

# 生物統計学

— データに基づく「よりよい推論」のために —

三中信宏

MINAKA Nobuhiro

独立行政法人 農業環境技術研究所 生態系計測研究領域 上席研究員  
東京大学大学院 農学生命科学研究科 生物・環境工学専攻 教授 [生態系計測学]  
東京農業大学大学院 農学研究科 客員教授 [応用昆虫学]

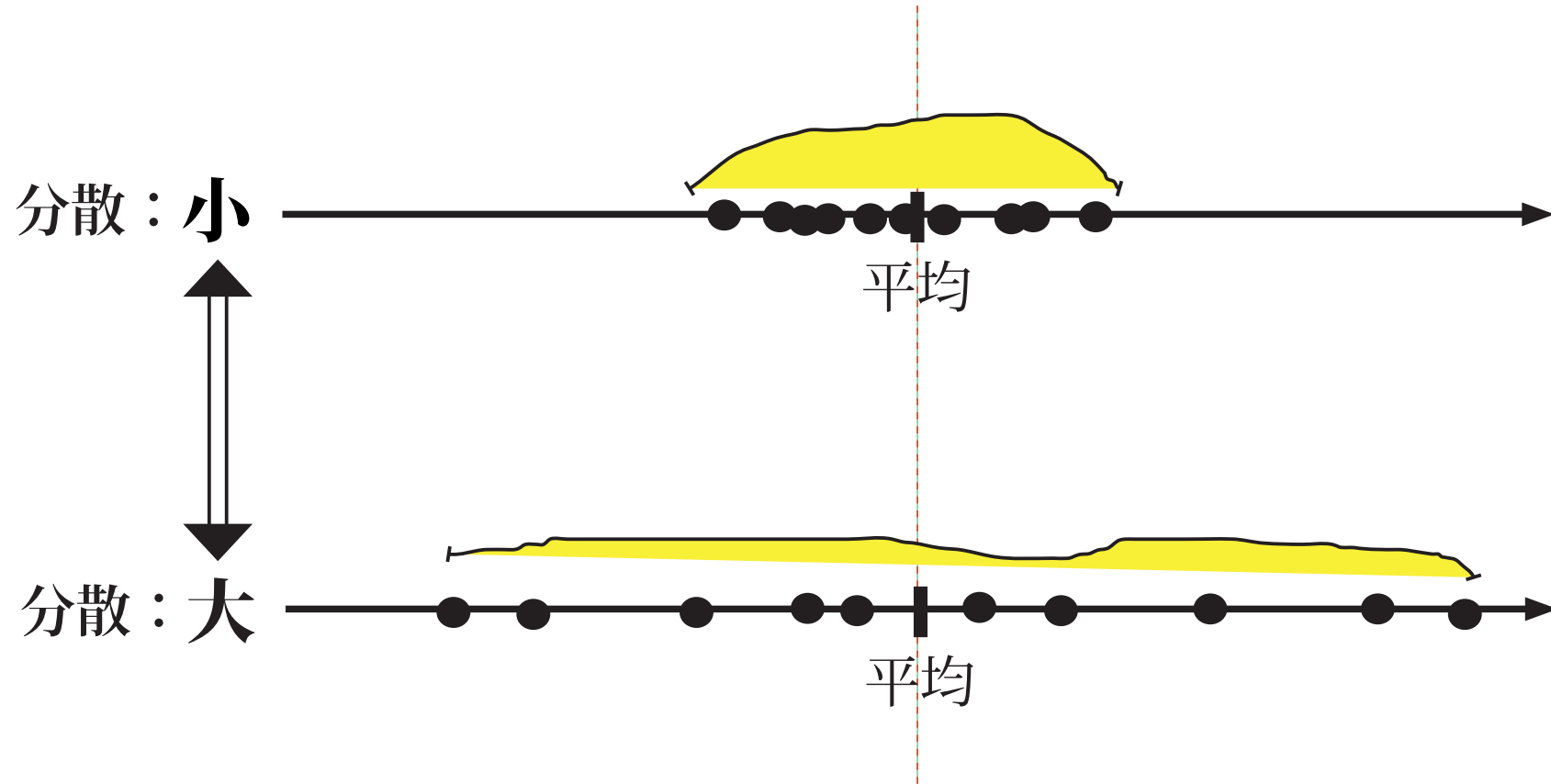
[minaka@affrc.go.jp](mailto:minaka@affrc.go.jp) (メール)

