Hypothesis testing

Lectures organization

In order to ease the learning process, some portions of this handout will be completed during the lectures. The complete version of this handout will be posted online after each lecture.
The intended plan for our time together is the following:

- **Lecture 1:** Monday, May 16, 2022, 4PM-5.30PM
  Statistical inference (overview), likelihood ratio test.

- **Lecture 2:** Tuesday, May 17, 2022, 11AM-12.30PM
  Goodness-of-fit.

- **Exercises:** Tuesday, May 17, 2022, 2PM-3.30PM
  To be held jointly with the exercises for Glen and Harrison's lectures.

1 A brief overview on statistical inference

Statistical inference is the area of statistics which aims to develop and study reliable tools to make conclusion on the phenomenon/population under study based on what has been observed on a data sample. Among the main inferential tools we have *tests of hypotheses* or *goodness-of-fit tests*. Confidence intervals and upper limits are also widely used but more often than not, they can be obtained by simply inverting tests of hypotheses. Hence, we focus on

- **Tests of hypotheses:** Given a postulated model for the data (e.g., background only), we test it against an alternative model (e.g., background+ signal we are looking for).

- **Goodness-of-fit tests:** Given a postulated model for the data we test it against all possible alternatives (e.g., do we have only background events or is there something else?).

Which type of test to use typically depends on the problem under study but, in both cases, the classical formulation of a statistical test entails the following steps:

1. **Specification of the null and the alternative hypotheses, namely H_0 and H_1**

   - Tests of hypotheses: e.g., we expect that $X \sim N(\mu, 1)$, we test

   $$H_0 : \mu = 0 \quad \text{versus} \quad H_1 : \mu \neq 0.$$

   - Goodness-of-fit: e.g., we expect that $X \sim N(\mu, 1)$, we test

   $$H_0 : X \sim N(\mu, 1) \quad (\text{versus} \quad H_1 : X \nsim N(\mu, 1)).$$

   we call the model specified under H_0 the "null model" and the model specified under H_1 the "alternative model".

2. **Specification of a test statistic**

   - Tests of hypotheses: e.g., the Likelihood Ratio Test (LRT) statistic.

   - Goodness-of-fit: e.g., Pearson χ^2 test statistic.

   (We will see both of them later in this handout.)

3. **Derivation of the distribution of the test statistic under H_0**
   This is typically done by relying on asymptotic results or via Monte Carlo or parametric bootstrap simulations. *Do you know what is the difference between the two?*

   - Monte Carlo:

   - Parametric bootstrap:

4. **Computation of p-values or quantiles of the null distribution.**

2 How do we assess if a test is good or not?

In principle we could construct an infinite number of different test statistics for any model under study and perform a test of hypothesis by simulating their null distribution. However, not all tests are good tests and indeed, we can quantify how "good" a test is by looking at two main quantities the *probability of type I error* and the *power*.

- **The probability of type I error** is the probability of incorrectly rejecting H_0 when it is true. It is typically denoted by α and is also called "significance level", i.e.,

  $$P(\text{Probability of Type I error}) = \alpha = 1 - \text{coverage probability}.$$

  In physics, it is typically expressed in "number of sigmas", i.e., $\#\sigma = \Phi(1-\alpha/2)$, with $\Phi(\cdot)$ being the cumulative density function of a standard normal.

- **The power** is the probability of correctly rejecting H_0 when H_1 is true.

In order to provide an explicit visualization of what these two are, suppose we are testing two hypothesis H_0 and H_1 using a test statistic T which we know is asymptotically distributed as a χ^2_3. Suppose we have collected a sample of size $N = 500,000$ and so we are pretty sure we can trust the asymptotics. Call T_{obs} the value of the test statistic T calculated on such sample...

Clearly, as α increases, the power increases. Moreover, we say that a test is *admissible* or *unbiased*, if its power is higher than its probability of type I error.

Finally, power and probability of type I error allow us to tell a bit more about the differences between goodness-of-fit tests and tests of hypotheses. Specifically, the former will have some power against all possible alternative models. Whereas, the latter will have high power only against the alternative model we have specified under H_1. This is the main distinction among these two classes of tests.

3 Testing hypotheses using the Likelihood Ratio Test

Given two plausible models for the data, an hypothesis testing procedure aims to select the one which is more consistent with the data collected. But what does “more consistent” mean exactly? In statistics, we often refer to the *likelihood function* to quantify how plausible a model is for the data being collected.

