

確率統計

Probability and Statistics

第7回講義資料

Lecture notes 7

区間推定と仮説検定

Interval Estimation and Hypothesis Testing

豊橋技術科学大学

Toyohashi University of Technology

電気・電子情報工学系

Department of Electrical and Electronic Information Engineering

准教授 竹内啓悟

Associate Professor Keigo Takeuchi

区間推定の目的(Purpose of interval estimation)

推定誤差(Estimation error)

標本の大きさは有限なので、母数の点推定値には誤差が生じている。

A finite sample size results in an error for point estimation of a parameter.

推定誤差をどのように評価すべきか？

Should we evaluate the estimation error?

平均二乗誤差による評価(Evaluation based on mean-square error (MSE))

推定値の平均二乗誤差を評価して、理論的な下界と比べる。

Compare the MSE in estimation with a theoretical lower bound.

欠点(Disadvantages)

真の母数が**既知**でないと、平均二乗誤差を評価できない。

It is impossible to evaluate the MSE, unless the true parameter is **known**.

真の母数が既知であったとしても、評価するには**多数の標本**が必要である。

Even if the true parameter were known, we would need **many samples** to evaluate the MSE.

区間推定の目的(Purpose of interval estimation)

真の母数が**未知**の場合に、**単一の標本**から点推定の誤差を評価する。

Evaluation of errors in point estimation for **unknown** parameters with a **single sample**.

例(Example)

問題(Problem)

大きさ1の標本 X の実現値 $x = 0.5$ から、母平均 θ の推定誤差を評価せよ。

Evaluate the error in estimation of the population mean based on a realization of a sample with size 1.

母平均の推定値(Estimate of the population mean)

予備知識がないので、母平均の推定値を $\hat{\theta} = x = 0.5$ とするしかない。

What we can do with no a priori knowledge is to select $\hat{\theta} = x$ as an estimate of the population mean.

誤差の評価(Evaluation of the error)

不可能である。(Impossible)

標本 X が分散 $\sigma^2=1$ の正規分布に従うことを知っていたとしたら、

If we knew that the sample θ followed a normal distribution with variance $\sigma^2=1$,

母平均 θ は区間 $[0.5 - 3\sigma, 0.5 + 3\sigma]$ に入っていそうだと言える。

we could say that the population mean was likely to be included into the interval $[0.5 - 3\sigma, 0.5 + 3\sigma]$.

区間推定の手順(Procedure of interval estimation)

仮定(Assumption)

母集団分布族 $\{P(X; \theta) \mid \theta \in \Theta\}$ は既知である。

The family of population distributions is known.

1 分布が**未知の母数 θ に依存しない**統計量 T を設定せよ。

Select a statistic T such that the distribution of T is **independent of the unknown parameter**.

2 T の分布を評価し、 $100(1 - \alpha)\%$ **信頼水準**に基づく**信頼区間**を導出せよ。

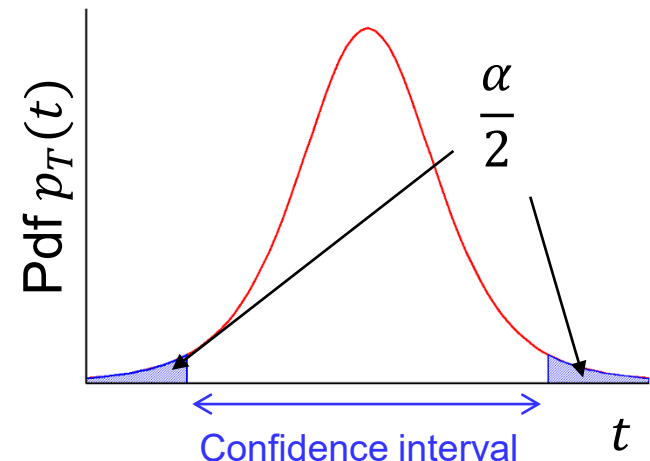
Evaluate the distribution of T , and derive a **confidence interval** based on a **confidence level** of $100(1 - \alpha)\%$.

