

CHAPTER 4

HYPOTHESIS TESTING

If a new form of therapy is tested on 100 patients and gives good results in 64 of them, while this only holds for 50% of the patients with the old therapy, is the new therapy truly better than the old one, or were we just "lucky"?

This is a typical question that concerns a decision between two conflicting hypotheses.

[Set-up]:

$$\underline{x} \sim \text{Model } p_{\theta}(\underline{x}) \quad \theta \in \Theta$$

Null hypothesis $\theta \in \Theta_0$ or

Alternative $\theta \in \Theta_1 = \Theta \setminus \Theta_0$.

Often the 2 hypotheses are not treated symmetrically

- Like in trials we want to know if H_1 is correct.
- If there is not enough evidence to reject H_0 .
 $\Rightarrow \begin{cases} H_1 \text{ incorrect} \\ \text{and } H_0 \text{ correct} \end{cases}$

because it can also mean that there are not sufficient proofs for either of the hypotheses.

Possible Conclusions

- Reject H_0
And accept H_1 as **strong** correct.
- Do not reject H_0
but do not accept H_1 as **weak** being correct.
(more investigation needed)

Possible errors

- type I : Reject H_0 when correct
 \Rightarrow accept strong conclusion that is wrong.
- type II : Maintain H_0 when incorrect.
 \Rightarrow accept the weak conclusion

ASYMMETRY	
type I	VERY UNDESIRABLE
Type II	not as bad.

Wisely choose the hypotheses : —
In principle, choose the statement we want to "prove" as H_1 .

Example 4.1 $p = \Pr(\text{success})$ of a new therapy

$$H_0 : p \leq 0.5 \quad H_1 : p > 0.5$$

Example 4.2 Playing dice

$$\theta = (p_1 \dots p_6) \quad H_0: p_i = \frac{1}{6} \quad \forall i$$

$$H_1: p_i \neq \frac{1}{6} \text{ at least one } i$$

Example 4.3 Two samples model

$$(x_1, \dots, x_m) \quad (y_1, \dots, y_n) \quad \begin{matrix} \text{all} \\ \text{independent.} \end{matrix}$$

\downarrow \downarrow
 $N(\mu, \sigma^2)$ $N(\nu, \tau^2)$

See Notebook 4 Box-plot not helpful

$$\left\{ \begin{array}{ll} H_0: \mu = \nu & \text{formal test.} \\ H_1: \mu \neq \nu & \end{array} \right.$$

$$\theta = (\mu, \nu, \sigma^2, \tau^2) \in \mathbb{R}^2 \times (0, \infty)^2$$

④. for the null is

$$\Theta_0 = \{(\mu, \mu) : \mu \in \mathbb{R}\} \times (0, \infty)^2.$$

4.3 Sample size and critical region

The decision on rejection of H_0 is based on the data $\underline{X} = (X_1, \dots, X_n)$

A statistical test on the hypothesis H_0 is defined by the critical region K , a subset of the sample space

if $\underline{X} \in K \Rightarrow$ Reject H_0

if $\underline{X} \notin K \Rightarrow$ don't reject H_0 .

Typically K is defined by a test statistic $T(\underline{X})$, real valued, that defines the distance between the data and the hypothesis.

Example 4.6 - Gauss test

$(X_1, \dots, X_n) \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$

σ^2 known.
(unrealistic
but see later)

Test of

$H_0: \mu \leq \mu_0$ against $H_1: \mu > \mu_0$

Example: Quality control based
on a sample \underline{X} and a Test statistic

$$T(\underline{X}) = \bar{X}$$

$$K = \{ \underline{x} : \bar{x} > c \}.$$

c chosen to get a sufficiently small
probability of type I error.

Notes: - Two different test statistics
may give the same K.

$$K = \{ \underline{x} : \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} > z \}$$

- A different system of the hypotheses implies a different K .

$$H_0: \mu \geq \mu_0 \quad H_1: \mu < \mu_0$$

$$K = \{ \underline{x} : \bar{x} \leq c \}.$$

4.3.1

Size and power function

What we want

if $\theta \in H_0$, $P(\underline{x} \in K)$ small

if $\theta \notin H_0$, $P(\underline{x} \in K)$ high

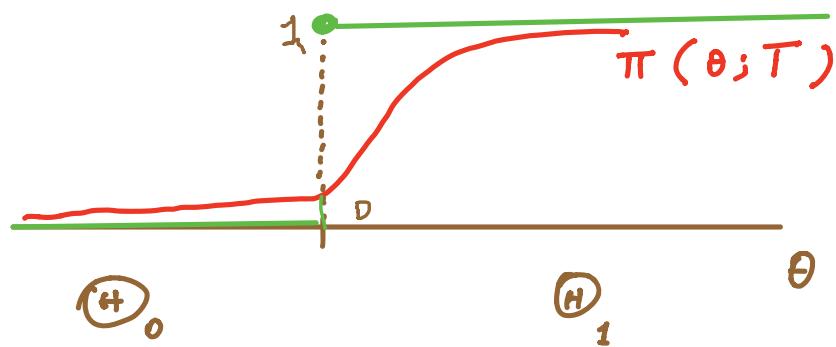
A function that measures the quality (or sensitivity) of a test is

$$\theta \mapsto P_{\theta}(\underline{x} \in K) = \pi(\theta, T)$$

called

Power function

Example of a real power function
and ideal



Size of a test T

$$\alpha = \sup_{\theta \in \mathbb{H}_0} \pi(\theta; T)$$

A test has level α_0 if $\alpha_0 \geq \alpha$.

Convention in testing (A)

Choose first a level α_0 and
then use only tests of
size $\alpha \leq \alpha_0$.

- It's appealing to choose tests with a very small level

- But this implies that the probability

$$P(\underline{II}) = P_{\theta}(\underline{x} \notin K) \quad \theta \in \Theta,$$

tends to become very large.

- We don't try to minimize the sum of the two errors $P(I) + P(\underline{II})$

- The level α_0 should be chosen depending on the consequences of the decision, but...

- in practice we choose only tests with a level 0.05

- Of the tests of level 0.05 we prefer those having the smallest $P(\underline{II})$.

