W13 Notes

Significance testing

06 Theory - Significance testing

Significance test

Ingredients of a significance test (unary hypothesis test):

- H_0 — Null hypothesis event
  - Identify a Claim
  - Then: H_0 is background assumption (supposing Claim isn't known)
  - Goal is to *invalidate H_0* in favor of Claim
- R — Rejection Region event (decision rule)
  - R is written in terms of **decision statistic** X and **significance level** α
  - R is *unlikely* assuming H_0. R is *more likely* if Claim
- $P[R \mid H_0]$ — Able to compute this
  - Usually: inferred from $f_{X|H_0}$ or $P_{X|H_0}$
  - Adjust R to achieve $P[R \mid H_0] = \alpha$

Significance level

Suppose we are given a null hypothesis H_0 and a rejection region R.

The **significance level of R** is:

$$\alpha = P[R \mid H_0]$$

$$= P\big[\text{reject } H_0 \mid H_0 \text{ is true}\big]$$

Sometimes the condition is dropped and we write $\alpha = P[R]$, e.g. when a background model without assuming H_0 is not known.

Null hypothesis implies a distribution

Usually S is unspecified, yet H_0 determines a known *distribution*.

In this case H_0 will *not* take the form of an event in a sample space, $H_0 \subset S$.

At a minimum, H_0 must determine $P[R \mid H_0]$.

We do NOT need these details:

- Background sample space S
- Non-conditional distribution (full model): f_X or P_X

- Complement conditionals: $f_{X|H_0^c}$ or $P_{X|H_0^c}$

In basic statistical inference theory, there are two kinds of error.

- Type I error concludes with rejecting H_0 when H_0 is true.
- Type II error concludes with maintaining H_0 when H_0 is false.

Type I error is usually a bigger problem. We want to consider H_0 as "innocent until proven guilty."

	H_0 **is true**	H_0 **is false**
Maintain null hypothesis	Made right call	Wrong acceptance Type II Error
Reject null hypothesis	Wrong rejection Type I Error	Made right call

---

To *design a significance test at α,* we must identify H_0, and specify R with the property that $P[R \mid H_0] = \alpha$.

When R is written using a variable X, we must choose between:

- One-tail rejection region: x with $R(x) \leq r$ or x with $R(x) \geq r$
- Two-tail rejection region: x with $|R(x) - \mu| \geq c$

07 Illustration

Example - One-tail test: Weighted die

Your friend gives you a single regular die, and say she is worried that it has been weighted to prefer the outcome of 2. She wants you to test it.

Design a significance test for the data of 20 rolls of the die to determine whether the die is weighted. Use significance level $\alpha = 0.05$.

Solution

Let X count the number of 2s that come up.

The Claim: "the die is weighted to prefer 2"
The null hypothesis H_0: "the die is normal"

Assuming H_0 is true, then $X \sim \text{Bin}(20, 1/6)$, and therefore:

$$P_{X|H_0}(k) \;=\; \binom{20}{k}(1/6)^k(5/6)^{20-k}$$

⚠️ Notice that “prefer 2” implies the claim is for *more* 2s than normal.

Therefore: Choose a one-tail rejection region.

Need r such that:

$$P[X \geq r \mid H_0] \;=\; 0.05$$

$$\iff \quad P[X < r \mid H_0] \;=\; 0.95$$

Solve for r by computing conditional CDF values:

k :	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
$F_{X\mid H_0}(k)$:	0.026	0.130	0.329	0.567	0.769	0.898	0.963	0.989

Therefore, choose $r = 6$:

$P[X \geq 6 \mid H_0] < 0.04$, but $P[X \geq 5 \mid H_0] > 0.05$. Final answer:

$$R \;=\; \{x \mid x \geq 6\}$$

Two-tail test: Circuit voltage

A boosted AC circuit is supposed to maintain an average voltage of 130 V with a standard deviation of 2.1 V. Nothing else is known about the voltage distribution.

