Report worksheet 1

Joaquín Rapela

January 22, 2023

Exercise 1: t-test for non-Gaussian distributions

Under the null hypothesis, p-values should follow a uniform distribution (see proof in Apendix 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) Please refer to Figure 1. 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 Figure 1 shows.

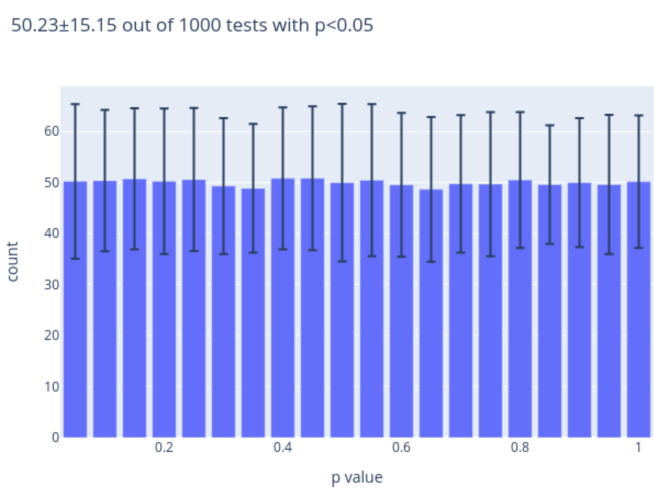
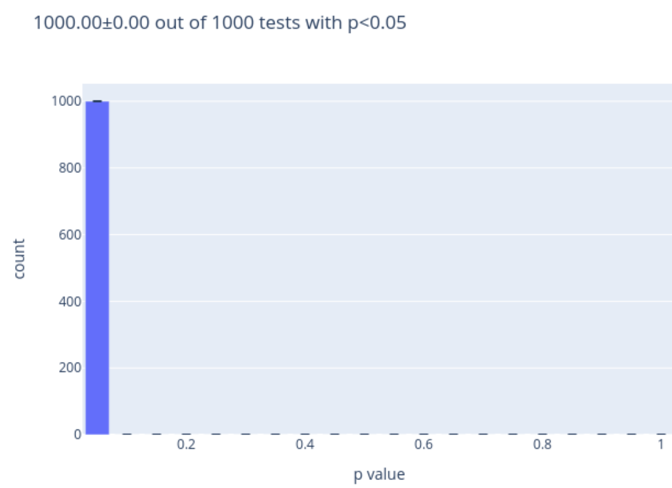
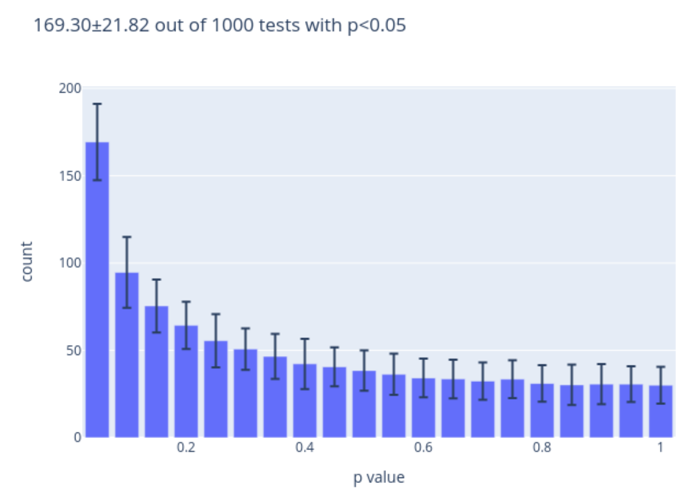


Figure 1: Exercise 1a. Histogram of p-values of 1.000 t-tests evaluating if the mean of 10.000 samples from a $\mathcal{N}(0, 1)$ is equal to zero. Click on the figure to see its interactive version. The code to generate this figure appears here and the parameters used for this script appear here.

(b) Please refer to Figure 2.



