

# 確率・統計 B

## 標本分布

若木宏文

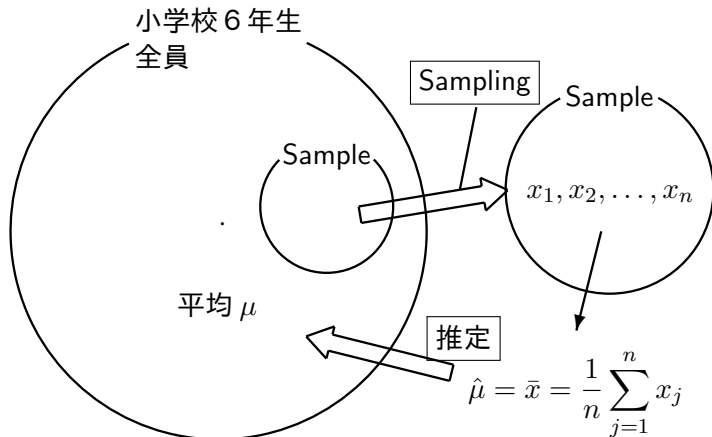
wakaki@math.sci.hiroshima-u.ac.jp

<http://home.hiroshima-u.ac.jp/wakaki/lecture/probstatB16/index.shtml>

2016.10.21

## 母集団分布と統計モデル (1/15)

例 1 (日本の小学 6 年生の平均身長が知りたい)



小学 6 年生全員の身長を調べるのは大変！

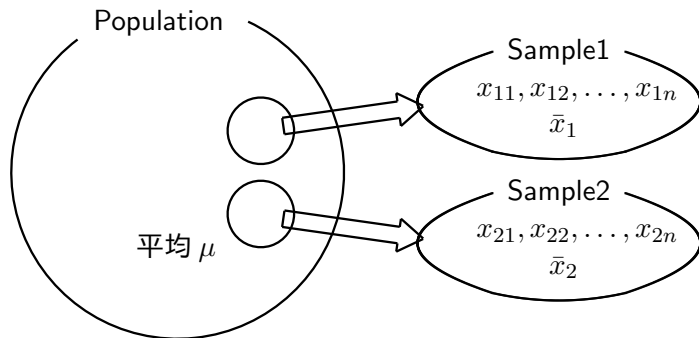
## 母集団分布と統計モデル (2/15)

### 重要な語句

- **母集団 (population)** : 実験や調査の対象
  - 有限母集団 : 対象の個数が有限 (例. 日本人全員)
  - 無限母集団 : 対象の個数が無限 (例. 生産システムでの不良品率の調査)
- **標本抽出 (sampling)** : 対象の一部を取り出すこと
  - **標本 (sample)** : 取り出された対象
  - 標本数 (sample size) : 取り出された対象の数 ( $n$  で表すことがほとんど)
  - 実現値 : 観測された値  $x_1, \dots, x_n$
  - 標本値 : 抽出された対象の特性値 (ここでは  $\hat{\mu} = \bar{x}$ )

## 母集団分布と統計モデル (3/15)

## 疑問点



- $\hat{\mu} = \bar{x}$  は母集団全体での平均  $\mu$  に近い値をとるのか？
- 標本を取り直すと  $\bar{x}$  の値も変わるが、どのくらい値が変わるのか？

## 母集団分布と統計モデル (4/15)

### 定義 6.1 (無作為抽出)

母集団に含まれるどの対象も同じ確率で選ばれるような方法で標本を取り出すこと

1. **復元抽出** : 母集団に含まれる対象をひとつずつ選んでゆくとき, 前に選ばれた対象も何度も選べる抽出法 (標本は独立)
2. **非復元抽出** : どの対象も一度しか選べない抽出法 (標本は独立でないが, 標本数に比べ母集団が大きいときは独立と見なしてもよい)

## 母集団分布と統計モデル (5/15)

### 定義 6.2 (母集団分布)

**母集団分布** : 母集団から対象を1つ無作為抽出して, 特性  $X$  の値を測定するときの, 確率変数  $X$  の分布を**母集団分布**,  
**母平均**, **母分散** :  $X$  の平均を**母平均**, 分散を**母分散**という.  
**実現値** : 実際に測定することによって得られた  $X$  の値  $x$

