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 hypoteses.

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

$$\not\Rightarrow \begin{cases} H_1 \text{ incorrect} \\ \text{and } H_0 \text{ correct} \end{cases}$$

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 correct.

  strong

- Do not reject H_0 but do not accept H_1 as being correct.

  weak (more investigation needed)

Possible errors

- type I : Reject H_0 when correct

$\Rightarrow$ accept strong conclusion what is wrong.

- type II : Mantain H_0 when incorrect.

$\Rightarrow$ accept the weak conclusion

ASYMMETRY

type I	VERY UNDESIRABLE
type II	not as bad.

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

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

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

Example 4.2 Playing dice

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

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

Example 4.3 Two samples model

$(X_1 \dots X_m)$ $(Y_1 \dots Y_n)$ all independent.

$\downarrow \qquad \downarrow$

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

See Notebook 4 Box-plot not helpful

$$\begin{cases} H_0 : \mu = \nu \\ H_1 : \mu \neq \nu \end{cases}$$ formal test.

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

Θ_0 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

$(X_1, \dots, X_n) \overset{iid}{\sim} N(\mu, \sigma^2)$ σ^2 known.
(unrealistic but see later)

Test of

$$H_0 : \mu \leq \mu_0 \text{ against } H_1 : \mu > \mu_0$$

Example: Quality control based on a sample $\underline{X}$ and a Teststatistic

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

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

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

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

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

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

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

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

$\boxed{4.3.1}$ Size and power function

$\boxed{\text{what we want}}$

if $\theta \in \Theta_0$ $P(\underline{x} \in K)$ small

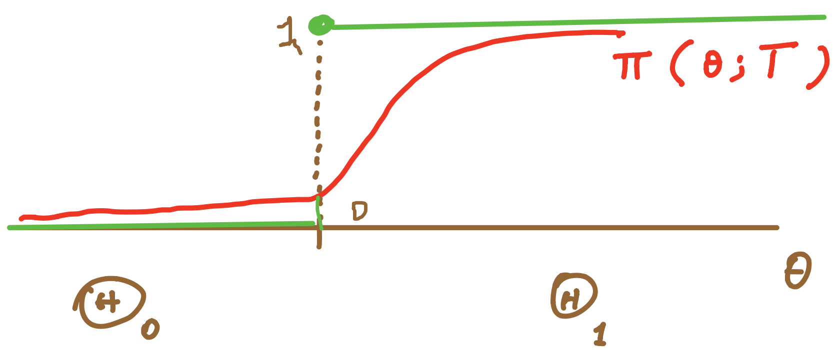
if $\theta \notin \Theta_0$ $P(\underline{X} \in K)$ high

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

$$\theta \longmapsto P_\theta(\underline{X} \in K) = \pi(\theta, T)$$

called $\boxed{\text{Power function}}$

Example of a real power function and ideal



Size of a test T

$$\alpha = \sup_{\theta \in \Theta_0} \pi(\theta; T)$$

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

$$P(\mathrm{II}) = P_\theta(\underline{X} \notin K) \quad \theta \in \Theta_1$$

tends to become very large.

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

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 new therapy to 100 patients.
- Let $X = \#$ successes $\sim Bin(100, p)$
- 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