$100(1 - \alpha)\%$ 信頼水準の信頼区間:

Confidence interval with a confidence level of $100(1 - \alpha)\%$

統計値が区間に含まれない確率が、左右均等に合計で α に等しくなるような区間

Interval that does not include a realization of the statistic with probability α , which is equally allocated on both sides.



信頼区間の意義 (Significance of confidence interval)

母数を確率変数とみなしていないので、区間推定は、信頼水準に等しい確率で母数は信頼区間の中に入るという定式化にはなっていない。

Since the parameter is **not regarded** as a random variable, the formulation of interval estimation **does not indicate** that the parameter is included into the confidence interval with probability that is equal to the confidence level.

例: 今現在の日本人の男女比 (Male-to-female ratio in Japan just now)

ある標本調査を行った結果、95%信頼水準における男女比の信頼区間は $[0.985, 1.015]$ であったとする。この結果を真の男女比は95%の確率でこの区間に入ると解釈するのはおかしい。

For the male-to-female ratio, suppose that we obtained the confidence interval $[0.985, 1.015]$ with a confidence level of 95% from a sample. The following interpretation of this result is wrong: The true ratio is included into the interval with a probability of 95%.

正しい解釈 (Correct interpretation)

同じ標本調査を独立に1000回行って男女比の信頼区間を計算した結果、約950個の信頼区間は真の男女比を含んでいる。

Compute 1000 confidence intervals for the male-to-female ratio from independent trials of the identical survey. Then, approximately 950 intervals would include the true ratio.

分散既知の正規母集団の母平均

Population mean of the normal population with known variance

分散 σ^2 が既知の正規母集団から無作為抽出された大きさ N の標本 $\mathbf{X} = \{X_1, \dots, X_N\}$ から、標本値 $\mathbf{x} = \{x_1, \dots, x_N\}$ が得られたとする。未知の母平均 μ を95%の信頼水準で区間推定せよ。

Suppose that we obtained a realization x of an independent and identically distributed (i.i.d.) sample \mathbf{X} from the normal population with known variance σ^2 . Interval-estimate the unknown population mean at a confidence level of 95%.

統計量(Statistic) $T = \frac{1}{\sqrt{N}} \sum_{n=1}^N \frac{X_n - \mu}{\sigma} = \frac{\bar{X} - \mu}{\sigma/\sqrt{N}}, \quad \bar{X}: \text{標本平均(Sample mean)}$

定理7.1(Theorem 7.1)



正規確率変数の線形結合は、正規分布に従う。

A linear combination of normal random variables follows a normal distribution.

定理7.1から、 T は任意の標本の大きさ N に対して標準正規分布に従う。

From Theorem 7.1, T follows the standard normal distribution for any sample size N .

分散既知の正規母集団の母平均

Population mean of the normal population with known variance



統計量(Statistic) $T = \frac{1}{\sqrt{N}} \sum_{n=1}^N \frac{X_n - \mu}{\sigma} = \frac{\bar{X} - \mu}{\sigma/\sqrt{N}}, \quad \bar{X}: \text{標本平均(Sample mean)}$

数値計算により、 $P(|T| \leq t_0) = 0.95$ を満たす閾値は $t_0 \approx 1.95996$ である。

Numerical computation implies that the threshold $t_0 \approx 1.95996$ results in $P(|T| \leq t_0) = 0.95$.

$|T| \leq t_0$ を μ に関して解くと、信頼区間 $\mu \in [\bar{X} - N^{-1/2}t_0\sigma, \bar{X} + N^{-1/2}t_0\sigma]$ を得る。

Solving $|T| \leq t_0$ with respect to μ yields the confidence interval $\mu \in [\bar{X} - N^{-1/2}t_0\sigma, \bar{X} + N^{-1/2}t_0\sigma]$.

$$\begin{aligned} \because \quad \left| \frac{\bar{X} - \mu}{\sigma/\sqrt{N}} \right| \leq t_0 & \iff -t_0 \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{N}} \leq t_0 \\ & \iff \bar{X} - \frac{t_0\sigma}{\sqrt{N}} \leq \mu \leq \bar{X} + \frac{t_0\sigma}{\sqrt{N}}. \end{aligned}$$