### 例 2 (不良品の調査)

**母集団** : ある製品の在庫. 不良品が  $100 \times p$  パーセント含まれているとする

母集団から無作為に製品を1個取り出して,  
不良品なら  $X = 1$ , 良品なら  $X = 0$  とする

$X$  の母集団分布は, 繰り返し数 1, 成功確率  $p$  の二項分布  
母平均は  $p$ , 母分散は  $p(1 - p)$

## 母集団分布と統計モデル (6/15)

### 例 2 (不良品の調査 (つづき))

復元抽出で無作為に  $n$  個標本を取り出し, 不良品かどうかを表す変数を

$$X_1, \dots, X_n$$

とすると,  $X_1, \dots, X_n$  は独立に, 母集団分布と同じ分布 ( $B(1, p)$ ) に従う.

非復元抽出で無作為に  $n$  個標本を取り出す場合  
母集団に含まれる製品の個数を  $N$  個とし, その内,  $m$  個が不良品  
であるとすると母集団分布は  $B(1, \frac{m}{N})$

$X_1 = \dots = X_{n-1} = 1$  のときの  $X_n$  の条件付分布:  $B(1, \frac{m-n+1}{N-n+1})$

$X_1 = \dots = X_{n-1} = 0$  のときの  $X_n$  の条件付分布:  $B(1, \frac{m}{N-n+1})$

$n$  が  $N$  に比べて十分小さいときは, 復元抽出の場合とほぼ同じ

## 母集団分布と統計モデル (7/15)

### 母集団分布の分布関数 (有限母集団の場合)

母集団に属する対象の個数を  $N$  個とする

$x_1, \dots, x_N$  : 各対象の特性  $X$  の値

$X$  の母集団分布の分布関数は

$$F(x) = P(X \leq x) = \frac{x_i \leq x \text{ である } x_i \text{ の個数}}{N}$$

⇒ テキスト 113 ページ, 例 6.2



## 母集団分布と統計モデル (8/15)

### 定義 6.3 (標本と母集団分布の数学的な定義)

確率変数  $X_1, \dots, X_n$  が互いに独立で, 同一の確率分布  $P$  に従うとき,  $X_1, \dots, X_n$  は, 分布  $P$  から大きさ  $n$  のランダム標本, あるいは無作為標本といい,  $P$  を母集団分布という.

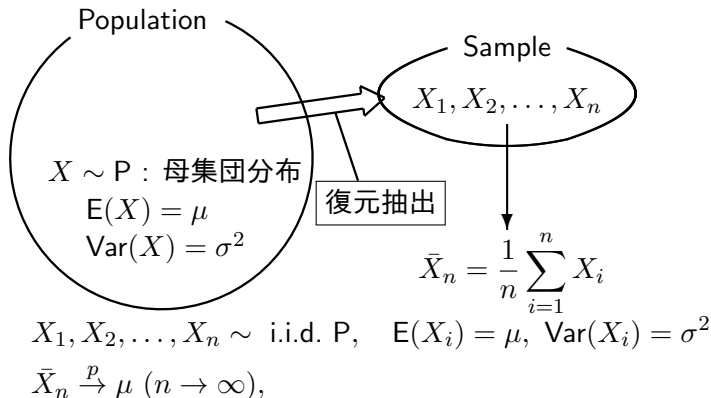
### 注 6.1 (i.i.d.)

$X_1, \dots, X_n$  が独立 (independently) に, 同一の (identically) 分布  $P$  に従う (distributed) とき,

$$X_1, \dots, X_n \sim \text{i.i.d. } P$$

と書く

## 母集団分布と統計モデル (9/15)



$$P(|\bar{X}_n - \mu| \leq \delta) \approx P(|Z| \leq \frac{\sqrt{n}\delta}{\sqrt{\sigma^2}}), \quad Z \sim N(0, 1)$$

$$\left( \frac{\sqrt{n}\delta}{\sqrt{\sigma^2}} = 1.96 \Rightarrow P(|\bar{X}_n - \mu| \leq \delta) \approx 0.95 \right)$$

## 母集団分布と統計モデル (10/15)

### 定義 6.4

(統計モデル) 母集団分布が, ある確率分布の集まり

$$\mathcal{P} = \{P_\theta; \theta \in \Theta\}, \quad \Theta \subset \mathbb{R}^p \text{ (} p \text{ はある自然数)}$$

に属すると想定されるとき,  $\mathcal{P}$  を母集団分布の統計モデルという.  
このとき,  $\theta$  を母数,  $\Theta$  を母数空間とよぶ.

