

PH2282 part 3: Hypothesis test and p-values

Applied Multi-Messenger Astronomy 2:
Statistical and Machine Learning Methods in Particle and Astrophysics

Matteo Agostini

TUM - summer term 2019

Tasks of statistical inference

Task name	Task description	Some Frequentist tools	Some Bayesian tools
Point Estimation	what is the best estimate for a parameter of the model?	Maximum likelihood estimator	Median or mode of posterior distribution
Hypothesis Testing	which (model) hypothesis can be accepted or rejected given the data?	likelihood ratios	Bayes factors
Interval Estimation	which range of values is plausible for a given parameter of the model?	inverse hypothesis test	intervals of the posterior distribution
Goodness of Fit	are my data compatible with a model?	chi-square test, likelihood ratio test	posterior predictive p-values

Summary of key concepts

- A **random phenomenon** is a phenomenon whose outcome is unpredictable
- The **Frequentist probability** is the frequency of an outcome for a large (infinite) number of trials
- A **random variable** X is a variable whose possible values x are the outcomes of a random phenomenon. It is a variable in the sense that the frequency of its outcomes depends on the properties of the phenomenon (aka the parameters of a models). It is random in the sense that the outcome of the process is random, ergo unpredictable;
- The **probability function** of a random variable $X \sim f(x)$ is a function that associates a probability value for each possible outcome to occur. If the function is continuous, it is called probability density function (PDF) if it is discrete it is called probability mass function (PMF)
- A **statistical model** is defined by a set of parameters, a set of basic random variables, and their probability functions
- The parameters of a model can be grouped in:
 - **parameters of interest:** those on which we want to do statistical inference
 - **nuisance parameters:** those derived from the data and known with a limited accuracy
 - **fixed parameters:** universal constants or predictions from theory without uncertainties)

Summary of key concepts

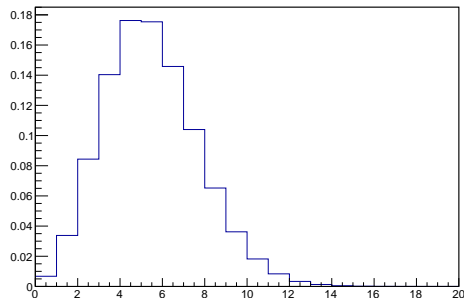
- The probability of a composite outcome (i.e. the simultaneous occurrence of outcomes of independent random processes) is the product of the probability of each outcome:
 $P(X, Y) = P(X) \cdot P(Y)$. The combined probability is called **joint probability**
- The **joint probability function of a model** is the composite probability of the probability functions of all random variables and represent the overall probability of an event given the model
- The **likelihood** is formally exactly the joint probability function but it is viewed as a function in which the outcome is fixed (defined by an experimental result) and the model parameters are a variable
- Before doing the experiment
 - what is the probability that I get a particular outcome?
 - the probability for an outcome is given by the probability function
- after running the experiment and having observed a data set
 - what is the model that gives my observed data with the largest probability?
 - the one that has a higher likelihood

Joint probability function vs likelihood

Consider a Poisson process and a counting experiment (random variable is N). The expected number of counts in the data set is $\lambda = 5$, the observed number of counts is 5:

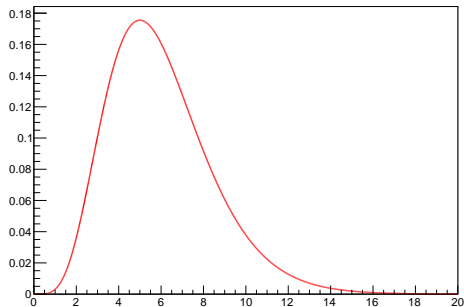
$$f_N(n; \lambda) = \mathcal{L}(\lambda; n) = \frac{e^{-\lambda} \lambda^n}{n!}$$

$f_N : n \rightarrow P(n)$ for $\lambda = 5$



what is the probability of an outcome given the model?

$\mathcal{L}(\lambda) : \lambda \rightarrow \mathcal{L}(\lambda)$ for $n = 5$



what is the model that gives my observed data with the largest probability?