Design a two-tail test incorporating the data of 40 independent measurements to determine if the expected value of the voltage is truly 130 V. Use $\alpha = 0.02$.

Solution

Use $M_{40}(V)$ as the decision statistic, i.e. the sample mean of 40 measurements of V.

The Claim to test: $E[V] \neq 130$

The null hypothesis H_0: $E[V] = 130$

Rejection region:

$$|M_{40} - 130| \geq c$$

where c is chosen so that $P\big[\,|M_{40} - 130| \geq c\,\big] = 0.02$

Assuming H_0, we expect that:

$$E[M_{40}] = 130, \qquad \sigma^2_{M_{40}} = \frac{2.1^2}{40} \approx 0.110$$

Recall Chebyshev’s inequality:

$$P\left[\ |M_{40} - 130| \geq c\ \right] \leq \frac{\sigma^2_{M_{40}}}{c^2} \approx \frac{0.110}{c^2}$$

Now solve:

$$\frac{0.110}{c^2} = 0.02 \quad \gg\gg \quad c \approx 2.348$$

Therefore the rejection region should be:

$$M_{40} < 127.65 \quad \cup \quad 132.35 < M_{40}$$

One-tail test with a Gaussian: Weight loss drug

Assume that in the background population in a specific demographic, the distribution of a person's weight W satisfies $W \sim \mathcal{N}(190, 24^2)$. Suppose that a pharmaceutical company has developed a weight-loss drug and plans to test it on a group of 64 individuals.

Design a test at the $\alpha = 0.01$ significance level to determine whether the drug is effective.

Solution

Since the drug is tested on 64 individuals, we use the sample mean $M_{64}(W)$ as the decision statistic.

The Claim: "the drug is effective in reducing weight"

The null hypothesis H_0: "no effect: weights on the drug still follow $\mathcal{N}(190, 24^2)$"

Assuming H_0 is true, then $W \sim \mathcal{N}(190, 24^2)$.

⚠️ One-tail test because the drug is expected to *reduce* weight (unidirectional). Rejection region:

$$M_{64}(W) \leq r$$

Calculate:

$$\sigma^2_{M_{64}} \quad \gg\gg \quad \frac{24^2}{64} \quad \gg\gg \quad 9$$

⚠️ Standardized $M_{64}(W)$ is approximately normal!

(The standardization of $M_{64}(W)$ removes the effect of $\frac{1}{n}$. As if it's the summation.)

So, standardize and apply CLT:

$$\frac{M_{64}(W) - 190}{\sqrt{9}} \sim \mathcal{N}(0,1),$$

$$\gg\gg \quad P[M_{64}(W) \le r] \approx P\left[Z \le \frac{r-190}{3}\right] = \Phi\left(\frac{r-190}{3}\right)$$

Solve:

$$P[M_{64}(W) \le r] = 0.01$$

$$\gg\gg \quad \Phi\left(\frac{r-190}{3}\right) = 0.01$$

$$\gg\gg \quad \Phi\left(\frac{190-r}{3}\right) = 0.99$$

$$\gg\gg \quad \frac{190-r}{3} = 2.33$$

$$\gg\gg \quad r = 183.01$$

Therefore, the rejection region:

$$M_{64}(W) \le 183.01$$

Binary hypothesis testing

01 Theory - Binary testing, MAP and ML

Binary hypothesis test

Ingredients of a binary hypothesis test:

- H_0 and H_1 — Complementary hypotheses
  - Maybe also know the **prior probabilities** $P[H_0]$ and $P[H_1]$
  - Goal: determine which case we are in, H_0 or H_1
- A_0 and A_1 — Complementary events of the Decision Rule
  - Directionality: given H_0, A_0 is likely; given H_1, A_1 is likely.
  - Decision Rule: outcome A_0, accept H_0; outcome A_1, accept H_1
  - Usually: A_i written in terms of **decision statistic** X using a **design**
  - We cover three **designs**:
    - MAP and ML (minimize 'error probability')
    - MC (minimizes 'error cost')
  - Designs use $P_{X|H_0}$ and $P_{X|H_1}$ (or $f_{X|H_0}$, $f_{X|H_1}$) to construct A_0 and A_1