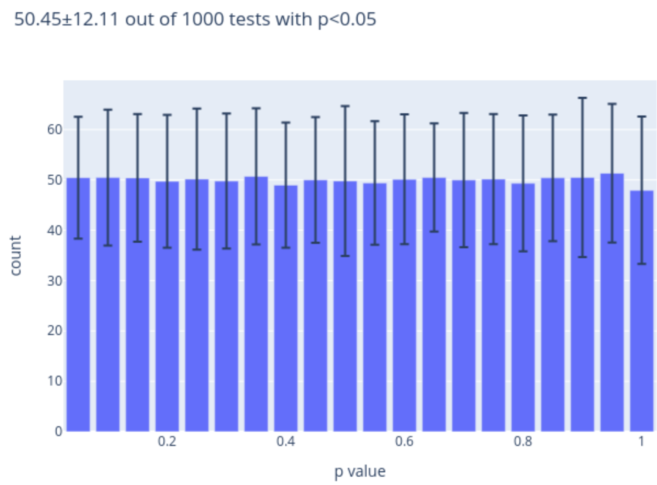
(a) Histogram of p-values of 1.000 t-tests evaluating if the mean of 10.000 samples from a $\mathcal{N}(0.1, 1)$ is equal to zero. Click on the figure to see its interactive version.



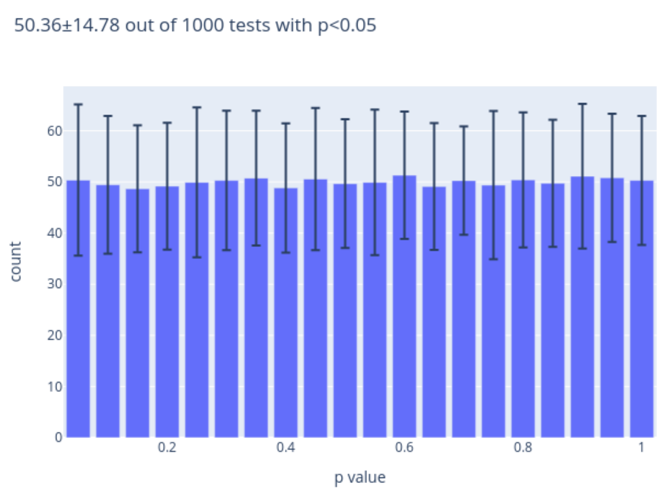
(b) Histogram of p-values of 1.000 t-tests evaluating if the mean of 10.000 samples from a $\mathcal{N}(0.01, 1)$ is equal to zero. Click on the figure to see its interactive version.

Figure 2: Exercise 1b. The code to generate this figure appears here and the parameters used for this script appear here.

(c) Please refer to Figure 3.



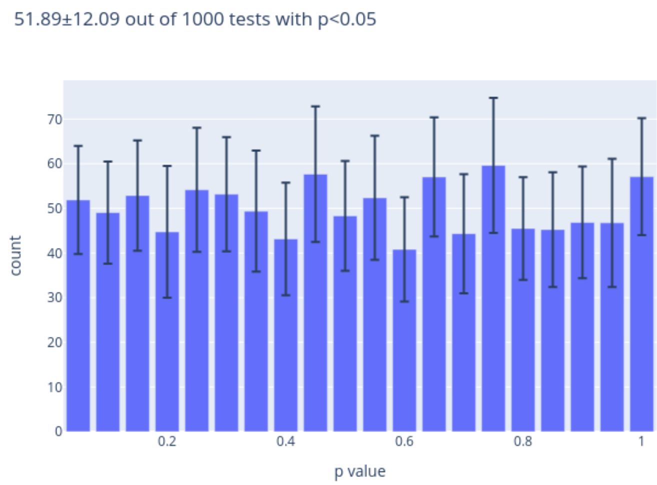
(a) Histogram of p-values of 1.000 t-tests evaluating if the mean of 10.000 samples from a standard Cauchy distribution is equal to zero. Click on the figure to see its interactive version.