Joint probability function vs likelihood

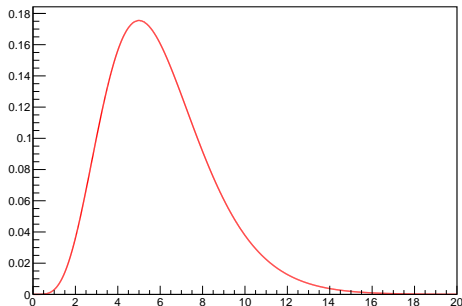
- Be careful to the pitfalls: the likelihood is not a probability function!
- A probability function is defined for a random variable, i.e. the outcome of a random process
- The parameters of a model are not random variables:
 - either the Higgs boson exists or it does not
 - if it exist its mass value is either one or an other
- That's why it's wrong to use the likelihood has the probability function of a model parameter
- Still, it is okay to ask what is probability of an outcome given a value of the model parameters.
This is what is used to perform statistical inference

Joint probability vs likelihood

If we compare the likelihood at two points and find that

$$P_{\vec{\theta}_1}(\vec{X} = \vec{x}) = L(\vec{\theta}_1; \vec{x}) > L(\vec{\theta}_2; \vec{x}) = P_{\vec{\theta}_2}(\vec{X} = \vec{x})$$

then the observed data are more likely to have occurred if $\vec{\theta} = \vec{\theta}_1$ than if $\vec{\theta} = \vec{\theta}_2$, which can be interpreted as saying that $\vec{\theta}_1$ is a more plausible value for the true value of $\vec{\theta}$ than $\vec{\theta}_2$.



Summary of our statistical model

Parameters of models:

λ_s and λ_b , i.e. the expectation for the total numbers of signal and background events

Elementary random variables:

N number of events (signal and background); X energy of an event

Probability density function for a single event energy:

$$X \sim f_X(x; \mu, \sigma, \lambda_s, \lambda_b) = \frac{1}{\lambda_s + \lambda_b} \left(\lambda_s \cdot f_X^s(x; \mu, \sigma) + \lambda_b \cdot f_X^b(x) \right) = \frac{1}{\lambda_s + \lambda_b} \left[\lambda_s \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} + \lambda_b \cdot \frac{1}{20} \right]$$

Probability mass function for the number of events: background events:

$$N \sim f_N(n; \lambda_s + \lambda_b) = \frac{e^{-(\lambda_s + \lambda_b)} (\lambda_s + \lambda_b)^n}{n!}$$

Likelihood function for our reference example

For a given data set in which $N = n$ and $\vec{X} = \vec{x}$:

$$\begin{aligned}\mathcal{L}(\lambda_s, \lambda_b; n, \vec{x}) &= f_{N, \vec{X}}(n, \vec{x}; \mu, \sigma, \lambda_s, \lambda_b) \\ &= \frac{e^{-(\lambda_s + \lambda_b)} (\lambda_s + \lambda_b)^n}{n!} \cdot \prod_{i=1}^N \left\{ \frac{1}{\lambda_s + \lambda_b} \left[\lambda_s \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}} + \lambda_b \cdot \frac{1}{20} \right] \right\}\end{aligned}$$

this is called an “extended unbinned” likelihood function

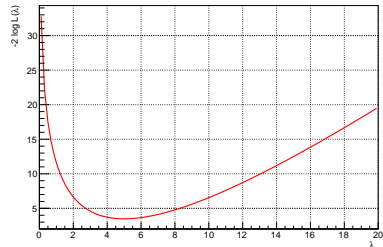
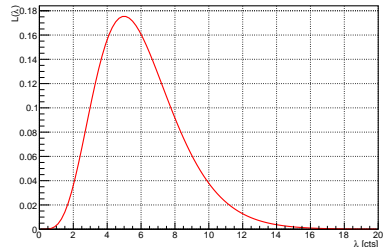
- it is unbinned in the sense that the energy of each events is used
- it is extended in the sense that the total number of events observed is not fixed and it is described by the Poisson term

Maximum likelihood Point Estimators

- Point Estimators are rules for calculating a single value which is to serve as a "best guess" or "best estimate" of an unknown parameter
- For a given outcome \vec{x} , the maximum likelihood estimator (MLE) of a parameter θ is the parameter value that maximizes $\mathcal{L}(\theta; \vec{x})$ considered as a function of θ with \vec{x} held fixed. We will denote the MLE with $\hat{\theta}(\vec{x})$
- the maximum can be found with computational methods. To this purpose, the problem of finding the maximum of the likelihood is converted into finding the minimum of the negative log-likelihood:

$$\hat{\theta} = \left\{ \max_{\theta \in \Theta} \mathcal{L}(\theta; x) \right\} = \left\{ \min_{\theta \in \Theta} -2 \log \mathcal{L}(\theta; x) \right\}$$

The logarithm transforms products in sums that reduce numerical approximations



Questions?

Hypothesis testing - intro

- new task of statistical inference!
- The goal of hypothesis testing is to decide, based on an observed outcome and a model, which of two complementary hypothesis is true.
- ingredients needed:
 - definition of hypothesis
 - a test o test statistic
 - an acceptance/rejection rule

What's an hypothesis?

