Report worksheet 1

Joaquín Rapela

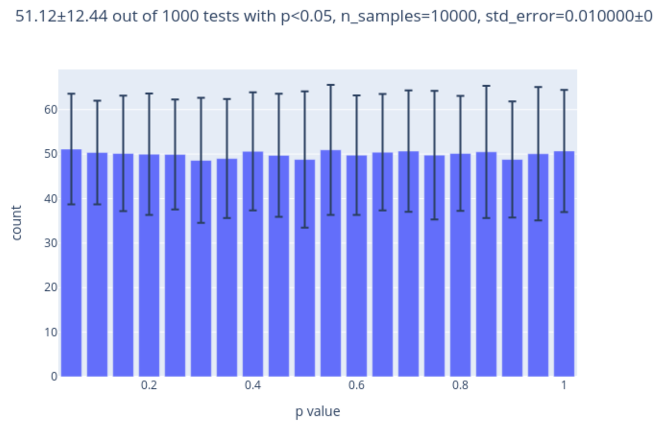
January 30, 2023

Exercise 1: t-test for non-Gaussian distributions

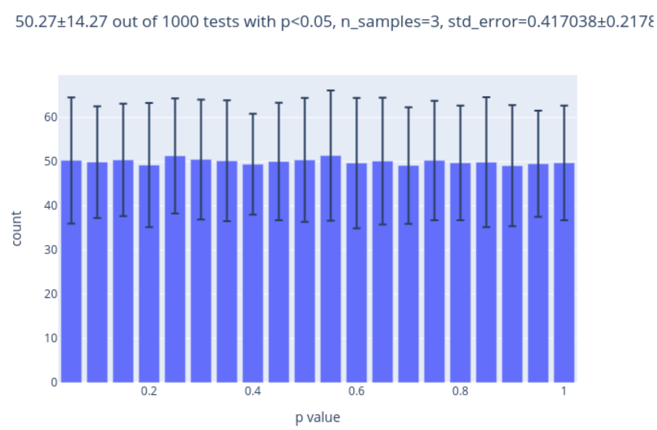
Under the null hypothesis, p-values should follow a uniform distribution (see proof in Appendix A). Therefore, when sampling from a normal distribution with zero mean, we should observe 0.05*`n_repeats`=50 tests with pvalues in the range $[p, p + 0.05]$, $\forall p \in [0, 0.95]$. In particular, when sampling from a Normal distribution with zero mean, we should observe 50 tests with p_value<0.05.

(a) Here we are sampling from a normal distribution with zero mean. Thus, all histogram bins of length 0.05 should have around 50 counts. And this is what Figs. 1a and 1b show.

For a probability of type I error less than 0.05, the null hypothesis is rejected 5% of the time when the null hypothesis is true.



(a) sample size = 10,000



(b) sample size = 3

Figure 1: Exercise 1a. Histogram of p-values of 1.000 t-tests evaluating if the mean of 10.000 samples (a) or 3 samples (b) from a $\mathcal{N}(0,1)$ is equal to zero. The t-test works well for both small (a) and large (b) samples. Click on the figure to see its interactive version. The script to generate this figure appears here and the parameters used for this script appear here.

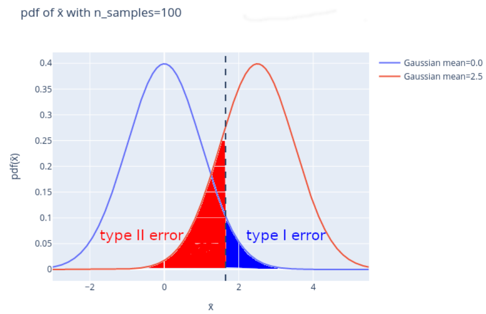
(b) The previous item, and all the following ones, study the probability Type I error of a hypothesis test. This is the probability of rejecting the null hypothesis when this hypothesis is true. This item studies the probability of Type II error, which is the probability of accepting the null hypothesis when this hypothesis is false.

Fig. 2a illustrates Type I and II errors. Here the null hypothesis is that the observed samples come from a standard Gaussian distribution (mean 0 and standard deviation 1). The probability density function of the sample mean under the null hypothesis is the blue trace in Fig. 2a. If we accept a probability of type I error of at most 5% (i.e., take $\alpha = 0.05$), a type I error will occur if the observed sample mean is to the right of the dashed vertical line. The probability of this error is the area of the region coloured in blue in Fig. 2a.