<http://twitter.com/leeswijzer> (ツイッター)

<http://cse.niaes.affrc.go.jp/minaka/> (ウェブサイト)

<http://d.hatena.ne.jp/leeswijzer/> (書評ブログ)

# 「分散」のイメージは……



「分散」を数字で表すには……

→ 「平均値」からの差をとる

○単純な引き算をする

データ： $x_1, x_2, x_3$  → 平均： $\bar{x} = \frac{x_1 + x_2 + x_3}{3}$

偏差

$x_1 - \bar{x}$
$x_2 - \bar{x}$
$x_3 - \bar{x}$

「偏差」を集計するには？

「**偏差**」の正負にかかわらず、  
その大きさだけを集計する

- 案1：「絶対値」を集計――

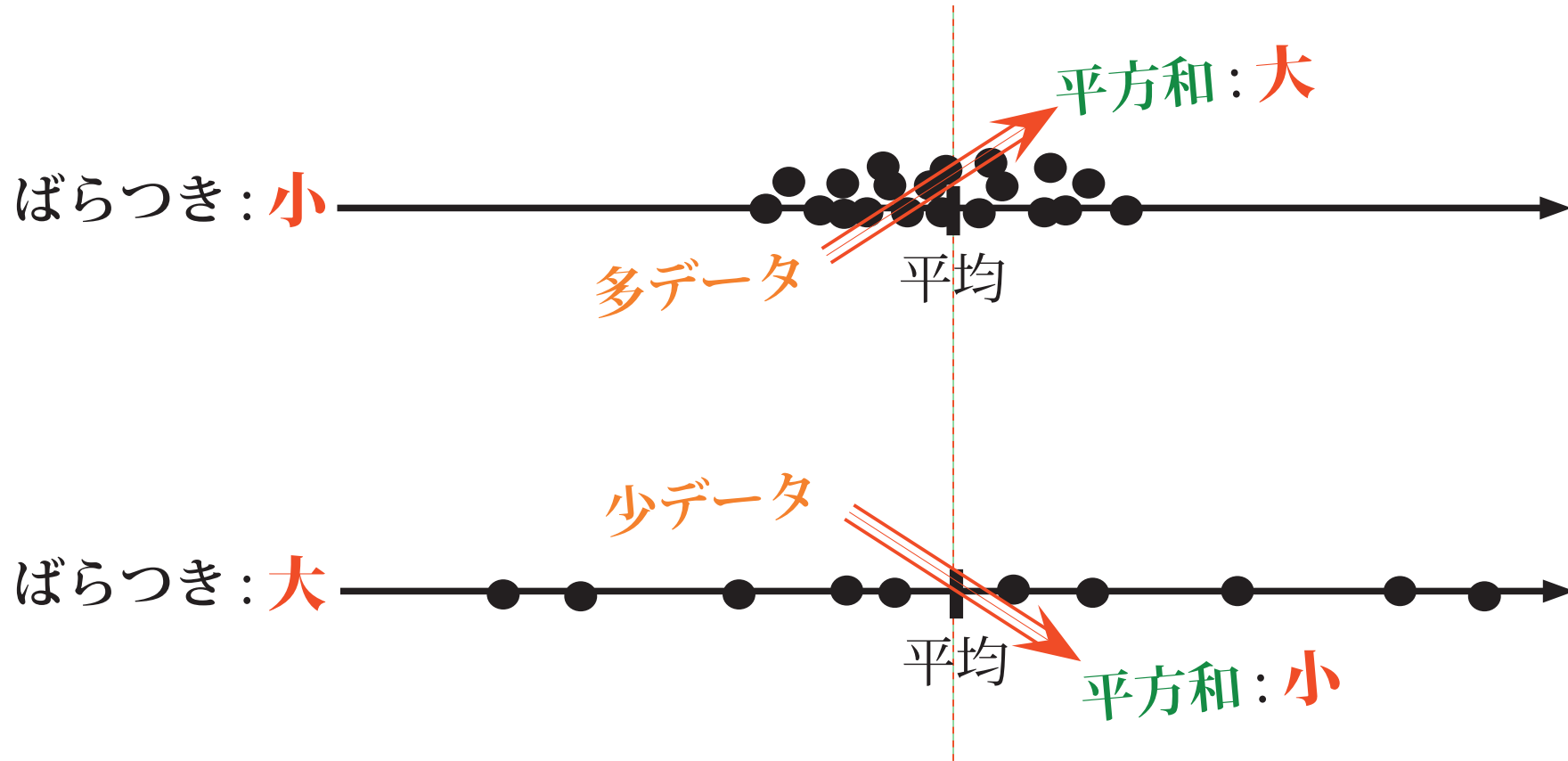
✗  $|x_1 - \bar{x}| + |x_2 - \bar{x}| + |x_3 - \bar{x}|$