- A hypothesis is a statement about a model or about a parameter of a model
- The two complementary hypotheses in a hypothesis testing problem are called:
 - null hypothesis H_0
 - alternative hypothesis H_1
- **complementary** hypotheses means that
 - either H_0 is true and H_1 is false
 - or H_0 is false and H_1 is true
- if θ denotes a parameter of a model, the general form of the null and alternative hypothesis is:
 - $H_0 : \theta \in \Theta_0$
 - $H_1 : \theta \in \Theta_0^c$

where Θ_0 is some subset of the parameter space allowed for θ and Θ_0^c is its complement

$$P(\theta \in \Theta_0) + P(\theta \in \Theta_0^c) = 1$$

Examples

Example 1: experimentation of a new drug

- θ denotes the average change in a patients blood pressure after taking an experimental drug
- question of interest:
 - $H_0 : \theta = 0$ (the treatment has no effect)
 - $H_1 : \theta \neq 0$ (the treatment has an effect)
- this common situation in which H_0 states that a treatment has no effect led to the term null hypothesis

Example 2: proportion of defective items

- question of interest:
 - $H_0 : \theta \geq \theta_0$
 - $H_1 : \theta < \theta_0$
- the value θ_0 is the maximum proportion of defective items acceptable by the producer

Hypothesis testing procedure

- A Hypothesis testing procedure is a rule that specifies:
 - for which set of outcomes the decision is made to accept H_0 as true
 - for which set of outcomes H_0 is rejected and H_1 is accepted as true
- The subset of the outcome space for which H_0 will be rejected is called **rejection region**. Its complement is called the **acceptance region**
- A hypothesis test is typically specified in terms of a test statistic

$$T(X_1, \dots, X_n) : \vec{X} \rightarrow \mathbb{R}$$

i.e. a function that maps an outcome into a scalar number

Hypothesis testing procedure

- A Hypothesis testing procedure is a rule that specifies:
 - for which set of outcomes the decision is made to accept H_0 as true
 - for which set of outcomes H_0 is rejected and H_1 is accepted as true
- The subset of the outcome space for which H_0 will be rejected is called **rejection region**. Its complement is called the **acceptance region**
- A hypothesis test is typically specified in terms of a test statistic

$$T(X_1, \dots, X_n) : \vec{X} \rightarrow \mathbb{R}$$

i.e. a function that maps an outcome into a scalar number

- Example 1: an example of rule to reject H_0 is when the mean change of blood pressure is above 10 mmHg (mean over K patients). Indicating with x_i the measured pressure change for the i -th patient, this rule can be formally expressed giving a test statistic:

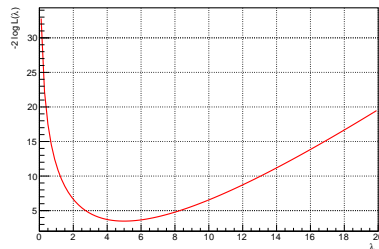
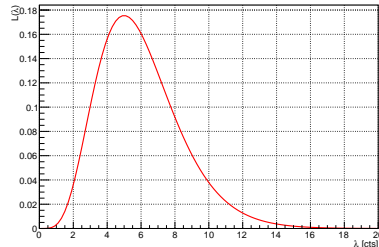
$$T(X_1, \dots, X_k) = \vec{X}_{\text{mean}} = \sum_{i=1}^K X_i / K$$

and a rejection region:

$$\{(x_1, \dots, x_k) : T(x_1, \dots, x_k) = \vec{X}_{\text{mean}} > 10 \text{ mmHg}\}$$

Likelihood Ratio Test Statistics

- there are different methods of finding test procedures and test statistics
- comparing the likelihoods given different model parameters seems pretty natural considering the relationship with the joint probability of the model
- let's consider a simple experiment that measures a number of counts N from a process that according to your theoretical colleague can have an expectation $\lambda = 1$ or $\lambda = 5$. The probability function is $N \sim \text{Poisson}(n; \lambda)$. After observing $n = 5$ we test:
 - $H_0 : \lambda = 1$
 - $H_1 : \lambda = 5$



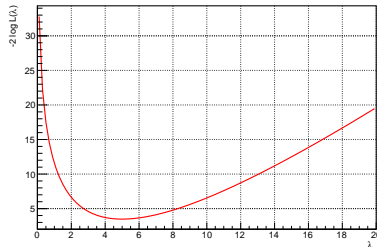
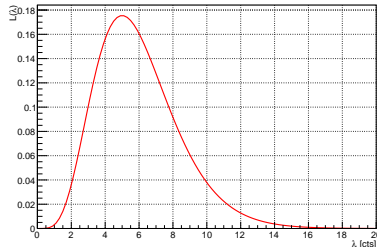
Likelihood Ratio Test Statistics

- there are different methods of finding test procedures and test statistics
- comparing the likelihoods given different model parameters seems pretty natural considering the relationship with the joint probability of the model
- let's consider a simple experiment that measures a number of counts N from a process that according to your theoretical colleague can have an expectation $\lambda = 1$ or $\lambda = 5$. The probability function is $N \sim \text{Poisson}(n; \lambda)$. After observing $n = 5$ we test:

- $H_0 : \lambda = 1$
- $H_1 : \lambda = 5$

$$T(n = 5) = \frac{\mathcal{L}(\lambda = 1, N = 5)}{\mathcal{L}(\lambda = 5, N = 5)}$$

$$t(n = 5) = -2 \log T = -2 \log \frac{\mathcal{L}(\lambda = 1, N = 5)}{\mathcal{L}(\lambda = 5, N = 5)} = -2 \log \mathcal{L}(\lambda = 1, N = 5) - (-2 \log \mathcal{L}(\lambda = 5, N = 5))$$



LRT: counting with known signal and background

Let's make the previous example a bit more realistic:

- The experiment measures a number of counts
 $N \sim \text{Poisson}(n; \lambda_s + \lambda_b)$
- λ_s is the signal expectation
- λ_b is the background expectation
- $\mathcal{L}(\lambda_s, \lambda_b; n) = e^{-(\lambda_s + \lambda_b)} (\lambda_s + \lambda_b)^n / n!$
- hypotheses (both signal and background expectations are fixed):
 - $H_0 : \lambda_s = 0, \lambda_b = 5$
 - $H_1 : \lambda_s = 3, \lambda_b = 5$
- $t(n) = -2 \log \mathcal{L}(\lambda_s = 0, \lambda_b = 5; n) - (-2 \log \mathcal{L}(\lambda_s = 3, \lambda_b = 5; n))$

LRT: counting with known signal and background

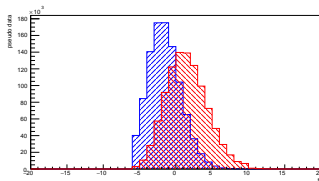
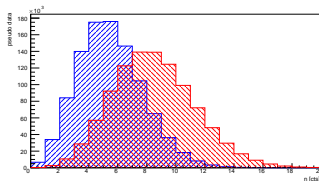
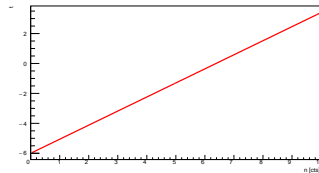
Let's make the previous example a bit more realistic:

- The experiment measures a number of counts
 $N \sim \text{Poisson}(n; \lambda_s + \lambda_b)$
- λ_s is the signal expectation
- λ_b is the background expectation
- $\mathcal{L}(\lambda_s, \lambda_b; n) = e^{-(\lambda_s + \lambda_b)} (\lambda_s + \lambda_b)^n / n!$
- hypotheses (both signal and background expectations are fixed):
 - $H_0 : \lambda_s = 0, \lambda_b = 5$
 - $H_1 : \lambda_s = 3, \lambda_b = 5$
- $t(n) = -2 \log \mathcal{L}(\lambda_s = 0, \lambda_b = 5; n) - (-2 \log \mathcal{L}(\lambda_s = 3, \lambda_b = 5; n))$
 $= -2(-5 + n \log(5) - \log(n!)) + 2(-8 + n \log(8) - \log(n!))$

LRT: counting with known signal and background

Let's make the previous example a bit more realistic:

- The experiment measures a number of counts
 $N \sim \text{Poisson}(n; \lambda_s + \lambda_b)$
- λ_s is the signal expectation
- λ_b is the background expectation
- $\mathcal{L}(\lambda_s, \lambda_b; n) = e^{-(\lambda_s + \lambda_b)} (\lambda_s + \lambda_b)^n / n!$
- hypotheses (both signal and background expectations are fixed):
 - $H_0 : \lambda_s = 0, \lambda_b = 5$
 - $H_1 : \lambda_s = 3, \lambda_b = 5$
- $t(n) = -2 \log \mathcal{L}(\lambda_s = 0, \lambda_b = 5; n) - (-2 \log \mathcal{L}(\lambda_s = 3, \lambda_b = 5; n))$
 $= -2(-5 + n \log(5) - \log(n!)) + 2(-8 + n \log(8) - \log(n!))$
 $= -2n(\log(5) - \log(8)) - 6$



LRT: counting with unknown signal and known background

- The experiment still measures a number of counts
 $N \sim \text{Poisson}(n; \lambda_s + \lambda_b)$ where λ_s is the signal expectation and λ_b is the background expectation
- $\mathcal{L}(\lambda_s, \lambda_b; n) = e^{-(\lambda_s + \lambda_b)} (\lambda_s + \lambda_b)^n / n!$
- hypotheses (no signal vs some signal):
 - $H_0 : \lambda_s = 0, \lambda_b = 5$
 - $H_1 : \lambda_s \geq 0, \lambda_b = 5$