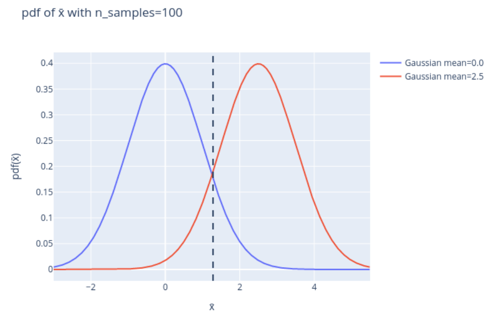
To calculate the probability of type II error we need to specify the distribution of an alternative hypothesis. In Fig. 2a the distribution of the observations under the alternative hypothesis is a Gaussian with mean 2.5 and standard deviation 1.0. If we take $\alpha = 0.05$, a type II error will occur if the observed sample mean is to the left of the dashed vertical line. The probability of this error is the area of the region coloured in red in Fig. 2a.

To reduce the probability of Type II error we can:

1. accept a larger probability of type I error (i.e., reduce α). Fig. 2b illustrates the reduction of type II error when accepting a probability of type I error of 10%. Compare with Fig. 2a.
2. use an alternative hypothesis with a mean more different from that of the null hypothesis (i.e., increase the effect size). Fig. 2c illustrates the reduction of type II error when samples under the alternative hypothesis have a mean of 3.5 and standard deviation of 1. Compare with Fig. 2a.
3. increase the sample size. If n samples come from a Gaussian distribution with mean μ and standard deviation σ (i.e., $x_i \sim \mathcal{N}(x_i|\mu, \sigma^2)$), then the sample mean will follow another Gaussian distribution with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$. Therefore, as the sample size increases, the variability of the sample mean decreases, both under the null and alternative hypothesis. Then the Gaussian distributions under the null and alternative hypothesis will overlap less, which will decrease the probability of type II error. Fig. 2d shows the distributions of the sample means under the same null and alternative hypothesis as in Fig. 2a, but with an increased sample size of 250.

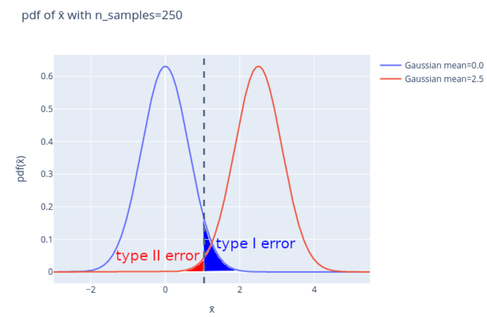


(a) Baseline distributions of $\bar{x}$ under the null (blue trace) and alternative (red trace) hypothesis.



(b) Decreasing the probability of type II error by increasing the probability of type I error.

(c) Decreasing the probability of type II error by increasing the effect size.

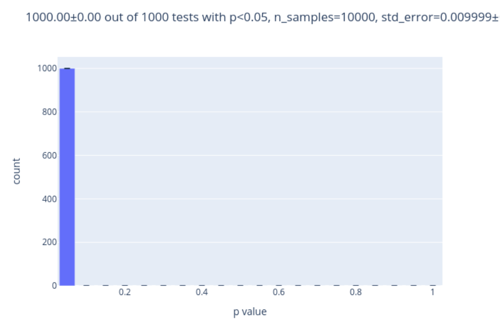


(d) Decreasing the probability of type II error by increasing the sample size.

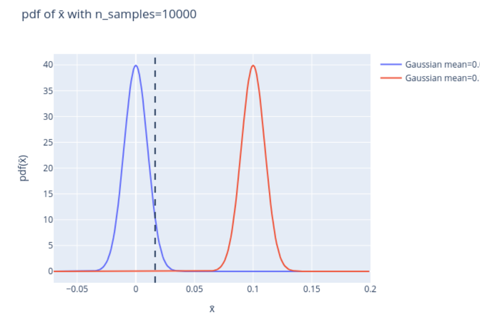
Figure 2: Exercise 1b. Type II errors in hypothesis testing and ways to decrease it. Click on any of the figures to see its interactive version. The script to generate these figures appears here and the parameters used for this script appear here.

Fig. 3a shows a histogram of p-values for 1000 one-sample t-tests checking if the mean of a sample of size 1000 from the $\mathcal{N}(0.1, 1)$ distribution is different from zero. For all 1000 tests the null hypothesis was rejected at the 0.05 criterion. Fig. 3b shows the pdf of the sample mean under the null (blue trace) and the alternative (red trace) hypothesis. That the t-test rejected the null hypothesis for all samples is explained by the fact that the area of the pdf of the sample mean under the alternative hypothesis (red trace) to the left of the vertical dashed line (i.e., the probability of type II error) is almost zero.

