Hypothesis testing and statistical decision theory

Lirong Xia

Rensselaer

Fall, 2016

Schedule

- Hypothesis testing
- Statistical decision theory
  - a more general framework for statistical inference
  - try to explain the scene behind tests
- Two applications of the minimax theorem
  - Yao's minimax principle
  - Finding a minimax rule in statistical decision theory

An example

- The average GRE quantitative score of
  - RPI graduate students vs.
  - national average: 558(139)
- Randomly sample some GRE Q scores of RPI graduate students and make a decision based on these

Simplified problem: one sample location test

- You have a random variable X
  - you know
    - the shape of X: normal
    - the standard deviation of X: 1
  - you don't know
    - the mean of X

The null and alternative hypothesis

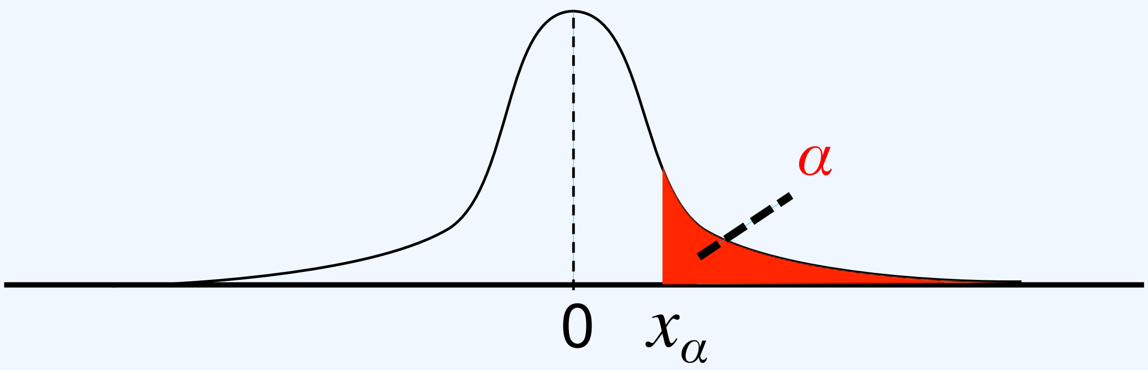
- Given a statistical model
  - parameter space: Θ
  - sample space: S
  - $\Pr(s|\theta)$
- H_1: the alternative hypothesis
  - $H_1 \subseteq \Theta$
  - the set of parameters you think contain the ground truth
- H_0: the null hypothesis
  - $H_0 \subseteq \Theta$
  - $H_0 \cap H_1 = \varnothing$
  - the set of parameters you want to test (and ideally reject)
- Output of the test
  - reject the null: suppose the ground truth is in H_0, it is unlikely that we see what we observe in the data
  - retain the null: we don't have enough evidence to reject the null

One sample location test

- Combination 1 (one-sided, right tail)
  - H_1: mean>0
  - H_0: mean=0 (why not mean<0?)
- Combination 2 (one-sided, left tail)
  - H_1: mean<0
  - H_0: mean=0
- Combination 3 (two-sided)
  - H_1: mean≠0
  - H_0: mean=0
- A hypothesis test is a mapping f: S⟶{reject, retain}