MAP design

Suppose we know:

- $P[H_0]$ and $P[H_1]$
  - Both prior probabilities
- $P_{X|H_0}(x)$ and $P_{X|H_1}(x)$ (or $f_{X|H_0}(x)$ and $f_{X|H_1}(x)$)
  - Both conditional distributions

The **maximum a posteriori probability (MAP)** design for a decision statistic X:

$$A_0 \;=\; \text{set of } x \text{ for which:}$$

Discrete case:

$$P_{X|H_0}(x) \cdot P[H_0] \;\geq\; P_{X|H_1}(x) \cdot P[H_1]$$

Continuous case:

$$f_{X|H_0}(x) \cdot P[H_0] \;\geq\; f_{X|H_1}(x) \cdot P[H_1]$$

And $A_1 \;=\; \{x \in \mathbb{R} \mid x \notin A_0\}$.

The MAP design minimizes the total probability of error.

ML design

Suppose we don't know the priors, we know only:

- $P_{X|H_0}(x)$ and $P_{X|H_1}(x)$ (or $f_{X|H_0}(x)$ and $f_{X|H_1}(x)$)
  - Both conditional distributions

The **maximum likelihood (ML)** design for X:

$$A_0 \;=\; \text{set of } x \text{ for which: } \begin{array}{ll} P_{X|H_0}(x) \;\geq\; P_{X|H_1}(x) & \text{(discrete)} \\ f_{X|H_0}(x) \;\geq\; f_{X|H_1}(x) & \text{(continuous)} \end{array}$$

ML is a simplified version of MAP. (Set $P[H_0]$ and $P[H_1]$ to 0.5.)

The probability of a *false alarm*, a Type I error, is called P_{FA}.

The probability of a *miss*, a Type II error, is called P_{Miss}.

$$P_{FA} \;=\; P[A_1 \mid H_0]$$

$$P_{\text{Miss}} \;=\; P[A_0 \mid H_1]$$

Total probability of error:

$$P_{\text{ERR}} \;=\; P[A_1 \mid H_0] \cdot P[H_0] + P[A_0 \mid H_1] \cdot P[H_1]$$

⚠ Wrong meanings of P_{FA}

Suppose A_1 sets off a smoke alarm, and H_0 is 'no fire' and H_1 is 'yes fire'.

Then P_{FA} is the odds that we get an alarm *assuming there is no fire*.

This is *not* the odds of *experiencing* a false alarm (no context). That would be $P[A_1 H_0]$.

This is *not* the odds of a *given* alarm being a false one. That would be $P[H_0 \mid A_1]$.

02 Illustration

Example - ML test: Smoke detector

Suppose that a smoke detector sensor is configured to produce 8 V when there is smoke, and 0 V otherwise. But there is background noise with distribution $\mathcal{N}(0, 3^2 \,\text{V})$.

Design an ML test for the detector electronics to decide whether to activate the alarm.

What are the three error probabilities? (Type I, Type II, Total.)

Solution

First, establish the conditional distributions:

$$X \mid H_0 \sim \mathcal{N}(0, 3^2) \qquad X \mid H_1 \sim \mathcal{N}(8, 3^2)$$

Density functions:

$$f_{X|H_0} = \frac{1}{\sqrt{2\pi 9}} e^{-\frac{1}{2}\left(\frac{x-0}{3}\right)^2} \qquad f_{X|H_1} = \frac{1}{\sqrt{2\pi 9}} e^{-\frac{1}{2}\left(\frac{x-8}{3}\right)^2}$$