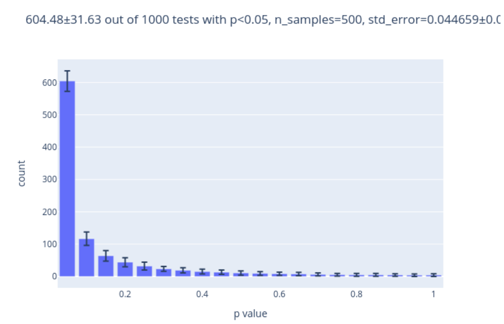
From the discussion above we see that, if we decrease the sample size, the Gaussian distributions under the null and alternative hypothesis should overlap more, which will increase the probability of a type II error. Fig. 3c is identical to Fig. 3a, but uses a reduced sample size of 500. We observe an increase in the counts for bins corresponding to p-values greater than 0.05, which implies that the probability of type II error increased. Fig. 3d shows the probability density functions of the sample means under the null and alternative hypothesis. These pdfs overlap substantially more than those in Fig. 3b, which indicates a larger probability of type II error, in agreement with Fig. 3c.



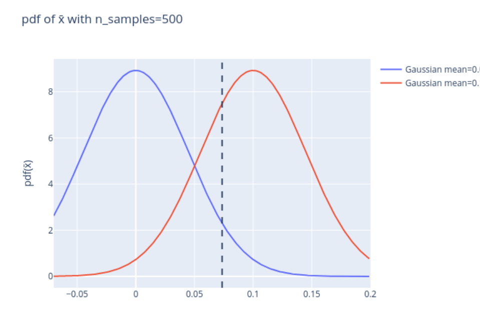
(a) Histogram of p-values of 1.000 t-tests evaluating if the mean of 10.000 samples from a $\mathcal{N}(0.1, 1)$ is equal to zero. Click on the figure to see its interactive version.



(b) Distributions of samples means under the null hypothesis that samples come from a $\mathcal{N}(0, 1)$ (blue trace) and under the alternative hypothesis that samples come from a $\mathcal{N}(0.1, 1)$ (red trace), for a sample size of 10000.



(c) Histogram of p-values of 1.000 t-tests evaluating if the mean of 500 samples from a $\mathcal{N}(0.1, 1)$ is equal to zero. Click on the figure to see its interactive version.

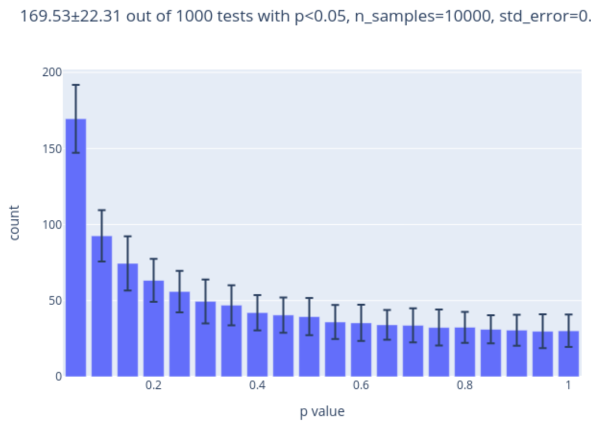


(d) Distributions of samples means under the null hypothesis that samples come from a $\mathcal{N}(0, 1)$ (blue trace) and under the alternative hypothesis that samples come from a $\mathcal{N}(0.1, 1)$ (red trace), for a sample size of 500.

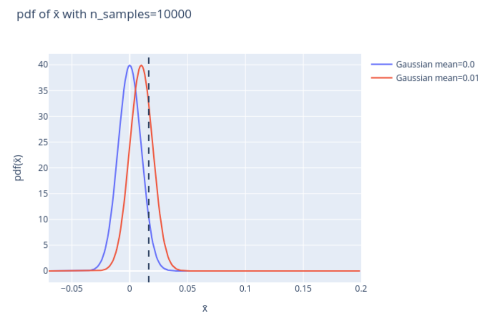
Figure 3: Exercise 1b. Histogram of p-values from 1000 one-sample t-tests used with samples from a $\mathcal{N}(0.1, 1)$, and pdfs of the sample mean under the null and alternative hypothesis. The script to generate this figure appears here and the parameters used for this script appear here.

The second part of this exercise asks to run t-tests using samples with a mean that is an order or magnitude more similar to the mean under the null hypothesis than the mean of the samples in the first part of this exercise. That is, in this second part we want to study the probability of type II error when the difference between the true and alternative hypothesis (i.e., the effect size) is smaller. As illustrated in Figs. 2a and 2c, decreasing the effect size should increase the probability of type II error. This increase is reflected in the larger number of counts for bins with p-values larger than 0.05 in Fig. 4a than in Fig. 3a, and by the larger overlap between the pdfs of the sample means under the null and alternative hypothesis in Fig. 4b than in Fig. 3b.