Convention in testing (B)

Given level α_0 , we prefer a test with level α_0 that has the highest $\pi(\theta; k)$ for $\theta \in \Theta_1$.

Example 4.11 } Binomial test.

- Experiment: give a) therapy to 100 patients.
 - Let $X = \# \text{ successes} \sim \text{Bin}(100, p)$
 - Test if the new therapy has probability of success > 0.5.
 - Test statistic $T(\underline{x}) = X$
 - Critical region $K = \{ T(x) \geq c \}$
- 
CRITICAL VALUE
TO BE CHOSEN
- So: $K = \{ c, c+1, \dots, 100 \}$.

A large value of X gives an indication that $H_0: p \leq 0.5$ is incorrect.

Assume that we choose $c = 59$

$$K = \{59, 60, \dots, 100\}$$

The size of the test is

$$\alpha = \sup_{p \leq 0.5} P_p(X \geq 59)$$

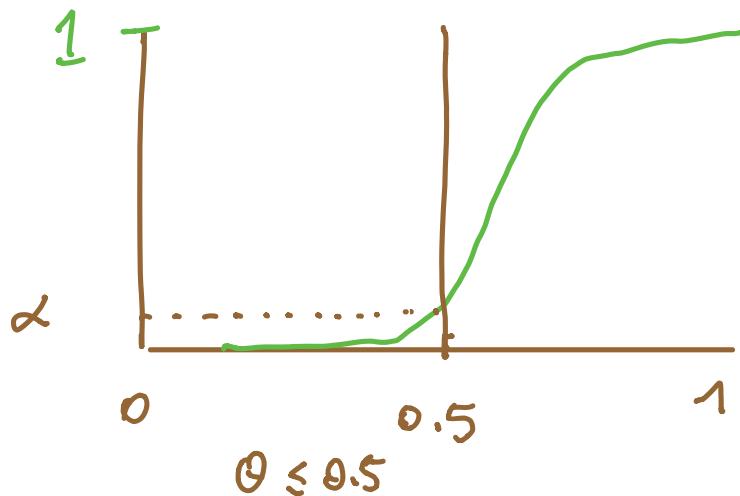
$$\begin{aligned}\alpha &= P_{0.5}(X \geq 59) = \\ &= 1 - \text{binocdf}(58, 100, 0.5) \\ &= 0.0443\end{aligned}$$

* because $P_p(X \geq 59)$ is increasing

The power of the test is

$$p \mapsto P_p(X \geq 59)$$

→ See Notebook



- If we choose the level $\alpha_0 = 0.05$ we have

$$\text{size} = P_{0.5}(X \geq 59) = 0.0443 \quad \text{OK}$$

$$\text{size} = P_{0.5}(X \geq 58) = 0.067 \quad \text{NOT OK}$$

If we find 64 successes on 100 trials with the new therapy we reject H_0 at level 5%.

Example 4.12 - Gauss test

$\underline{X} = (X_1, \dots, X_n)$ i.i.d. $N(\mu, \sigma^2)$
↳ Known

$$H_0: \mu \leq \mu_0 \quad H_1: \mu > \mu_0.$$

Test statistic $T(\underline{X}) = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$

Equivalent to $T(\underline{X}) = \bar{X}$.

Critical region $K = \left\{ \underline{X} : \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \geq c \right\}$

Note that

$$T(\underline{X}) \sim N(0, 1) \text{ under } \mu = \mu_0$$

So we look for a test of level at most α_0

$$\sup_{\mu \leq \mu_0} P_\mu \left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \geq c \right) \leq \alpha_0 \quad (*)$$

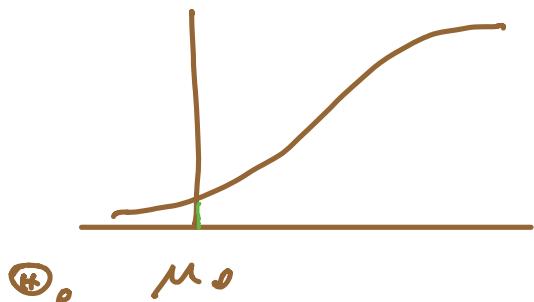
Result

The critical region of level at most α_0 is obtained for $c > \xi_{1-\alpha_0}$ where $\xi_{1-\alpha}$ is the $1-\alpha$ -quantile of the $N(0, 1)$

Proof:

$$\begin{aligned} P_\mu \left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{m}} \geq c \right) &= P_\mu \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{m}} \geq c + \frac{\mu_0 - \mu}{\sigma/\sqrt{m}} \right) \\ &\approx 1 - \Phi \left(c + \frac{\mu_0 - \mu}{\sigma/\sqrt{m}} \right). \end{aligned}$$

is an increasing function of μ .



So the sup is attained
 $\mu \leq \mu_0$
for $\mu = \mu_0$.

$$\text{Condition } (*) \Leftrightarrow P_{\mu_0} \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \geq c \right) \leq \alpha_0$$

$$\Leftrightarrow 1 - \Phi(c) \leq \alpha_0$$

$$\Leftrightarrow 1 - \alpha_0 \leq \Phi(c)$$

$$\Leftrightarrow \Phi^{-1}(1 - \alpha_0) \leq c$$

and $c \geq \xi_{1-\alpha_0}$ is the solution! CHOOSE
C = $\xi_{1-\alpha_0}$

Critical region for the Gauss test —

$$H_0: \mu \geq \mu_0 \quad H_1: \mu < \mu_0$$

$$K = \left\{ \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \leq -\xi_{1-\alpha_0} \right\}$$

Critical region for the Gauss test —

$$H_0: \mu = \mu_0 \quad H_1: \mu \neq \mu_0$$

combining two one-sided regions with level $\alpha_0/2$ each

$$K = \left\{ \left| \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \right| \geq \xi_{1-\alpha_0/2} \right\}$$

See Notebook 4

Example 4.13

Binomial test

In general

Test $H_0: p \leq p_0$ $H_1: p > p_0$

1) reject the null if $X = \# \text{ Successes}$
is large

$$X \in \{c, c+1, \dots, n\}$$

2) Choose c such that

• $P_{p_0}(X \geq c) \leq \alpha_0$ level.