The ML condition becomes:

$$\frac{1}{\sqrt{2\pi 9}} e^{-\frac{1}{2}\left(\frac{x-0}{3}\right)^2} \overset{?}{\geq} \frac{1}{\sqrt{2\pi 9}} e^{-\frac{1}{2}\left(\frac{x-8}{3}\right)^2}$$

$$\gg\gg \quad -\frac{1}{2}\left(\frac{x-0}{3}\right)^2 \overset{?}{\geq} -\frac{1}{2}\left(\frac{x-8}{3}\right)^2$$

$$\gg\gg \quad x^2 \overset{?}{\leq} (x-8)^2$$

$$\gg\gg \quad x \leq 4$$

Therefore, A_0 is $x \leq 4$, while A_1 is $x > 4$.

The decision rule is: activate alarm when $x > 4$.

Type I error:

$$P_{FA} = P[A_1 \mid H_0] \quad \gg\gg \quad P[X > 4 \mid H_0]$$

$$\gg\gg \quad 1 - P\left[\frac{X-0}{3} \leq \frac{4}{3} \,\middle|\, H_0\right]$$

$$\gg\gg \quad 1 - P[Z \leq 1.3333] \quad \gg\gg \quad \approx \boxed{0.0912}$$

Type II error:

$$P_{\text{Miss}} = P[A_0 \mid H_1] \quad \gg\gg \quad P[X \leq 4 \mid H_1]$$

$$\gg\gg \quad P\left[\frac{X-8}{3} \leq \frac{4-8}{3} \,\middle|\, H_1\right]$$

$$\gg\gg \quad P[Z \leq -1.3333] \quad \gg\gg \quad \approx \boxed{0.0912}$$

Total error:

$$P_{\text{ERR}} = P_{FA} \cdot 0.5 + P_{\text{Miss}} \cdot 0.5 \quad \gg\gg \quad \approx \boxed{0.0912}$$

Example - MAP test: Smoke detector

Suppose that a smoke detector sensor is configured to produce $8\,\text{V}$ when there is smoke, and $0\,\text{V}$ otherwise. But there is background noise with distribution $\mathcal{N}(0, 3^2\,\text{V})$.

Suppose that the background chance of smoke is 5%. Design a MAP test for the alarm.

What are the three error probabilities? (Type I, Type II, Total.)

Solution

First, establish priors:

$$P[H_0] = 0.95 \qquad P[H_1] = 0.05$$

The MAP condition becomes:

$$\frac{1}{\sqrt{2\pi 9}} e^{-\frac{1}{2}\left(\frac{x-0}{3}\right)^2} \cdot 0.95 \quad \overset{?}{\geq} \quad \frac{1}{\sqrt{2\pi 9}} e^{-\frac{1}{2}\left(\frac{x-8}{3}\right)^2} \cdot 0.05$$

$$\gg\gg \quad e^{-\frac{1}{2}\left(\frac{x-0}{3}\right)^2} \quad \overset{?}{\geq} \quad e^{-\frac{1}{2}\left(\frac{x-8}{3}\right)^2} \cdot \frac{0.05}{0.95}$$

$$\gg\gg \quad -\frac{1}{2}\left(\frac{x-0}{3}\right)^2 \quad \overset{?}{\geq} \quad -\frac{1}{2}\left(\frac{x-8}{3}\right)^2 + \ln\left(\frac{0.05}{0.95}\right)$$

$$\gg\gg \quad x^2 \quad \overset{?}{\leq} \quad (x-8)^2 - 18\ln\left(\frac{0.05}{0.95}\right)$$

$$\gg\gg \quad x \leq 7.31$$

Therefore, A_0 is $x \leq 7.31$, while A_1 is $x > 7.31$.

The decision rule is: activate alarm when $x > 7.31$.