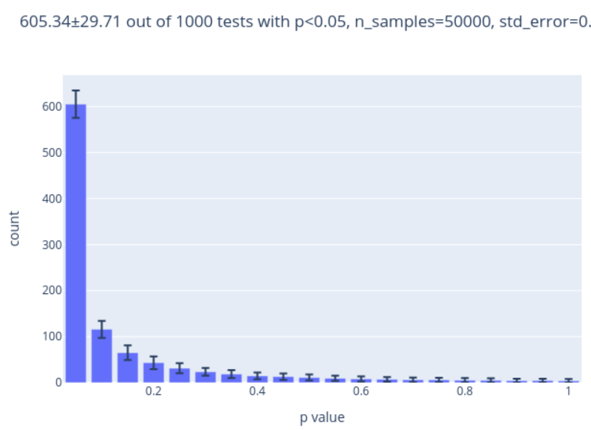
As illustrated in Fig. 2d, increasing the sample size reduces the probability of type II error. In Figs. 4c and 4d we increased the sample size from 10000 to 50000. Fig. 4d shows that the overlap between the pdf of the sample mean under the null and alternative hypothesis decreased, and Fig. 4d shows that the proportion of type II errors decreased from 1-170/10000=83% to 1-610/1000=39%.



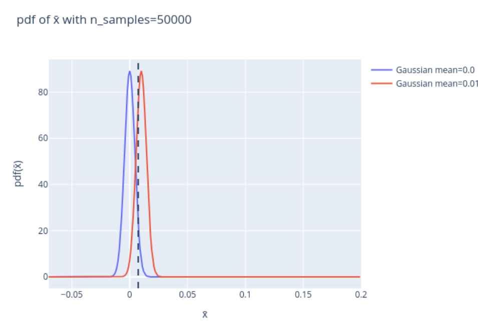
(a) Histogram of p-values of 1.000 t-tests evaluating if the mean of 10.000 samples from a $\mathcal{N}(0.01, 1)$ is equal to zero.



(b) Distributions of samples means under the null hypothesis that samples come from a $\mathcal{N}(0, 1)$ (blue trace) and under the alternative hypothesis that samples come from a $\mathcal{N}(0.01, 1)$ (red trace), for a sample size of 10000.



(c) Histogram of p-values of 1.000 t-tests evaluating if the mean of 50,000 samples from a $\mathcal{N}(0.01, 1)$ is equal to zero.



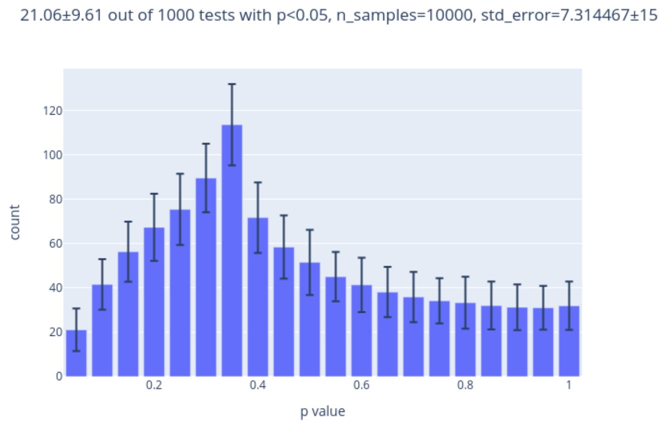
(d) Distributions of samples means under the null hypothesis that samples come from a $\mathcal{N}(0, 1)$ (blue trace) and under the alternative hypothesis that samples come from a $\mathcal{N}(0.01, 1)$ (red trace), for a sample size of 50000.

Figure 4: Exercise 1b. The script to generate this figure appears here and the parameters used for this script appear here.

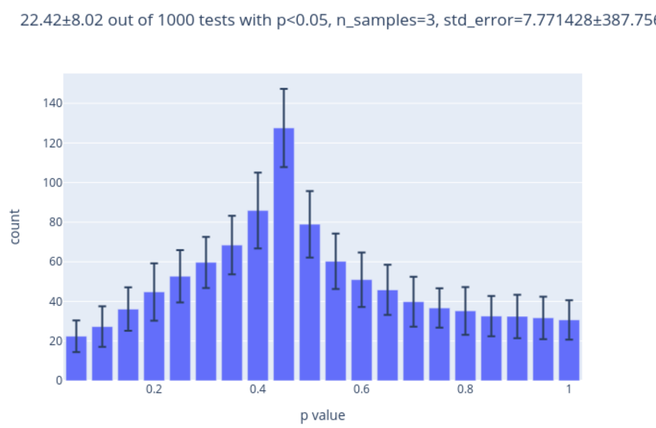
(c) Fig. 5 shows histograms of p-values for samples from a standard Cauchy distribution.