## 母集団分布と統計モデル (11/15)

### 例 6.3 (くり返し測定モデル)

$x_1, \dots, x_n$  : 未知の量  $\mu$  をくり返して測定して得られる観測値

母集団 : 測定対象を無限回測定して得られるすべての値

$\Rightarrow x_1, \dots, x_n$  は, 無限母集団からの大きさ  $n$  の標本  
 $X_1, \dots, X_n$  の実現値

くり返し測定に対してよく想定される統計モデル :

$$X_i = \mu + \varepsilon_i, \quad i = 1, \dots, n$$

## 母集団分布と統計モデル (12/15)

### 例 6.3 (くり返し測定モデル(つづき))

$$X_i = \mu + \varepsilon_i, \quad i = 1, \dots, n$$

#### 繰り返しモデルに関する仮定

- (1)  $\varepsilon_1, \dots, \varepsilon_n$  は互いに独立である.
- (2)  $\varepsilon_1, \dots, \varepsilon_n$  は、それぞれ、同一分布に従う.
- (3)  $E\{\varepsilon_i\} = 0, i = 1, \dots, n$ .
- (4)  $\text{Var}[\varepsilon_i] = \sigma^2, i = 1, \dots, n$ .

ここに、 $-\infty < \mu < \infty, \sigma^2 > 0$ .

- (5)  $\varepsilon_1, \dots, \varepsilon_n$  は正規分布に従う.

$$\Rightarrow X_1, \dots, X_n \sim \text{i.i.d. } N(\mu, \sigma^2).$$

## 母集団分布と統計モデル (13/15)

### 定義 6.5 (経験分布関数)

$X_1, \dots, X_n$  : ある分布  $P$  からの大きさ  $n$  のランダム標本

$x_1, \dots, x_n$  : その実現値

$x_1, \dots, x_n$  から定義される関数

$$F_n(x) = \frac{x_i \leq x \text{ である } x_i \text{ の個数}}{n}$$

を**経験分布関数**という.

$n$  個のランダム標本  $X_1, \dots, X_n$  の経験分布関数は形式上, 全く同じ記号を用いて

$$F_n(x) = \frac{X_i \leq x \text{ である } X_i \text{ の個数}}{n}$$

と定義される.

## 母集団分布と統計モデル (14/15)

確率変数  $X^*$  : 離散型確率変数,  $P(X^* = x_i) = 1/n, i = 1, \dots, n$   
 $\Rightarrow X^*$  の分布関数は  $F_n(x)$

経験分布関数によって定まる離散型分布, すなわち  $X^*$  の分布を **経験分布** とよぶ.

## 母集団分布と統計モデル (15/15)

### 定理 6.1 (経験分布関数の性質)

$X_1, \dots, X_n$  : ある分布  $P$  からの大きさ  $n$  のランダム標本

$F(x)$  : 分布  $P$  の分布関数

$F_n$  : ランダム標本から定義される経験分布関数

$\Rightarrow$

$$(1) nF_n(x) \sim B(n, F(x)).$$

$$(2) F_n(x) \xrightarrow{p} F(x).$$

$$(3) \sqrt{n}\{F_n(x) - F(x)\} \xrightarrow{d} N(0, F(x)(1 - F(x))).$$

(証明は板書で)



# メモ用紙

# メモ用紙

## 統計量の分布 (1/8)

### 定義 6.6

#### 統計量

$t(x_1, \dots, x_n)$  :  $n$  変数のボレル可測関数,  
未知パラメータは含まれないものとする.