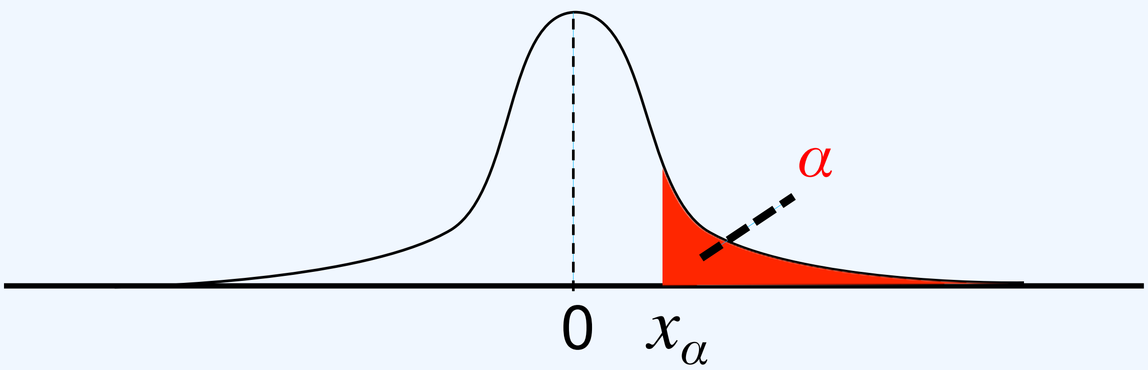
One-sided Z-test

- H_1: mean>0
- H_0: mean=0
- Parameterized by a number $0<\alpha<1$
  - is called the level of significance
- Let x_α be such that $\Pr(X>x_\alpha|H_0)=\alpha$
  - x_α is called the critical value



- Output reject, if
  - $x>x_\alpha$, or $\Pr(X>x|H_0)<\alpha$
    - $\Pr(X>x|H_0)$ is called the p-value
- Output retain, if
  - $x\leq x_\alpha$, or p-value$\geq\alpha$

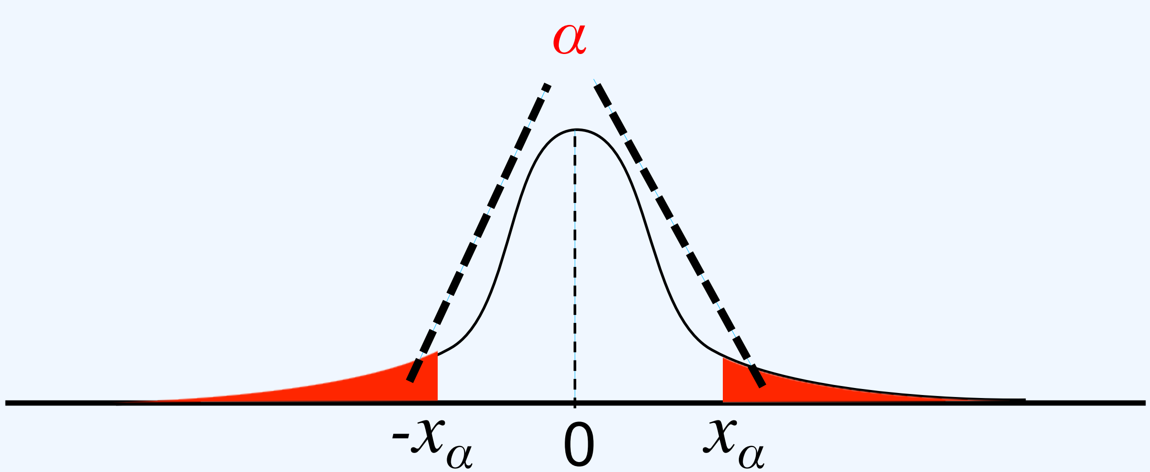
Interpreting level of significance



- Popular values of α:
  - 5%: x_α= 1.645 std (somewhat confident)
  - 1%: x_α= 2.33 std (very confident)
- α is the probability that given mean=0, a randomly generated data will leads to “reject”
  - Type I error

Two-sided Z-test

- H_1: mean≠0
- H_0: mean=0
- Parameterized by a number $0<\alpha<1$
- Let x_α be such that $2\Pr(X>x_\alpha|H_0)=\alpha$



- Output reject, if
  - $x>x_\alpha$, or $x<x_\alpha$
- Output retain, if
  - $-x_\alpha \le x \le x_\alpha$

Evaluation of hypothesis tests

- What is a “correct” answer given by a test?
  - when the ground truth is in H_0, retain the null (≈saying that the ground truth is in H_0)
  - when the ground truth is in H_1, reject the null (≈saying that the ground truth is in H_1)
  - only consider cases where $\theta \in H_0 \cup H_1$
- Two types of errors
  - Type I: wrongly reject H_0, false alarm
  - Type II: wrongly retain H_0, fail to raise the alarm
  - Which is more serious?