- 案2：「平方値」を集計――

$$(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2$$

平方和

# でも「平方和」ではまだ不十分……



→ データの「サイズ」で補正しよう

# 「平方和」をサイズ補正するには……

- 案1 : 「データ数」で割り算してはどうか——

✗ 平方和 / データ数

- 案2 : 「自由度」で割り算するべきだ——

○ 平方和 / 自由度

……「自由度」って何？

# 「自由度」とは何か？

3つの偏差の間には.

$$(x_1 - \bar{x}) + (x_2 - \bar{x}) + (x_3 - \bar{x}) = 0$$

という制約があるので、その自由度は「3」ではない.

実際は、3偏差のうちのいずれかひとつは、他の2偏差によって完全に決定されるので：

$$(x_3 - \bar{x}) = -(x_1 - \bar{x}) - (x_2 - \bar{x})$$

偏差の自由度は「 $3 - 1 = 2$ 」となる.

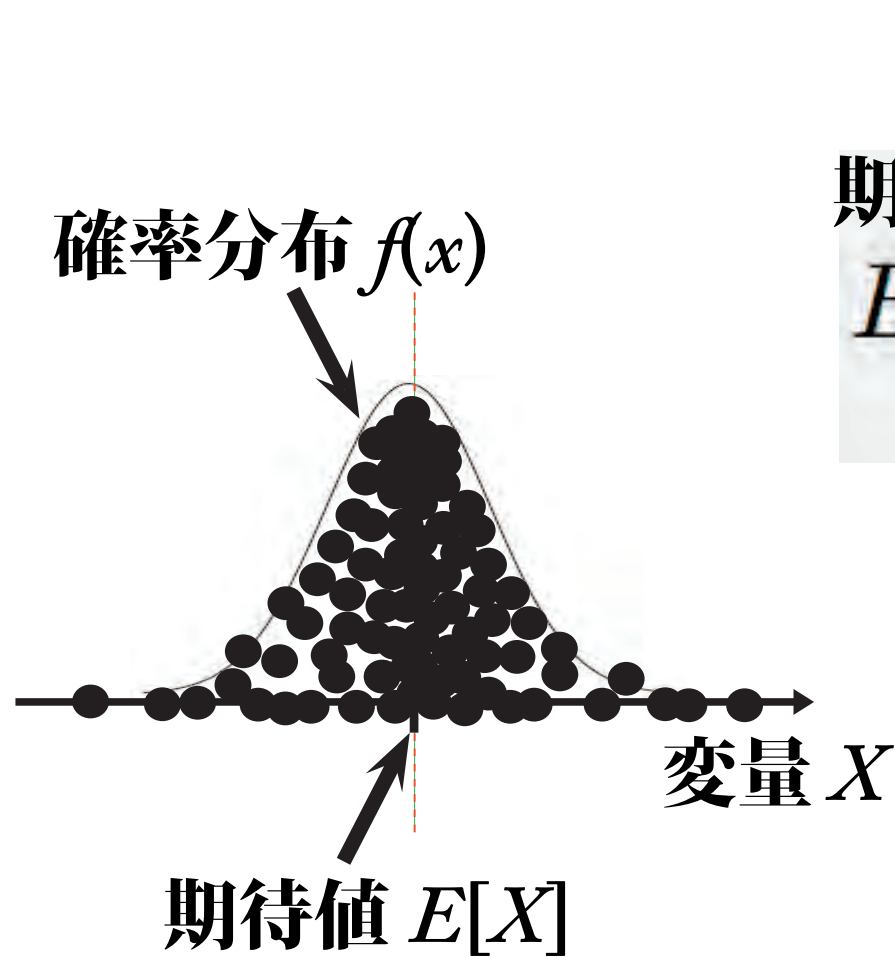
# 「分散」を定義する

平方和を構成する  $n$  個の偏差の自由度は  $(n - 1)$  だから、平方和もまた  $(n - 1)$  の自由度をもつことになる。

$$\text{分散} = \text{平方和} / \text{自由度}$$



# 確率分布と「分散」の関係



期待値

$$E[X] = \int_{-\infty}^{+\infty} x f(x) dx$$

変量  $X$  の値

期待値演算子  $E[.]$

# 確率分布と「分散」の関係

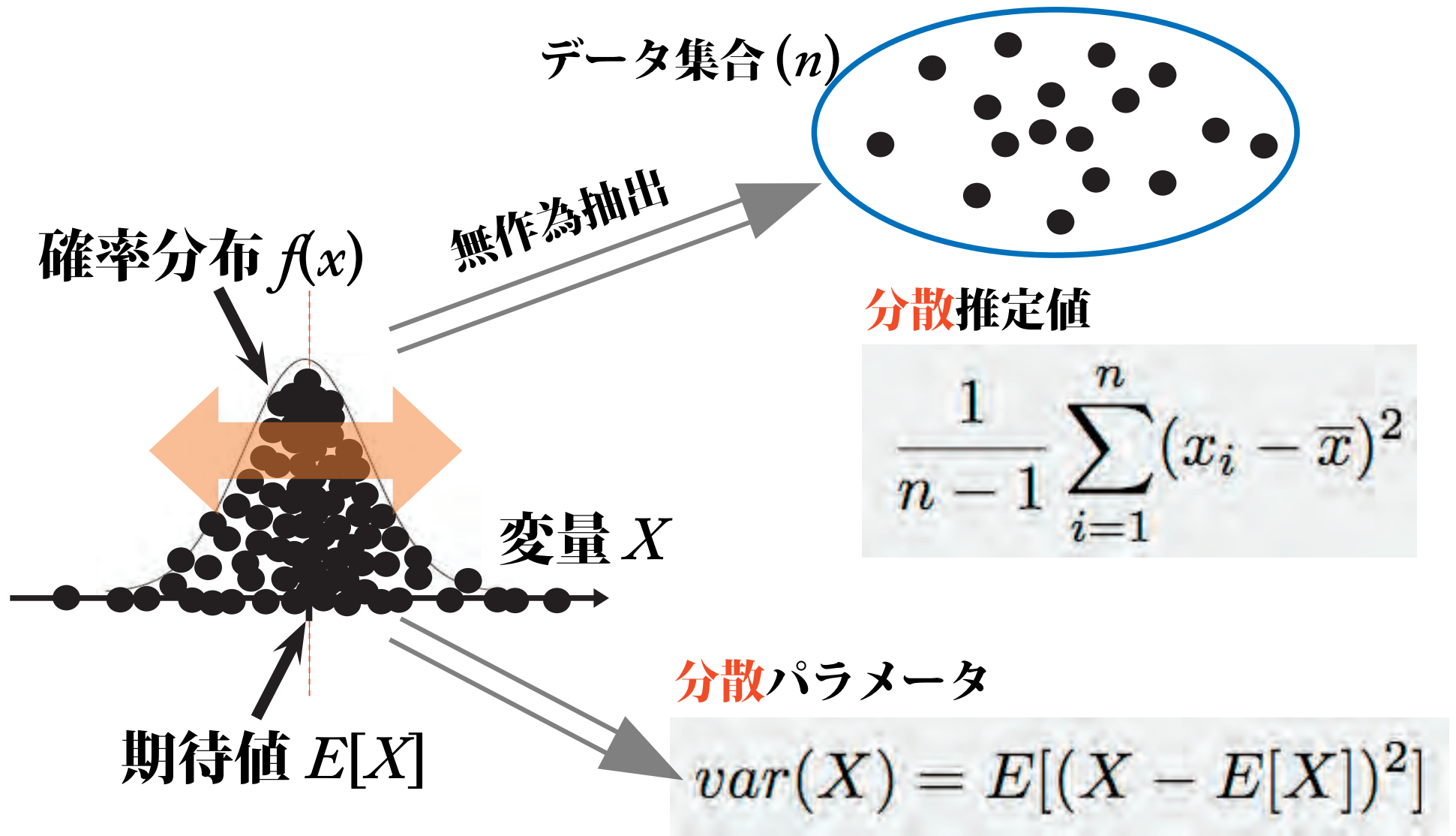
ある変量  $X$  の確率分布の密度関数が  $f(x)$  で与えられ、その期待値を  $E[X]$  とするとき、 $X$  の「分散」  $var(X)$  は：

