

Exercise session 12 – Prob. & Stat. 2 — Dec 19, 2023

Entropy

Recall that *entropy* of a discrete random variable X is

$$H(X) = \sum_{x \in \text{Im}X} -P(X = x) \log_2 P(X = x).$$

- (a) What is the entropy of a random variable uniform on a set $\{1, 2, \dots, n\}$?
(b) What is the entropy of Bernoulli (two-valued) random variable $Ber(p)$?
(c) What is smallest possible entropy?
(d) * Show that if X takes on n values, then $H(X) \leq \log_2 n$. (Hint: use Jensen's inequality.)
- (a) Let X, Y be independent random variables. Then

$$H(X, Y) = H(X) + H(Y).$$

- (b) Find some X, Y (not independent) where

$$H(X, Y) \neq H(X) + H(Y).$$

- (c) * Show that always

$$H(X) \leq H(X, Y) \leq H(X) + H(Y).$$

Nonparametric tests

Possible choices:

- for 2-sample tests: Permutation test (`scipy.stat.permutation_test`), Mann-Whitney U-test, also known as Wilcoxon sum test (`scipy.stats.mannwhitneyu`) (we didn't cover the latter one)
- for 1-sample & paired tests: Sign test (easiest using `scipy.stats.binomtest`), Wilcoxon signed rank test (`scipy.stats.wilcoxon`)

3. Students coming for an oral exam had to wait the following number of minutes before being examined: 17, 15, 20, 20, 32, 28, 12, 26, 25, 25, 35, 24. Test at the 0.05 level of significance the professor's claim that the median waiting time is at most 20 minutes. Use the sign test and the Wilcoxon signed rank test.

4. Another warm-up: use the permutation test and/or Mann-Whitney U-test on data 1, 3 in one group and 2, 4 in second group.

5. Two brands of cell-phones are compared in terms of how many years they last before failure.

A: 2.1, 4.0, 6.3, 5.4, 4.8, 3.7, 6.1, 3.3

B: 4.1, 0.6, 3.1, 2.5, 4.0, 6.2, 1.6, 2.2, 1.9, 5.424

Test the hypothesis that the durability of both brands is the same (against the alternative that they are unequal). Use Mann-Whitney U-test and/or permutation test.