• $c = \min_t \{t: P_{p_0}(X \geq t) \leq \alpha_0\}$

this implies a maximum power -

For large values of n :

$$(np(1-p) > 5)$$

the Binomial can be approximated by a normal.

$$\bullet P_{P_0}(X \geq c) = P_{P_0}\left(\frac{X - np_0}{\sqrt{np_0(1-p_0)}} \geq \frac{c - np_0}{\sqrt{np_0(1-p_0)}}\right)$$

$$\text{under } H_0 \approx P\left(Z \geq \frac{c - np_0}{\sqrt{np_0(1-p_0)}}\right)$$

$$= 1 - \Phi\left(\frac{c - np_0}{\sqrt{np_0(1-p_0)}}\right).$$

$$\bullet \alpha_0 = 1 - \Phi\left(\frac{c - np_0}{\sqrt{np_0(1-p_0)}}\right)$$

$$\Phi^{-1}(1 - \alpha_0) = (c - np_0) / \sqrt{np_0(1-p_0)}$$

$$c = np_0 + \xi_{1-\alpha_0} \sqrt{np_0(1-p_0)}$$

Example $X = 64$ $p_0 = 0.58$

$$\alpha_0 = 0.05 \quad \xi_{1-\alpha_0} = 1.645$$

$\text{norminv}(1 - 0.05)$

$$c = 58 + 1.645 \sqrt{100 \cdot 0.58 \cdot 0.42}$$

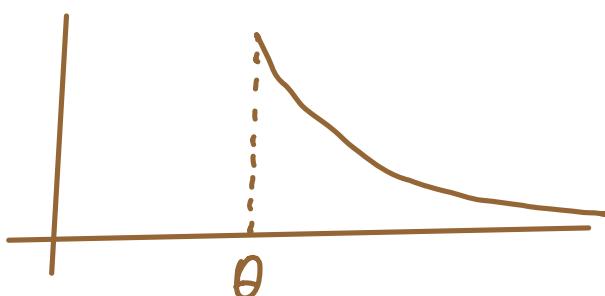
$$= 66.12$$

Not reject $H_0: p \leq 0.58$.

Example 4.14 - Shifted exp.

$$(X_1, \dots, X_m) \stackrel{\text{i.i.d}}{\sim} P_\theta(x) = e^{\theta-x}$$

for $x \geq \theta$



if $\theta = 0$ is $\mathcal{E}xp(\lambda = 1)$

Test: $H_0: \theta \leq 0$ vs $H_1: \theta > 0$.

at level α_0 .

$$\hat{\theta}_{ML} = X_{(1)}.$$

Reasonable to use $X_{(1)}$ as test statistic.

$$K = \{ \underline{x} : X_{(1)} \geq c \}$$

Determine the critical value c for which

- size $\leq \alpha_0$
- power maximal.

$$\text{size} = \sup_{\theta \leq 0} P_\theta (X_{(1)} \geq c)$$

→ Then $P_{\theta} (X_{(1)} \geq c) =$

$$= P_{\theta} (X_1 \geq c, X_2 \geq c, \dots, X_n \geq c)$$

$$= [P_{\theta} (X \geq c)]^n = (e^{\theta-c})^n.$$

$= e^{n(\theta-c)}$. is increasing in θ .

→ The sup is taken in $\theta = 0$.

So

$$\alpha_0 \geq e^{-mc} \Leftrightarrow \boxed{c \geq -\frac{1}{m} \log \alpha_0}$$

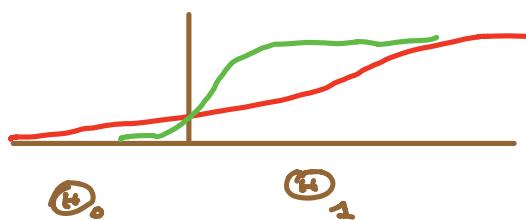
→ The critical region is

$$K = \left\{ \underline{x} : X_{(1)} \geq -\frac{1}{m} \log \alpha_0 \right\}.$$

| It can be proved that the test is most powerful among all tests of level α_0 . |

4.3.2 Sample size (x_1, \dots, x_n)

For $n \rightarrow \infty$ the power function tends to the ideal one.



Can we influence the slope of $\pi(\mu; \tau)$?

Example Gauss test (continued)

Reject $H_0: \mu \leq \mu_0$ at level α_0 if

$$T = (\bar{X} - \mu_0) / (\sigma/\sqrt{n}) \geq \xi_{1-\alpha_0}$$

Power function

$$\pi(\mu; \tau) = 1 - \Phi\left(\xi_{1-\alpha_0} - \frac{\mu - \mu_0}{\sigma/\sqrt{n}}\right)$$

- the greater n , the greater the power in $\mu > \mu_0$
- the greater μ the greater the power

- the greater σ the smaller the power
- the greater α_0 , the greater the $P(I)$ and the greater the power.

$\rightarrow \underline{\text{notebook 4}}$

Now suppose that we want to select a sample size n such that $P(II) \leq \beta$

$$P(II) = 1 - \pi(\mu, T) = \Phi\left(\xi_{1-\alpha_0} - \frac{\mu - \mu_0}{\sigma/\sqrt{n}}\right)$$

Thus we solve:

$$\Phi\left(\xi_{1-\alpha_0} - \frac{\mu - \mu_0}{\sigma/\sqrt{n}}\right) \leq \beta$$

$$\xi_{1-\alpha_0} - \frac{\mu - \mu_0}{\sigma/\sqrt{n}} \leq \Phi^{-1}(\beta) = \xi_\beta.$$

$$(\xi_{1-\alpha_0} - \xi_\beta) \leq \frac{\mu - \mu_0}{\sigma} \sqrt{n}$$

$$\sigma \frac{\xi_{1-\alpha_0} - \xi_\beta}{\mu - \mu_0} \leq \sqrt{n}$$

Example let $H_0: \mu \leq 0$
 $H_1: \mu > 0$

$$X_i \sim N(\mu, \sigma^2 = 4) \quad \beta = 0.4 \\ \alpha_0 = 0.05.$$

$$\sqrt{n} \geq \frac{1.645 - (-0.2533)}{\mu - 0} = \frac{1.8983}{\mu}$$

If we would like to detect a
mean $\mu = 0.1$

$$\sqrt{n} \geq \frac{1.8983}{0.1} \Rightarrow n \geq 361$$

If it's enough to detect a mean

$$\mu = 0.5$$

$$\sqrt{n} \geq \frac{1.8983}{0.5} \Rightarrow n \geq 15.$$

Example 4.16 - Binomial test

$$X \sim \text{Bin}(n, p) \quad H_0: p \leq p_0$$

Reject H_0 at approximate level α_0 :

if

$$X \geq c = np_0 + \xi_{1-\alpha_0} \cdot \sqrt{np_0(1-p_0)}$$

Approximation better with

$$\frac{X - np_0 - \frac{1}{2}}{\sqrt{np_0(1-p_0)}} \stackrel{H_0}{\approx} N(0, 1)$$