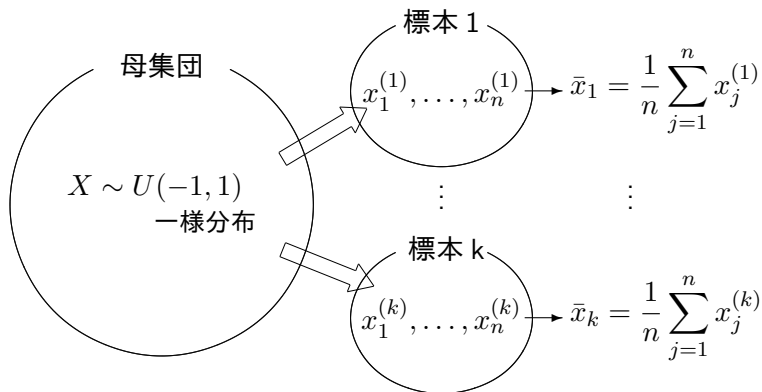
$X_1, \dots, X_n$  : ランダム標本

$\Rightarrow T = t(X_1, \dots, X_n)$  を統計量とよぶ.

統計量の分布や, その特性値が知りたい.

## 統計量の分布 (2/8)

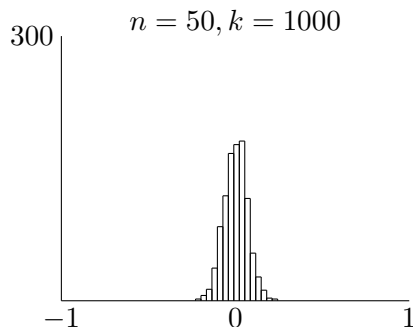
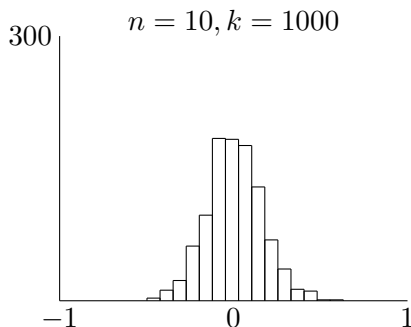
例. 統計量の分布 (数値実験)



## 統計量の分布 (3/8)

例. 統計量の分布 (数値実験)(つづき)

標本平均の分布



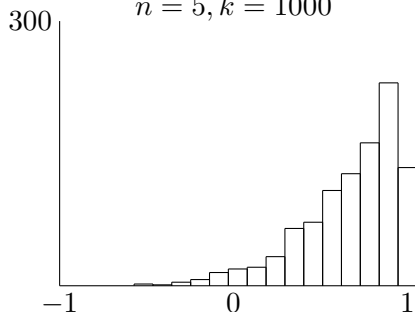
$\bar{x} = 0$  を中心にほぼ左右対称に分布している  
標本数 ( $n$ ) が大きくなると, 分布の「幅」が狭くなる

## 統計量の分布 (4/8)

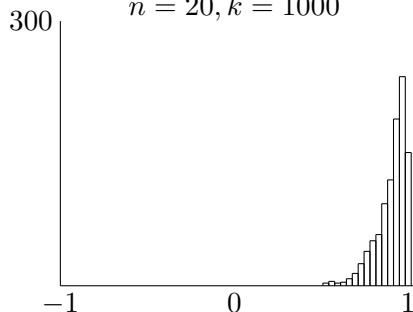
例. 統計量の分布 (数値実験)(つづき)

最大値の分布 ( $t = \max\{x_1, \dots, x_n\}$ )

$n = 5, k = 1000$



$n = 20, k = 1000$



右側に偏った分布

標本数 ( $n$ ) が大きくなると, 分布の「幅」が狭くなる

## 統計量の分布 (5/8)

### 定義 6.7

標本平均, 標本分散

$X_1, \dots, X_n$  : ランダム標本

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

をそれぞれ, 標本平均, 標本分散とよぶ.

## 統計量の分布 (6/8)

$P_n$  : ランダム標本  $X_1, \dots, X_n$  の経験分布

$X^*$  :  $P_n$  に従う確率変数

$\Rightarrow$

$$E(X^*) = \sum_{i=1}^n X_i P(X^* = X_i) = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}_n$$

$$\text{Var}(X^*) = \sum_{i=1}^n (X_i - \bar{X}_n)^2 P(X^* = X_i) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

標本平均は、経験分布の平均

標本分散は、経験分布の分散の  $\frac{n}{n-1}$  倍



## 統計量の分布 (7/8)

### 定理 6.2

$X_1, \dots, X_n$  をランダム標本とし

$$E(X_i) = \mu, \quad \text{Var}(X_i) = \sigma^2, \quad i = 1, \dots, n$$

とする. このとき次が成り立つ.