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

$$* \quad = P_{0.5}(X \geq 59) =$$

$$= 1 - \text{binocdf}(58, 100, 0.5)$$

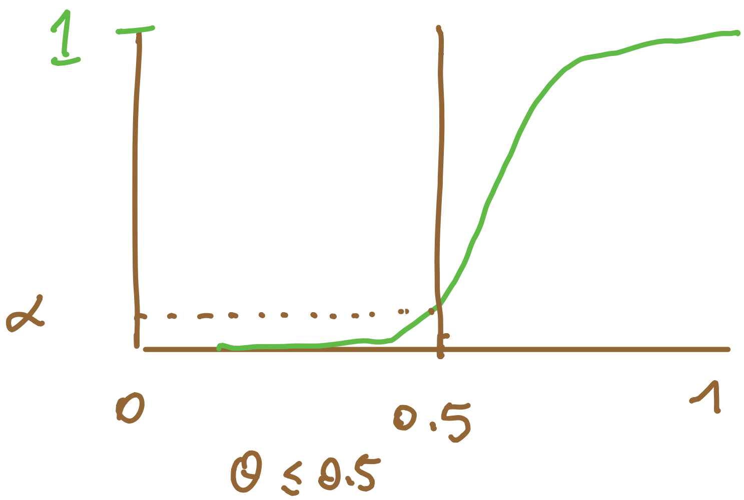
$$= 0.0443$$

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

The power of the test is

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

→ See Notebook



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

we have

size $= P_{0.5}(X \geqslant 59) = 0{,}0443$ OK

size $= P_{0.5}(X \geqslant 58) = 0.067$ 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

$\underset{\sim}{X} = (X_1, \dots, X_n) \overset{iid}{\sim} N(\mu, \sigma^2)$ ↳ known

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

Test statistic $T(\underset{\sim}{x}) = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$

Equivalent to $T(\underset{\sim}{x}) = \bar{x}$.

Critical region $K = \left\{ \underset{\sim}{x} : \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \geq c \right\}$

Note that

$$\boxed{T(\underset{\sim}{x}) \sim N(0,1) \text{ under } \mu = \mu_0}$$

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 $c \geq \xi_{1-\alpha_0}$ where $\xi_{1-\alpha}$ is the $1-\alpha_0$-quantile of the $N(0,1)$

Proof:

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

$$= 1-\Phi\left(c + \frac{\mu_0-\mu}{\sigma/\sqrt{n}}\right).$$

is an increasing function of μ.

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

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

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

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

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

and $c \geq \xi_{1-\alpha_0}$ is the solution.

CHOOSE $c = \xi_{1-\alpha_0}$

Critical region for the Gauss test

$H_0\, \mu \geq \mu_0 \qquad H_1\, \mu < \mu_0$

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

critical region for the Gauss test

$H_0\, \mu = \mu_0 \qquad H_1\, \mu \neq \mu_0$

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

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

See Notebook 4

Example 4.13

Binomial test

In general

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

1) reject the null if $X = \#$ 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.

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

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

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

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

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

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

Example $\quad X = 64 \quad p_0 = 0.58$

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

norminv $(1-0.05)$

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

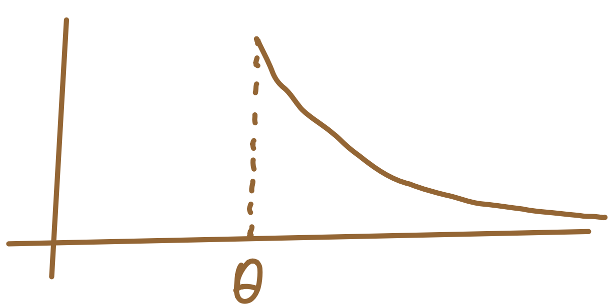
$= 66.12$

Not reject $H_0 : p \leq 0.58$.

Example 4.14 - Shifted exp.

$(X_1, \dots, X_n) \overset{iid}{\sim} p_\theta(x) = e^{\theta - x}$

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

$$= P_\theta(X_1 \geq c, X_2 \geq c, \ldots X_n \geq c)$$

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

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

→ The sup is taken in $\theta = 0$.

So

$$\alpha_0 \geq e^{-nc}. \Longleftrightarrow \boxed{c \geq -\frac{1}{n}\log\alpha_0}$$

→ The critical region is

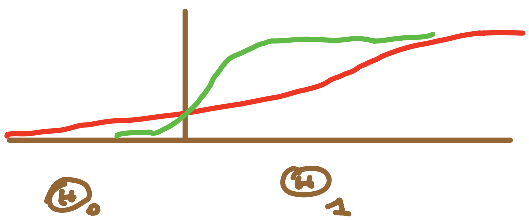
$$K = \left\{ \underline{X} \; ; \; x_{(1)} \geq -\frac{1}{n}\log\alpha_0 \right\}.$$

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



Can we influence the slope of $\pi(\mu, T)$?