Type I and Type II errors

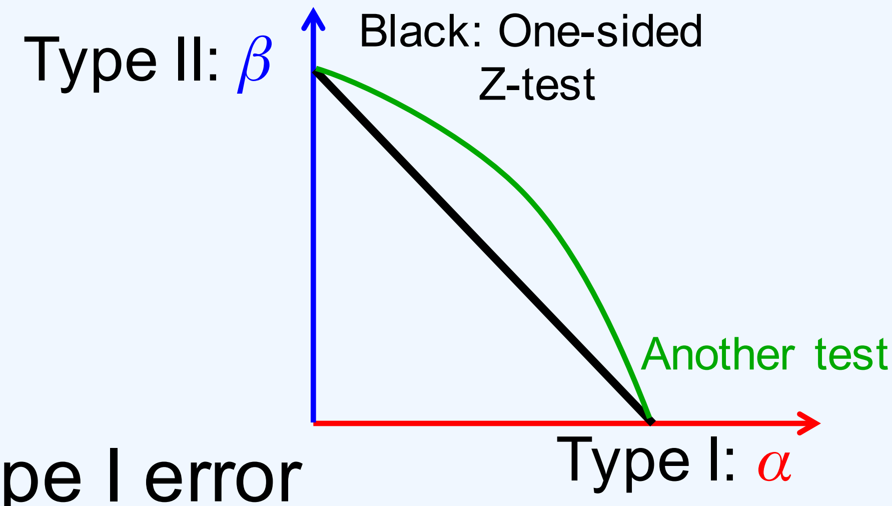
		Output	
		Retain	Reject
Ground truth in	H_0	size: $1-\alpha$	Type I: α
	H_1	Type II: β	power: $1-\beta$

- Type I: the max error rate for all $\theta \in H_0$

$$\alpha = \sup_{\theta \in H_0} \Pr(\text{false alarm} | \theta)$$

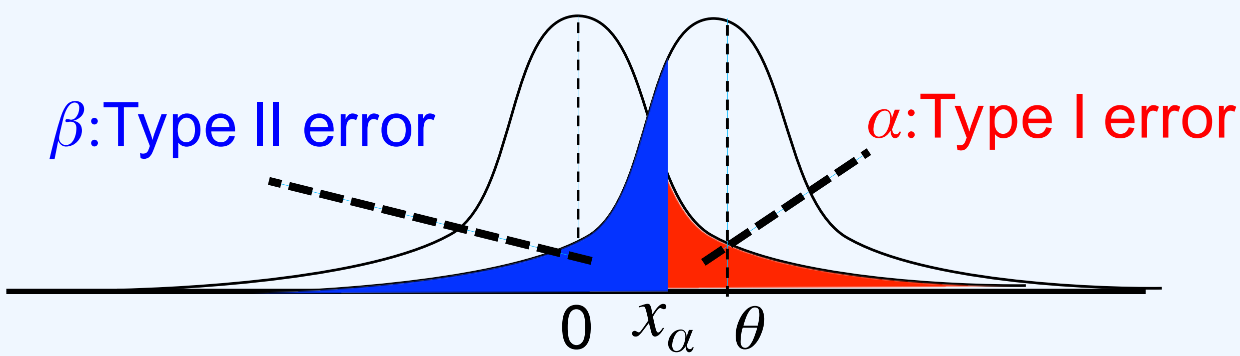
- Type II: the error rate given $\theta \in H_1$
- Is it possible to design a test where $\alpha=\beta=0$?
  - usually impossible, needs a tradeoff

