推定量の評価について

担当： 長倉 大輔
（ながくらだいすけ）

パラメーターの推定

- パラメーターとは？

パラメーターとは**母集団の分布を特徴付ける値**の事である。

例えば、$N(\mu, \sigma^2)$ の正規分布では μ と σ^2 がパラメーターと見なされる。

$U(a, b)$ の一様分布であれば a と b がパラメーターと見なされる。

$B(n, p)$ の2項分布であれば n と p がパラメーターと見なされる。

パラメーターの推定

- パラメーターとは？

確率分布が未知の時には、平均 μ や、分散 σ^2 をパラメーターと見なす事ができる。

通常、パラメーターは未知であるので(これを**未知パラメーター**と呼ぶ)、標本よりこれらを推定する事を**未知パラメーターを推定する**という。

パラメーターの推定

- 推定量とは？

ある統計量が未知パラメーターを推定する目的で使われた場合に、**推定量**と呼ばれる。
（例: 標本平均は母平均 μ の推定量である）。

統計量は確率変数であるので、**推定量も確率変数**。

- 推定値とは?

推定量に標本の実現値が与えられると**推定値**になる。

パラメーターの推定

- 推定量の評価

大きさ n の標本 $\{X_1, \ldots, X_n\}$ が与えられた時に母平均の推定量としてどのようなものが考えられるだろうか?

まず真っ先に標本平均 $\overline{X}$ があげられる。さらに分布が左右対称であればメディアンも母平均の推定量として使える(この時母メディアンと母平均は等しいので)。

さらに適当な X_i (例えば X_5) を持ってきて、これを母平均の推定量という事もできる。

さらにさらに適当な定数 a を持ってきてこれを母平均の推定量という事もできる。

パラメーターの推定

■ 推定量の評価

このように推定量は無数に考える事ができる。

無数の推定量の中から、どのような推定量を選べばよいのだろうか？何かを比較し選択する場合には何らかの基準がいる。推定量の比較については次の3つの基準が代表的なものである。

1. **一致性**
2. **不偏性**
3. **効率性(有効性)**

パラメーターの推定

- 一致性

1つ目の推定量の比較の基準に**一致性**がある。

これは言葉でいうならば、推定量がパラメーターの真の値から離れる確率が標本が大きくなるにつれて、どんどん小さくなるという事を意味している。正確には T_n が一致性を持つとは、ある $\varepsilon > 0$ に対して