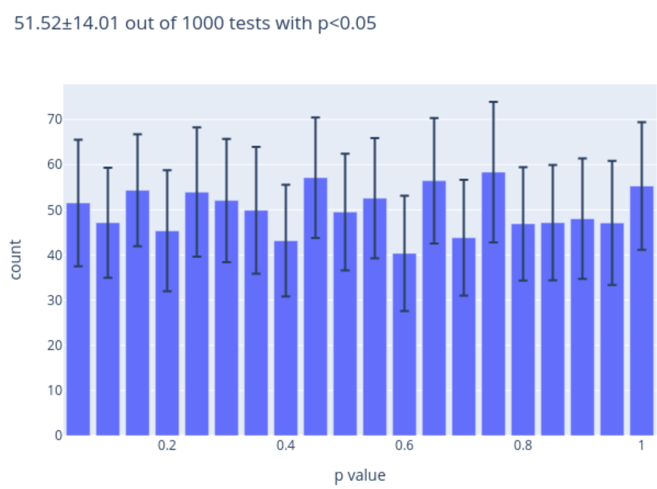
(b) Histogram of p-values of 1.000 t-tests evaluating if the mean of 3 samples from a standard Cauchy distribution is equal to zero. Click on the figure to see its interactive version.

Figure 3: Exercise 1c. The code to generate this figure appears here and the parameters used for this script appear here.

(d) Please refer to Figure 4.



(a) Histogram of p-values of 1.000 t-tests evaluating if the mean of 10.000 samples from a Rademacher distribution is equal to zero. Click on the figure to see its interactive version.

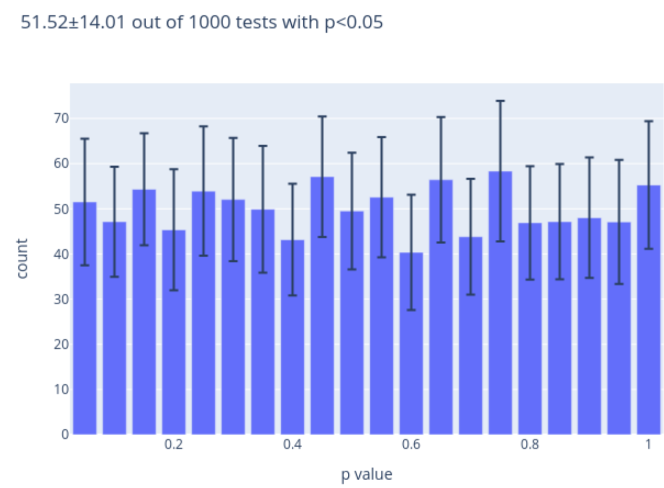


(b) Histogram of p-values of 1.000 t-tests evaluating if the mean of 3 samples from a Rademacher distribution is equal to zero. Click on the figure to see its interactive version.

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

(e) Please refer to Figure 5.

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



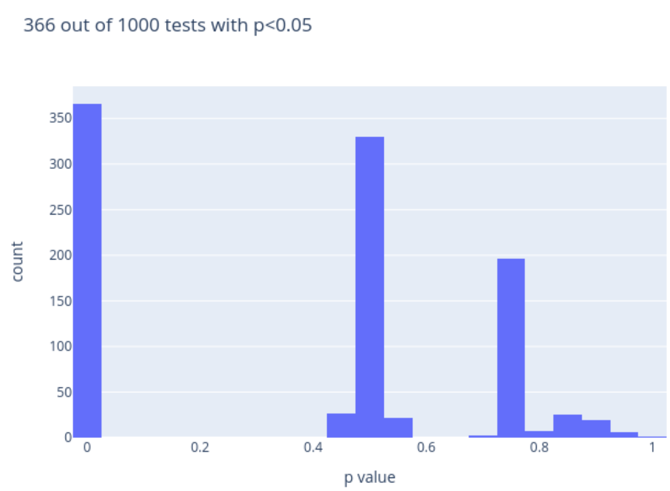
(b) Histogram of p-values of 1.000 t-tests evaluating if the mean of 3 samples from the very skewed distribution is equal to zero. Click on the figure to see its interactive version.

Figure 5: Exercise 1e. The code to generate this figure appears here and the parameters used for this script appear here.

Exercise 2: randomization test

Please refer to Figure 6.

(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.