Type I error:

$$P_{FA} \;=\; P[A_1 \mid H_0] \quad \ggg \quad P[X > 7.31 \mid H_0]$$

$$\ggg \quad 1 - P[Z \leq 2.4367] \quad \ggg \quad \approx 0.007411$$

Type II error:

$$P_{\text{Miss}} \quad = \quad P[A_0 \mid H_1] \quad \ggg \quad P[X \leq 7.31 \mid H_1]$$

$$\ggg \quad P[Z \leq -0.23] \quad \ggg \quad \approx 0.4090$$

Total error:

$$P_{\text{ERR}} \;=\; P_{FA} \cdot 0.95 + P_{\text{Miss}} \cdot 0.05 \quad \approx \quad 0.02749$$

03 Theory - MAP criterion proof

Explanation of MAP criterion - discrete case

First, we show that the MAP design selects for A_0 all those x which render H_0 more likely than H_1. This will be used in the next step to show that MAP minimizes probability of error.

Observe this calculation:

$$P[H_i \mid X = x] \;=\; P[X = x \mid H_i] \cdot \frac{P[H_i]}{P[X]} \qquad \text{(Bayes' Rule)}$$

$$= \; P_{X|H_i}(x) \cdot \frac{P[H_i]}{P[X]} \qquad \text{(Conditional PMF)}$$

Recall the MAP criterion:

$$P_{X|H_0}(x) \cdot P[H_0] \;\geq\; P_{X|H_1}(x) \cdot P[H_1]$$

Divide both sides by $P[X]$ and apply the above Calculation in reverse:

$$\ggg \quad P[H_0 \mid X = x] \;\geq\; P[H_1 \mid X = x]$$

This is what we sought to prove.

Next, we verify that the MAP design minimizes the total probability of error.

The total probability of error is:

$$P_{\text{ERR}} \quad = \quad P[A_1 \mid H_0] \cdot P[H_0] + P[A_0 \mid H_1] \cdot P[H_1]$$

Expand this with summation notation (assuming the discrete case):

$$\gg\gg \quad \sum_{x\in A_1} P_{X|H_0}(x)\cdot P[H_0] + \sum_{x\in A_0} P_{X|H_1}(x)\cdot P[H_1]$$

Now, how do we choose the set $A_0 \subset \mathbb{R}$ (and thus $A_1 = A_0^c$) in such a way that this sum is minimized?

Since all terms are positive, and any $x \in \mathbb{R}$ may be placed in A_1 or in A_0 freely and independently of all other choices, the total sum is minimized when we minimize the impact of placing each x.

So, for each x, we place it in A_0 if:

$$P_{X|H_0}(x)\cdot P[H_0] \;\geq\; P_{X|H_1}(x)\cdot P[H_1]$$

That is equivalent to the MAP criterion.

04 Theory - MC design

- Write C_{10} for cost of false alarm, i.e. cost when H_0 is true but decided H_1.
  - Probability of incurring cost C_{10} is $P_{FA}\cdot P[H_0]$.
- Write C_{01} for cost of miss, i.e. cost when H_1 is true but decided H_0.
  - Probability of incurring cost C_{01} is $P_{\text{Miss}}\cdot P[H_1]$.

Expected value of cost incurred

$$E[C] \;=\; P[A_1 \mid H_1]\cdot P[H_0]\cdot C_{10} + P[A_0 \mid H_1]\cdot P[H_1]\cdot C_{01}$$

MC design

Suppose we know:

- Both prior probabilities $P[H_0]$ and $P[H_1]$
- Both conditional distributions $P_{X|H_0}(x)$ and $P_{X|H_1}(x)$ (or $f_{X|H_0}(x)$ and $f_{X|H_1}(x)$)

The **minimum cost (MC)** design for a decision statistic X:

$$A_0 \quad = \quad \text{set of } x \text{ for which:}$$

Discrete case:

$$P_{X|H_0}(x)\cdot P[H_0]\cdot C_{10} \quad \geq \quad P_{X|H_1}(x)\cdot P[H_1]\cdot C_{01}$$

Continuous case:

$$f_{X|H_0}(x)\cdot P[H_0]\cdot C_{10} \quad \geq \quad f_{X|H_1}(x)\cdot P[H_1]\cdot C_{01}$$

Then $A_1 \;=\; \{x \in \mathbb{R} \mid x \notin A_0\}$.