- (1)  $E(\bar{X}_n) = \mu.$
- (2)  $\text{Var}(\bar{X}_n) = \frac{1}{n}\sigma^2.$
- (3)  $E(S_n^2) = \sigma^2.$

(証明 は (3) のみ板書で)

# メモ用紙

## 統計量の分布 (8/8)

### 定理 6.3

定理 6.2 と同じ仮定のもとで次が成り立つ.

$$(1) \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0, 1).$$

$$(2) S_n^2 \xrightarrow{p} \sigma^2.$$

$$(3) \frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \xrightarrow{d} N(0, 1).$$

(証明は板書で)

## メモ用紙

## メモ用紙

## 正規母集団からの統計量の分布 (1/10)

母集団分布が正規分布の場合

標本平均と標本分散の関数の分布を導出する.

平均や分散に関する「区間推定」や「検定」で必要となる.

定理 6.8

$X_1, \dots, X_n \stackrel{i.i.d.}{\sim} N(\mu, \sigma^2)$  とする. このとき, 次が成り立つ.

- (1)  $\bar{X}_n \sim N(\mu, \sigma^2/n)$ .
- (2)  $(n-1)S_n^2/\sigma^2 \sim \chi_{n-1}^2$ .
- (3)  $\bar{X}_n$  と  $S_n$  は互いに独立である.

(証明には, 定義といくつかの定理が必要)

## 正規母集団からの統計量の分布 (2/10)

### 定義 6.1

カイ 2 乗分布 確率変数  $X$  の確率密度関数が

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

で与えられるとき,  $X$  は自由度  $n$  のカイ 2 乗分布に従うといい,  $X \sim \chi_n^2$  とかく.

## 正規母集団からの統計量の分布 (3/10)

$\mathbf{X} = (X_1, \dots, X_n)'$  の確率密度関数が

$$f(x_1, \dots, x_n) = (2\pi)^{-n/2} |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}$$

と表せるとき、 $n$ 次元正規分布に従うといい、 $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \Sigma)$  と書く。ここで、

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \cdots & \sigma_{nn} \end{pmatrix}.$$

パラメータ  $\boldsymbol{\mu}$  は任意の定数、すなわち、 $-\infty < \mu_i < \infty$  ( $i = 1, \dots, n$ ) で、 $\Sigma$  は正定値行列である。



## 正規母集団からの統計量の分布 (4/10)

## 命題 1

$\mathbf{Z} = (Z_1, \dots, Z_n)'$ ,  $Z_1, \dots, Z_n \stackrel{i.i.d.}{\sim} N(0, 1)$  とする.

(1)  $\mathbf{Z} \sim N_n(\mathbf{0}, I_n)$

ただし,  $I_n$  は  $n$  次単位行列.

(2)  $A$  :  $n$  次正則行列,  $\mathbf{b}$  :  $n$  次元ベクトル

$$A\mathbf{Z} + \mathbf{b} \sim N_n(\mathbf{b}, AA')$$

(3)  $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \Sigma)$

$$\Rightarrow E(\mathbf{X}) = \boldsymbol{\mu}, \text{Var}(\mathbf{X}) = \Sigma,$$

$$\varphi_{\mathbf{X}}(\mathbf{t}) := E(e^{i\mathbf{t}'\mathbf{X}}) = \exp(i\mathbf{t}'\boldsymbol{\mu} - \mathbf{t}'\Sigma\mathbf{t}/2)$$

(4)  $B$  :  $m \times n$  行列,  $\text{rank} B = m$ ,  $\mathbf{b}$  :  $m$  次元ベクトル,

$$\mathbf{X} \sim N_n(\boldsymbol{\mu}, \Sigma), \quad \mathbf{Y} = B\mathbf{X} + \mathbf{b}$$

$$\Rightarrow \mathbf{Y} \sim N_m(B\boldsymbol{\mu} + \mathbf{b}, B\Sigma B')$$

((3), (4) はテキストの定理 6.5, 証明は板書で)

# メモ用紙

## メモ用紙

# メモ用紙

# メモ用紙

## 正規母集団からの統計量の分布 (5/10)

### 定理 6.7

$X_1, \dots, X_n \stackrel{i.i.d.}{\sim} N(0, 1)$  とし,  $V = X_1^2 + \dots + X_n^2$  とする. このとき,  $V \sim \chi_n^2$  である.