3.1 The likelihood function and maximum likelihood estimation

Let $F(y;\boldsymbol{\theta})$ be the postulated cumulative distribution function (cdf) for the sample $\boldsymbol{y} = (y_1,\dots,y_N)$, and denote with $f(y;\boldsymbol{\theta})$ the respective probability density function (pdf). The vector $\boldsymbol{\theta}$ collects the p parameters which characterize the distribution F. For instance, if our sample was generated from a Normal distribution with mean μ and variance σ^2 we would have $\boldsymbol{\theta} = (\mu,\sigma^2)$, $p = 2$, and

$$f(y;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}(y-\mu)^2\right\}$$

whereas, $F(y;\mu,\sigma^2) = \Phi(y;\mu,\sigma^2) = \int_{-\infty}^{y} f(t;\mu,\sigma^2)\mathrm{d}t$. The (continuous/unbinned) likelihood function specifies as:

$$L(\boldsymbol{\theta};\boldsymbol{y}) = \prod_{i=1}^{N} f(y_i;\mu,\sigma^2).$$

We will see in Section 4 how the likelihood specifies if the data are binned. The likelihood function is particularly useful for the purpose of estimating the parameters in $\boldsymbol{\theta}$. Specifically, if the parameters in $\boldsymbol{\theta}$ are unknown, we can estimate them by maximizing the likelihood function, and obtain the so-called *Maximum Likelihood Estimate* (MLE), i.e.:

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} l(\boldsymbol{\theta};\boldsymbol{y})$$

where $l(\boldsymbol{\theta};\boldsymbol{y}) = \log\{L(\boldsymbol{\theta};\boldsymbol{y})\}$ is the *log-likelihood* function, and we rely on it just because it is typically much easier to maximize $l(\boldsymbol{\theta};y)$ rather than $L(\boldsymbol{\theta};\boldsymbol{y})$.

Finally, under suitable regularity conditions we have that

$$\widehat{\boldsymbol{\theta}} \approx N(\boldsymbol{\theta}, I^{-1}(\boldsymbol{\theta})), \quad \text{as } n\to\infty, \tag{1}$$

where $I^{-1}(\boldsymbol{\theta})$ is the inverse of the so-called *Fisher's Information Matrix*, i.e., $I(\boldsymbol{\theta}) = -E\left[\frac{\partial^2}{\partial^2\boldsymbol{\theta}} l(\boldsymbol{\theta};\boldsymbol{y})\right]$.

3.2 The Likelihood Ratio Test

The likelihood ratio test consists in comparing the null and the alternative models specified under H_0 and H_1 in terms of their log-likelihood.

- Specification of the null and alternative hypotheses.
  Typically, we consider tests such as:

$$H_0:\boldsymbol{\theta}=\boldsymbol{\theta}_0 \quad \text{(simple hypothesis)} \quad \text{versus} \quad H_1:\boldsymbol{\theta}>\boldsymbol{\theta}_0 \quad \text{(one sided hypothesis)}$$
$$H_0:\boldsymbol{\theta}=\boldsymbol{\theta}_0 \quad \text{(simple hypothesis)} \quad \text{versus} \quad H_1:\boldsymbol{\theta}\neq\boldsymbol{\theta}_0 \quad \text{(two sided hypothesis)}$$

- Specification of the test statistic:

$$\Lambda(\boldsymbol{\theta}_0) = -2\log\frac{L(\boldsymbol{\theta}_0;\boldsymbol{y})}{L(\widehat{\boldsymbol{\theta}};\boldsymbol{y})} = -2\sum_{i=1}^{N}[l(y_i;\boldsymbol{\theta}_0) - l(y_i;\widehat{\boldsymbol{\theta}})]$$

- Derivation of the distribution of the test statistic under H_0.
  Under suitable regularity conditions, as $N \to \infty$, under H_0,
  $$\Lambda(\boldsymbol{\theta}_0) \approx \chi^2_p.$$
  Alternatively, one can simulate the distribution of Λ under H_0 via Monte Carlo.

- Computation of the p-value.
  Let Λ_{obs} be the value of the test statistic Λ evaluated on the data,
  $$\text{p-value} = P(\chi^2_p \geq \Lambda_{obs}).$$

- What do we do after we have computed our p-value?
  - If p-value$\geq \alpha$: we "fail to reject" $H_0 \Rightarrow$ our null model fits the data well.
  - If p-value$< \alpha$: we reject $H_0 \Rightarrow$ our null model does not fit the data well, the alternative model is better.