By the central limit theorem, I was expecting that for large samples the t-test was robust to the use of standard Cauchy samples. However, in Fig. 5a the distribution of p-values is not $\mathcal{U}[0, 1]$, indicating that the t-test is not robust to the use of standard Cauchy samples. Note, from Fig. 6, that the t-test is robust the use of Rademacher samples. The reason for this difference in robustness between standard Cauchy and

Rademacher samples is explained in Appendix B.



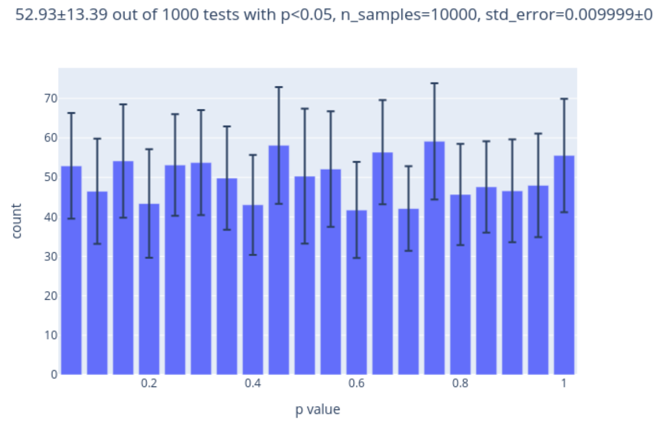
(a) sample size = 10,000



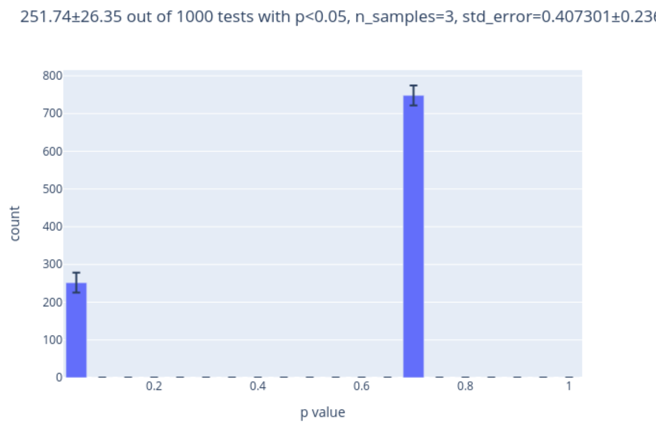
(b) sample size = 3

Figure 5: Exercise 1c. Histogram of p-values of 1.000 t-tests evaluating if the mean of 10.000 samples (a) or 3 samples (b) from a standard Cauchy distribution is equal to zero. The t-test fails for both small (a) and large (b) samples. Click on the figure to see its interactive version. The script to generate this figure appears here and the parameters used for this script appear here.

(d) Please refer to Fig. 6.



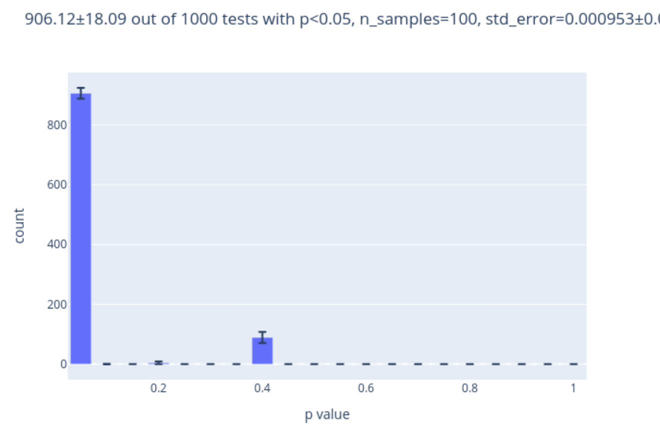
(a) sample size = 10,000



(b) sample size = 3

Figure 6: Exercise 1d. Histogram of p-values of 1.000 t-tests evaluating if the mean of 10.000 samples (a) or 3 samples (b) from a standard Rademacher distribution is equal to zero. For Rademacher samples, the t-test is robust to the assumption that its samples should come from a Normal distribution. It works well for both small (a) and large (b) samples. Click on the figure to see its interactive version. The script to generate this figure appears here and the parameters used for this script appear here.

(e) Please refer to Fig. 7.



(a) Histogram of p-values of 1.000 t-tests evaluating if the mean of 100 samples from the very skewed distribution distribution with mean 0.001 is equal to 0.001. Click on the figure to see its interactive version.