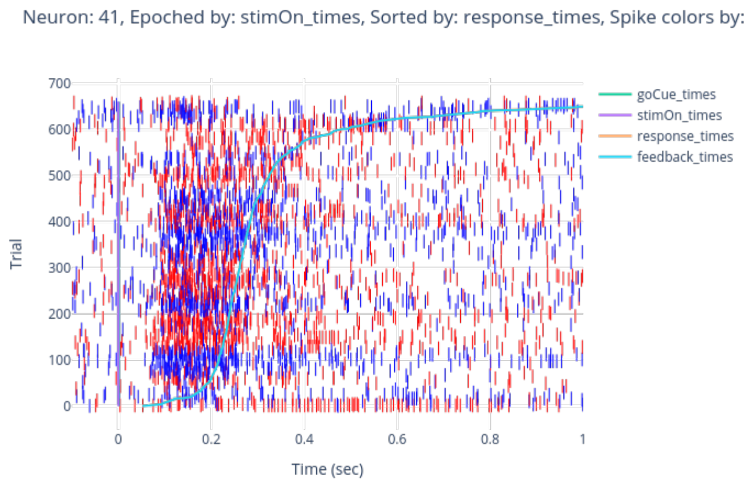


(b) Histogram of p-values of 1.000 randomization tests evaluating if the mean of 10 samples from the skewed distribution is equal to zero. Click on the figure to see its interactive version.

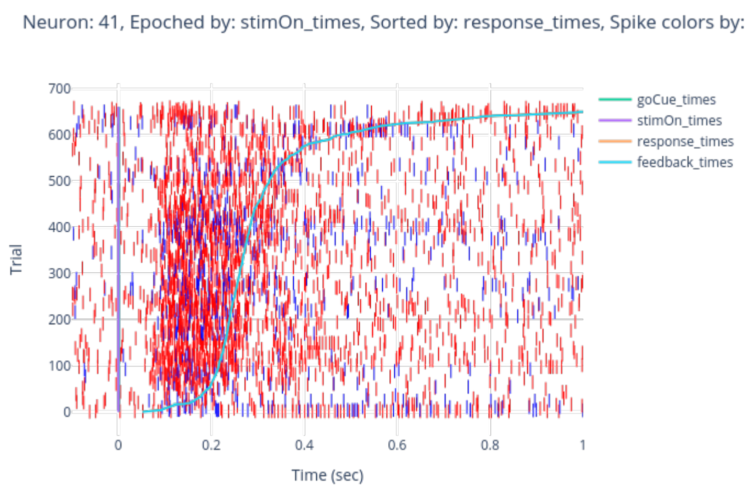
Figure 6: Exercise 2. The code to generate this figure appears here and the parameters used for this script appear here.

Exercise 3: raster plots

Please refer to Figure 7.



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



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

Figure 7: Exercise 3. The code 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).

Figure 8 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 hipothesis is true.

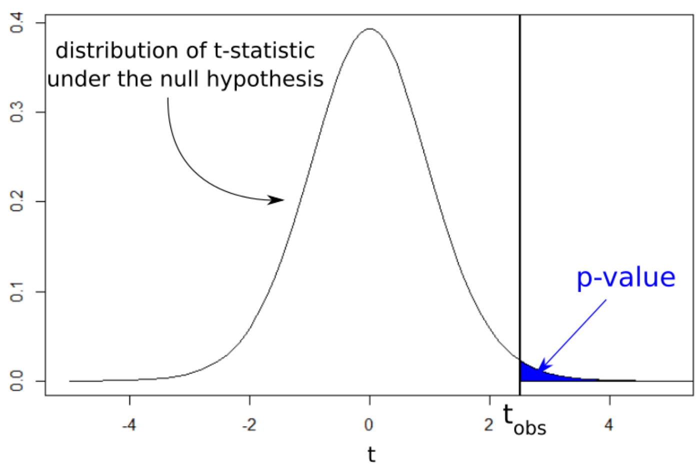


Figure 8: Illustration of the p-value concept. A p-value is the probability of observing a statistic, t, greater than the observed one, t_{obs}, when the null hypothesis is true.

Claim 1. *Let P be the p-values random variable, T be the t-statistic random variable, and F_T be the cummulative 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 8). Because $p_i = 1 - F_T(t_{\text{obs},i})$ holds for any pair of samples p_i and $t_{\text{obs},i}$, then $P = 1 - F_T(T)$. $\square$

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 cummulative distribution function is $F_U(u) = P(U < u) = u$, for $u \in [0,1]$.*

Proof.

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

$\square$

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 $F_P(p) = p$.

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

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

$\square$