Essential Mathematics for Global Leaders I

Spring 2019

Statistics

Lecture 7: 2019 June 10th – 17th

Xavier DAHAN
Ochanomizu Graduate Leading Promotion Center

Office:理学部2号館503
mail: dahan.xavier@ocha.ac.jp

Where are we ? Today's plan

PART II: Statistical inference (推計統計学)

4. Null Hypothesis Significant Test (NHST) 帰無仮説検定

4.1 Concepts and 1st example: z-test

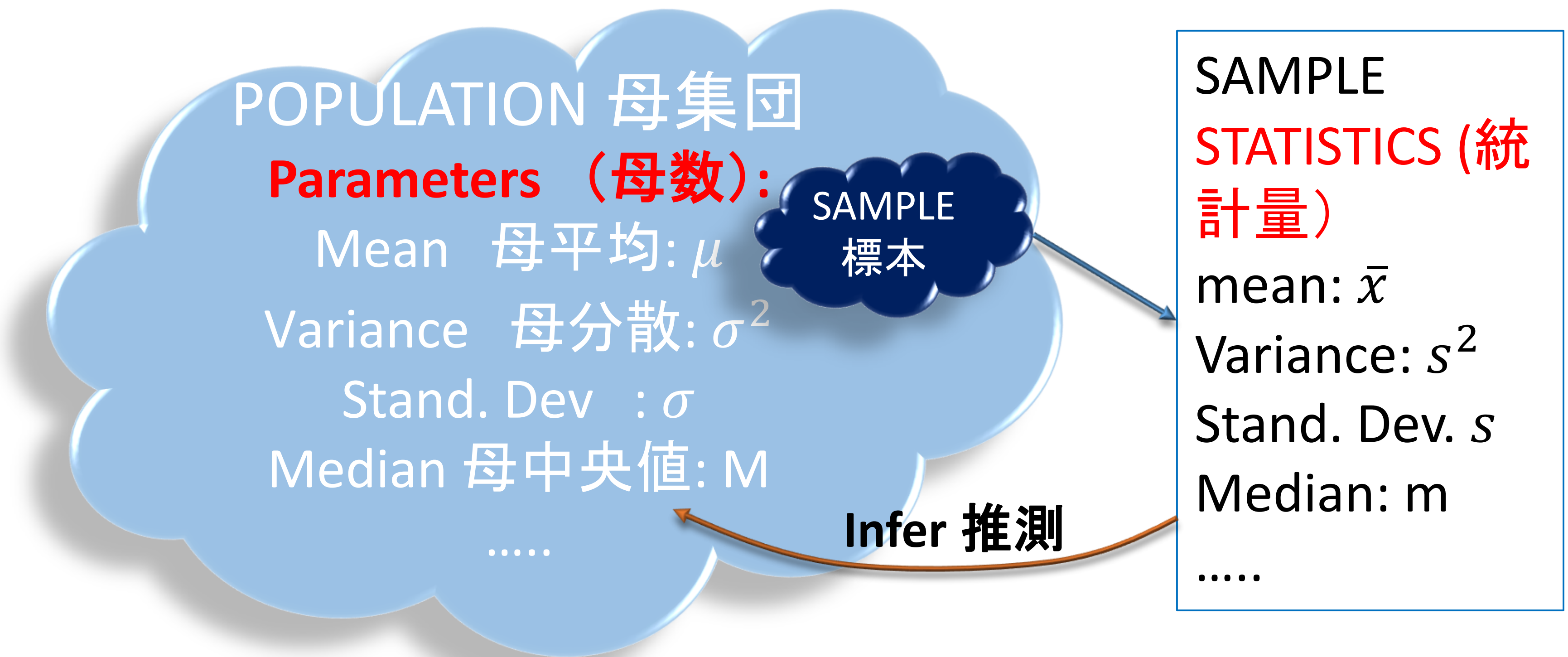
4.2 chi2 and sample variance （カイ二乗と標本分散）
4.3 the Student t-test （Student t-検定)
4.4 Two-sample t-test (=variance) 対応のないt検定
4.5 Paired difference sampling 対応のあるデータ
4.6 Comparing two population variances: F-test

4.7 chi2-test (goodness-of-fit)カイ二乗(簡単な適合度検定)
4.8 chi2-test of independence 独立性のカイ二乗検定
4.9 One-way ANOVA (F-test) 一元配置分散分析 (F検定)

Chapter 4: NHST

4.1 Concept & 1st example: the *z*-test

Review of sampling (Chapter 3)



Introduction – Goals of this section
章4.1の目的

1. Know the definitions of the significance testing terms 有意性検定の用語:
   - NHST, null hypothesis 帰無仮説,
   - alternative hypothesis 対立仮説,
   - simple hypothesis 単純仮説,
   - composite hypothesis 複合仮説,
   - Types of error 種過誤
   - significance level 有意水準
   - power 検出力

2. Be able to design and run a significance test for Bernoulli or binomial data.
   ベルヌーイと二項分布に従うデータ用の有意検定を設計して行うことができる。

3. Be able to compute a p-value for a normal hypothesis and use it in a significance test.
   正規分布の仮定におけるp値を算出して、有意性検定の際に利用することができる。