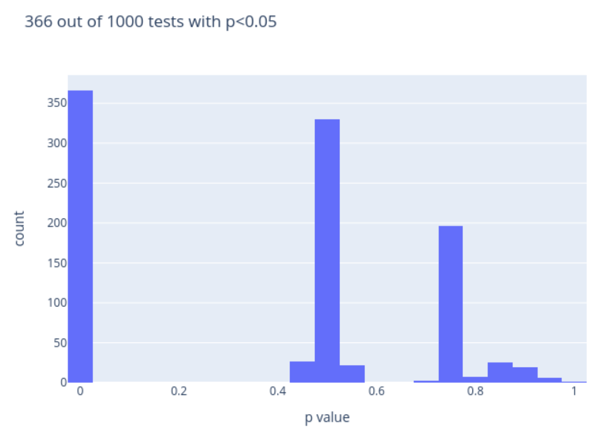
(b) Histogram of p-values of 1.000 t-tests evaluating if the mean of 100 samples from the very skewed distribution distribution with mean 0.001 is equal to 0.0. Click on the figure to see its interactive version.

Figure 7: Exercise 1e. The script to generate this figure appears here and the parameters used for this script appear here.

Exercise 2: randomization test

Please refer to Fig. 8.

(a) Histogram of p-values of 1.000 randomization tests evaluating if the mean of 10 samples from the Rademacher distribution distribution is equal to 0.0. Click on the figure to see its interactive version.

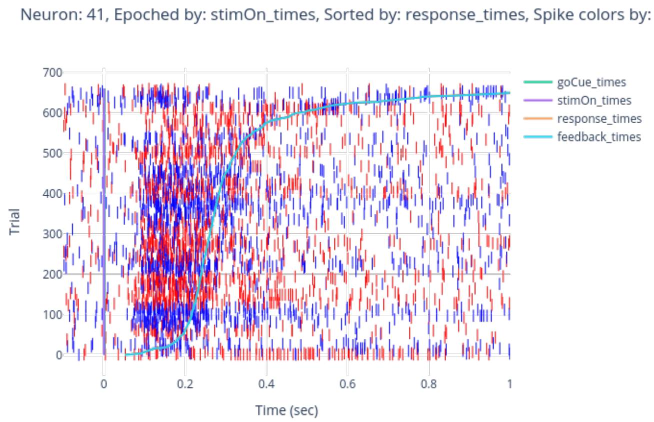


(b) Histogram of p-values of 1.000 randomization tests evaluating if the mean of 10 samples from the skewed distribution is equal to zero. Click on the figure to see its interactive version.

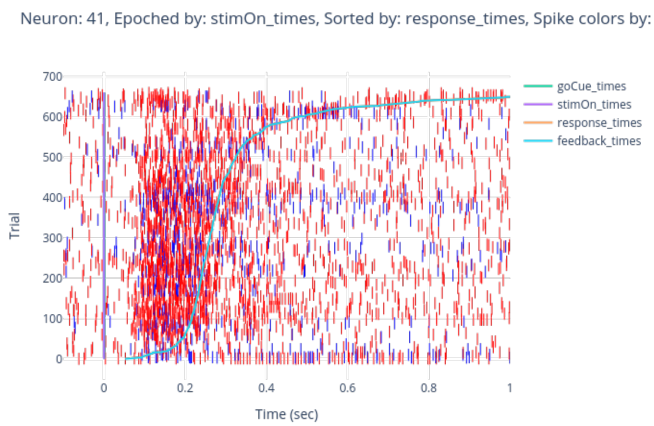
Figure 8: Exercise 2. The script to generate this figure appears here and the parameters used for this script appear here.

Exercise 3: raster plots

Please refer to Fig. 9.



(a) Rasterplot of neuron 41 aligned to `stimOn_times`, sorted by `response_times` and coloured by



(b) Rasterplot of neuron 41 aligned to `stimOn_times`, sorted by `response_times` and coloured by `feedbackType`.

Figure 9: Exercise 3. The script to generate this figure appears here and the parameters used for this script appear here.

Appendix A Under the null hypothesis, p-values are uniformly distributed in $[0, 1]$

In Exercise 1, for $i = 1, \ldots, 1000$, we generated samples from random variables $\{x_{(i,1)}, \ldots, x_{(i,10000)}\}$, from these samples we computed a t-statistic $t_i = f(x_{(i,1)}, \ldots, x_{(i,10000)})$, and from this statistic we calculated a p-value, $p_i = g(t_i)$. Because the t-statistic is a function, f, of random variables, it can be considered as a random variable, T. Because the p-value is a function, g, of a random variables, it can also be considered as a random variable, P. The goal of this section is to prove that the p-value random variable is uniformly distributed in [0,1]; i.e., $P \sim \mathcal{U}[0, 1]$. This proof is given in Lemma 1. Before giving this proof we prove two auxiliary claims (Claims 1 and 2).

Fig. 10 illustrates the concept of a p-value. It is the probability of observing a statistic, t, greater than the observed one, t_{obs}, when the null hypothesis is true.