Example Gauss test (continued)

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

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

Power function

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

- 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.

→ 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 $H_0: \mu \leq 0$

$H_1: \mu > 0$

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

$\alpha_0 = 0.05$.

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

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

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

Reject H_0 at approximate level α_0:
if

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

(normal Approximation better with

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

Power function

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

Example

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

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

2) power 0.8 in $p = 0.6$

$P_{0.6}(X \geqslant c) = 0.8$

System :

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

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

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

Similarly :

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

Thus :

$$\begin{cases} c - \frac{n}{2} - \frac{1}{2} \simeq 1.645 \sqrt{\frac{n}{4}} \\ \\ c - n \times 0.6 - \frac{1}{2} \leq -0.842 \sqrt{n \times 0.24} \end{cases}$$

The first gives

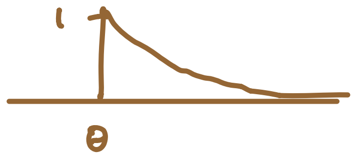
$$c \cong \frac{n+1}{2} + 1.645 \frac{\sqrt{n}}{2}$$

Substituting in the second gives

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

$\Rightarrow$ $n > 152$. See notebook 4

Example 4.17 - Shifted exponential

$X_i \sim$ 

$H_0 : \theta \leq 0$

$\boxed{\text{Reject } H_0 \text{ if } \quad X_{(1)} \geq -\frac{1}{n} \log \alpha_0 \quad \text{at level } \alpha_0}$

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

Power function $\pi(\theta; K)$

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

So we impose

$\pi(0.1; K) \geq 0.8 \iff \alpha_0 e^{-n/10} \geq 0.8$

and as $\alpha_0 = 0.05$ $\quad 0.05\, e^{-n/10} \geq 0.8$

implies $\quad n \geq 27.7$.

Example – Contaminated water

See → 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

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

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

$Y_i = 1(X_i > 0)$

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

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

where

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

→ Transformation of the H_0.

$$H_0 : \; p_\mu \leq 0.05$$

$$\Updownarrow$$

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

$$\Updownarrow$$

$$\mu \leq 85.05$$

→ see notebook

$$\Updownarrow$$

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

$$\Updownarrow$$

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

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

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

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

$$\approx 57.28 + 0.5 + 1.645 \sqrt{24.47} = 65{,}91.$$

$\boxed{\text{p-values}}$

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

$$p = \sup_{\theta \in \Theta_0} P_\theta (T \geq t_{obs})$$

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

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

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

→ He takes a sample of 9 boards

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

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

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

→ Sample mean : 6.07 mm.

→ Observed test statistic : $t_{oss} = \dfrac{6.07 - 6}{0.1/3} = 2.1$

→ P-value

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

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

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

= sup obtained for $\mu = 6$.

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

Thus $\boxed{p = 0.0179.}$

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 $\leq \alpha_0$ 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 wether one of the one-sided p-values is $\leq \alpha_0/2$

Since $2 \min(a,b) \leq \alpha_0 \iff a \leq \alpha_0/2$ or $b \leq \frac{\alpha_0}{2}$

For a collection of tests that contains a test of level α for every α ∈ (0, 1), the observed significance level or p-value is the smallest value of α for which the corresponding test rejects H0.

Example 4.24

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

$n = 100 \qquad X = 64.$

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

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

$p = 1 - \text{binocdf}(64)$ see Notebook 4

$= 0.0033.$

For 64 successes the null is rejected at all levels $\alpha_0 \geq 0.0033.$

$\Rightarrow$ 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 approximation