$$\lim_{n\to\infty} \Pr(\,|T_n - \mu| \geq \varepsilon) = 0$$

という事である。これを T_n が μ に**確率収束**するという。これは

$$\operatorname*{plim}_{n\to\infty} T_n = \mu$$

と書く。

パラメーターの推定

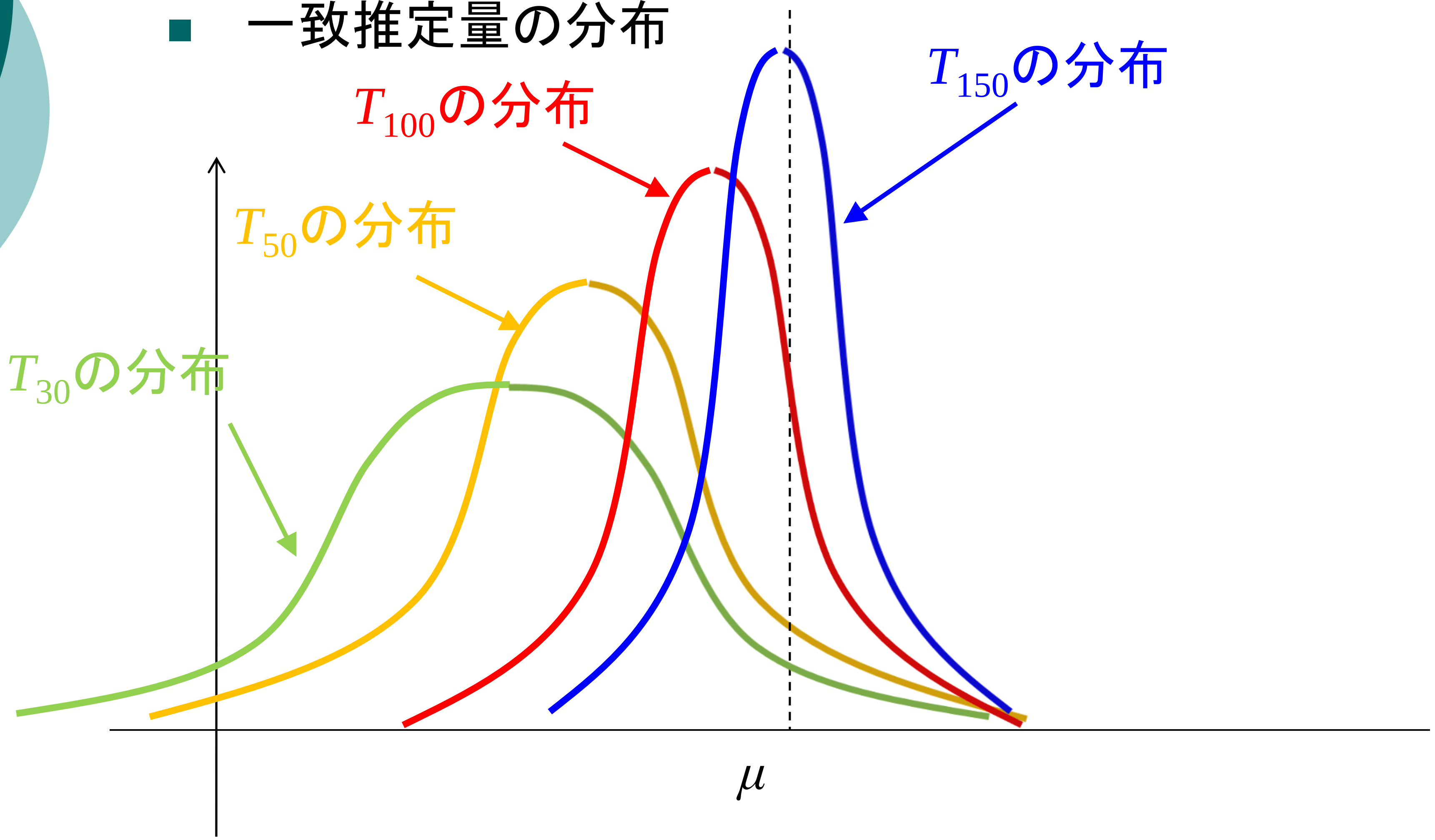
- 一致推定量

一致性のある推定量を**一致推定量**という。

一致性は推定量が推定量として意味があるために最低限満たさないといけない性質といってよいだろう。

パラメーターの推定

- 一致推定量の分布



パラメーターの推定

- 不偏性

2 つめの基準は**不偏性**と呼ばれるものである。
例えば母平均 μ の推定量 $T_n = T(X_1, \ldots, X_n)$ が不偏性を持っているとは

