規分布で近似することによる補正（図 7-7 を参照）

X〜$B(20, 0.3)$のとき，

$$P(X \geq 7) = P(X \geq 7 - 0.5) = P\left(Z \geq \frac{6.5-6}{\sqrt{4.2}}\right) = P(Z \geq 0.244) \cong 0.40$$

（正確な値は 0.392）

上の例題 (b) の implication: 試行回数が多くなると，標本平均（結果）が母集団分布の平均（実力）から離れる確率が 0 に収束することを意味する.

大数の法則

平均 μ，分散 σ^2 の母集団からのランダム標本を $X_1, \cdots, X_n$ とする．このとき，標本平均 $\overline{X}$ が母集団平均 μ から離れる確率は n が大きくなるにつれて 0 に近づく.

$\Leftrightarrow$ **任意の正数 ε に対して $n \to \infty$ のとき，** $P\left(\left|\overline{X} - \mu\right| > \varepsilon\right) \to 0$

$\Leftrightarrow$ $\underset{n \to \infty}{p\lim} \overline{X} = \mu$ $\Leftrightarrow$ $\overline{X}$ **は μ に確率収束する**

（問題 1）大数の法則をチェビシェフの不等式（102 ページ）を使って証明せよ,

（問題 2）成功の確率が p，試行回数 n の独立試行において，成功する回数を X とする．このとき，X/n が p に確率収束することを示せ．また，X-np は 0 に確率収束しないことを示せ.