  This is a legitimate decision rule because, if you think about, the p-value is nothing but the probability of obtaining a value of our test statistic Λ which is more extreme than the value we have observed on the data, Λ_{obs}, when H_0 is true. So if this event is very unlikely (small p-value), that means that probably, the true distribution of Λ is different from the one we expect under H_0, and consequently, it is unlikely that H_0 is valid.

3.3 The Likelihood Ratio Test using the profile likelihood

We may encounter situations where, in addition to the parameter $\boldsymbol{\theta}$ which we are interested in, we may also have another set of parameters, namely $\boldsymbol{\tau}$, which are of no interest to us and play the role of nuisance parameters. The latter may correspond, for instance, to systematic uncertainties. The question then is: how do we deal with those?
Interestingly, we can proceed exactly as we did before but this time, instead of the likelihood functions under H_0 and H_1 we consider the respective *profile likelihood* functions. The latter consist of replacing $\boldsymbol{\tau}$ with its MLE obtained by fixing $\boldsymbol{\theta}$ to its value specified under the hypotheses,

$$\Lambda_{\boldsymbol{\tau}}(\boldsymbol{\theta}_0) = -2\log\frac{L(\boldsymbol{\theta}_0; \widehat{\boldsymbol{\tau}}_{\boldsymbol{\theta}_0}, y)}{L(\widehat{\boldsymbol{\theta}}; \widehat{\boldsymbol{\tau}}_{\widehat{\boldsymbol{\theta}}}, y)} = -2\sum_{i=1}^{N}\left[l(y_i; \widehat{\boldsymbol{\tau}}_{\boldsymbol{\theta}_0}, y) - l(y;\widehat{\boldsymbol{\tau}}_{\widehat{\boldsymbol{\theta}}}, y)\right].$$

The asymptotic distribution of Λ under H_0 remains unchanged, i.e., we still have $\Lambda \approx \chi^2_p$ (provided that the regularity conditions needed hold). Alternatively, one can proceed by simulating the null distribution of $\Lambda_{\boldsymbol{\tau}}(\boldsymbol{\theta}_0)$ under H_0, that is, when $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ and $\boldsymbol{\tau} = \widehat{\boldsymbol{\tau}}_{\boldsymbol{\theta}_0}$ via the *parametric bootstrap.*

3.4 Some warnings

In the previous sections, when deriving the asymptotic distribution of $\widehat{\boldsymbol{\theta}}$ and $\Lambda(\boldsymbol{\theta}_0)$ we have said that, those results are valid "under some regularity conditions". Despite the latter are quite technical, we can identify a few conditions necessary for the validity of our χ^2 approximation. Specifically,

C1. The true value of the parameters must not be on the boundaries of their parameter space.

C2. The parameters are identifiable. That is, different values of the parameters specify distinct models.

C3. The model under H_0 is a special case of the model under H_1.

C4. The true model is one between those specified under H_0 or under H_1.

3.5 Discussion: Identifying situations of non-regularity

Suppose X is the number of events detected by our instrument (e.g., photon emissions over a given range of energy) and let's assume that the resolution of the latter is sufficiently high to allow us to treat the data as a continuous stream of data (no binning). The probability density function of X specifies as follows:

$$f(x;\eta,\phi,\theta) = (1-\eta)b(x,\phi) + \eta s(x,\theta), \quad \eta \in [0,1) \tag{2}$$

where $b(x,\phi)$ is the density of the background and ϕ is a parameter characterizing it. For instance, if b is a power-law, ϕ is its slope. Whereas, $s(x,\theta)$ is the density of the signal and θ the parameter characterizing it. For instance, if we may consider s to be the pdf of a Gaussian with variance 1 and mean θ. Finally, η is the relative intensity of the signal, i.e., the proportion of events that we expect to come from the signal and it is assumed to be non-negative. Finally, you are given a sample of $N = 10,000$ observations which you may use to assess the validity of the model in (2). A few questions for you...

1. **How would you specify your hypotheses in order to assess if a signal is present or not?**

2. **Suppose θ is known to be 125. Which among the necessary conditions needed for the χ^2-approximation of the LRT fail (see Section 3.4)? Why?**

3. **Suppose θ is unknown. Which among the necessary conditions needed for the χ^2-approximation of the LRT fail (see Section 3.4)? Why?**

4. **Suppose you decided to test $H_0 : b(x, \phi)$ versus $H_1 : s(x, \theta)$. Which among the necessary conditions needed for the χ^2-approximation of the LRT fail?**

You can find a self-contained review on solutions for the problems above in

> Algeri et al., 2020. *Searching for new phenomena with profile likelihood ratio tests.* Nature Reviews Physics. `https://www.nature.com/articles/s42254-020-0169-5`.