$$var(X) = \int_{-\infty}^{+\infty} (x - E[X])^2 f(x) dx$$

ただし、

$$E[X] = \int_{-\infty}^{+\infty} x f(x) dx$$

# 「分散」パラメータとその推定値



# 正規分布

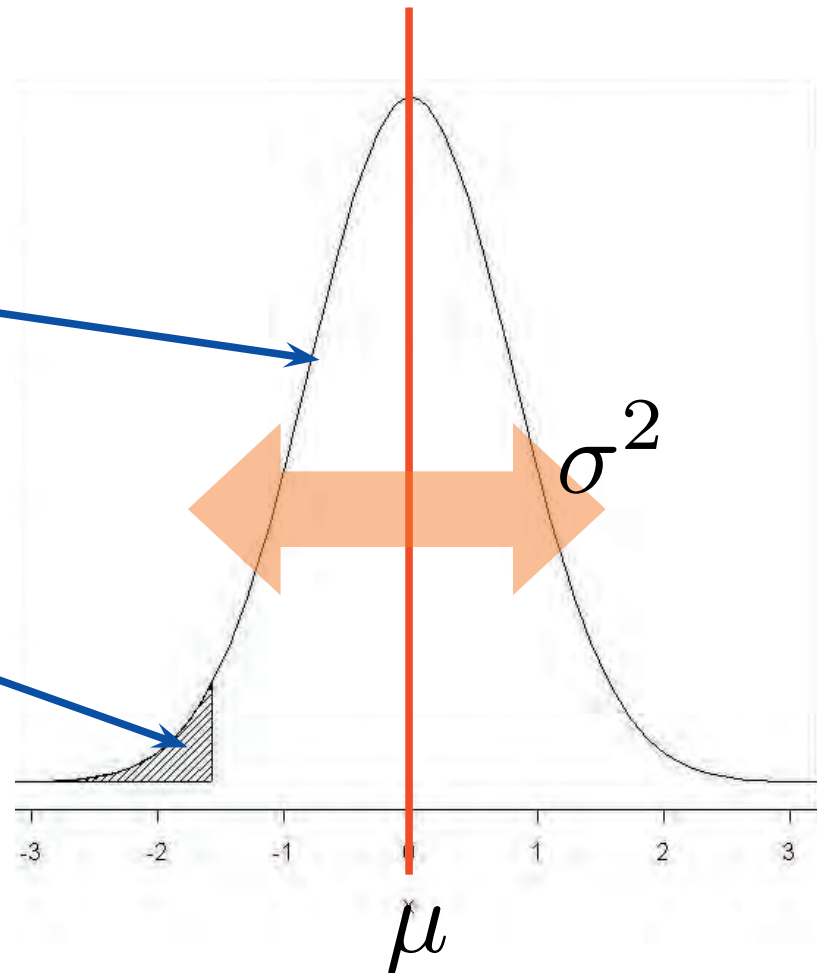
平均  $\mu$ , 分散  $\sigma^2$  の正規分布  $N(\mu, \sigma^2)$  の確率密度関数