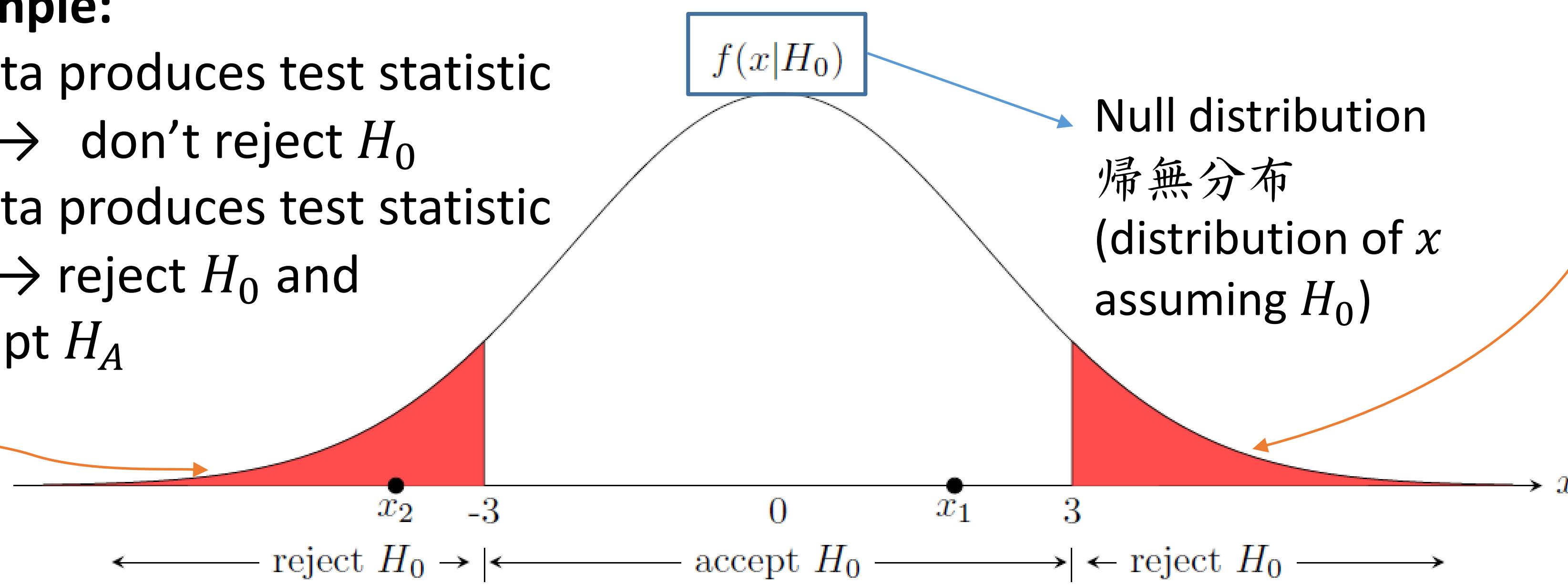
NHST ingredients (NHSTの材料)

- Null hypothesis (帰無仮説) : H_0
- Alternative hypothesis (対立仮説): H_A
- Test statistic: 統計量検定 x
  (function of the sample サンプルに付随するもの)
- **Rejection (critical) region**: (棄却域)
  reject H_0 in favor of H_A if x is in this region.
  xはこの域にあれば、H_0の代わりにH_Aを選ぶ。

Example:

● data produces test statistic x_1 → don't reject H_0

● data produces test statistic x_2 → reject H_0 and accept H_A



Null distribution
帰無分布
(distribution of x assuming H_0)

NHST for Bernoulli: coin is fair?

- Coin with probability of heads θ. ($\sim Bernoulli(\theta)$)
- Test statistic x = the number of heads in 10 tosses
  統計量検定x=コイン投げ１０回の中、表の個数
  (population: all (infinite) tosses.　　Sample: 10 tosses)
- H_0: "the coin is fair", → $\theta = .5$
- H_A: "the coin is biased", → $\theta \neq .5$

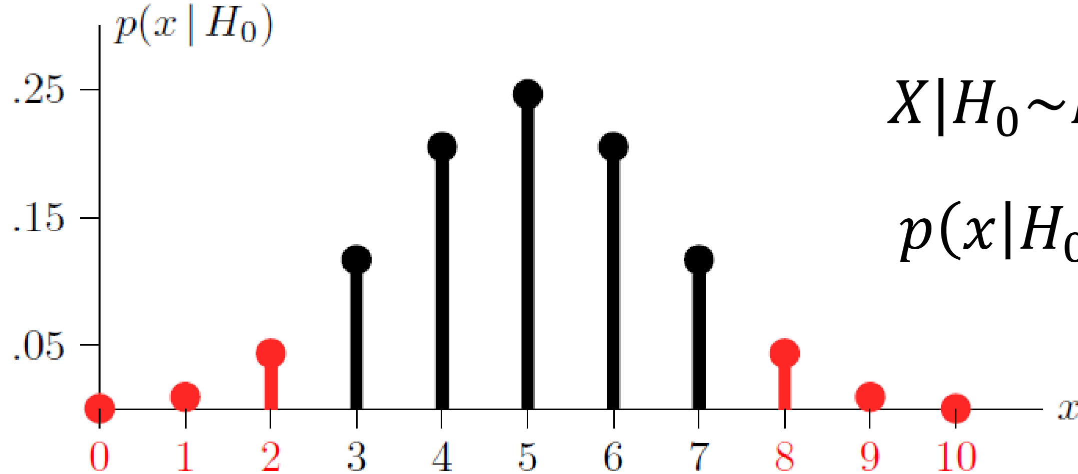
Two strategies: (作戦が二つ)

1. Choose rejection region then compute significance level
   棄却域を選択してから有意水準を算出する。
2. Choose significance level then determine rejection region
   有意水準を選択してから棄却域を確定する。

Everything is done assuming H_0

Is coin fair? The two strategies (I)

1. The rejection region is bordered in red, what's the significance level (=sum of probability in the rejection region)?
棄却域は赤い部分に囲まれる。有意水準(=棄却域にある確率の和)は何か？



$$X|H_0 \sim Binomial(10,0.5)$$

$$p(x|H_0 = k) = \binom{10}{k}\frac{1}{2^k}$$

x	0	1	2	3	4	5	6	7	8	9	10
$p(x\|H_0)$	.001	.010	.044	.117	.205	.246	.205	.117	.044	.010	.001

Significance level (有意水準) $\alpha = 0.11$

Is coin fair? The two strategies (II)