Power function

$$P_p(X > c) \approx 1 - \Phi \left(\frac{c - np - \frac{1}{2}}{\sqrt{np(1-p)}} \right)$$

Example

$$H_0 : p \leq p_0 = \frac{1}{2}$$

$$1) \quad \alpha_0 = 0.05 : \quad P_{0.5}(X > c) = 0.05$$

$$2) \quad \text{power } 0.8 \text{ im } p = 0.6$$

$$P_{0.6}(X > c) = 0.8$$

System :

$$1) \quad 1 - \Phi \left(\frac{c - 0.5 \cdot n - \frac{1}{2}}{\sqrt{n \cdot 0.5 \cdot 0.5}} \right) = 0.05$$

$$\therefore \Phi \left(\frac{c - 0.5 \cdot n - \frac{1}{2}}{\sqrt{n \cdot 0.5 \cdot 0.5}} \right) = 0.95$$

$$\frac{c - n \cdot 0.5 - \frac{1}{2}}{\sqrt{n \cdot 0.5 \cdot 0.5}} = \xi_{0.95} = 1.645$$

Similarly :

$$2) \frac{c - m \cdot 0.6 - \frac{1}{2}}{\sqrt{m \cdot 0.6 \cdot 0.4}} \leq \xi_{0.2} = -0.842$$

Thus :

$$\left\{ \begin{array}{l} c - \frac{m}{2} - \frac{1}{2} = 1.645 \sqrt{\frac{m}{4}} \\ c - m \times 0.6 - \frac{1}{2} \leq -0.842 \sqrt{m \times 0.24} \end{array} \right.$$

The first gives

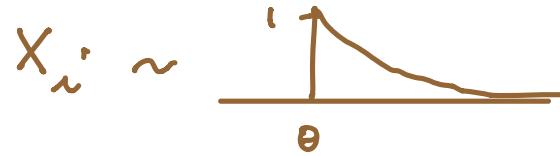
$$c \approx \frac{m+1}{2} + 1.645 \frac{\sqrt{m}}{2}$$

Substituting in the second gives

$$\frac{m+1}{2} + 1.645 \frac{\sqrt{m}}{2} - m(0.6) - \frac{1}{2} + 0.842 \sqrt{m(0.24)} \leq 0$$

$\Rightarrow m > 152$. See notebook 4

Example 4.17 - Shifted exponential



$$H_0: \theta \leq 0$$

Reject H_0 if $X_{(1)} \geq -\frac{1}{n} \log \alpha_0$
at level α_0 .

minimal sample size for power
at least 0.8 in $\theta = 0.1$.

Power function $\Pi(\theta; k)$

$$\begin{aligned} P_\theta(X_{(1)} \geq -\frac{1}{n} \log \alpha_0) &= \left(P_\theta(X_1 \geq -\frac{1}{n} \log \alpha_0) \right)^n \\ &= \alpha_0 e^{-n\theta}. \end{aligned}$$

So we impose

$$\Pi(0.1; k) \geq 0.8 \Leftrightarrow \alpha_0 e^{-n/10} \geq 0.8$$

and as $\alpha_0 = 0.05$ $0.05 e^{-n/10} \geq 0.8$
implies $n \geq 27.7$.

Example - Contaminated water

See \rightarrow Example 3.19.

Is the water contaminated by Legionella bacteria?

we use as norm that the probability of more than 100 colony-forming units of bacteria per liter should be at most 5%.

Let X be the number of colony-forming units of Legionella bacteria in a 1 liter sample of pool water. We assume that X has a Poisson distribution with unknown parameter μ .

Now consider the probability

$$P_\mu = P(X > 100)$$

The quality of water can be tested comparing

$$H_0 : P_\mu \leq 0.05 \text{ v.s } H_1 : P_\mu > 0.05$$

We partition the sample mixed with 100 liters of pure water over 100 Petri dishes of 1 liter each

Define X_i = # colonies in the i -th Petri dish

$$Y_i = \mathbb{1}(X_i > 0)$$

$$(X_1, \dots, X_n) \text{ iid } \text{Poi}(\mu/100)$$

$$(Y_1, \dots, Y_n) \text{ iid } \text{Bern}(q_\mu)$$

where

$$q_\mu = P(X > 0) = 1 - e^{-\frac{\mu}{100}}$$

→ Transformation of the H_0 .

$$H_0 : P_\mu \leq 0.05$$



$$P_\mu(X > 100) \leq 0.05$$



$$\mu \leq 85.05$$

→ see notebook

$$\uparrow \downarrow \\ 1 - e^{-\frac{\mu}{100}} \leq 1 - e^{-\frac{85.05}{100}}$$

$$\uparrow \downarrow \\ H_0': q_\mu \leq 0.5728$$

Can be tested with test statistic

$T = \sum Y_i$ = # Petri dishes
with presence of
bacteria.

where $T \sim \text{Bin}(100; q_\mu)$

We use the Binomial test

Reject at level α_0 if

$$\sum Y_i \geq n q_0 + \frac{1}{2} + \xi_{1-\alpha_0} \sqrt{n q_0 (1-q_0)}$$

$$= 100 \times 0.5728 + \frac{1}{2} + 1.645 \sqrt{100 \times 0.5728 \times 0.4272}$$

$$\approx 57.28 + 0.5 + 1.645 \sqrt{24.47} = 65.91 .$$

P-values

In Fisherian inference testing
is done using p-values.

p-value

For a test that rejects H_0 for
large values of a test statistic T
is

$$p = \sup_{\theta \in \mathcal{H}_0} P_\theta(T \geq t_{obs})$$

where t_{obs} is the observed value of T .