Chi-square χ^2_m

A r.v. W has a χ^2_m distribution with m degrees of freedom if

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

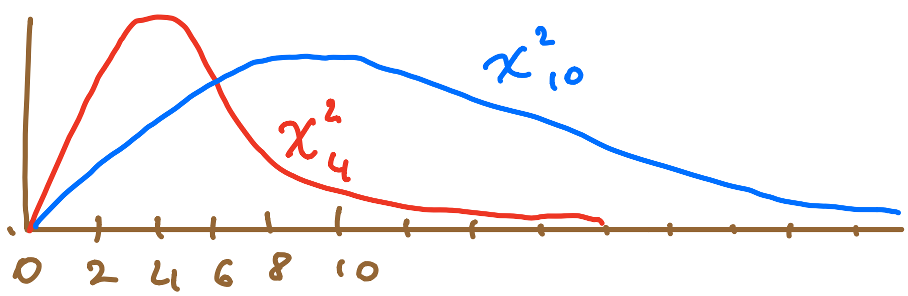
for $(X_1 \dots X_m) \overset{iid}{\sim} N(0,1)$

χ^2_n is a special case of Gamma

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

$$\boxed{\frac{\alpha}{\lambda} = n \qquad \frac{\alpha}{\lambda^2} = 2n} \iff \boxed{\alpha = \frac{n}{2} \; ; \; \lambda = \frac{1}{2}}$$

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



Student's T distribution

$T \sim t_n$ is a t with n degrees of freedom if T has the same distribution as

$$\frac{Z}{\sqrt{W/n}} \qquad \text{with } Z \sim N(0,1), \; W \sim \chi^2_n \text{ independent}$$

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

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

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) $(n-1) S_x^2/\sigma^2 \sim \chi^2_{n-1}$

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

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

Proof (i) ok.

Assume that $\mu = 0$ and $\sigma^2 = 1$ and we prove first that:

$$\boxed{\sum (X_i - \bar{X})^2 \sim \chi^2_{n-1} \perp\!\!\!\perp \bar{X}} \qquad (*)$$

The density of $(X_1, \dots X_n)^T = \underline{X}$ is

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

- First we transform $X \to Y = OX$

with an orthogonal matrix O

constructed in this way:

Let $\underline{f}_1 = (1, \dots, 1)^T / \sqrt{n}$ $\quad \|f_1\|^2 = 1$

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

Define the matrix $O = \begin{pmatrix} f_1^T \\ \vdots \\ f_n^T \end{pmatrix} \in \mathbb{R}^{n \times n}$

Thus:

$$\boxed{\begin{gathered} OO^T = I_n \iff O^{-1} = O^T. \\ \|O\underline{x}\|^2 = \underline{x}^T O^T O \underline{x} = \|\underline{x}\|^2. \end{gathered}}$$

Let $Y = OX = \begin{bmatrix} f_1^T \\ \vdots \end{bmatrix} \begin{bmatrix} X_1 \\ \vdots \end{bmatrix}$

$$\begin{bmatrix} f_n^T \end{bmatrix} \quad \begin{bmatrix} X_n \end{bmatrix}$$

so that $Y_1 = f_1^T \cdot X = \sqrt{n}\, \bar{X}$

and $$\sum_{i=2}^{n} Y_i^2 = \|\underline{Y}\|^2 - Y_1^2$$
$$= \|\underline{X}\|^2 - n\bar{X}^2$$
$$= \sum_{i=1}^{n} (X_i - \bar{X})^2 = (n-1)S_X^2 .$$

- Now we show that $(Y_1, \dots, Y_n)$ are i.i.d $N(0,1)$.

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

substitution: $O\underline{x} = \underline{u} \Rightarrow \|x\| = \|Ox\| = \|u\|$

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

So $\underline{Y}$ has the same joint density of $\underline{X}$ and $Y_1 \dots Y_n$ are $\sim$ i.i.d $N(0,1)$.