2. Given significance level $\alpha = .05$ find a two-sided rejection region.
有意水準を$\alpha = .05$として、両側棄却域を求めよ。

x	0	1	2	3	4	5	6	7	8	9	10
$p(x \mid H_0)$	.001	.010	.044	.117	.205	.246	.205	.117	.044	.010	.001

$\alpha = .05$ Sum of the probability of the two-sided rejection region is $\leq \alpha$
両側棄却域にある確率の和$\leq \alpha$

x	0	1	2	3	4	5	6	7	8	9	10
$p(x\|H_0)$	.001	.010	.044	.117	.205	.246	.205	.117	.044	.010	.001

$2 \times 0.001 + 2 \times 0.01 = 0.022 \leq 0.05$ but
$2 \times 0.001 + 2 \times 0.01 + 2 \times 0.044 > 0.05$

Simple and composite hypotheses
単純と複合仮説

- **Definition:** A simple hypothesis is one for which we can specify its distribution completely. A typical simple hypothesis is that a parameter of interest takes a specific value.

  Example: H_0 : Coin is fair? → $Bernoulli(\theta), \theta = \frac{1}{2}$

- **Definition:** If its distribution cannot be fully specified, we say that the hypothesis is composite. A typical composite hypothesis is that a parameter of interest lies in a range of values.

  Example: H_A: Coin is not fair→ $Bernoulli(\theta), \theta \neq \frac{1}{2}$ but what is θ ?

Simple and Composite hypotheses (II)

- **Example:** Suppose we have data $x_1, \dots, x_n$. Suppose also that our hypotheses are
  H_0: the data is drawn from $N(0,1)$
  H_A: the data is drawn from $N(1,1)$.

A. Only H_0 is simple B. Only H_A is simple
C. Both H_0 and H_A are simple D. Only H_0 is composite
E. Only H_A is composite F. Both H_0 and H_A are composite

- **Example:** Now suppose that our hypotheses are
  H_0: the data follow $Exponential(\lambda)$, λ unknown
  H_A: the data do not follow an $Exponential$ distribution

A. Only H_0 is simple B. Only H_A is simple
C. Both H_0 and H_A are simple D. Only H_0 is composite
E. Only H_A is composite F. Both H_0 and H_A are composite

Representative example: **z-tests**, p-values

- **Assumption:** Population has distribution $Normal(\boldsymbol{\mu}, \sigma^2)$

  know σ, don't know μ. **Aim: find $\boldsymbol{\mu}$**

- **Data**: $x_1, \ldots, x_n$ (taken from a population of $N(\mu, \sigma^2)$)
- **Hypotheses**: H_0: $x_i \sim N(\mu_0, \sigma^2)$

  H_A: Two-sided if $\mu \neq \mu_0$, or one sided if $\mu > \mu_0$
- **Test statistic**: $\bar{x}$ (sample mean)
- **Null Distribution**: $\bar{x} \sim N(\mu_0, \sigma^2/n)$ (Lecture 5, slide 10)
- **z-value**: standardized $\bar{x}$: $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ (Lecture 4, Slide 18)
- **p-values**: Two-sided p-**value**: $p = P(|Z| > z|H_0)$

  Right-sided p-**value**: $p = P(Z > z|H_0)$
- **Significance level** α: For $p \leq \alpha$ we reject H_0 in favor of H_A.

Method 1: find the p-value

- $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} = \frac{112-100}{15/\sqrt{9}} = \frac{12}{5} = 2.4.$ (z-value)
- Right-sided p-value: $p = P(Z > 2.4|H_0) \leq \alpha$
- $Z|H_0 \sim N(0,1)$. Table gives: $p \approx 0.008 \leq 0.05$
- p-value < significance level, therefore the z-test rejects H_0 in favor of $H_A: \mu > 100$.

Most common method

Method 2: find the rejection region

- Sample of size 9: $\bar{X}_9 \sim Normal\left(\mu_0, \frac{\sigma^2}{9}\right)$
- Rejection region: $P(\overline{X_9} \geq z_\alpha) \leq \alpha$
- $\alpha = 0.05$ significance level. $z_\alpha = z_{.05} \approx 108.2$
- Rejection region: $[108.2, +\infty)$.
- 112 is in the rejection region so the z-test rejects H_0