Example: A furniture manufacturer
gets wooden boards of 6 mm. thickness.
To reject the boards makes the test of

$$H_0: \mu \leq 6 \text{ vs } H_1: \mu > 6.$$

→ He takes a sample of 9 boards

$$\underline{X} = (X_1, \dots, X_9)$$

assuming that $X_i \sim N(\mu, 0.1 \text{ mm})$

→ Test statistic : $T = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$

→ Sample mean ; 6.07 mm.

→ Observed test statistic : $t_{\text{obs}} = \frac{6.07 - 6}{0.1/\sqrt{3}} = 2.1$

→ P-value

$$\sup_{\mu \leq 6} P_\mu \left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \geq 2.1 \right) =$$

$$= \sup_{\mu \leq 6} P_\mu \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} + \frac{\mu - \mu_0}{\sigma/\sqrt{n}} \geq 2.1 \right)$$

$$= \sup_{\mu \leq 6} 1 - \Phi \left(2.1 - \frac{(\mu - 6)}{0.1/\sqrt{3}} \right)$$

= sup obtained for $\mu = 6$.

$$= 1 - \Phi(2.1) = 1 - 0.9821 = 0.0179.$$

Thus $P = 0.0179$

Interpretation.	Evidence against H_0
$p \leq 0.01$	highly significant
$0.01 < p \leq 0.05$	significant
$0.05 < p$	not significant

When the p-value is $\leq \alpha$, we reject H_0 . otherwise we don't.

In words,

The p-value is the maximum probability under the null that an identical experiment gives a value of the test statistic larger than that observed.

Extending the p-value to other critical regions

- For a test that rejects the null for small values of T

$$P = \sup_{\theta \in \Theta_0} P_\theta (T \leq t_{obs})$$

- For a test that rejects the null for small and large values of T

$$P = 2 \min \left\{ \sup_{\theta \in \Theta_0} P(T \leq t_{obs}), \sup_{\theta \in \Theta_0} P_\theta (T \geq t_{obs}) \right\}$$

Again, when p-value $\leq \alpha_0$ reject H_0 .

This amounts to checking whether one of the one-sided p-values is $\leq \alpha_0/2$

$$\text{since } 2 \min(a, b) \leq \alpha_0 \Leftrightarrow a \leq \alpha_0/2 \text{ or } b \leq \frac{\alpha_0}{2}$$

For a collection of tests that contains a test of level α for every $\alpha \in (0, 1)$, the observed significance level or p-value is the smallest value of α for which the corresponding test rejects H_0 .

Example 4.24

$$H_0: p \leq 0.5 \quad H_1: p > 0.5$$

$$n = 100 \quad X = 64.$$

$$\text{p-value} = \sup_{p \leq 0.5} P_p(X \geq 64) = P_{0.5}(X \geq 64)$$

As $X \sim \text{Bin}(p=0.5; n=100)$ we compute

$$P = 1 - \text{bincdf}(64) \quad \text{see } \underline{\text{Notebook 4}}$$

$$= 0.0033.$$

| For 64 successes the null is rejected
at all levels $\alpha > 0.0033$. |

| \Rightarrow p-value is more informative. |

4.6 Some standard Tests

the general idea is to find a test statistic that is "reasonable" (often based on a good estimator for the parameter) and for which we can easily compute a critical value or p-value

Chi-square distribution

→ Important when testing parameters related to the normal and for large sample approximations

Chi-square χ_m^2
A r.v. W has a χ_m^2 distribution with m degrees of freedom if

$$W \equiv \sum_{i=1}^m X_i^2$$

for $(X_1, \dots, X_m) \stackrel{iid}{\sim} N(0,1)$

χ^2_n is a special case of Gamma2

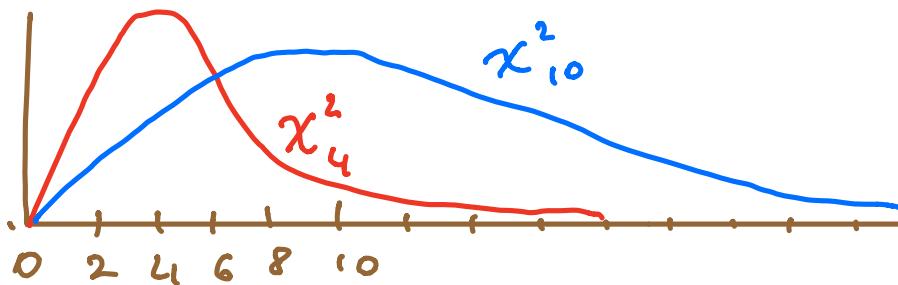
$\chi^2_n \equiv \text{Gamma}(\alpha, \lambda)$ with

$$\frac{\alpha}{\lambda} = n \quad \frac{\alpha}{\lambda^2} = 2n$$

\Leftrightarrow

$$\alpha = \frac{n}{2}; \lambda = \frac{1}{2}$$

$$\text{So } E(W) = n; \text{ var}(W) = 2n.$$



Student's T distribution

$T \sim t_m$ is a t with m degrees of freedom if T has the same distribution

as

$$\frac{Z}{\sqrt{W/m}}$$

with $Z \sim N(0,1)$

$W \sim \chi^2_m$
independent

$t_1 = \text{Cauchy}$ $t_\infty = N(0, 1)$

$$E(\tau) = \begin{cases} \text{undefined for } n=1 \\ 0 \text{ for } n>1. \end{cases}$$

The following result is important:

Theorem 4.29 — if $\underline{x} \sim N(\mu, \sigma^2)$

(i) $\bar{x} \sim N(\mu, \sigma^2/n)$

(ii) $(n-1)S_x^2/\sigma^2 \sim \chi_{n-1}^2$

(iii) \bar{x} and S_x^2 are independent

(iv) $\frac{\bar{x} - \mu}{S_x / \sqrt{n}} \sim t_{n-1}$

Proof (i) ok.