仮説検定の目的(Purpose of hypothesis testing)

データのみから、ある仮説の真偽を検定したい。

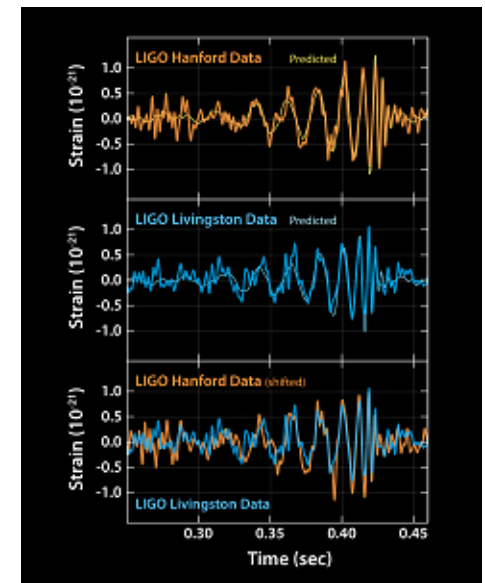
Test true or false for a hypothesis by using only data.

例: 2015年9月14日重力波の直接観測

(Example: Direct observation of gravitational waves on 14 Sep. 2015)

観測信号が重力波であることを立証したい。

Verify that the observed signals are a gravitational wave.



© LIGO

<http://mediaassets.caltech.edu/gwave>

問題: ただの観測ノイズである可能性をどのように否定するか？

Problem: How should we deny that the signals are just noise in observations?

困難: ノイズではないことの証明は原理的に不可能である。

Difficulty: It is impossible in principle to prove that they are not noise.

回避策: ノイズである可能性(正確には**p値**)が十分に低いことを立証する。

Solution: Verify that a possibility of noise (more precisely, **p-value**) is very rare.

本例の場合、p値は約 $p = 0.00006\%$ である。

The p-value is approximately $p = 0.00006\%$ in this example.

仮説検定の手順(Procedure of hypothesis testing)

1 帰無仮説 H_0 を立てる—注目している事象は偶然起こったという仮説

(Postulate a **null hypothesis** H_0 —An event we focus on occurred by chance.)

観測信号はノイズである。(The observed signals are noise.)

2 対立仮説 H_1 を立てる—帰無仮説を論理否定した仮説

(Set the **alternative hypothesis** H_1 —Logical negation of the null hypothesis.)

観測信号はノイズではない。(The observed signals are not noise.)

3 検定統計量 T を設定して、帰無仮説 H_0 を有意水準 α で棄却する。

Define a **test statistic** T , and reject the null hypothesis H_0 at a **significance level** α .

p値が α を下回ったとき、ただのノイズである可能性を棄却する。

さもなければ、この可能性を棄却はしない。

When the p-value is below α , reject a possibility that they are just noise. Otherwise, do not reject this possibility

用語の定義(Definitions of terminology)

棄却域(Rejection region) : $\mathcal{R} \subset \mathbb{R}$

検定統計量 T が領域 \mathcal{R} に含まれるならば、帰無仮説を棄却する。

Reject the null hypothesis if the test statistic is in the rejection region.

第1種の誤り(Type I error)

帰無仮説が真のときに帰無仮説を棄却してしまう誤り

Incorrect rejection of a true null hypothesis

第1種誤り確率(Type I error probability) : $P(T \in \mathcal{R} | H_0)$

第2種の誤り(Type II error)

帰無仮説が偽のときに帰無仮説を棄却しない誤り

Incorrect non-rejection of a false null hypothesis

第2種誤り確率(Type II error probability) : $P(T \notin \mathcal{R} | H_1)$

有意水準(Significance level) α :

第1種誤り確率が有意水準以下になるように、棄却域を設定する。

Define a rejection region such that the type I error probability is below the significance level.

用語の定義(Definitions of terminology)

p値(p-value): 帰無仮説が真のときに、得られた統計値 t 以上に稀な統計値が、検定統計量 T が従う分布から発生する確率 p

Probability p with which the distribution of the statistic T generates rarer realizations than an obtained realization t of the statistic.