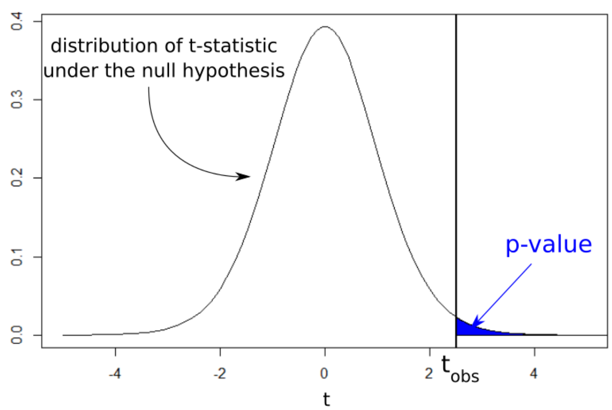


Figure 10: Illustration of the p-value concept. A p-value is the probability of observing a statistic, t, greater than the observed one, t_{obs}, when the null hypothesis is true.

Claim 1. *Let P be the p-values random variable, T be the t-statistic random variable, and F_T be the cumulative distribution function of T. Then $P = 1 - F_T(T)$*

Proof. Let $t_{\text{obs},i}$ and p_i be an observed statistic and associated p-value, respectively. Then,

$$p_i = P(T > t_{\text{obs},i}) = 1 - P(T < t_{\text{obs},i}) = 1 - F_T(t_{\text{obs},i}) \tag{1}$$

The first equality in Eq. 1 follows from the definition of a p-value (Fig 10). Because $p_i = 1 - F_T(t_{\text{obs},i})$ holds for any pair of samples p_i and $t_{\text{obs},i}$, then $P = 1 - F_T(T)$. □

Claim 2. *A random variable U is uniformly distributed in [0,1]; i.e., $U \sim \mathcal{U}[0, 1]$, if and only if its cumulative distribution function is $F_U(u) = P(U < u) = u$, for $u \in [0, 1]$.*

Proof.

$$
\begin{aligned}
U \sim \mathcal{U}[0,1] &\iff f_U(u) = 1 \text{ for } u \in [0,1] \text{ and } f_U(u) = 0 \text{ elsewhere} \qquad (2)\\
&\iff F_U(u) = P(U < u) = \int_0^p f_U(u)dp = u, \text{ for } u \in [0,1]. \qquad (3)
\end{aligned}
$$

□

Lemma 1. *When the null hypothesis holds, p-values are uniformly distributed in* $[0,1]$*; i.e.,* $P \sim \mathcal{U}[0,1]$.

Proof. By Claim 2 it suffices to show that $F_P(p) = p$.

$$
\begin{aligned}
F_P(p) &= P(P < p) = P(1 - F_T(T) < p) = P(1 - p < F_T(T) \qquad (4)\\
&= P(T > F_T^{-1}(1-p)) = 1 - P(T < F_T^{-1}(1-p))\\
&= 1 - F_T(F_T^{-1}(1-p)) = 1 - (1-p) = p
\end{aligned}
$$

Note: the second equality in Eq. 4 follows from Claim 1.

□

Appendix B Central limit theorem

The central limit theorem demonstrates that, under certain conditions, the distribution of the mean of a sufficiently large number of random samples is Normally distributed. Precisely:

Theorem 1 (Central Limit Theorem)**.** *If* $X_1, X_2, \ldots, X_n, \ldots$ *are random samples drawn from a population with overall mean* μ *and finite variance* σ^2*and if* $\bar{X}_n$ *is the sample mean of the first* n *samples, then the limiting form of the distribution,* $Z = \lim_{n\to\infty}\left(\frac{\bar{X}_n - \mu}{\sigma_{\bar{X}}}\right)$*, with* $\sigma_{\bar{X}} = \sigma/\sqrt{n}$*, is a standard normal distribution*[1].

For $n = 10,000$ we expect that the distribution of the sample mean closely follows a Normal distribution.

The t statistic, i.e., $t = \frac{\bar{X}_n - \mu}{S_n}$, depends on the samples only through the sample mean, i.e., $\bar{X}_n$, and the sample standard deviation, i.e., $S_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{X}_n\right)^2}$. For simplicity of explanation, we ignore the dependence on the sample standard deviation.