Equation (*) follows because:

$$Y_1 \sim N(0,1) \perp\!\!\!\perp (Y_2 \dots Y_n) \overset{iid}{\sim} N(0,1)$$

$$\Downarrow \qquad\qquad\qquad \Downarrow$$

$$\sqrt{n}\,\bar{X} \sim N(0,1) \perp\!\!\!\perp \sum_{i=2}^{n} Y_i^2 = (n-1) S_X^2 \sim \chi^2_{n-1}$$

- In the general case $X_i \sim N(\mu, \sigma^2)$

we apply (*) to $Z_i = \dfrac{X_i - \mu}{\sigma} \sim N(0,1)$

So that $Y = OZ$

- $Y_1 = \sqrt{n}\,\bar{Z} = \sqrt{n} \displaystyle\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)\frac{1}{n}$

$$= \frac{\sqrt{n}}{n} \; \frac{n\bar{X} - n\mu}{\sigma} = \frac{(\bar{X} - \mu)}{\sigma/\sqrt{n}} \sim N(0,1)$$

$$\sum_{i=2}^{n} Y_i^2 = \sum_{i=1}^{n} (Z_i - \bar{Z})^2$$

$$= \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} - \frac{\bar{X} - \mu}{\sigma} \right)^2$$

$$= \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{\sigma^2} = (n-1) S_x^2 \sim \chi^2_{n-1}$$

independent of $\bar{X}$.

So (ii) and (iii) are proved.

(iv) Follows by

$$\frac{\bar{X} - \mu}{S_x / \sqrt{n}} = \frac{\dfrac{\bar{X} - \mu}{\sigma / \sqrt{n}}}{\sqrt{\dfrac{(n-1) S_x^2 / \sigma^2}{n-1}}}$$

//

$\boxed{\text{"Student" } t \text{ test.}}$

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

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

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

The Gauss test cannot be used because the test statistic

$$T = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$ depends on σ^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_0 we must have

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

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

$$p = P_{\mu_0 \sigma^2}\left(T \geq \frac{\bar{x}_{oss} - \mu_0}{s_{x,oss}} \right)$$

- 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

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

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

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

with statistic T.

Example → Notebook 4

4.6.3 Two sample test

$\underline{X}$, $\underline{Y}$ two samples

Distinction → Paired obs. (1)

Distinction → Independent obs. (2)

(1) PAIRED

X_i = state of patient BEFORE Treatment

Y_i = state of patient AFTER Treatment.

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

X_i and Y_i may be related

(X_i, Y_i) are independent.

(2) INDEPENDENT

$(X_1, \dots, X_m) \overset{iid}{\sim} F_X$

$(Y_1, \dots, Y_n) \overset{iid}{\sim} F_Y$

Example 4.33 Paired observations

Effect of treatment $= Z_i := X_i - Y_i$

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

with
$$\Delta = EX_i - EY_i.$$

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 :

Reject if $\dfrac{|\bar{z}|}{S_z/\sqrt{n}} \geq t_{n-1, 1-\alpha_0/2}$.

The power depends strongly on $\sigma^2 = \mathrm{var}(Z_i) = \sigma_x^2 + \sigma_y^2 - 2\sigma_{xy}$

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

$(X_1, \dots X_m) \sim iid\ N(\mu, \sigma^2)$

$(Y_1, \dots, Y_n) \sim iid\ N(\nu, \sigma^2)$

same σ^2.

Test

$H_0 : \mu - \nu \leq 0$ vs $H_1 : \mu - \nu > 0$

- 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

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

wth $$S^2_{x,y} = \frac{1}{m+n-2} \cdot \left(\sum_{1}^{m} (X_i - \bar{X})^2 + \sum_{1}^{n} (Y_i - \bar{Y})^2 \right)$$

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

The MLE for σ^2 is

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

If $\mu = \nu$

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

$$T = \frac{\dfrac{(\bar{X} - \bar{Y})}{\sqrt{\left(\frac{1}{m} + \frac{1}{n}\right)\sigma^2}}}{\sqrt{\dfrac{(m+n-2)\, S_{x,y}^2 / \sigma^2}{m+n-2}}} \qquad \begin{matrix} N(0,1) \\ \\ \dfrac{\chi^2_{m+n-2}}{m+n-2} \end{matrix}$$

Note that

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

$$\chi^2_{m-1} \quad \perp\!\!\!\perp \quad \chi^2_{n-1}$$

$$\perp\!\!\!\perp \qquad\qquad \perp\!\!\!\perp$$

$$\bar{X} \qquad\qquad \bar{Y}$$

- 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. → 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

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

Thus this an asymptotic test.