検出力: 帰無仮説が偽のときに帰無仮説を棄却する確率 $1 - \beta$

(Power) Probability $1 - \beta$ with which a false null hypothesis is rejected.

第2種誤り確率(Type II error probability): $\beta = P(T \notin \mathcal{R} | H_1)$

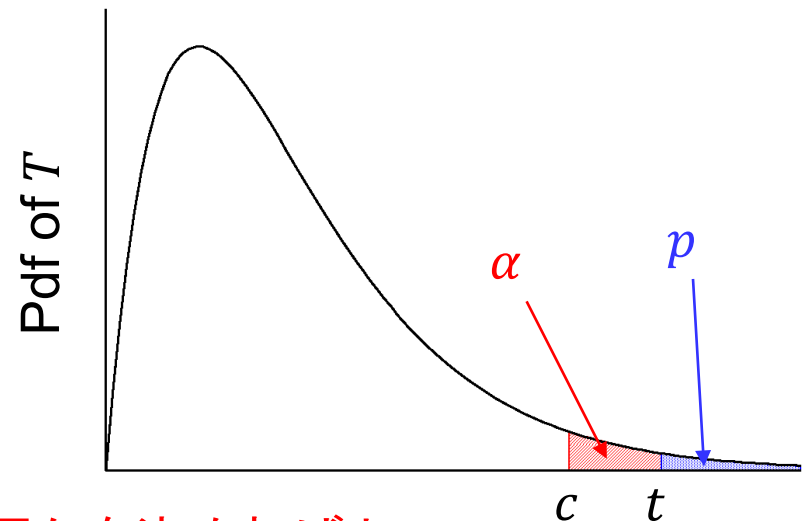
片側検定の場合(Case of a one-sided test)

第1種誤り確率が有意水準 α と等しくなるように、棄却域の閾値 c を設定

Select a threshold c such that the type I error probability is equal to the significance level.

$\mathcal{R} = [c, \infty)$ s.t. $\alpha = P(T \in \mathcal{R} | H_0)$.

p値(p-value): $p = P(T \geq t | H_0)$



有意水準とp値を比べて、棄却するか否かを決めればよい。

Compare the significance level and p-value to decide whether rejection should be selected.

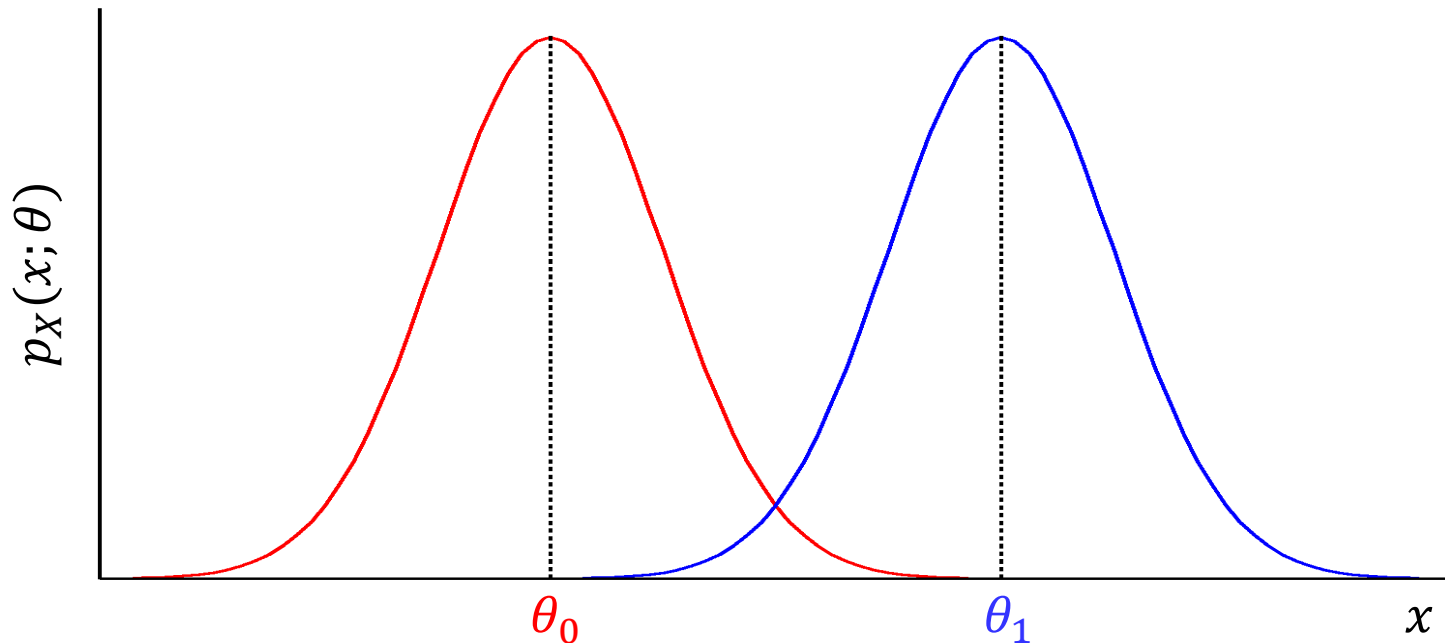
例：母集団分布の検定(Example: Testing for population distributions)