LRT: counting with unknown signal and known background

- The experiment still measures a number of counts
 $N \sim \text{Poisson}(n; \lambda_s + \lambda_b)$ where λ_s is the signal expectation and λ_b is the background expectation
- $\mathcal{L}(\lambda_s, \lambda_b; n) = e^{-(\lambda_s + \lambda_b)} (\lambda_s + \lambda_b)^n / n!$
- hypotheses (no signal vs some signal):
 - $H_0 : \lambda_s = 0, \lambda_b = 5$
 - $H_1 : \lambda_s \geq 0, \lambda_b = 5$
- $t(n) = -2 \log \mathcal{L}(\lambda_s = 0, \lambda_b = 5; n) - (-2 \log \mathcal{L}(\lambda_s = \hat{\lambda}_s, \lambda_b = 5; n))$

LRT: counting with unknown signal and known background

- The experiment still measures a number of counts
 $N \sim \text{Poisson}(n; \lambda_s + \lambda_b)$ where λ_s is the signal expectation and λ_b is the background expectation
- $\mathcal{L}(\lambda_s, \lambda_b; n) = e^{-(\lambda_s + \lambda_b)} (\lambda_s + \lambda_b)^n / n!$
- hypotheses (no signal vs some signal):
 - $H_0 : \lambda_s = 0, \lambda_b = 5$
 - $H_1 : \lambda_s \geq 0, \lambda_b = 5$
- $t(n) = -2 \log \mathcal{L}(\lambda_s = 0, \lambda_b = 5; n) - (-2 \log \mathcal{L}(\lambda_s = \hat{\lambda}_s, \lambda_b = 5; n))$

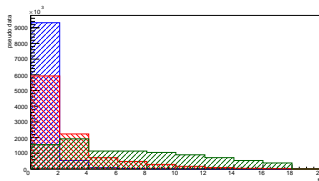
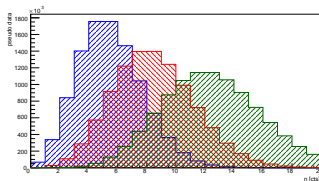
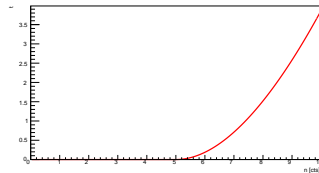
$$\hat{\lambda}_s = \begin{cases} 0 & \text{if } n \leq \lambda_b \\ n - \lambda_b & \text{if } n > \lambda_b \end{cases}$$

LRT: counting with unknown signal and known background

- The experiment still measures a number of counts
 $N \sim \text{Poisson}(n; \lambda_s + \lambda_b)$ where λ_s is the signal expectation and λ_b is the background expectation
- $\mathcal{L}(\lambda_s, \lambda_b; n) = e^{-(\lambda_s + \lambda_b)} (\lambda_s + \lambda_b)^n / n!$
- hypotheses (no signal vs some signal):
 - $H_0 : \lambda_s = 0, \lambda_b = 5$
 - $H_1 : \lambda_s \geq 0, \lambda_b = 5$
- $t(n) = -2 \log \mathcal{L}(\lambda_s = 0, \lambda_b = 5; n) - (-2 \log \mathcal{L}(\lambda_s = \hat{\lambda}_s, \lambda_b = 5; n))$

$$\hat{\lambda}_s = \begin{cases} 0 & \text{if } n \leq \lambda_b \\ n - \lambda_b & \text{if } n > \lambda_b \end{cases}$$

$$t(n) = \begin{cases} 0 & \text{if } n \leq \lambda_b \\ 2\lambda_b - 2n(\log \lambda_b - 1 + \log n) & \text{if } n > \lambda_b \end{cases}$$



LRT formal definition

In general the two hypotheses tested are expressed as $H_0 : \vec{\theta} \in \Theta_0$ and $H_1 : \vec{\theta} \in \Theta_0^c$ where $\vec{\theta} = \{\theta_1, \dots, \theta_k\}$ is a vector of model parameters, Θ_0 is some subset of the parameter space allowed for $\vec{\theta}$ and Θ_0^c is its complement. In this case the test statistic can be expressed as

$$t(x) = -2 \log \frac{\mathcal{L}(\hat{\theta}_0; x)}{\mathcal{L}(\hat{\theta}; x)}$$

where $\hat{\theta}$ is the MLE for $\vec{\theta} \in \Theta$ and $\hat{\theta}_0$ is the MLE for $\vec{\theta} \in \Theta_0$.