$$E(T_n) = \mu$$

となる事である(推定量は確率変数である事に注意)。

つまり、「推定量の期待値が推定したい未知パラメーターと等しい」という事である。このような推定量を**不偏推定量**という。

パラメーターの推定

- 一致性と不偏性の関係

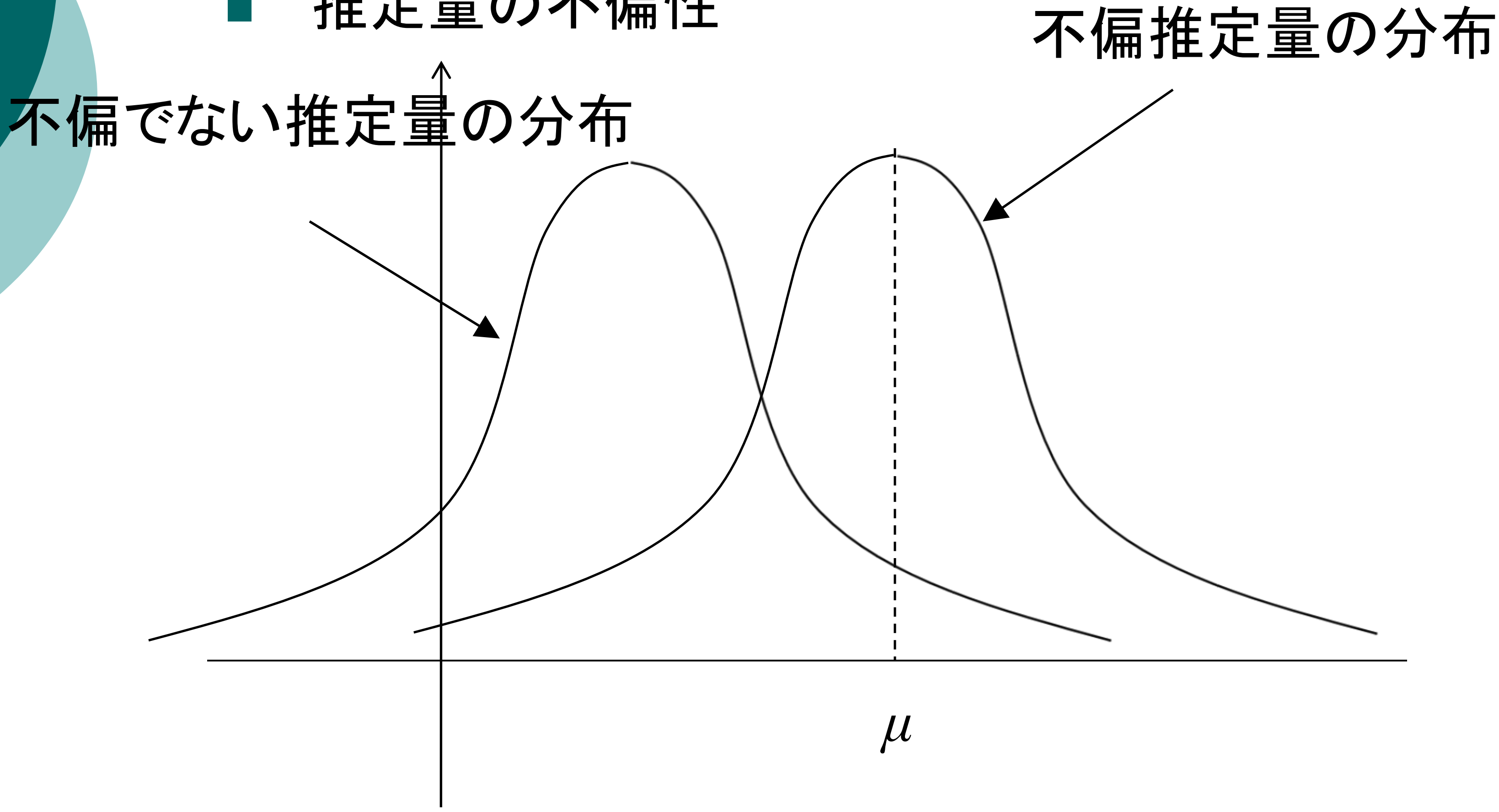
一般的に一致推定量は必ずしも不偏推定量ではなく、また逆に不偏推定量も必ずしも一致推定量ではない。

(例: 不偏推定量だが一致推定量ではないもの)
例えば母平均 μ の推定量 T_n として $T_n = X_5$ (標本の中の5番目の値)とすると、$E(X_5) = \mu$ であるので T_n は不偏推定量であるが、一致推定量ではない。