母集団分布は $p_X(x; \theta_0)$ か $p_X(x; \theta_1)$ のどちらかであることを突き止めた。

標本値 x が得られたときに、 $p_X(x; \theta_1)$ が真の分布であることを立証したい。

We have found the fact that the population distribution is $p_X(x; \theta_0)$ or $p_X(x; \theta_1)$. Verify that the latter is the true distribution, when a sample realization x has been obtained.

帰無仮説(Null hypothesis) $H_0 : p_X(x; \theta_0)$ 対立仮説(Alternative hypothesis) $H_1 : p_X(x; \theta_1)$



どのように検定統計量を設定すべきか？(How should a test statistic be set?)

尤度比検定 (Likelihood ratio test)

帰無仮説 (Null hypothesis) $H_0 : p_X(x; \theta_0)$ 対立仮説 (Alternative hypothesis) $H_1 : p_X(x; \theta_1)$

検定統計量 T として、尤度比を取る。

$$T(x) = \frac{p_X(x; \theta_1)}{p_X(x; \theta_0)}$$

(Select the likelihood ratio as a test statistic.)

有意水準 α に対して、棄却域 $\mathcal{R} = [c, \infty)$ の閾値 c を $\alpha = P(T \in \mathcal{R} | H_0)$ を満たすように定める。

For a significance level α , take a threshold c such that $\alpha = P(T \in \mathcal{R} | H_0)$ is satisfied for the rejection region $\mathcal{R} = [c, \infty)$.

補題 7.1: ネイマン・ピアソンの補題 (Lemma 7.1: Neyman-Pearson Lemma)



任意の有意水準 α に対して、尤度比検定は検出力 $1 - \beta$ を最大にする。

For any significance level, the likelihood ratio test is the most powerful test.)

目標とする第1種誤り確率 α を定めたときに、あらゆる検定の中で、尤度比検定は第2種誤り確率 β を最小にする最良な検定法である。

For any target type I error probability α , the likelihood ratio test is the best test that minimizes the type II error probability β among all possible tests.

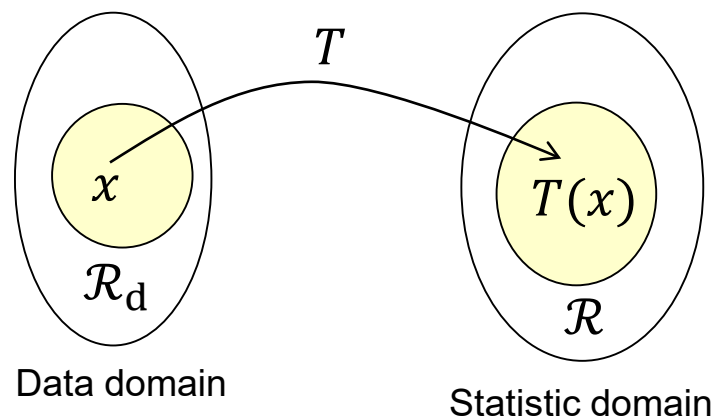
証明(Proof)

最初に記号の準備を行う。

We first introduce the notation.