確率密度関数

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

累積確率関数

$$P(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$



# 正規分布

## 線形変換における正規性の保存

1)  $X \sim N(\mu, \sigma^2)$  のとき,  $aX + b \sim N(a\mu + b, a^2\sigma^2)$

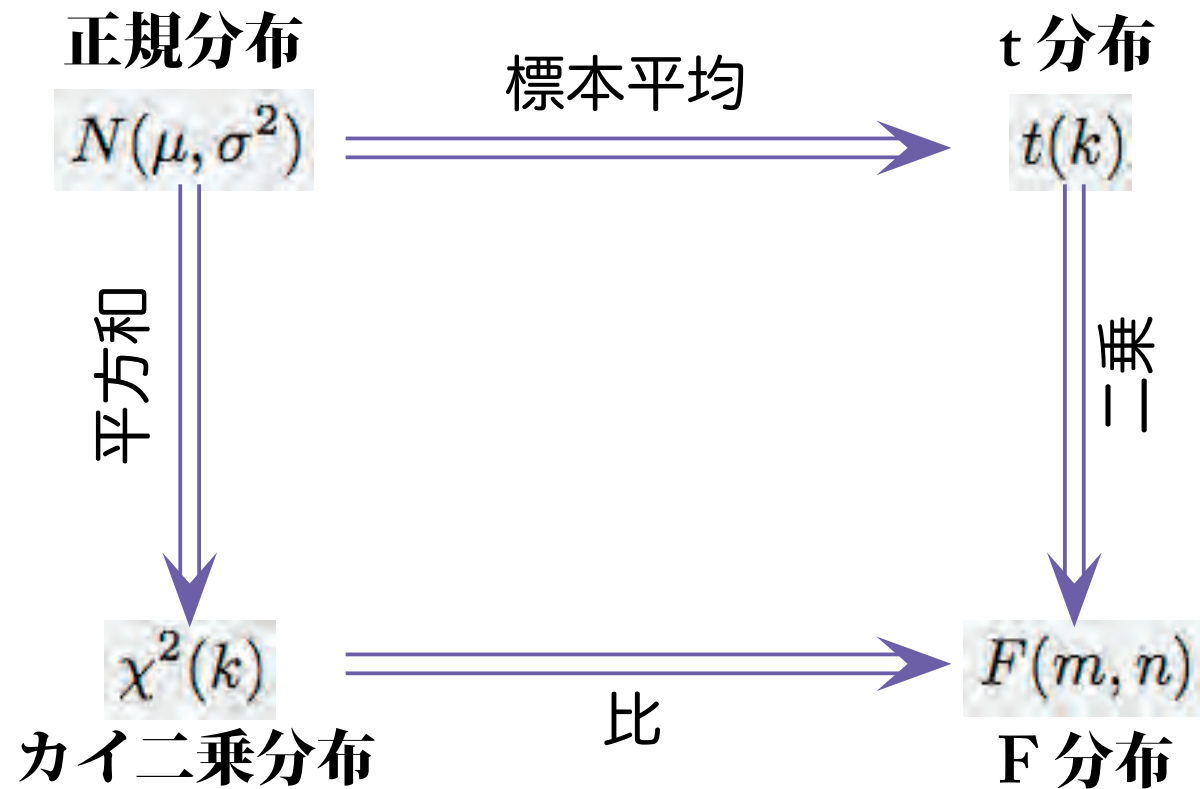
とくに,  $Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$  を標準正規分布と呼ぶ.

2)  $X_i \sim N(\mu_i, \sigma_i^2) (i = 1, 2, \dots, n)$  のとき,

$$\sum_{i=1}^n a_i X_i \sim N\left(\sum_{i=1}^n a_i \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2\right)$$

# 正規分布

## 関連する他の確率分布の導出



# 正規分布

## 関連する他の確率分布の導出

カイ二乗分布

$$\chi^2(k) = \frac{1}{\Gamma(\frac{k}{2})} \left(\frac{1}{2}\right)^{\frac{k}{2}} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}$$

t 分布

$$t(k) = \frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})} \frac{1}{\sqrt{k\pi}} \frac{1}{\left(1 + \frac{x^2}{k}\right)^{\frac{k+1}{2}}}$$

F 分布

$$F(m, n) = \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \left(\frac{m}{n}\right)^{\frac{m}{2}} \frac{x^{\frac{m-2}{2}}}{\left[1 + \left(\frac{m}{n}\right)x\right]^{\frac{m+n}{2}}}$$



# 正規分布

## 「中心極限定理」

任意の確率分布変量  $X$

平均  $\mu$

分散  $\sigma^2$

$n$  無作為標本

統計量

$$Z_n = \frac{\overline{X_n} - \mu}{\frac{\sigma}{n}}$$

$$\lim_{n \rightarrow \infty} Z_n$$

標準正規分布

$$N(0, 1)$$

正規分布

$$N(\mu, \sigma^2)$$