You may watch a seminar on this at `https://www.youtube.com/watch?v=eOY9QszyO1E`.

4 Classical goodness-of-fit tests

4.1 The binned data regime

Binned data may originate from the fact that the resolution of the detector is not sufficiently high to produce a continuous stream of data. Alternatively, researchers may decide to artificially bin the data. This can be done, for instance, for the sake of simplifying the estimation process or introducing Poisson uncertainties.

Suppose that our search region $\mathcal{X}$ is the energy interval $1-50$GeV and it has been divided into $i=1,\dots,n$ bins. Denote with x_i the value of the energy at the center of bin i. Moreover, let Y_i be the number of events in bin i. We typically assume that

$$Y_i \sim \text{Multinomial}\big[N, p(x_i,\boldsymbol{\theta})\big] \qquad \text{or} \qquad Y_i \sim \text{Poisson}\big[m(x_i,\boldsymbol{\theta})\big]$$

In the first case, the total number of events observed N, is treated as fixed, and $p(x_i,\boldsymbol{\theta})$ is the probability to observe an event in bin i, whereas, the expected number of events in bin i is $m(x_i,\boldsymbol{\theta}) = Np(x_i,\boldsymbol{\theta})$. In the second case, the total number of events observed is itself random and $m(x_i,\boldsymbol{\theta})$ is tells us how many event we expect to observe in bin i. Finally, $\boldsymbol{\theta}$ is a set of p (potentially unknown) parameters characterizing the shape of our mean function $m(x_i,\boldsymbol{\theta})$. Our goal is to assess if the our model $m(x_i,\boldsymbol{\theta})$ for the number of expected events in each of the bins is valid. Notice that, if $\boldsymbol{\theta}$ is unknown, we may proceed by estimating it via maximum likelihood (see Section 3.1). In this setting, however, we must consider the binned data likelihood, i.e.,

$$L(\boldsymbol{\theta};\boldsymbol{y}) = \frac{N!}{\prod_{i=1}^{n} y_i!}\prod_{i=1}^{n}\Big[p(x_i,\boldsymbol{\theta})\Big]^{y_i} \quad \text{for } Y_i \sim\text{Multinomial.}$$

$$L(\boldsymbol{\theta};\boldsymbol{y}) = \prod_{i=1}^{n}\frac{\big[m(x_i,\boldsymbol{\theta})\big]^{y_i}}{y_i!}\exp\Big\{m(x_i,\boldsymbol{\theta})\Big\} \quad \text{for } Y_i \sim\text{Poisson,}$$

4.1.1 Pearson X^2 test

- Specification of the null hypothesis.

$$H_0 : E[Y_i] = m(x_i,\boldsymbol{\theta}) \qquad\qquad (\text{versus } H_1 : E[Y_i] \neq m(x_i,\boldsymbol{\theta}))$$

  for all $i=1,\dots,n$.

- Specification of the test statistic.

$$X^2 = \sum_{i=1}^{n}\frac{\big[Y_i - m(x_i,\boldsymbol{\theta})\big]^2}{m(x_i,\boldsymbol{\theta})}$$

- Derivation of the distribution of the test statistic under H_0.

  - If , $\boldsymbol{\theta}$ known:
    If $m(x_i, \boldsymbol{\theta}) > 7$ for all $i = 1, \ldots, n$, as $N \to \infty$, under H_0,

    $$X^2 \approx \chi^2_{n-1}$$

  - If , $\boldsymbol{\theta}$ unknown:
    If $m(x_i, \boldsymbol{\theta}) > 7$ for all $i = 1, \ldots, n$, as $N \to \infty$, under H_0,

    $$X^2 \approx \chi^2_{n-p-1}$$

- Computation of the p-value.
  Let X^2_{obs} be the value of the test statistic X^2 evaluated on the data,

  - If , $\boldsymbol{\theta}$ known: p-value=$P(\chi^2_{n-1} \geq X^2_{obs})$

  - If , $\boldsymbol{\theta}$ unknown: p-value=$P(\chi^2_{n-p-1} \geq X^2_{obs})$

- What do we do after we have computed our p-value?
  We check if it is larger or smaller than α.