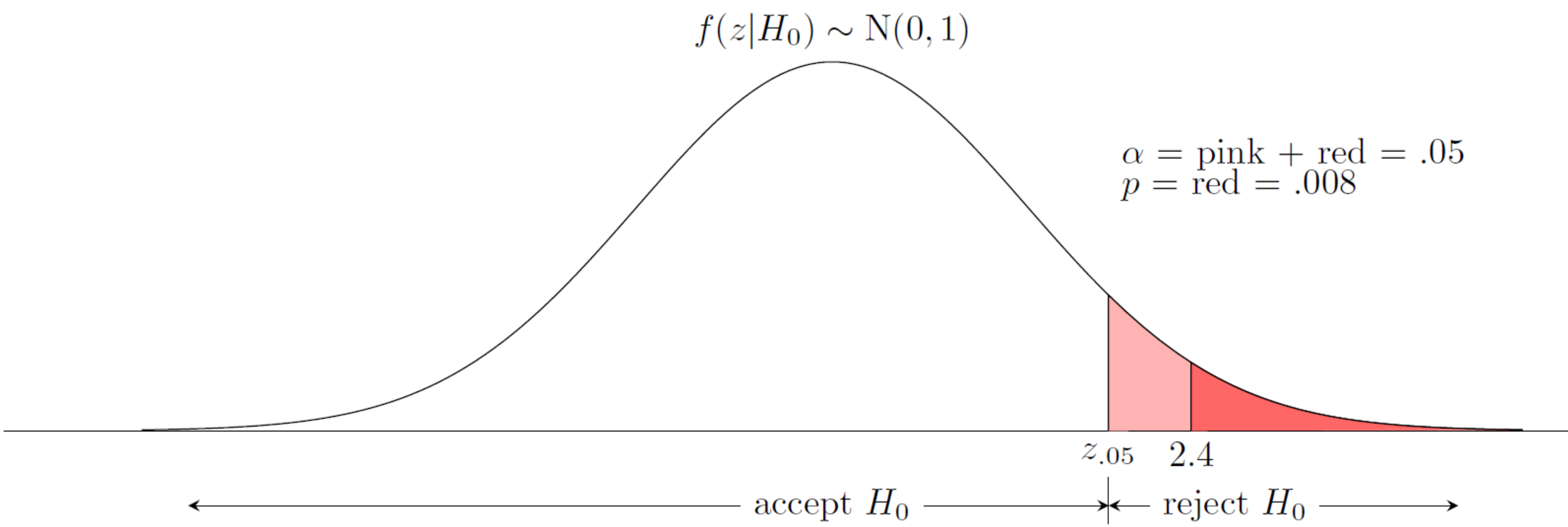
Found with a computer

Visualization (Example)

- Population data follows a normal distribution $N(\mu, 15^2)$ where μ is unknown.

  H_0: $\mu = \mu_0 = 100$
  H_A: $\mu > 100$ (one-sided)

- Collect 9 data points $x_1, \ldots, x_9$: sample mean is $\bar{x} = 112$.
- Can we reject H_0 at significance level 0.05?



Exercise (z-test)

- H_0: data follows a $N(5, 10^2)$
- H_A : data follows a $N(\mu, 10^2)$ where $\mu \neq 5$.
- Test statistic: $\bar{x}$ the average of the data (sample mean).
- Data: 64 data points $x_1, \ldots, x_{64}$ with $\bar{x}$ = 6.25.
- Significance level set to α = .05.
  i. What is the null distribution $f(\bar{x}|H_0)$?
  ii. Find the z-value.
  iii. Find the (one-sided or two-sided?) p-value for this z-value.
  iv. Decide whether or not to reject H_0 in favor of H_A
  v. Find the rejection region.

- Answer

  i. $f(x_0 = t|H_0) = \frac{1}{10\sqrt{2\pi}} e^{-\frac{(t-5)^2}{200}}$

  ii. $z = \frac{\bar{x}-5}{10/\sqrt{n}} = \frac{\bar{x}-5}{10/8} = 1$

  iii. H_A is $\mu \neq 5$ so it is a two-sided p-value.
  $P(|Z| > z|H_0) = P(|Z| > 1) = 2\big(1 - P(Z \leq 1)\big)$
  Table gives: $2(1 - 0.8413) \approx 0.316$

  iv. 2-sided p-value $0.316 > \alpha = 0.05$
  So we do not reject H_0

  v. Rejection region. Need to find critical values c_α (see page 27 & 29).
  $P(|Z| > c_\alpha) \leq 0.05 \quad P(Z \leq c_\alpha) \geq 0.975$
  Table gives $\mathrm{c}_\alpha \approx 2$.

$$\left|\frac{\bar{x} - 5}{1.25}\right| \geq 2 \;\Rightarrow\; |\bar{x} - 5| \geq 2.5 \;\Rightarrow (-\infty, 2.5] \cup [7.5, \infty)$$

Exercise (NHST for binomial)

- Two coins C_1 and C_2:
  probability of heads is .5 for C_1; and .6 for C_2.

➢ We pick one at random, flip it 8 times and get 6 heads.

1. H_0 = 'The coin is C_1' H_A = 'The coin is C_2'
   Do you reject H_0 at the significance level $\alpha = .05$?
2. H_0 = 'The coin is C_2' H_A = 'The coin is C_1'
   Do you reject H_0 at the significance level $\alpha = .05$?
3. Do your answers to (1) and (2) seem paradoxical?

Here are $Binomial(8, \theta)$ tables for $\theta = .5$ and .6.

k	0	1	2	3	4	5	6	7	8
$p(k\|\theta = .5)$	.004	.031	.109	.219	.273	.219	.109	.031	.004
$p(k\|\theta = .6)$	.001	.008	.041	.124	.232	.279	.209	.090	.017

Answer:

1. Since $0.5 < 0.6$ the alternative distribution $X|H_A$ is on the right of the null distribution $X|H_0$.
   - We consider a __right__ -sided rejection region.
   - (line $\theta = 0.5$ of the table) Sum of probabilities from the right is smaller than 0.05
     We find the rejection region: __{7,8}__
     Since__$6 \notin \{7,8\}$__, we ~~reject~~/do not reject H_0

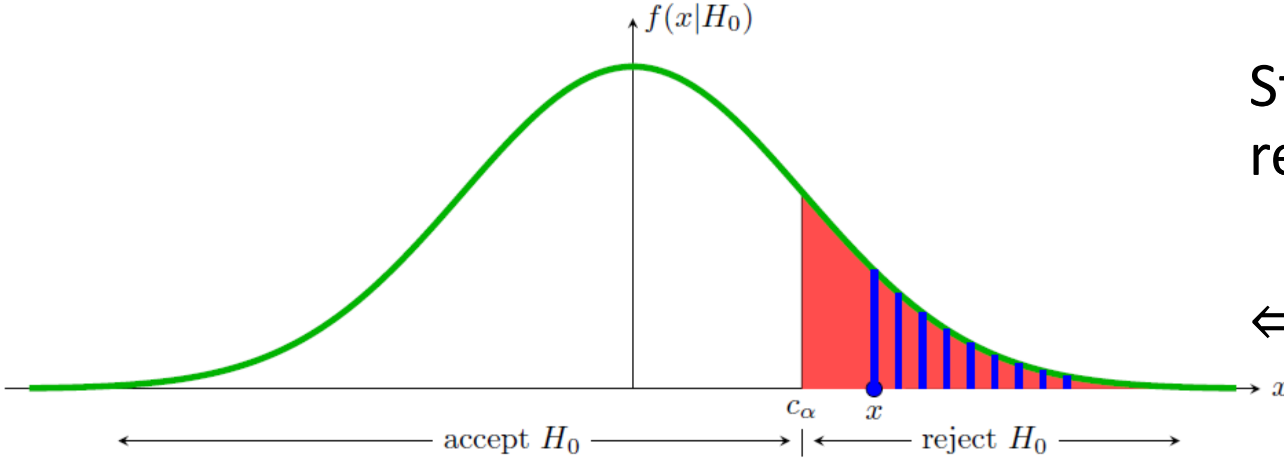
2. This time the alternative distribution $X|H_A \sim Binomial(8,0.5)$ is on the left of the null hypothesis $X|H_0 \sim Binomial(8,0.6)$.
   - We consider a __left__ -sided rejection region.
   - We find rejection region (on the line $\theta = 0.6$)__{0,1,2}__
   - Since__$6 \notin \{0,1,2\}$__, we ~~reject~~/do not reject H_0.