尤度比統計量 T の棄却域 \mathcal{R} に入る標本値の集合を $\mathcal{R}_d = \{x : T(x) \in \mathcal{R}\}$ と書く。

We write the set of sample realizations that fall in the rejection region for the likelihood ratio statistic as \mathcal{R}_d .



母数 θ に対して、標本値が棄却域 \mathcal{R}_d に入る確率は以下となる。

For a parameter θ , the probability of data falling in the rejection region \mathcal{R}_d is given by

$$\Pr_{\theta}(\mathcal{R}_d) = \int_{\mathcal{R}_d} p_X(x; \theta) dx$$

$\alpha = P(\tilde{T} \in \tilde{\mathcal{R}} | H_0)$ を満たす任意の検定統計量 \tilde{T} と棄却域 $\tilde{\mathcal{R}}$ を取り、上記と同様に $\tilde{\mathcal{R}}_d$ と $\Pr_{\theta}(\tilde{\mathcal{R}}_d)$ を定義する。

Let \tilde{T} and $\tilde{\mathcal{R}}$ denote any test statistic and the corresponding rejection region that satisfy $\alpha = P(\tilde{T} \in \tilde{\mathcal{R}} | H_0)$, and define $\tilde{\mathcal{R}}_d$ and $\Pr_{\theta}(\tilde{\mathcal{R}}_d)$ in the same manner.

検定統計量 T と \tilde{T} の検出力は、それぞれ $\Pr_{\theta_1}(\mathcal{R}_d)$ と $\Pr_{\theta_1}(\tilde{\mathcal{R}}_d)$ である。

The powers of the test statistics T and \tilde{T} are $\Pr_{\theta_1}(\mathcal{R}_d)$ and $\Pr_{\theta_1}(\tilde{\mathcal{R}}_d)$, respectively.

証明(Proof)

二つの検出力の差を評価すると、(Evaluating the difference between the two powers, we have)

$$\begin{aligned}\Pr_{\theta_1}(\mathcal{R}_d) - \Pr_{\theta_1}(\tilde{\mathcal{R}}_d) &= \Pr_{\theta_1}(\mathcal{R}_d \cap \tilde{\mathcal{R}}_d) + \Pr_{\theta_1}(\mathcal{R}_d \cap \tilde{\mathcal{R}}_d^c) \\ &\quad - \{\Pr_{\theta_1}(\mathcal{R}_d \cap \tilde{\mathcal{R}}_d) + \Pr_{\theta_1}(\mathcal{R}_d^c \cap \tilde{\mathcal{R}}_d)\} \\ &= \Pr_{\theta_1}(\mathcal{R}_d \cap \tilde{\mathcal{R}}_d^c) - \Pr_{\theta_1}(\mathcal{R}_d^c \cap \tilde{\mathcal{R}}_d).\end{aligned}\quad (7.1)$$

尤度比検定の棄却域の定義 $\mathcal{R}_d = \{x: p_X(x; \theta_1) \geq c p_X(x; \theta_0)\}$ を使って、

Using the definition of the rejection region for the likelihood ratio test yields

$$\Pr_{\theta_1}(\mathcal{R}_d \cap \tilde{\mathcal{R}}_d^c) = \int_{\mathcal{R}_d \cap \tilde{\mathcal{R}}_d^c} p_X(x; \theta_1) dx \geq c \Pr_{\theta_0}(\mathcal{R}_d \cap \tilde{\mathcal{R}}_d^c).$$

式(7.1)の議論を母数 θ_0 に関して行くと、 $\alpha = \Pr_{\theta_0}(\mathcal{R}_d) = \Pr_{\theta_0}(\tilde{\mathcal{R}}_d)$ から