The MC design minimizes the expected value of the cost of error.

MC minimizes expected cost

Inside the argument that MAP minimizes total probability of error, we have this summation:

$$P_{\text{ERR}} \;=\; \sum_{x\in A_1} P_{X|H_0}(x)\cdot P[H_0] + \sum_{x\in A_0} P_{X|H_1}(x)\cdot P[H_1]$$

The expected value of the cost has a similar summation:

$$E[C] \;=\; \sum_{x\in A_1} P_{X|H_0}(x)\cdot P[H_0]\cdot C_{10} + \sum_{x\in A_0} P_{X|H_1}(x)\cdot P[H_1]\cdot C_{01}$$

Following the same reasoning, we see that the cost is minimized if each x is placed into A_0 precisely when the MC design condition is satisfied, and otherwise it is placed into A_1 .

05 Illustration

Example - MC Test: Smoke detector

Suppose that a smoke detector sensor is configured to produce 8 V when there is smoke, and 0 V otherwise. But there is background noise with distribution $\mathcal{N}(0, 3\,\text{V})$.

Suppose that the background chance of smoke is 5%. Suppose the cost of a miss is $50\times$ the cost of a false alarm. Design an MC test for the alarm.

Compute the expected cost.

Solution

We have priors:

$$P[H_0] \;=\; 0.95 \qquad P[H_1] \;=\; 0.05$$

And we have costs:

$$C_{10} = 1 \qquad C_{01} = 50$$

(The ratio of these numbers is all that matters in the inequalities of the condition.)

The MC condition becomes:

$$\frac{1}{\sqrt{2\pi 9}} e^{-\frac{1}{2}\left(\frac{x-0}{3}\right)^2}\cdot 0.95\cdot 1 \quad \overset{?}{\geq} \quad \frac{1}{\sqrt{2\pi 9}} e^{-\frac{1}{2}\left(\frac{x-8}{3}\right)^2}\cdot 0.05\cdot 50$$

$$\gg\gg \quad e^{-\frac{1}{2}\left(\frac{x-0}{3}\right)^2} \quad \overset{?}{\geq} \quad e^{-\frac{1}{2}\left(\frac{x-8}{3}\right)^2}\cdot\frac{2.5}{0.95}$$

$$\gg\gg \quad -\frac{1}{2}\left(\frac{x-0}{3}\right)^2 \quad \overset{?}{\geq} \quad -\frac{1}{2}\left(\frac{x-8}{3}\right)^2 + \ln\left(\frac{2.5}{0.95}\right)$$

$$\gg\gg \quad x^2 \quad \overset{?}{\leq} \quad (x-8)^2 - 18\ln\left(\frac{2.5}{0.95}\right)$$

$$\gg\gg \quad x \leq 2.91$$

Therefore, A_0 is $x \leq 2.91$, while A_1 is $x > 2.91$.

The decision rule is: activate alarm when $x > 2.91$.

Type I error:

$$P_{FA} \;=\; P[A_1 \mid H_0]$$

$$\gg\gg \quad P[X > 2.91 \mid H_0] \quad \gg\gg \quad \approx 0.1660$$

Type II error:

$$P_{\text{Miss}} \quad = \quad P[A_0 \mid H_1]$$

$$\gg\gg \quad P[X \leq 2.91] \quad \gg\gg \quad \approx 0.04488$$

Total error:

$$P_{\text{ERR}} \;=\; P_{FA} \cdot 0.95 + P_{\text{Miss}} \cdot 0.05 \quad \approx \quad 0.1599$$

PMF of total cost:

$$P_C(c) \quad = \quad \begin{cases} 0.002244 & c = 50 \\ 0.1577 & c = 1 \\ 0.840056 & c = 0 \end{cases}$$

Therefore $E[C] \;=\; 0.27$.