3. No paradox____. We ~~can~~/ cannot say which coin is it at this level

   → need more than 8 tosses.

p-values and critical values

- p-values are not only defined for the z-test but for any simple null hypothesis H_0.
- Area (面積) in red $= P(rejection\ region|H_0) = \alpha$



Statistic x inside rej. region

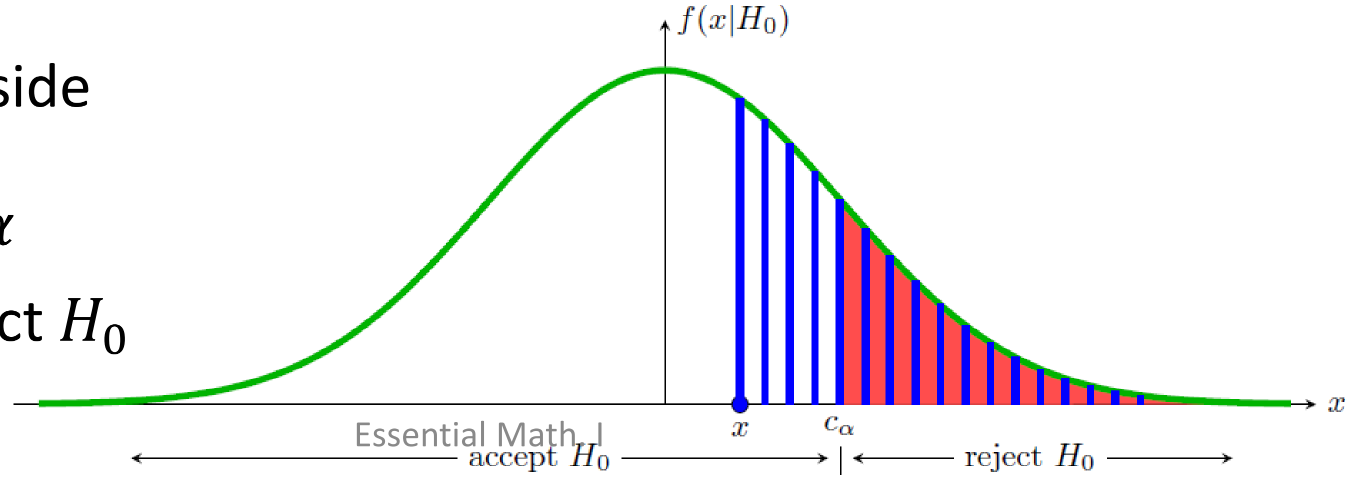
$\Leftrightarrow p < \alpha$

$\Leftrightarrow$ Reject H_0

Statistic x outside rej. region

$\Leftrightarrow p > \alpha$

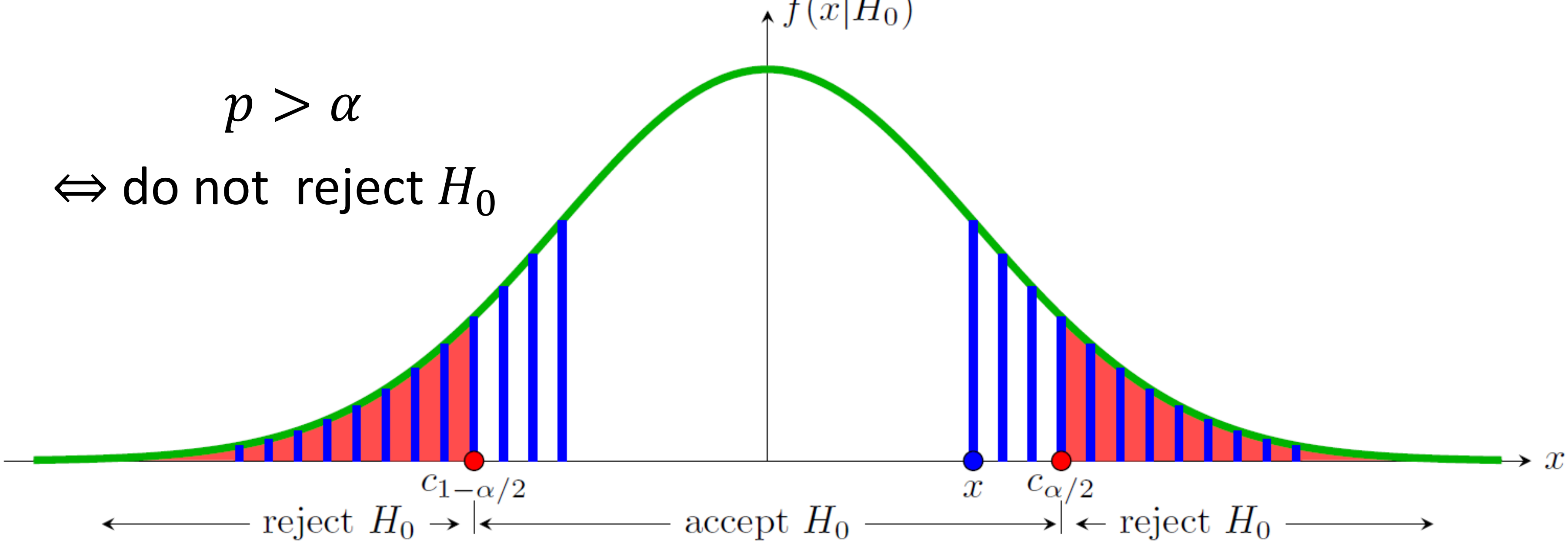
$\Leftrightarrow$ do not reject H_0



Two-sided p-values and critical values

$p > \alpha$

$\Leftrightarrow$ do not reject H_0



Critical values

- The boundary of the rejection region are called critical values. 棄却値=棄却域の境界点
- Critical values are labeled by the probability to their right.
  $P(X > c_\alpha) \leq \alpha$ **(left-sided)**
  $P(X > c_{1-\alpha}) \leq 1 - \alpha \quad \Rightarrow \quad P(X < c_{1-\alpha}) \leq \alpha$ **(right-sided)**
  2-sided
  $$P(|X| > c_{\alpha/2}) = P(X > c_{\alpha/2}) + P(X < c_{1-\alpha/2}) \leq \alpha$$

Type I and II error of a NHST
第一種過誤　と　第二種過誤