Assume without loss of generality
that $\mu=0$ and $\sigma^2=1$. The density
of $(x_1, \dots, x_n)^T = \underline{x}$ is

$$\underline{x} = (x_1, \dots, x_n) \mapsto (2\pi)^{-\frac{m}{2}} \exp\left(-\frac{1}{2} \|\underline{x}\|^2\right)$$

$$\text{Let } \underline{f} = (1, \dots, 1)^T / \sqrt{n} \quad \|\underline{f}\|^2 = 1$$

and extend it arbitrarily to an orthonormal basis $(f_1, \dots, f_n) \in \mathbb{R}^n$.

$$\text{Define the matrix } O = \begin{pmatrix} \underline{f}^T \\ \vdots \\ \underline{f}_n^T \end{pmatrix} \in \mathbb{R}^{n \times n}$$

Thus:

$$O \underline{O}^T = I_n \Leftrightarrow O^{-1} = O^T.$$

$$\|\underline{Ox}\|^2 = \underline{x}^T O^T O \underline{x} = \|\underline{x}\|^2.$$

$$\text{Let } Y = O \underline{x} = \begin{bmatrix} \underline{f}_1^T \\ \vdots \\ \underline{f}_n^T \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix}$$

$$\text{so that } y_1 = \underline{f}_1^T \cdot \underline{x} = \sqrt{n} \bar{x}$$

$$\text{and } \sum_{i=2}^m y_i^2 = \|\underline{Y}\|^2 - y_1^2 \\ = \|\underline{X}\|^2 - m \bar{x}^2$$

$$= \sum_{i=1}^m (x_i - \bar{x})^2 = (m-1)S_x^2$$

(ii) and (iii) follow if we prove
that y_1, \dots, y_n are iid $N(0, 1)$.

$$P(\underline{Y} \leq \underline{y}) = \int \dots \int_{\underline{x}: O\underline{x} \leq \underline{y}} (2\pi)^{-n/2} \exp\left(-\frac{1}{2} \|\underline{x}\|^2\right) d\underline{x}$$

substitution: $O\underline{x} = \underline{u}$

$$\|\underline{x}\| = \|O\underline{x}\| = \|\underline{u}\|$$

$$\text{Jacobian} = \det O = 1$$

$$= \int \dots \int_{\underline{u}: \underline{u} \leq \underline{y}} (2\pi)^{-n/2} \exp\left(-\frac{1}{2} \|\underline{u}\|^2\right) d\underline{u}$$

So \underline{Y} has the same joint density of \underline{X}
and y_1, \dots, y_n are iid $N(0, 1)$.

(iv) Follows by

$$\frac{\bar{X} - \mu}{S_x/\sqrt{n}} = \frac{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1) S_x^2 / \sigma^2}{n-1}}} \sim \mathcal{N}(0, 1) \quad //$$



je test et di "Student"

used for tests about the mean
given samples from a $N(\mu, \sigma^2)$
with σ^2 unknown.

$$H_0: \mu \leq \mu_0 \quad H_1: \mu > \mu_0$$

$$\underline{\Theta} = (\mu, \sigma^2) \quad \Theta_0 = \{(\mu, \sigma^2) : \mu \leq \mu_0, \sigma^2 > 0\}$$

The Gauss test cannot be used because the test statistic

$$T = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \text{ depends on } \sigma^2$$

T-test statistic

$$T = \frac{\bar{X} - \mu_0}{S_x/\sqrt{n}}$$

with $S_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$

Reject $H_0: \mu \leq \mu_0$ for large values of T .

Under $\mu = \mu_0$

$$T \sim t_{n-1}$$

For a test of level α , we must have

$$\alpha = \sup_{\substack{\mu \leq \mu_0 \\ \sigma^2 > 0}} P_{\mu, \sigma^2} \left(\frac{\bar{X} - \mu_0}{S_x / \sqrt{n}} \geq c \right) \leq \alpha_0.$$

It can be proved that

$$\alpha = \sup_{\sigma^2 > 0} P_{\mu_0, \sigma^2} \left(\frac{\bar{X} - \mu_0}{S_x / \sqrt{n}} \geq c \right) \leq \alpha_0.$$

So that the distribution of

$$\frac{\bar{X} - \mu_0}{S_x / \sqrt{n}} \sim t_{n-1}$$

independently of (μ_0, σ^2)

To obtain the largest possible

power we take $c = t_{n-1, 1-\alpha_0}$

(the $(1-\alpha_0)$ -quantile of the t_{n-1})

The p-value is

$$p = P_{\mu_0, \sigma^2} \left(T \geq \frac{\bar{X}_{\text{obs}} - \mu_0}{S_{x, \text{obs}}} \right)$$

- for small values of n ($n \leq 10$)
the t-test is greatly different
from the Gauss test (that has a
size much greater than desired.)
- for $n \geq 20$ t-test gives results
practically identical to the Gauss
test.
- Adjusting the t-test to test problems
 $H_0: \mu \geq \mu_0$ or $H_0: \mu = \mu_0$
is analogous to the Gauss test.
- Data must be normal! Otherwise
we can transform the observations.

Example → see Notebook 4

Example 4.31 Sign test.

a test for the median μ

$$H_0: \mu \leq \mu_0$$

Use the test statistic

$$T = \#\{X_i > \mu_0\} \sim \text{Bin}(n, P_\mu)$$

where $P_\mu = P_\mu(X > \mu_0)$

- If $\mu = \mu_0$

$$P_\mu = \frac{1}{2}$$

- if $\mu \leq \mu_0$

$$P_\mu \leq \frac{1}{2}$$

- if $\mu > \mu_0$

$$P_\mu > \frac{1}{2}$$

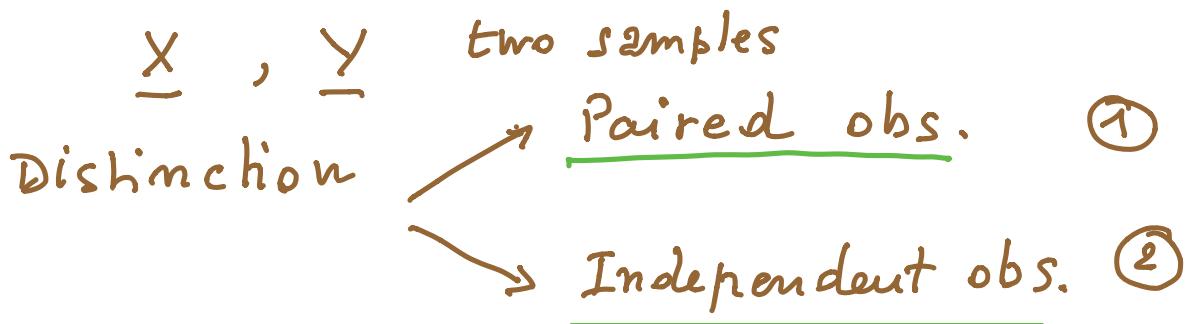
Thus the sign test is equivalent
to a binomial test of