Illustration



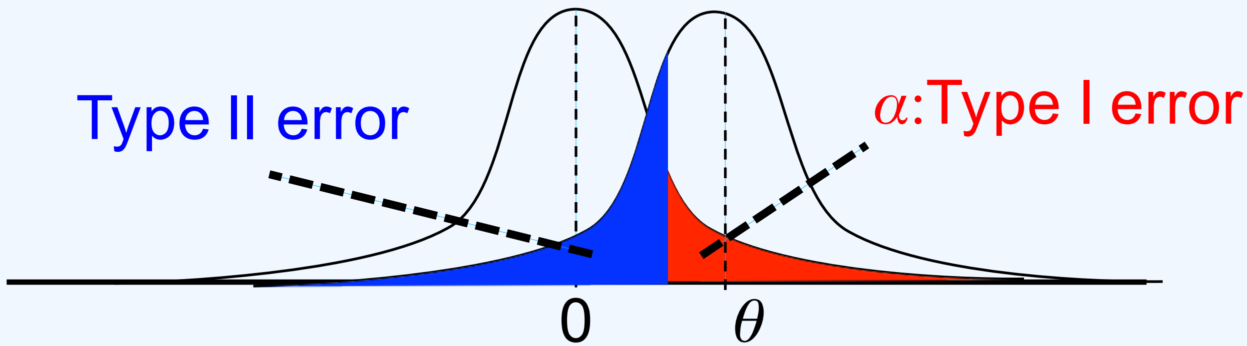
- One-sided Z-test
  - we can freely control Type I error
  - for Type II, fix some $\theta \in H_1$

		Output	
		Retain	Reject
Ground truth in	H_0	size: $1-\alpha$	Type I: α
	H_1	Type II: β	power: $1-\beta$