		True state of nature 正自然狀態	
		H_0	H_A
Our decision	Reject H_0	Type I error	Correct decision
	“accept” H_0	Correct decision	Type II error

Type I : false rejection of H_0 ⟷ False positive
第一種過誤：偽陽性 ⟷ Convincing an innocent
(無罪の人に有罪判定を下す)

Type II: false “acceptance” of H_0 ⟷ False negative
第二種過誤：偽陰性 ⟷ Acquitting a guilty person
(有罪の人を放免する)

Significance level and power of a NHST
有意水準と検出力

Significance level $= P(type\ I\ error\ \text{第一種過誤})$

$= \text{probability we incorrectly reject } H_0$

$= P(test\ statistic\ in\ rejection\ region\ |H_0)$

Power $= \text{probability we correctly reject } H_0$

$= P(test\ statistic\ in\ rejection\ region\ |H_A)$

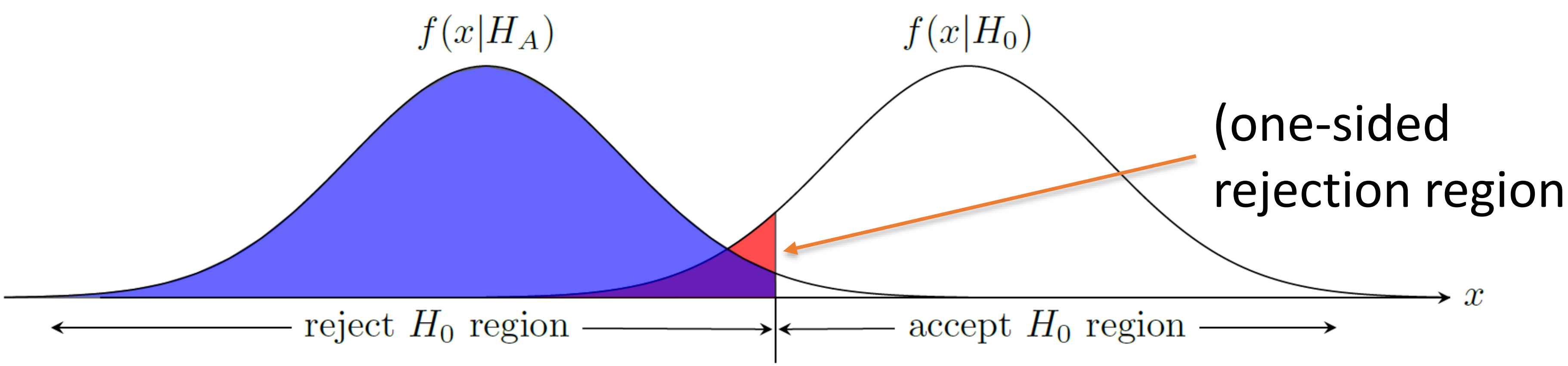
$= 1 - P(type\ II\ error)$

- Power can be computed for simple hypothesis H_A.
- Otherwise power varies in function of the (unknown) parameters.

Want significance level near 0 and power near 1

Question

Suppose that the null and alternative hypotheses H_0 and H_A are both simple.
帰無仮説 H_0 も対立仮説 H_A も単純であるとする。
Therefore they have a determined distribution $f(x|H_A)$ and $f(x|H_0)$, shown below.
ゆえに、両方は定めた分布 $f(x|H_0)$ と $f(x|H_A)$ を持つ