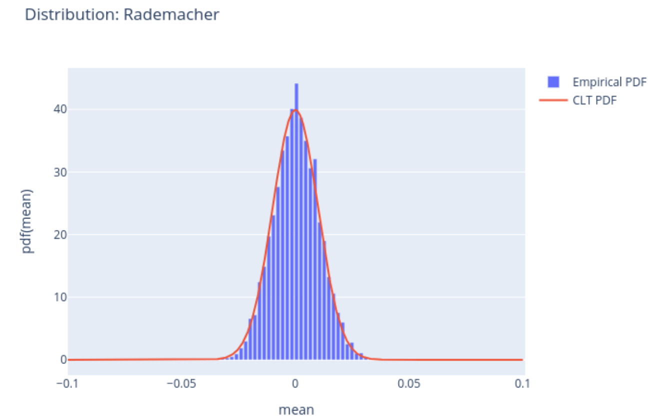
With this simplification, the t-statistic depends on the samples only though the sample mean. For sufficiently large n, by the central limit theorem, the mean of samples from most distributions will be Normally distributed, as required by the t-test. Thus, we expect that, for large samples, the t-test works correctly for samples from most distributions.

The central limit theorem holds to the Rademacher distribution and, as shown in Fig. 11a, the distribution of the mean of 10,000 Rademacer samples follows a Normal distribution. This explain why the t-test worked well for the samples of 10,000 Rademacher random

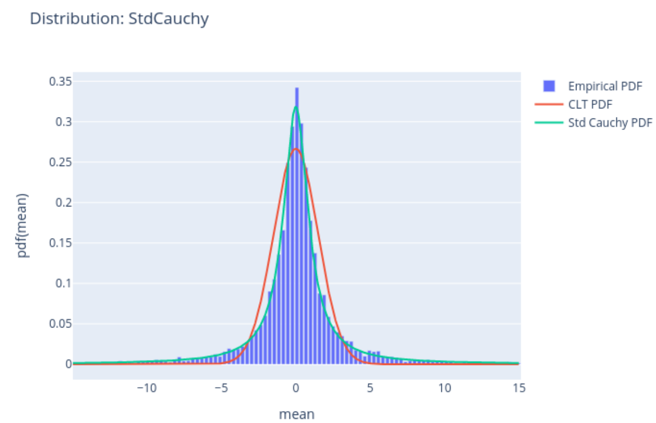
[1] https://en.wikipedia.org/wiki/Central_limit_theorem

variables in exercise 1d (Fig. 6a). However, the central limit theorem does not hold for the standard Cauchy distribution, since this theorem requires finite variance and variance is not defined for the standard Cauchy distribution. As shown in Fig. 11b, the distribution of the mean of 10,000 standard Cauchy samples does not follows a Normal distribution. This explains why the t-test was not robust to the use of 10,000 standard Cauchy samples in exercise 1c (Fig. 5a).

Fig. 11b shows that the distribution of the mean of 10,000 standard Cauchy samples is well approximated by the standard Cauchy distribition, as expected by Lemma 2.



(a) Rademacher samples



(b) Standard Cauchy

Figure 11: Central limit theorem examples. The central limit theorem applies to the Rademacher distribution, and panel (a) shows that a historgram of 10,000 samples from this distribution is well approximated by a Normal density with mean 0 and standard deviation $\sigma = \frac{1}{\sqrt{10{,}000}}$. The central limit theorem does not apply to the standard Cauchy distribution, and panel (b) shows that a histogram of 10,000 samples from this distribution is not well approximated by a Normal density with mean 0 and standard deviation equal to the sample standard deviation. This histogram is well approximated by a Standard Cauchy density, as expected by Lemma 2. The script to generate this figure appears here and the parameters used for this script appear here.

Lemma 2. *If $X_1, \ldots, X_n$ are iid Standard Cauchy random variables, then $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ is also Standard Cauchy.*

Proof. We will show that the characteristic function of the mean of Standard Cauchy iid random variables, i.e., $\varphi_{\bar{X}_n}(t)$, equals the characteristic function of a Standard Cauchy random variable, i.e., $\varphi_X(t) = e^{-|t|}$[2]. Then, because the characteristic function uniquely specficies a random variable, we conclude that this mean is a Standard Cauchy random variable.

$$\varphi_{\bar{X}_n}(t) = E\left\{e^{j\bar{X}_n t}\right\} = E\left\{\prod_{i=1}^{n} e^{\frac{jX_i t}{n}}\right\} = \prod_{i=1}^{n} E\left\{e^{\frac{jX_i t}{n}}\right\} = \prod_{i=1}^{n} e^{-\frac{|t|}{n}} = \left(e^{-\frac{|t|}{n}}\right)^n = e^{-|t|} \tag{5}$$

Notes:

1. the first equality follows from the definition of the characteristic function,
2. the second equality uses the property that the exponential of sums is a product of exponentials,
3. the third equality is a consequence of the independece of the random variables,
4. the fourth equality follows from the characteristic function of a Cauchy random variable, i.e., $\varphi_X(t) = e^{-|t|}$, evaluated at $\frac{t}{n}$.

□

[2] https://en.wikipedia.org/wiki/Cauchy_distribution