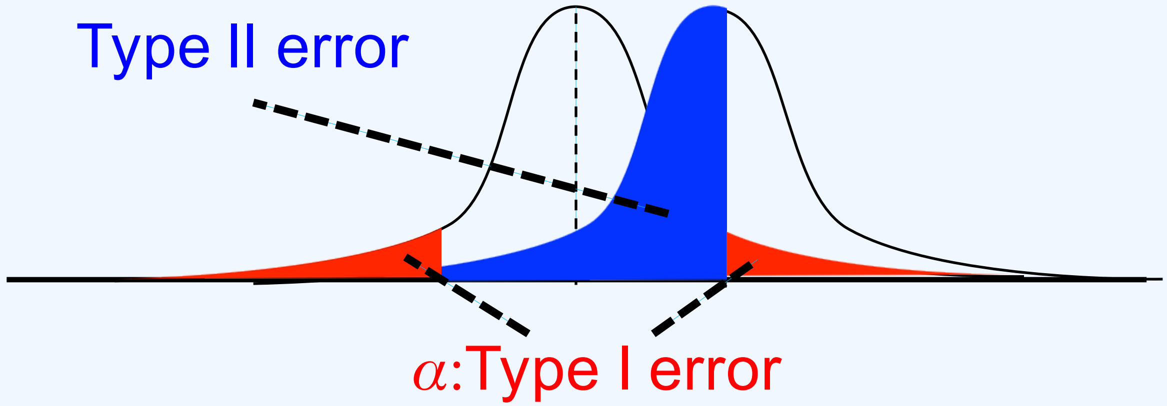


Using two-sided Z-test for one-sided hypothesis

- Errors for one-sided Z-test



- Errors for two-sided Z-test, same α



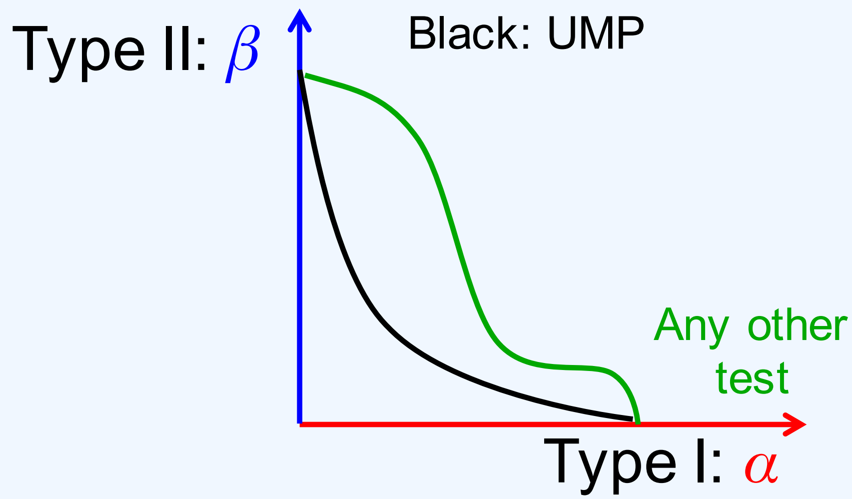
Using one-sided Z-test for a set-valued null hypothesis

- H_0: mean≤0 (vs. mean=0)
- H_1: mean>0
- $\sup_{\theta \leq 0}$Pr(false alarm|θ)=Pr(false alarm|θ=0)
  - Type I error is the same
- Type II error is also the same for any θ>0
- Any better tests?

Optimal hypothesis tests

- A hypothesis test f is uniformly most powerful (UMP), if
  - for any other test f' with the same Type I error
  - for any $\theta \in H_1$,

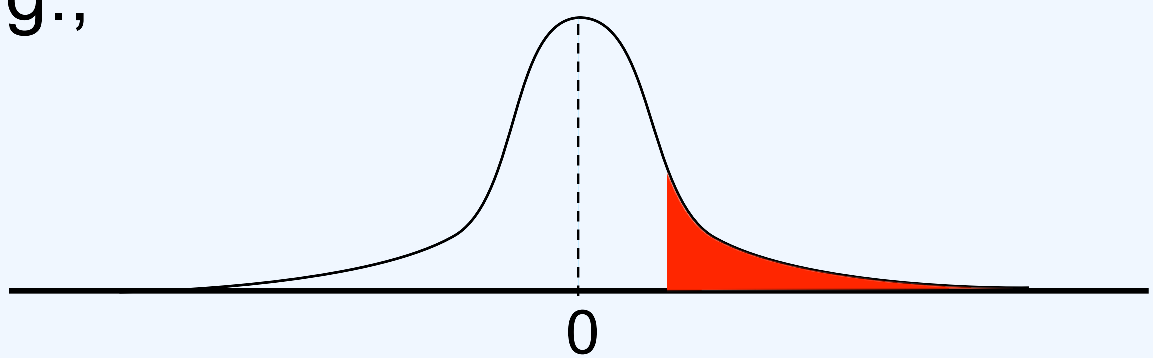
  Type II error of f < Type II error of f'



- Corollary of Karlin-Rubin theorem:
  One-sided Z-test is a UMP for H_0:≤0 and H_1:>0
  - generally no UMP for two-sided tests

Template of other tests

- Tell you the H_0 and H_1 used in the test
  - e.g., H_0:mean≤0 and H_1:mean>0
- Tell you the test statistic, which is a function from data to a scalar
  - e.g., compute the mean of the data
- For any given α, specify a region of test statistic that will leads to the rejection of H_0
  - e.g.,



How to do test for your problem?

- Step 1: look for a type of test that fits your problem (from e.g. wiki)
- Step 2: choose H_0 and H_1
- Step 3: choose level of significance α
- Step 4: run the test

Statistical decision theory

- Given
  - statistical model: Θ, S, Pr($s|\theta$)
  - decision space: D
  - loss function: L(θ, d)$\in\mathbb{R}$
- We want to make a decision based on observed generated data
  - decision function f: data⟶D

Hypothesis testing as a decision problem

- D={reject, retain}
- L(θ, reject)=
  - 0, if $\theta \in H_1$
  - 1, if $\theta \in H_0$ (type I error)
- L(θ, retain)=
  - 0, if $\theta \in H_0$
  - 1, if $\theta \in H_1$ (type II error)