ただし、もちろん不偏推定量であり、一致推定量でもある推定量は存在する。例えば標本平均 $\overline{X}$ は母平均 μ の不偏推定量であり、一致推定量でもある。

パラメーターの推定

- 推定量の不偏性



パラメーターの推定

- 効率性(有効性)

推定量を選ぶ3つ目の基準に**効率性**(**有効性**)がある。

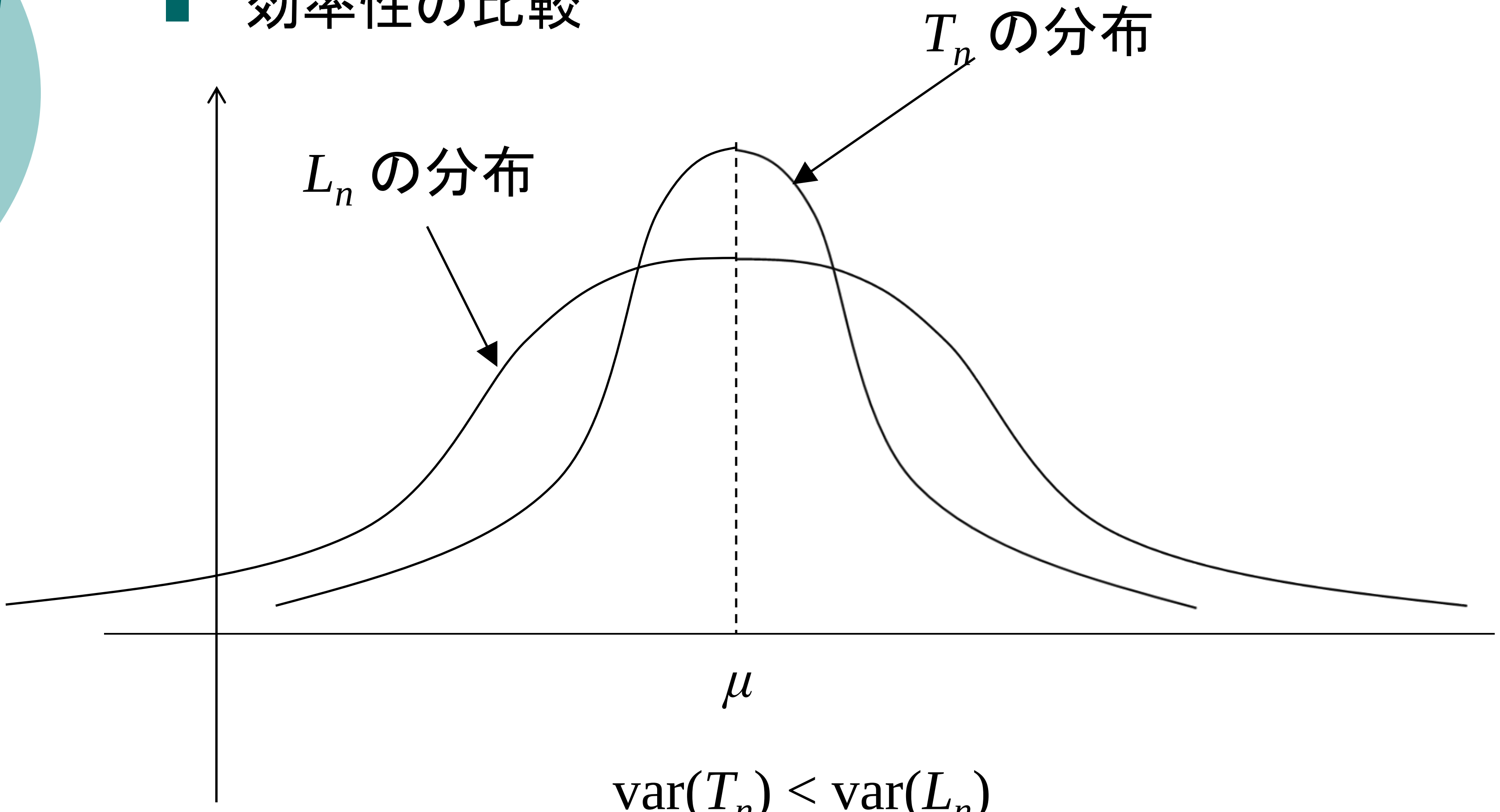
2つの不偏推定量 T_n , L_n があった時に、推定量 T_n の方が**効率的**であるとは T_n の方が分散が小さい、つまり