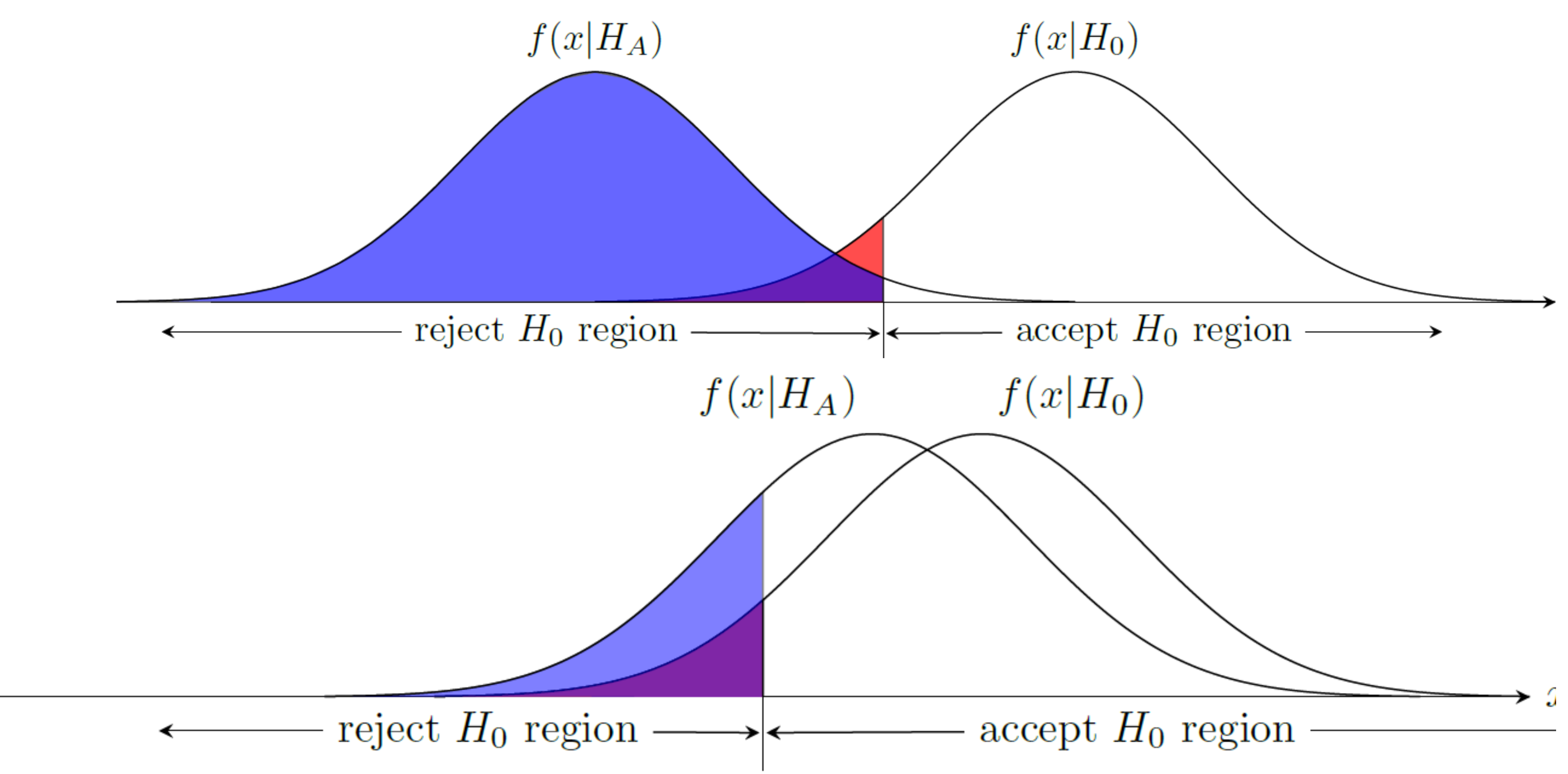


The significance level of the test is given by the area of which region?

1. red
2. purple
3. blue
4. white
5. blue + purple
6. red + purple
7. white + red + purple.

Which test has higher power ?

- Two tests have both simple hypotheses H_A and H_0

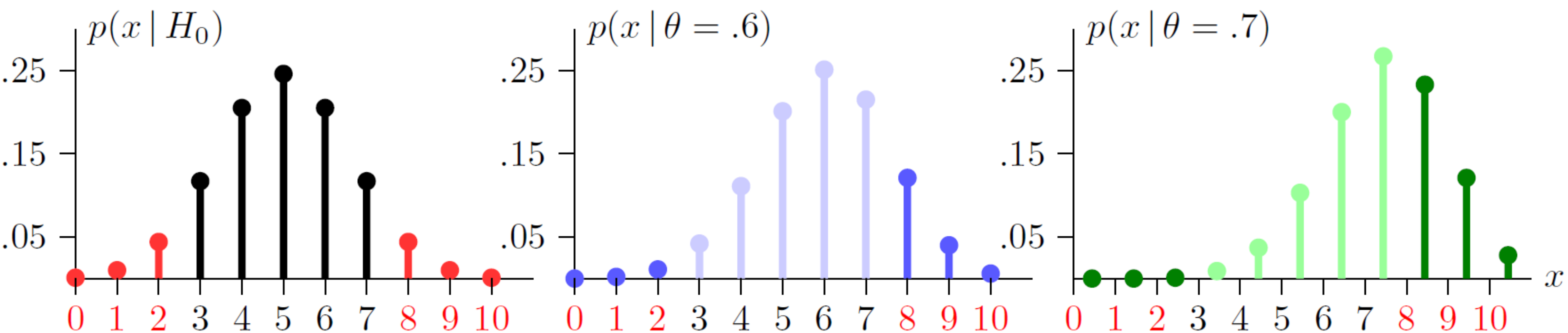


TOP BOTTOM

Back to testing a fair coin

- Composite alternative hypothesis H_A: 複合対立仮説
  coin is not fair $f(x|H_A) \sim Binomial(10, \theta),\ \ \theta \neq \frac{1}{2}$
- Let's try some simple alternative hypotheses: $\theta = 0.6,\ 0.7$
  単純対立仮説を試してみよう： $\theta = 0.6,\ 0.7$

x	0	1	2	3	4	5	6	7	8	9	10
H_0: $p(x\|\theta = .5)$	.001	.010	.044	.117	.205	.246	.205	.117	.044	.010	.001
H_A: $p(x\|\theta = .6)$	.000	.002	.011	.042	.111	.201	.251	.215	.121	.040	.006
H_A: $p(x\|\theta = .7)$	.000	.0001	.001	.009	.037	.103	.200	.267	.233	.121	.028



- Significance level　有意水準
  = probability we reject H_0 when it is true
  H_0は正しいときにH_0を棄却する確率.