Bayesian expected loss

- Given data and the decision d
  - $EL_B(\text{data}, d) = E_{\theta|\text{data}} L(\theta, d)$
- Compute a decision that minimized EL for a given the data

Frequentist expected loss

- Given the ground truth θ and the decision function f
  - $EL_F(\theta, f) = E_{data|\theta} L(\theta, f(\mathrm{data}))$
- Compute a decision function with small EL for all possible ground truth
  - c.f. uniformly most powerful test: for all $\theta \in H_1$, the UMP test always has the lowest expected loss (Type II error)
- A minimax decision rule f is $\text{argmin}_f \max_\theta EL_F(\theta, f)$
  - most robust against unknown parameter

Two interesting applications of game theory

The Minimax theorem

- For any simultaneous-move two player zero-sum game
- The value of a player's mixed strategy s is her worst-case utility against against the other player
  - Value(s)=min$_{s'}$ $U(s,s')$
  - s_1 is a mixed strategy for player 1 with maximum value
  - s_2 is a mixed strategy for player 2 with maximum value
- Theorem Value(s_1)=-Value(s_2) [von Neumann]
  - (s_1, s_2) is an NE
  - for any s_1' and s_2', Value(s_1') ≤ Value(s_1)= -Value(s_2) ≤ -Value(s_2')
  - to prove that s_1* is minimax, it suffices to find s_2* with Value(s_1*)=-Value(s_2*)

App1: Yao's minimax principle

- Question: how to prove a randomized algorithm A is (asymptotically) fastest?
  - Step 1: analyze the running time of A
  - Step 2: show that any other randomized algorithm runs slower for some input
  - but how to choose such a worst-case input for all other algorithms?
- Theorem [Yao 77] For any randomized algorithm A
  - the worst-case expected running time of A

    is more than
  - for any distribution over all inputs, the expected running time of the fastest deterministic algorithm against this distribution
- Example. You designed a $O(n^2)$ randomized algorithm, to prove that no other randomized algorithm is faster, you can
  - find a distribution π over all inputs (of size n)
  - show that the expected running time of any deterministic algorithm on π is more than $O(n^2)$

Proof

- Two players: you, Nature
- Pure strategies
  - You: deterministic algorithms
  - Nature: inputs
- Payoff
  - You: negative expected running time
  - Nature: expected running time
- For any randomized algorithm A
  - largest expected running time on some input
  - is more than the expected running time of your best (mixed) strategy
  - =the expected running time of Nature's best (mixed) strategy
  - is more than the smallest expected running time of any deterministic algorithm on any distribution over inputs

App2: finding a minimax rule?

- Guess a least favorable distribution π over the parameters
  - let f_π denote its Bayesian decision rule
  - Proposition. f_π minimizes the expected loss among all rules, i.e. $f_\pi = \text{argmin}_f \, E_{\theta \backsim \pi} EL_F(\theta, f)$
- Theorem. If for all θ, $EL_F(\theta, f_\pi)$ are the same, then f_π is minimax

Proof

- Two players: you, Nature
- Pure strategies
  - You: deterministic decision rules
  - Nature: the parameter
- Payoff
  - You: negative frequentist loss, want to minimize the max frequentist loss
  - Nature: frequentist loss $EL_F(\theta, f) = E_{data|\theta}L(\theta, f(\mathrm{data}))$, want to maximize the minimum frequentist loss
- Nee to prove that f_π is minimax
  - suffices to show that there exists a mixed strategy π^* for Nature
    - π^* is a distribution over Θ
  - such that
    - for all rule f and all parameter θ, $EL_F(\pi^*, f) \geq EL_F(\theta, f_\pi)$
  - the equation holds for $\pi^*=\pi$ *QED*

Recap

- Problem: make a decision based on randomly generated data
- Z-test
  - null/alternative hypothesis
  - level of significance
  - reject/retain
- Statistical decision theory framework
  - Bayesian expected loss
  - Frequentist expected loss
- Two applications of the minimax theorem