Considering that the MLE is the value of the parameter that maximized the likelihood in the allowed parameter space, the LRT can be in general expressed as

$$t(x) = -2 \log \frac{\sup_{\vec{\theta} \in \Theta_0} \mathcal{L}(\vec{\theta}; x)}{\sup_{\vec{\theta} \in \Theta} \mathcal{L}(\vec{\theta}; x)}$$

Examples

In our example with the unknown signal and known background:

- $\vec{\theta} = \{\lambda_s\}$
- $\Theta_0 : \{\lambda_s : \lambda_s = 0\}$
- $\Theta : \{\lambda_s : \lambda_s \geq 0\}$

$$\begin{aligned} t(n) &= -2 \log \mathcal{L}(\lambda_s = 0, \lambda_b = 5; n) - (-2 \log \mathcal{L}(\lambda_s = \hat{\lambda}_s, \lambda_b = 5; n)) \\ &= -2 \log \frac{\sup_{\lambda_s=0} \mathcal{L}(\lambda_s, \lambda_b = 5; n)}{\sup_{\lambda_s \geq 0} \mathcal{L}(\lambda_s, \lambda_b = 5; n)} \end{aligned}$$

This generalizes to the case in which there are more parameters. E.g. if the background expectation is considered as not known and becomes a nuisance parameters:

- $\vec{\theta} = \{\lambda_s, \lambda_b\}$
- $\Theta_0 : \{\lambda_s, \lambda_b : \lambda_s = 0, \lambda_b \geq 0\}$
- $\Theta : \{\lambda_s, \lambda_b : \lambda_s \geq 0, \lambda_b \geq 0\}$

$$t(n) = -2 \log \frac{\sup_{\lambda_s=0, \lambda_b \geq 0} \mathcal{L}(\lambda_s, \lambda_b; n)}{\sup_{\lambda_s \geq 0, \lambda_b \geq 0} \mathcal{L}(\lambda_s, \lambda_b; n)}$$

The definition of the test statistic extended to nuisance parameters is also called **profile likelihood ratio**

Questions?

How to define the acceptance/rejection regions for the test – 1

Given a LRT statistic:

$$t(x) = -2 \log \frac{\sup_{\vec{\theta} \in \Theta_0} \mathcal{L}(\vec{\theta}; x)}{\sup_{\vec{\theta} \in \Theta} \mathcal{L}(\vec{\theta}; x)}$$

the acceptance region is defined as $\{\vec{x} : t(\vec{x}) \leq c\}$ where c is any positive real number.

How to define the acceptance/rejection regions for the test – 1

Given a LRT statistic:

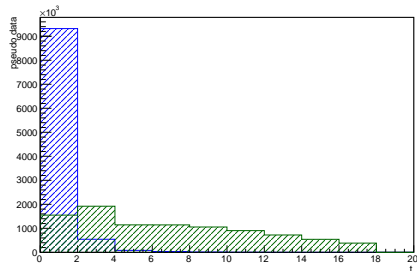
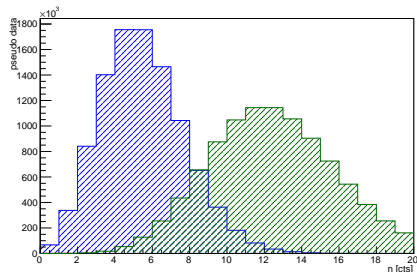
$$t(x) = -2 \log \frac{\sup_{\vec{\theta} \in \Theta_0} \mathcal{L}(\vec{\theta}; x)}{\sup_{\vec{\theta} \in \Theta} \mathcal{L}(\vec{\theta}; x)}$$

the acceptance region is defined as $\{\vec{x} : t(\vec{x}) \leq c\}$ where c is any positive real number.

But how to choose the value of c ? For any chosen value:

- the test will have some probability to reject H_0 even if it true (type I error)
- the test will have some probability to accept H_0 even if it is false (type II error)

Typically we fix the probability for a type I error. Test statistic based on likelihood ratios have the property to minimize the probability for type II errors in a large class of problems (Neyman-Pearson lemma)

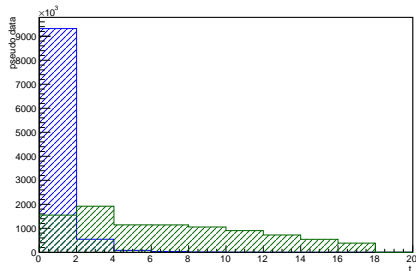
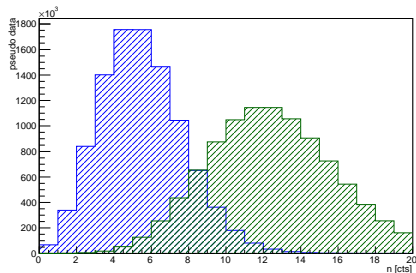


How to define the acceptance/rejection region for the test – 2

The probability of rejecting the null hypothesis when it is true is indicated with α and it is typically small e.g. 10% , 5%, 1%

Physicist use to call $1-\alpha$ the confidence level (C.L.) of a test. A test with type I error probability of 5% is called a 95% CL test.