$$\mathrm{var}(T_n) < \mathrm{var}(L_n)$$

という事である。

パラメーターの推定

- 効率性の比較



$$\mathrm{var}(T_n) < \mathrm{var}(L_n)$$

パラメーターの推定

- 効率性の比較

先ほどの例では不偏推定量どうしの分散を比べた。

これは、不偏推定量ではない場合、分散が小さいというだけでは、必ずしもよい推定量とは言えないからである。これは次の例を見ればわかる。

パラメーターの推定

- 効率性の比較

(例: 分散が 0 の推定量)
母平均 μ の推定量として、適当な定数 a に対して、推定量 T_n を $T_n = a$ とすれば、この推定量の分散は

$$\mathrm{var}(T_n) = \mathrm{var}(a) = 0 \quad (定数の分散は0)$$

であるので0 である。

このような推定量は、しかしながら、偶然 $a = \mu$ とでもならない限りはよい推定量でないのは明らかであろう。この推定量はもちろん一致性もない。

パラメーターの推定

- 不偏でない推定量の効率性の比較

推定量の良さを見る基準として「不偏性」と「効率性」を見てきたが、不偏性を満たさない推定量どうしを比較しないといけない場合もある。

　例えば、未知パラメーター μ の推定量として 2 つの**不偏でない**推定量 T_n と L_n があり、

$$|E(T_n) - \mu| < |E(L_n) - \mu|$$

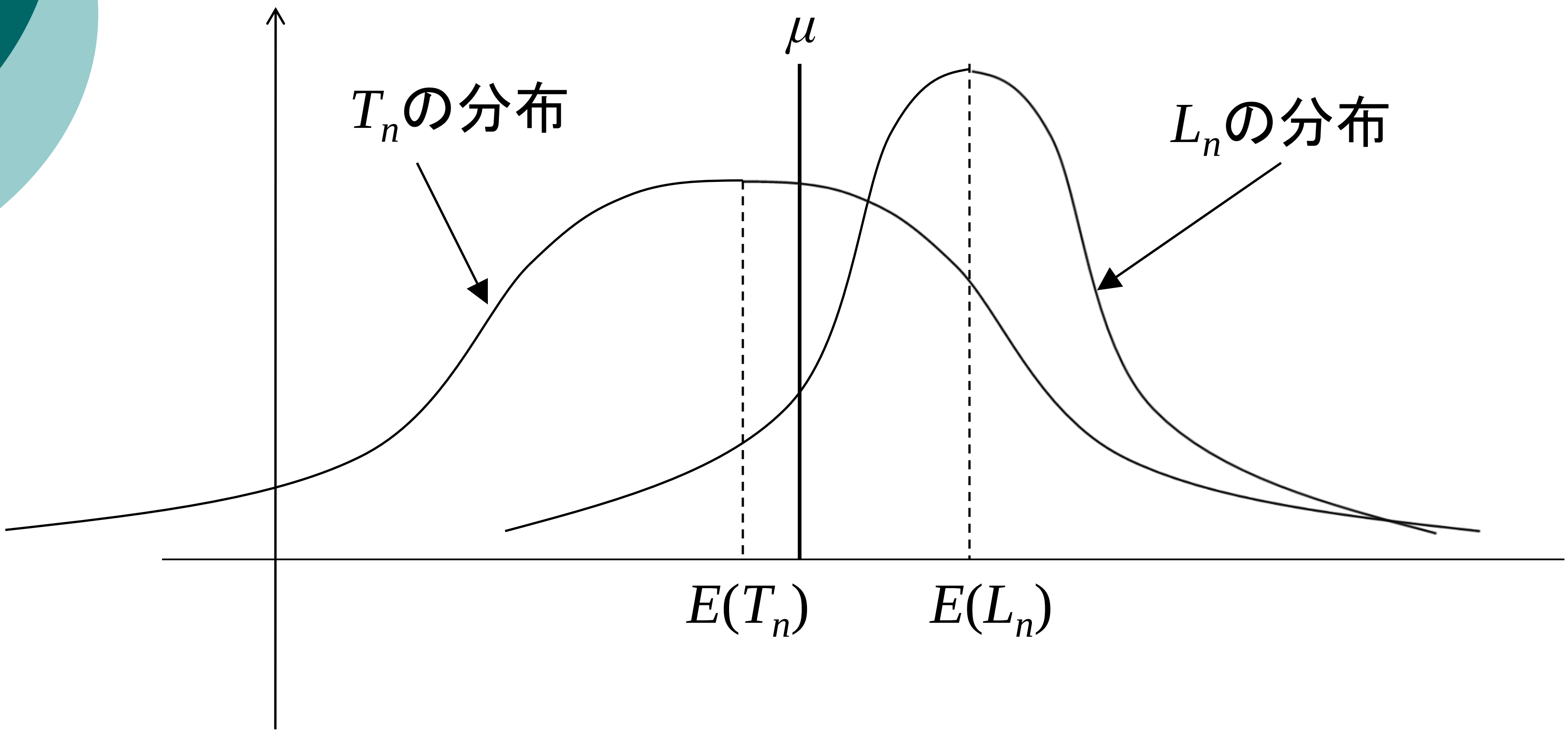
(推定量の期待値は T_n の方がμ に近い)、かつ

$$\mathrm{var}(T_n) > \mathrm{var}(L_n)$$

(推定量の分散は L_nの方が小さい)
のような場合、どちらがよい推定量といえるだろうか？

パラメーターの推定

- 不偏でない推定量の効率性の比較



パラメーターの推定

- バイアスと分散