Warning: If the number of events in the bins is small, X^2 may lead to biased (non-admissible) tests. That is true even if the χ^2 approximation hold!

4.1.2 G^2 test (also known as deviance test or likelihood ratio)

- Specification of the null hypothesis: same as Pearson X^2.

- Specification of the test statistic.

$$G^2 = 2\sum_{i=1}^{n}\left[Y_i \log\frac{Y_i}{m(x_i, \boldsymbol{\theta})} - (Y_i - m(x_i, \boldsymbol{\theta}))\right]$$

- Derivation of the distribution of the test statistic under H_0: same as Pearson X^2.

- Computation of the p-value: same as Pearson X^2.

Which one should be used?
In general, the χ^2 approximation is typically better for Pearson than it is for G^2, especially in the Poisson case. The χ^2 approximation may be better for G^2 if the counts are small. The literature on this is vast but a self-contained review can be found in

> Cressie and Read, 1989. *Pearson's X^2 and the Loglikelihood Ratio Statistic G^2: A Comparative Review.* International Statistical Review. `https://www.jstor.org/stable/1403582?seq=1#metadata_info_tab_contents`

4.2 The unbinned data regime

4.2.1 The univariate case

Let X be continuous random variable taking values over the real line and distributed according to the distribution function $F(x) = P(X \leq x)$, i.e., $X \sim F$. When F is unknown, we are typically interested in testing if, for a given distribution function G,

$$H_0 : F = G \qquad \text{versus} \qquad H_1 : F \neq G \tag{3}$$

To perform this test, given a sample of N i.i.d. random variables from F, it is sensible to work with the process

$$v_G(x) = \sqrt{N}[F_N(x) - G(x)] \tag{4}$$

where $F_N(x)$ is the so-called *empirical distribution function*, it provides an estimate of $F(x)$ and it is defined as

The stochastic process $v_G(x)$ is called *empirical process*. Let's focus for a moment on understanding our $v_G(x)$...

- As $N \to \infty$, we have that $F_N(x) \to F(x)$.
  So, what do you think will happen to $v_G(x)$ when $N \to \infty$ (under H_0 and H_1, respectively)?
  - Under H_0:
  - Under H_1:

When dealing with univariate distributions, we typically consider the so-called *Probability Integral Transform* (PIT), that is, we set $T = G(X)$.

- *What is the distribution of* $T = G(X)$ *under* H_0*?*

When applying such transformation, the empirical process $v_G(x)$ is transformed in the so called *uniform empirical process*, i.e.,

$$u(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} [\mathbb{1}_{\{t_i \leq t\}} - t]$$

How can we use $u(t)$ to test $H_0 : F = G$ versus $H_1 : F \neq G$? We simply take functionals of it, e.g.,

- **Kolmogorov statistic:** $\sup_t |u(t)|$
- **Cramer von Mises statistic:** $\int |u(t)|^2 dt$
- **Anderson-Darling statistic:** $\int \left| \frac{u(t)}{\sqrt{t(1-t)}} \right|^2 dt$

These are just some possibilities among an entire family of test statistics and they are all distribution free; that is, their distribution does not depend on the G that I am testing. Distribution-freeness is an important property of a test statistic as it allows us to avoid a case-by-case simulation/mathematical derivation.

- **Why is it useful to have an entire family of test statistics?**

5 The multivariate case

Now suppose that X takes value in $\mathbb{R}^D$ and suppose our G depends on a set of unknown parameters $\boldsymbol{\theta}$. The (multivariate parametric) empirical process

$$v_G(\boldsymbol{x},\boldsymbol{\theta}) = \sqrt{N}[F_N(\boldsymbol{x}) - G(\boldsymbol{x},\boldsymbol{\theta})], \quad \boldsymbol{x} = (x_1,\dots,x_D) \tag{5}$$

can still be used to test $H_0 : F = G$ versus $H_1 : F \neq G$. In this case, however we can no longer exploit the PIT so we lose distribution-freeness. That is because, in this case,

$$T = G(\boldsymbol{x},\widehat{\boldsymbol{\theta}}) \not\sim \text{Unif[0,1]}$$

where $\widehat{\boldsymbol{\theta}}$ is an estimator (e.g., the MLE) of $\boldsymbol{\theta}$. Nonetheless, we can still construct an entire family of test statistics which extend those we have seen for the univariate case, e.g.,