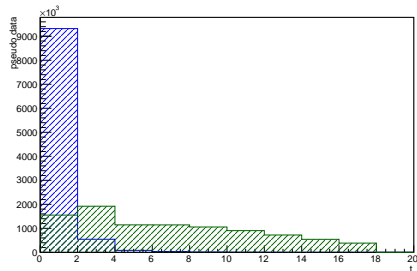
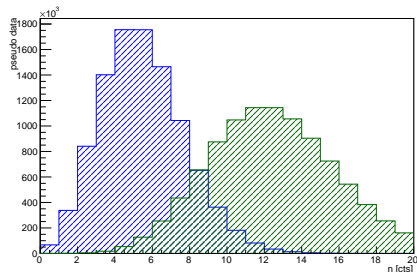
To map values of α in values of c , the probability distribution of the test statistic assuming the null hypothesis is needed. The probability distribution can be created by random sampling!



How to define the acceptance/rejection region for the test – 3

How to build the probability distribution of the statistic assuming the null hypothesis:

- 1) create a pseudo-data set by sampling a value for each random variables of the model according to the joint probability distribution for the null hypothesis parameters
- 2) compute and store the value of the test statistic for the pseudo-data set
- 3) repeat step 1 and 2 multiple times
- 4) plot a normalized histogram that will give the probability of observing a particular test statistic value given the model (the test statistic is actually a random variable by itself)

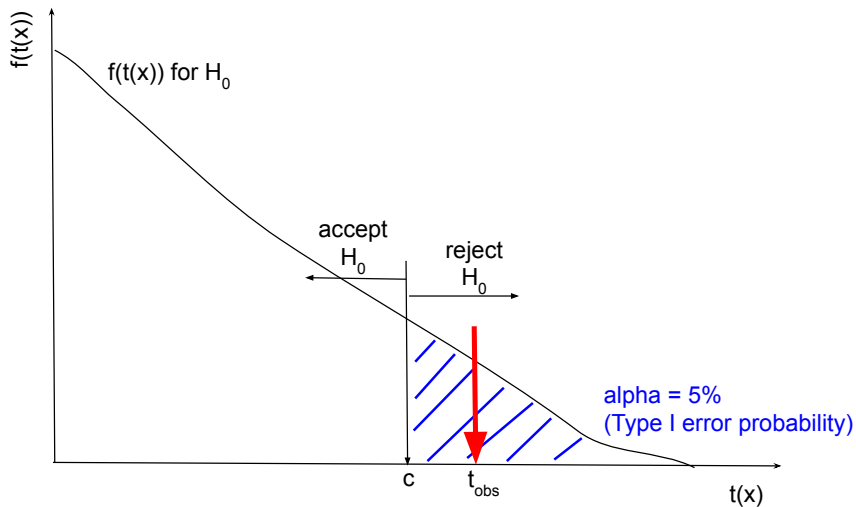


Summary of hypothesis testing

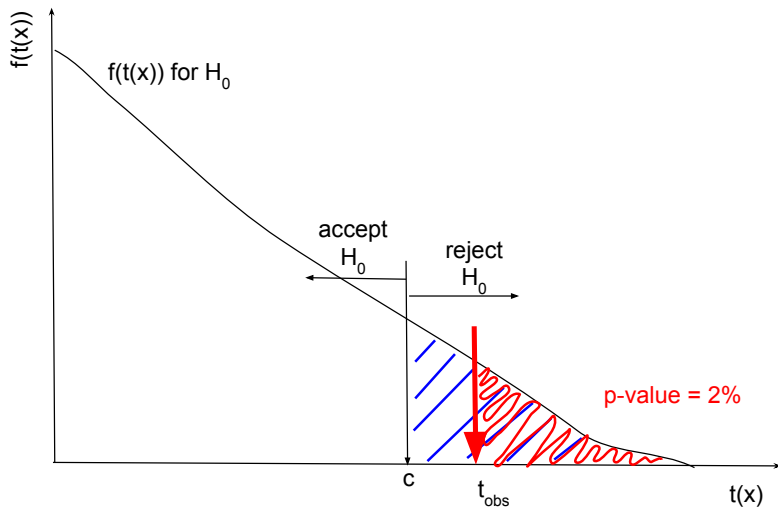
- 1) define the null and alternative hypothesis
- 2) once the two hypotheses are defined, a LRT statistic emerges naturally by maximizing the likelihood in the parameter space allowed by the two hypotheses. The maximization of the likelihood (i.e. minimize of the negative log likelihood) can be done using numerical techniques
- 3) construct the probability function of the test statistic assuming the null hypotheses by generating pseudo data and computing the test statistic value for each of them
- 4) define the C.L. of the test and calculate the rejection region
- 5) take the data from your experiment (i.e. the **observed** data), compute the test statistic on them and check whether its value is in the rejection or acceptance region

Questions?