T_n は、推定量の期待値がパラメータ μ により近いが、分散が大きいので、離れた値をとる確率も大きい。

L_n は推定量の期待値はパラメータ μ から離れているが、分散が小さいので、μ から離れた値をとる確率は小さい。

推定量の期待値とパラメーターの差を**バイアス**という。
T_n のバイアス： $E(T_n)-\mu$

パラメーターの推定

- 平均二乗誤差 (Mean Squared Error)

バイアスと分散の両方をバランスよく考えるひとつの基準として**平均二乗誤差**と呼ばれるものがある。以下のように定義される。

(未知パラメーター μ の推定量 T_n の平均2乗誤差)

$$\mathrm{MSE}(T_n) = E\,[(\,T_n - \mu\,)^2]$$

Mean Squared Error なので略して **MSE** と呼ばれる。

パラメーターの推定

- 平均 2乗誤差 (Mean Squared Error)

MSEは以下のようにバイアスと分散に分解できる。
$\mu_n = E(T_n)$ とおこう(これは**確率変数ではない**事に注意)。

$$
\begin{aligned}
\mathrm{MSE}(T_n) &= E\,[(T_n - \mu)^2] \\
&= E\,[(T_n - \mu_n + \mu_n - \mu)^2] \\
&= E\,[(T_n - \mu_n)^2 + 2(T_n - \mu_n)(\mu_n - \mu) + (\mu_n - \mu)^2] \\
&= E\,[(T_n - \mu_n)^2] + 2(\mu_n - \mu)E(T_n - \mu_n) + E\,[(\mu_n - \mu)^2] \\
&= \mathrm{var}\,(T_n) + (\mu_n - \mu)^2
\end{aligned}
$$

第 2 項は T_n の**バイアスの2乗**である。MSEが小さいということは分散とバイアスの両方が小さいということである。