(証明は板書)

### 定理 6.8

$X_1, \dots, X_n \stackrel{i.i.d.}{\sim} N(\mu, \sigma^2)$  とする. このとき, 次が成り立つ.

- (1)  $\bar{X}_n \sim N(\mu, \sigma^2/n)$ .
- (2)  $(n-1)S_n^2/\sigma^2 \sim \chi_{n-1}^2$ .
- (3)  $\bar{X}_n$  と  $S_n$  は互いに独立である.

## 正規母集団からの統計量の分布 (6/10)

(証明のポイント)

$$\mathbf{X} = (X_1, \dots, X_n)' \Rightarrow \mathbf{X} \sim N_n(\mu \mathbf{1}_n, \sigma^2 I_n) \quad (\mathbf{1}_n = (1, \dots, 1)')$$

$H$  : 第 1 行が  $(1/\sqrt{n}, \dots, 1/\sqrt{n})$  である  $n$  次の直交行列

$$\mathbf{Y} = (Y_1, \dots, Y_n)' = H\mathbf{X}$$

と定めると

$$\mathbf{Y} \sim N_n(\mu H\mathbf{1}_n, \sigma^2 I_n)$$

と定めると

$$E(\mathbf{Y}) = \mu \begin{pmatrix} 1/\sqrt{n} & \cdots & 1/\sqrt{n} \\ * & \cdots & * \\ \vdots & & \vdots \\ * & \cdots & * \end{pmatrix} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} \sqrt{n}\mu \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$Y_1, \dots, Y_n$  は独立,  $Y_1 \sim N(\sqrt{n}\mu, \sigma^2)$ ,  $Y_2, \dots, Y_n \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$

## メモ用紙



# メモ用紙

## メモ用紙

## メモ用紙

## 正規母集団からの統計量の分布 (7/10)

### 定義 6.10

( $t$ -分布) 確率変数  $T$  の確率密度関数が

$$f_T(t) = \frac{\Gamma\left(\frac{1}{2}(n+1)\right)}{\sqrt{n\pi}\Gamma\left(\frac{1}{2}n\right)} \left(1 + \frac{1}{n}t^2\right)^{-(n+1)/2}$$

で与えられるとき,  $X$  は自由度  $n$  の  $t$ -分布に従うといい,  $T \sim t_n$  とかく.

### 定理 6.9

$X \sim N(0, 1)$ ,  $Y \sim \chi_n^2$ ,  $X$  と  $Y$  は独立とする. このとき

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

(証明は省略, テキスト 125 ページ参照)

## 正規母集団からの統計量の分布 (8/10)

### 注 6.5

正規母集団  $N(\mu, \sigma^2)$  からの大きさ  $n$  の標本  $X_1, \dots, X_n$  に基づく標本平均, 標本分散を  $\bar{X}_n, S_n^2$  とする. このとき

$$T = t(X_1, \dots, X_n) = \frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \sim t_{n-1}^2$$

## 正規母集団からの統計量の分布 (9/10)

## 定義 6.11

$F$ -分布確率変数  $V$  の確率密度関数が

$$f_V(v) = \begin{cases} \frac{\Gamma(\frac{1}{2}(m+n))}{\Gamma(\frac{1}{2}m)\Gamma(\frac{1}{2}n)} \left(\frac{m}{n}\right)^{m/2} v^{m/2-1} \left(1 + \frac{m}{n}v\right)^{-(m+n)/2}, & v > 0 \\ 0, & v \leq 0 \end{cases}$$

で与えられるとき,  $V$  は自由度  $m, n$  の  $F$ -分布に従うといい,  
 $V \sim F_{m,n}$  とかく.

## 正規母集団からの統計量の分布 (10/10)

### 定理 6.10

$X \sim \chi_m^2$ ,  $Y \sim \chi_n^2$ ,  $X$  と  $Y$  は独立とする. このとき

$$V = \frac{X/m}{Y/n}$$

は自由度  $m, n$  の  $F$ -分布に従う.

(証明は省略, テキスト 127 ページ参照)