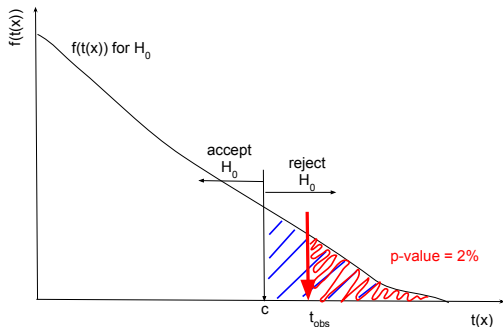
p-values



p-values



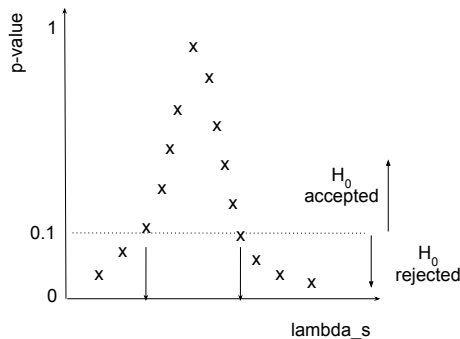
p-values



- A hypothesis test leads to either reject or accept the null hypothesis
- but the data can be more or less compatible with the null hypothesis
- while it's good practice to decide the value of α before running the experiment to avoid biases, the outcome of an experiment can be reported through its p-value
- the p-value gives the probability of an outcome more incompatible with the null hypothesis than the observed one. The p-value goes between 0 and 1, small values suggest incompatibility with the null hypothesis
- we can construct the hypothesis testing based on the p-value: if the p-value is smaller than α , the null hypothesis is rejected

From one hypothesis test to a set of tests

- Up to now we have been considering a test in which the null hypothesis was the background-only hypothesis. This is not necessarily true
- a set of tests is frequently done by varying the definition of the null hypothesis over a discrete sequence of tested values $\lambda_s = \{0, \lambda_1, \lambda_2, \dots\}$:
 - test 1: $H_0 : \lambda_s = 0$ and $H_1 : \lambda_s \geq 0$
 - test 2: $H_0 : \lambda_s = \lambda_1$ and $H_1 : \lambda_s \geq 0$
 - test 3: $H_0 : \lambda_s = \lambda_2$ and $H_1 : \lambda_s \geq 0$
 - ...
- a set of tests can be used to define intervals of parameter values accepted by the test given a certain confidence level.



This is the key concept behind the next task of the statistic inference: interval estimation. The interval constructed using a test with a given confidence level is called a confidence interval

Questions?

Exercise 3 - part 1

In Exercise 2 we created an infrastructure to generate pseudo-data given some parameters of the model and to compute the MLE for λ_s . The generation and MLE calculation was performed multiple times to plot the MLE distribution. The first part of Exercise 3 is:

- 1) consider a model in which the pseudo-data sets are generated for $(\lambda_s)_{injected} = 0$ and $(\lambda_b)_{injected} = 100$
- 2) instead of computing and plotting the MLE estimator compute and plot the test statistic distribution for a test in which:
 - $H_0 : \lambda_s = 0, \lambda_b \geq 0$
 - $H_1 : \lambda_s \geq 0, \lambda_b \geq 0$
 - $t(n, \vec{x}) = -2 \log \frac{\sup_{\lambda_s=0, \lambda_b \geq 0} \mathcal{L}(\lambda_s, \lambda_b; n, \vec{x})}{\sup_{\lambda_s \geq 0, \lambda_b \geq 0} \mathcal{L}(\lambda_s, \lambda_b; n, \vec{x})}$ (this is a profile likelihood!)
- 3) repeat steps 1 and 2 for 1000 times
- 4) compute the test statistic for a particular set of pseudo-data that are treated as the “observed data” of an experiment.
- 5) Calculate the p-value for the observed data. Check how the p-value changes by changing the parameters used to generate the observed data

Exercise 3 - part 2

- 1) Repeat the steps 1-2-3 of part 1 for different values of $(\lambda_s)_{\text{injected}} = \{1, 2, \dots, 20\}$. The hypothesis test will now be (e.g. for the second test)
 - $H_0 : \lambda_s = 1, \lambda_b \geq 0$
 - $H_1 : \lambda_s \geq 0, \lambda_b \geq 0$
 - $t(n, \vec{x}) = -2 \log \frac{\sup_{\lambda_s=1, \lambda_b \geq 0} \mathcal{L}(\lambda_s, \lambda_b; n, \vec{x})}{\sup_{\lambda_s \geq 0, \lambda_b \geq 0} \mathcal{L}(\lambda_s, \lambda_b; n, \vec{x})}$ (this is a profile likelihood!)
- 2) compute all test statistics for a particular set of pseudo-data that are treated as the “observed data” of an experiment.
- 3) Calculate the p-value for the observed data for all tests. Plot the p-values as a function of the tested value of λ_s