Repeating the argument in (7.1) for the parameter θ_0 , from $\alpha = \Pr_{\theta_0}(\mathcal{R}_d) = \Pr_{\theta_0}(\tilde{\mathcal{R}}_d)$ we have

$$\Pr_{\theta_0}(\mathcal{R}_d \cap \tilde{\mathcal{R}}_d^c) = \Pr_{\theta_0}(\mathcal{R}_d^c \cap \tilde{\mathcal{R}}_d).$$

尤度比検定の非棄却域 $\mathcal{R}_d^c = \{x: p_X(x; \theta_1) < c p_X(x; \theta_0)\}$ の定義から

Using the definition of the complement of the rejection region for the likelihood ratio test yields

$$c \Pr_{\theta_0}(\mathcal{R}_d^c \cap \tilde{\mathcal{R}}_d) \geq \Pr_{\theta_1}(\mathcal{R}_d^c \cap \tilde{\mathcal{R}}_d) \quad \text{等号(Equality): } \mathcal{R}_d^c = \emptyset.$$

これらを組み合わせると、 $\Pr_{\theta_1}(\mathcal{R}_d) \geq \Pr_{\theta_1}(\tilde{\mathcal{R}}_d)$ が成立する。

Combining these observations, we arrive at $\Pr_{\theta_1}(\mathcal{R}_d) \geq \Pr_{\theta_1}(\tilde{\mathcal{R}}_d)$. ■

例: 分散1の正規分布の場合

Example: Case of the normal distributions with unit variance

$$\text{帰無仮説 (Null hypothesis)} H_0: p_X(x; \theta_0 = 0) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

$$\text{対立仮説 (Alternative hypothesis)} H_1: p_X(x; \theta_1 = 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-1)^2}{2}}$$

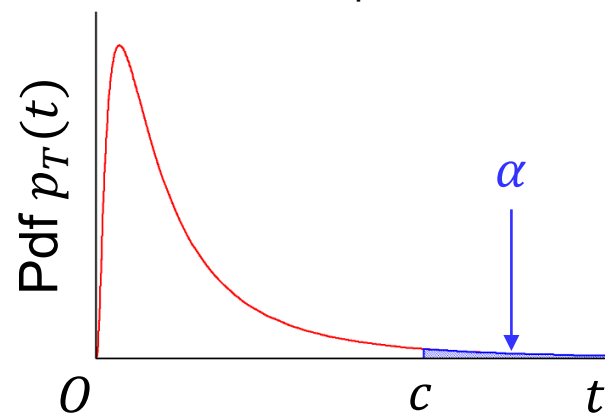
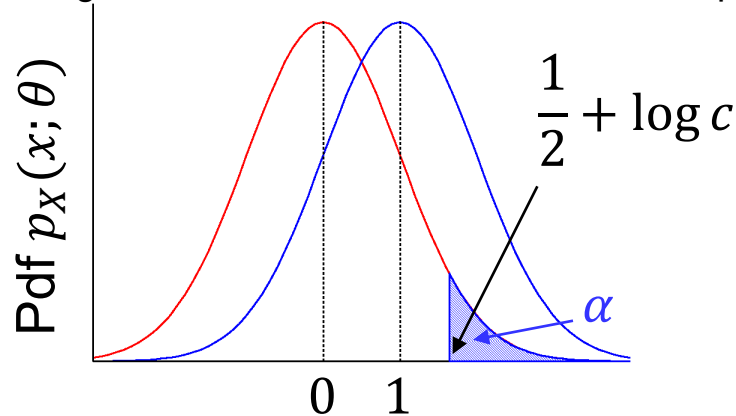
帰無仮説 H_0 の下で対数尤度比を計算すると、

Under the null hypothesis, we compute the log likelihood ratio as

$$\log T(X) = X - \frac{1}{2} \sim \mathcal{N}\left(-\frac{1}{2}, 1\right) \quad \text{確認せよ。 (Confirm it.)}$$

有意水準 $\alpha = 0.05$ に対して閾値を数値計算すると、 $c \approx 3.14198$ 。

For a significance level of 0.05, numerical computation of the threshold implies $c \approx 3.14198$.



標本値 x が閾値 1.64485 以上であれば、帰無仮説は棄却される。

When a realization of the sample is larger than the threshold, the null hypothesis is rejected.