$$H_0 : P_\mu \leq \frac{1}{2} \quad \text{vs} \quad H_1 : P_\mu > \frac{1}{2}$$

with statistic T .

Example → Notebook 4

4.63 Two sample test



① PAIRED

X_i = state of patient BEFORE treatment

Y_i = state of patient AFTER treatment.

$(X_i, Y_i) \quad i=1, \dots, n$ (same patient)

X_i and Y_i may be related

(X_i, Y_i) are independent.

② INDEPENDENT

$$(X_1, \dots, X_m) \stackrel{iid}{\sim} F_X$$
$$(Y_1, \dots, Y_m) \stackrel{iid}{\sim} F_Y$$

Example 4.33 Paired observations

Effect of treatment = $Z_i \cdot = X_i - Y_i$

Assumption $Z_1, \dots, Z_n \stackrel{iid}{\sim} N(\Delta, \sigma^2)$

with

$$\Delta = E X_i \cdot - E Y_i \cdot$$

The assumption is reasonable even if X_i and Y_i are dependent.

Test $H_0: \Delta = 0$ vs $H_1: \Delta \neq 0$

test $t: \frac{|z|}{S_z / \sqrt{n}} \geq t_{m-1, 1-\alpha/2}$.
Reject if

The power depends strongly

$$\text{on } \sigma^2 = \text{var}(Z_i) = \sigma_x^2 + \sigma_y^2 - 2 \sigma_{xy}$$

A small variance ensures a large power and the difference Δ can be detected easily.

A large $\sigma_{xy} \Rightarrow$ a small σ^2 .

→ Notebook 4

Example 4.34 Two sample t-test

$$(x_1, \dots, x_m) \sim \text{iid } N(\mu, \sigma^2) \quad \text{same } \sigma^2.$$

$$(y_1, \dots, y_n) \sim \text{iid } N(\nu, \sigma^2)$$

Test

$$H_0: \mu - \nu \leq 0 \quad \text{vs} \quad H_1: \mu - \nu > 0$$

- Estimate of $\mu - \nu \rightarrow \bar{x} - \bar{y}$
- Reject H_0 if $\bar{x} - \bar{y} > c$

Distribution

$$\bar{x} - \bar{y} \sim N\left(\mu - \nu; \frac{\sigma^2}{m} + \frac{\sigma^2}{n}\right)$$

depends on σ^2

So is not a test statistic.

test statistic

$$T = \frac{\bar{x} - \bar{y}}{\sqrt{s_{x,y}^2 \left(\frac{1}{m} + \frac{1}{n} \right)}}$$

with $s_{x,y}^2 = \frac{1}{m+n-2} \cdot \left(\sum_{i=1}^m (x_i - \bar{x})^2 + \sum_{j=1}^n (y_j - \bar{y})^2 \right)$

It will be proved that $s_{x,y}^2$ is
unbiased for the common variance σ^2 .

The MLE for σ^2 is

$$\hat{\sigma}_{ML}^2 = \frac{(m+n-2)}{m+n} s_{x,y}^2 .$$

If $\mu = \nu$

$$T \sim t_{m+n-2}$$

$$\frac{(\bar{x} - \bar{y})}{\sqrt{\left(\frac{1}{m} + \frac{1}{n}\right) \sigma^2}}$$

$N(0, 1)$

$$T = \sqrt{\frac{(m+n-2) S_{x,y}^2 / \sigma^2}{m+n-2}} \quad \frac{\chi^2_{m+n-2}}{m+n-2}$$

Note that

$$(m+n-2) \frac{S_{x,y}^2}{\sigma^2} = (m-1) \frac{S_x^2}{\sigma^2} + (n-1) \frac{S_y^2}{\sigma^2}$$

$$\frac{\chi^2_{m-1}}{\bar{x}} \perp \frac{\chi^2_{n-1}}{\bar{y}}$$

If the variances are $\sigma_x^2 \neq \sigma_y^2$

This is called the Behrens-Fisher

problem

There is no best test for this situation. \rightarrow See Book

- If the two populations are not normal, but the sample sizes m and n are large then we can use the test statistics

$$\frac{\bar{X}_m - \bar{Y}_m}{\sqrt{\frac{s_x^2}{m} + \frac{s_y^2}{n}}} \xrightarrow{L} N(0,1)$$

Thus this is an asymptotic test.

4.6.4 Goodness of fit tests

Is a test designed to check whether the distribution of observations belongs to a certain family.

Difference with standard tests:
we generally prefer not to reject H_0

Example Kolmogorov-Smirnov

Let $(x_1, \dots, x_n) \sim F$

The hypothesis is

$$H_0 : F = F_0$$

$$H_1 : F \neq F_0$$

Example : $F_0 = N(0,1)$

The KS test is based on the

Empirical distribution function

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{[x_i \leq x]}$$

= proportion of obs. $\leq x$

→ Notebook 4

$$\text{So } \hat{F}_n(x) \xrightarrow{P} E 1_{[x_i \leq x]} = F(x)$$

by the LLN.

If H_0 is True $F_n \approx F_0$ for sufficiently large n .

The KS statistic is

$$T = \sup_{x \in \mathbb{R}} |F_n(x) - F_0(x)|$$

and we reject the null for $T > c$

We need the distribution of T under H_0 . This is Tabulated and available in computer programs, and the same for every continuous F_0 .