  = probability the test statistic is in the rejection region when H_0 is true
  H_0は正しいときに棄却域に位置する検定統計量の確率

  = probability to be in the rejection region in the H_0 row of the table
  表の H_0行の棄却域にある確率

  = sum of red boxes in the $\theta = 0.5$ row
  $\theta = 0.5$行での 赤いボックスの和

  = .11

- Power when $\theta = 0.6$
  = probability we reject H_0 when $\theta = 0.6$
  $\theta = 0.6$のとき、H_0を棄却する確率。

  = probability the test statistic is in the rejection region when $\theta = 0.6$
  $\theta = 0.6$のとき、棄却域にある検定統計量の確率

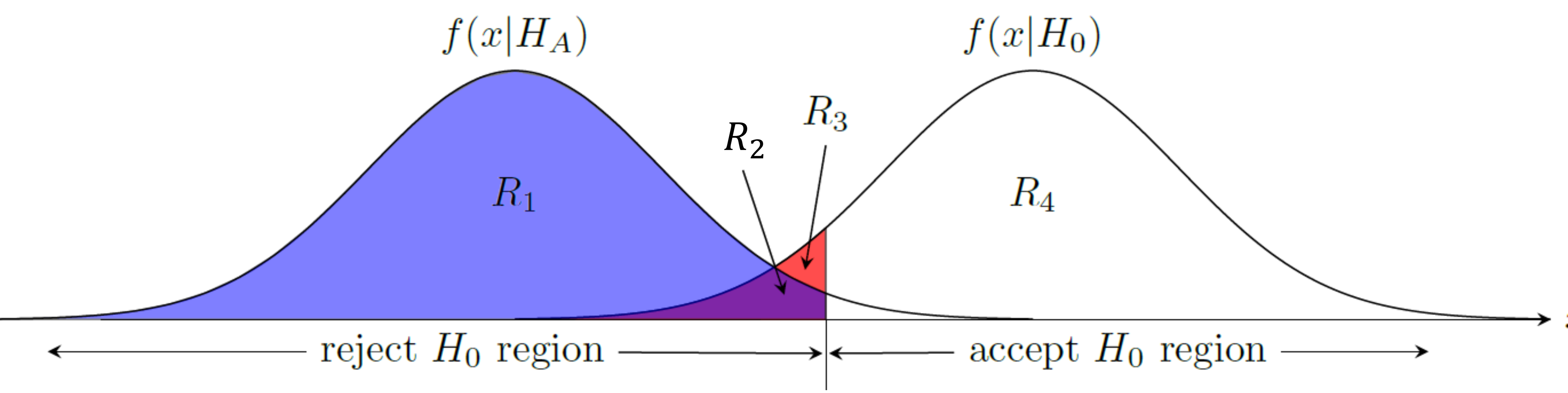
  = probability to be in the rejection region in the $\theta = 0.6$ row of the table　表の$\theta = 0.6$行にある確率。

  = sum of dark blue boxes in the $\theta = 0.6$ row
  $\theta = 0.6$行の紺青色ボックスの和
  = .180

- Power when $\theta = 0.7$
  ⋮
  = sum of dark green boxes in the $\theta = 0.7$ row
  $\theta = 0.7$行の暗緑色ボックスの和
  =0.384

Question

- A test has both simple hypotheses H_0 and H_A
- What is the area in the graph below which gives the power of the test?



1. R_1 2. R_2 3. $R_1 + R_2$ 4. $R_1 + R_2 + R_3$

Power=$P(\text{rejection region}|H_A)$

Question

- The null distribution for test statistic x is $N(4, 8^2)$.
  検定統計量帰無xに対する分布
  The rejection region is $\{x \ \geq \ 20\}$.
- What are the significance level and power of this test?

Answer:

- Significance level
  $P(x > 20|H_0) = P\left(\frac{x-4}{8} > \frac{20-4}{8}\right) = P(z > 2) \approx 0.0227.$
- Power: cannot compute it without alternative distribution (対立分布$P(x \leq \cdots |H_A)$ は必要)

Practice Exercise (Report)

Problem 1. Polygraph (=Lie detector) analogy. (うそ発見器)

In an experiment on the accuracy of polygraph tests, 140 people were instructed to tell the truth and 140 people were instructed to lie.

うそ発見器の正確さに対する実験では、140人に真実を言うよう、他の１４０人にうそつくように指示された。

Testers use a polygraph to guess whether or not each person is lying. 人がうそつくかどうか推測するためにテストを受けさせる人はうそ発見器を使う。

By analogy, let's say H_0 corresponds to the testee (テストを受ける人) telling the truth and H_A corresponds to the testee lying.

	Testee is truthful	Testee is lying
Tester thinks testee is truthful	131	15
Tester thinks tested is lying	9	125

a) Describe the meaning of type I and type II errors in this context, and estimate their probabilities based on the above table.
第一と第二種過誤の意味を書いて、上の表の元に基づいてそれぞれの確率を求めなさい。

b) In NHST, what relationships exist between the terms "significance level", "power", "type 1 error", and "type 2 error"?

Problem 2 (z-test)

Suppose we have 49 data points with sample mean $\bar{x} = 6.25$ and sample variance 100. We want to test the following hypotheses

H_0: the data is drawn from a $N(4, 10^2)$ distribution.

データが正規分布$N(4,10^2)$から抽出された。

H_A: the data is drawn from $N(\mu, 10^2)$ where $\mu \neq 4$.

データが正規分布$N(\mu, 10^2), \mu \neq 4$から抽出された

a) Test for significance at the $\alpha = 0.05$ level. (Use the z-table of $N(0,1)$) to compute the relevant p-value.

b) Draw (roughly) a picture showing the null pdf, the rejection region and the area used to compute the p-value.