- **Kolmogorov's statistic:** $\sup_x |v_G(\boldsymbol{x},\widehat{\boldsymbol{\theta}})|$
- **Cramer von Mises statistic:** $\int |v_G(\boldsymbol{x},\widehat{\boldsymbol{\theta}})|^2 dG(\boldsymbol{x},\widehat{\boldsymbol{\theta}})$
- **Anderson-Darling statistic:** $\int \left|\frac{v_G(\boldsymbol{x},\widehat{\boldsymbol{\theta}})}{\sqrt{G(\boldsymbol{x})\left[1-G(\boldsymbol{x},\widehat{\boldsymbol{\theta}})\right]}}\right|^2 dG(\boldsymbol{x},\widehat{\boldsymbol{\theta}})$

and we can simulate their distribution via the parametric bootstrap.

There are ways to recover distribution-free and even make the simulation faster but these would required an entire new lecture. A suitable reference is

> Algeri, 2022. *K-2 rotated goodness-of-fit for multivariate data.* Physical Review D. `https://journals.aps.org/prd/abstract/10.1103/PhysRevD.105.035030`.

You may watch a seminar on this at `https://indico.cern.ch/event/1130770/` (starting from minute 46:00 of the recording).

6 Exercises

6.1 Testing hypothesis via the Likelihood Ratio Test

Suppose our instrument has collected $N = 1000$ observations which appear to suggest that an excess of events at 125 GeV is present. For simplicity, we assume that the background was completely resolved and the goal is to understand what is the exact location of these signal events. The mass distribution is assumed to be a Gaussian with mean μ and variance σ^2 (both unknown). We want to test the hypotheses

$$H_0 : \mu = 125 \qquad \text{versus} \qquad H_1 : \mu \neq 125$$

For this scenario, specify the following.

1. **The (profile) log-likelihood function**.

2. **A formula for** $\Lambda(125)$.

3. **The asymptotic distribution of $\Lambda(125)$, when H_0 is true.**

4. **Suppose $\Lambda_{obs} = 2$, specify a formula for the p-value.**

5. **The p-value is 0.157, can we claim that $\mu = 125$GeV is likely to be the correct location of the signal?**

6.2 Performing Goodness-of-Fit via Pearson and Likelihood Ratio

Suppose we are given a sample of events over the energy range 1-50GeV, and the goal is to assess if a power law distribution is a reliable model for the expected number of counts in each bin. Moreover, we assume that the data is Poisson, i.e.,

$$Y_i \sim \text{Poisson}\left(\int_{x_i-\frac{\Delta}{2}}^{x_i+\frac{\Delta}{2}} \frac{\tau}{x^\theta}\mathrm{d}x\right).$$

For the moment, we assume that $\theta = 0.5$, $\tau = 500$ and that the data has been divided into $n = 10$ equally spaced bins, each of length $\Delta = 5$. The expected number of events in bin i is $m(x_i, \theta, \tau) = \int_{x_i-\frac{\Delta}{2}}^{x_i+\frac{\Delta}{2}} \frac{\tau}{x^\theta}\mathrm{d}x$.

Bin i	x_i	# of expected events	# of observed events
$[1, 5)$	2.5	1236.1	1249
$[5, 10)$	7.5	926.2	926
$[10, 15)$	12.5	710.7	712
$[15, 20)$	17.5	599.2	615
$[20, 25)$	22.5	527.9	530
$[25, 30)$	27.5	477.2	482
$[30, 35)$	32.5	438.9	437
$[35, 40)$	37.5	408.5	430
$[40, 45)$	42.5	383.7	363
$[45, 50]$	47.5	362.9	370

The values of our test statistics observed on these data are:

- Pearson: X^2_{obs}= 2.995
- Likelihood Ratio: G^2_{obs}=45.791

A few questions for you...

1. **What is H_0 here?**

2. **How were X^2_{obs} and G^2_{obs} calculated?**

3. **What is the asymptotic distribution of X^2 and G^2 under H_0?**

4. **Knowing that the quantile of order 95% of a χ^2_ν with $\nu =$ is 16.9 can we conclude that our power-law model fits the data well?**

5. **For which test between X^2_{obs} and G^2_{obs} would you expect the χ^2 approximation to be more reliable?**

6. **Specify the formula of the p-values to test H_0 using X^2 and G^2.**

7. **Suppose θ and τ are unknown. If we were to estimate them, would you expect the asymptotic distribution of X^2 and G^2 under H_0 to stay the same?**