→ Notebook 4

More generally we want to test

$H_0 : F \in \{F_\theta : \theta \in \Theta\}$.
for a given statistical model

Example:

$$H_0 : F \in \{N(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 > 0\}.$$

Test statistic

$$T^* = \sup_{x \in \mathbb{R}} |F_n(x) - \Phi\left(\frac{x - \bar{x}}{s_x}\right)|$$

Its distribution does not depend
on μ, σ^2 , and can be determi-
ned by simulation.

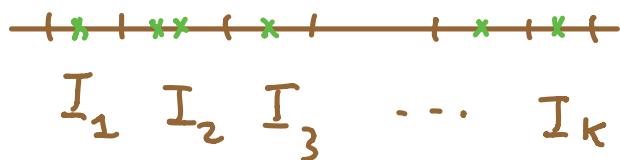
→ Notebook 4

Example 4.39 Chi-square test

Let $(X_1, \dots, X_n) \sim \text{iid } F$

Alternative to Kolmogorov-Smirnov test. for $H_0: F = F_0$ vs $H_1: F \neq F_0$

- Define a partition of the codomain of $X = X_1$



- Let $N_j := \#(1 \leq i \leq n : X_i \in I_j)$

$$P_j = P_{H_0}(X \in I_j)$$

$$= F_0(a_j) - F_0(a_{j-1})$$

$$E_{H_0}(N_j) = n P_j$$

chi-square test statistic

$$\underline{\chi}^2 = \sum_{j=1}^n \frac{(N_j - np_j)^2}{np_j}$$

Under $H_0 : F = F_0$

$$\underline{\chi}^2 \xrightarrow{L} \chi^2_{k-1} \text{ for } n \rightarrow \infty \\ k \text{ fixed}$$

4.7 LIKELIHOOD RATIO TEST

a general method for
finding tests.

Example : Is a coin fair ?

$$H_0 : p = \frac{1}{2} \quad H_1 : p \neq \frac{1}{2}$$

Data: $x = \# \text{ heads}$
 $\sim \text{Bin}(n, p)$

Likelihood

$$L(p; x) = \binom{n}{x} p^x (1-p)^{n-x}$$

Likelihood ratio

$$\lambda(x) = \frac{\sup_{\theta} L(\theta; x)}{L(p_0; x)} = \frac{\hat{L}}{L_0} .$$

for instance $n = 100, x = 60$

$$\hat{p}_{ML} = 0.6$$

$$\begin{aligned} \lambda(x) &= \frac{\hat{L}}{L_0} = \frac{\binom{100}{60} \hat{p}^{60} (1-\hat{p})^{40}}{\binom{100}{60} p_0^{60} (1-p_0)^{40}} \\ &= \frac{0.0812191}{0.0108439} = 7.49 \end{aligned}$$

The distribution of $\lambda(x)$ under the null is known asymptotically in this form :

$$2 \log \lambda_m(x) \xrightarrow{L} \chi^2_1 \quad \text{for } m \rightarrow \infty$$

\rightarrow See notebook 4

More generally we give the def:

+ Likelihood ratio statistic

for $H_0: \theta \in \Theta_0$

$$\lambda(x) = \frac{\sup_{\theta \in \Theta} P_\theta(x)}{\sup_{\theta \in \Theta_0} P_\theta(x)}$$

[Idep] maximize the likelihood twice

without restriction
under $H_0: \theta \in \Theta_0$

- The λ is always ≥ 1 .
- Larger values indicate that the space Θ_1 contains more likely parameters than the null.
- $K = \{\underline{x} : \lambda(\underline{x}) \geq c\}$

Theorem 4.43

Under suitable regularity conditions for $n \rightarrow \infty$

$$2 \log \lambda_n(\underline{x}) \xrightarrow{L} \chi^2_{K - K_0}$$

where $K = \dim \Theta$, $K_0 = \dim \Theta_0$.

The regularity conditions are discussed more in chapter 5.

Briefly, we need that

- the log-likelihood is continuous and differentiable for every x
- the derivative $\ell'_{\theta}(x)$ is such that $|\ell'_{\theta}(x)| \leq g(x)$ for all θ in a neighborhood of θ_0 and $g(x)$ is a function such that $E_{\theta_0} g(X)^2 < \infty$.
- θ_0 is not a boundary point of Θ or Θ^c .
- etc ...

Example: - Number of tropical cyclones

Season: 1 2 3 4 5 6 7 8 9 10 11 12 13
 x_i : 6 5 4 6 6 3 12 7 4 2 6 7 4

Assume that $X_i \stackrel{iid}{\sim} \text{Poisson}(\theta)$

We want to test that

$$\theta = 5 \text{ against } \theta \neq 5$$

Likelihood

$$L(\theta; \underline{x}) = \prod_{i=1}^{13} \frac{\theta^{x_i} e^{-\theta}}{x_i!}$$

log-lik

$$\begin{aligned} l(\theta; \underline{x}) &= \sum_{i=1}^{13} (x_i \log \theta - \theta - \log x_i!) \\ &= \text{const.} + n \bar{x} \cdot \log \theta - n\theta. \end{aligned}$$

lik equation

$$\begin{aligned} l'(\theta; \underline{x}) &= n \bar{x} - n = 0 \\ \hat{\theta} &= \bar{x}. \end{aligned}$$

$$\text{In the data } \bar{x} = \frac{72}{13} = 5.538$$

$$\text{So } \hat{\theta} = 5.538.$$

Likelihood ratio.

$$\begin{aligned}2 \log \lambda &= 2(\ell(\hat{\theta}) - \ell(\theta_0)) \\&= 2(n \bar{x} \log \hat{\theta} - n \hat{\theta} - n \bar{x} \log \theta_0 + n \theta_0) \\&= 2n \left(\bar{x} \log \frac{\hat{\theta}}{\theta_0} - (\hat{\theta} - \theta_0) \right)\end{aligned}$$

Gives $2 \log \lambda =$

$$\begin{aligned}&= 26 \left(5.538 \log \left(\frac{5.538}{5} \right) - (5.538 - 5) \right) \\&\approx 0.728\end{aligned}$$

P-value : $P(\chi^2_1 > 0.72) = 0.396.$

Not enough evidence.

NOTE the critical region is always one-sided even if the H_1 is two-sided.