4.6.4 Goodness of fit tests

Is a test devised 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

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

$= \text{proportion of obs.} \leq x$

→ Notebook 4

So $\mathbb{F}_n(x) \xrightarrow{P} E\, 1_{[X \leq x]} = F(x)$

by the LLN.

If H_0 is True $\mathbb{F}_n \approx F_0$ for sufficiently large n.

The KS statistic is

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

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.

$\rightarrow$ 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

$$\boxed{T^* = \sup_{x \in \mathbb{R}} \left| \mathbb{F}_n(x) - \Phi\left(\frac{x - \bar{X}}{S_X}\right) \right|}$$

It's distribution does not depend on μ, σ^2, and can be determined by simulation.

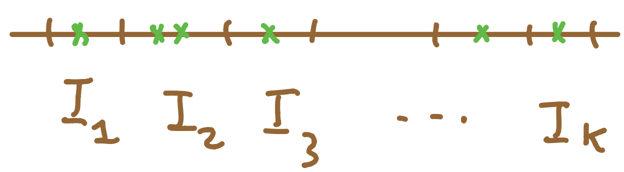
$\longrightarrow$ Notebook 4

Example 4.39 Chi-square test

Let $(X_1, \dots X_n) \sim$ 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

$$\underline{X}^2 = \sum_{j=1}^{k} \frac{(N_j - n p_j)^2}{n p_j}$$

Under $H_0 : F = F_0$

$$\underline{X}^2 \xrightarrow{L} \chi^2_{K-1} \quad \text{for } n \to \infty$$

K fixed

4.7 LIKELIHOOD RATIO TEST

a general method for finding tests.

Example : Is a coin fair ?

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

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

Likelihood

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

Likelihood ratio

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

for instance $n = 100$, $x = 60$

$\hat{p}_{ML} = 0.6$ $p_0 = .5$

$$\lambda(x) = \frac{\hat{L}}{L_0} = \frac{\binom{100}{60} \hat{p}^{60} (1-\hat{p})^{40}}{\binom{100}{60} p_0^{60} (1-p_0)^{40}}$$

$$= \frac{0.0812191}{0.0108439} = 7.49$$

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

$$\boxed{2 \log \lambda_n(X) \xrightarrow{L} \chi^2_1 \quad \text{for } n \to \infty}$$

→ See notebook 4

More generally we give the def:

Likelihood ratio statistic for $H_0 : \theta \in \Theta_0$

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

Idea: 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 $n \to \infty$

$$2 \log \lambda_n(\underline{X}) \xrightarrow{L} \chi^2_{k-k_0}$$

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_0
- 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 $X_i \overset{iid}{\sim}$ Poissom (θ)

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

$$\ell(\theta; \underline{x}) = \sum_{i=1}^{13} \left(x_i \log\theta - \theta - \log x_i!\right)$$

$$= \text{const.} + n\bar{x}\cdot\log\theta - n\theta.$$

lik equation

$$\ell'(\theta; \underline{x}) = \frac{n\bar{x}}{\theta} - n = 0$$

$$\hat{\theta} = \bar{x}.$$

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

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

Likelihood ratio.

$$2 \log \lambda = 2(\ell(\hat{\theta}) - \ell(\theta_0))$$

$$= 2(n\bar{x} \log \hat{\theta} - n\hat{\theta} - n\bar{x} \log \theta_0 + n\theta_0)$$

$$= 2n\left(\bar{x} \log \frac{\hat{\theta}}{\theta_0} - (\hat{\theta} - \theta_0)\right)$$

Gives $2 \log \lambda =$

$$= 26\left(5.538 \log\left(\frac{5.538}{5}\right) - (5.538 - 5)\right)$$

$$\simeq 0.728$$

P-value : $P(\chi^2_1 > 0.728) = 0.396.$

Not enough evidence.

<u>NOTE</u> the critical region is always one-sided even if the H_1 is two-sided.