Open Science Framework DOI:10.31219/osf.io/***

Guessing long-run mutual information [draft]

P.G.L. Porta Mana
<pgl@portamana.org>

C. Battistin
<>

S. Gonzalo Cogno
<@>

***; updated 12 July 2020

***

1 Introduction

We have two kinds of stimuli and, say, ten kinds of neural responses to these stimuli, under specific environmental and behavioural conditions, and for a specific kind of subjects. The responses could be, for example, ten different firing rates of a specific neuron, or ten different total-population activities of a brain region.

Imagine that a researcher told us the results of 10^{60} stimulus-response measurements, all performed for the specified conditions and subjects, and in which the two kinds of stimulus occurred equally often. From the long-run[1] joint frequencies of stimulus-response pairs we could calculate the long-run mutual information between stimulus and response.

But we don't have access to such a wealth of observations, and never will. Suppose we only have actual knowledge of a sample of few, say 20, measurements for each kind of stimulus, from such long-run sequence. We can calculate the joint frequencies of stimulus-response pairs in this sample and the resulting sample mutual information.

Our question is: how do the mutual information of the sample and of the long-run sequence differ?

Let's explore some possibilities.

[1] "But this *long run* is a misleading guide to current affairs. *In the long run* we are all dead." (**keynes1923_r2013**)

Case A. Suppose the long-run frequencies of responses conditional on each kind of stimulus are as given in the top panel of fig. 1[2]. In the long run all responses occur equally often for either stimulus. The long-run mutual information is 0 bit.

Now we consider a sample of 20 response measurements for each stimulus from such long-run distribution. From this sample we calculate the joint frequencies and their mutual information. The exact long-run sequence is unknown, so such a sample could yield different results. Considering all long-run sequences (having the same long-run distribution) as equally plausible, we plot some possible results for the sample mutual information in the bottom panel, with a histogram. The long-run mutual information is indicated by the red line. The sampled mutual information is most surely always higher than the long-run one; it is even outside the inner 95% interval range.

Case B. The long-run frequencies of responses are given in the top panel of fig. 2. The responses do not occur equally often in the long run, and the frequencies of some of them are different for the two kinds of stimulus. The long-run mutual information is 0.48 bit.

We again consider a sample of 20 response measurements for each stimulus. The histogram of the possible sampled mutual information is shown in the bottom panel. The possible values are almost symmetrically distributed around the long-run value, which is well within their inner 68% quantile range, close to the median, 0.50 bit, and the mean, 0.51 bit.

Case C. The long-run frequencies of responses are given in the top panel of fig. 3. There is some difference in the long-run conditional frequency distributions of the responses . The long-run mutual information is 0.38 bit.

The histogram for the possible values of the mutual information from 20 samples is shown in the bottom panel. The long-run mutual information is outside the inner 75% quantile.

Case D. The long-run frequencies are shown in the top panel of fig. 4. The conditional frequency distributions of the responses have almost no overlap. The long-run mutual information is 0.92 bit.

The histogram of possible values of the sampled mutual information is shown in the bottom panel. The histogram is very skewed, with an average of 0.97 bit (the histogram's support is almost divided in distinct

[2] **panzerietal2007**.

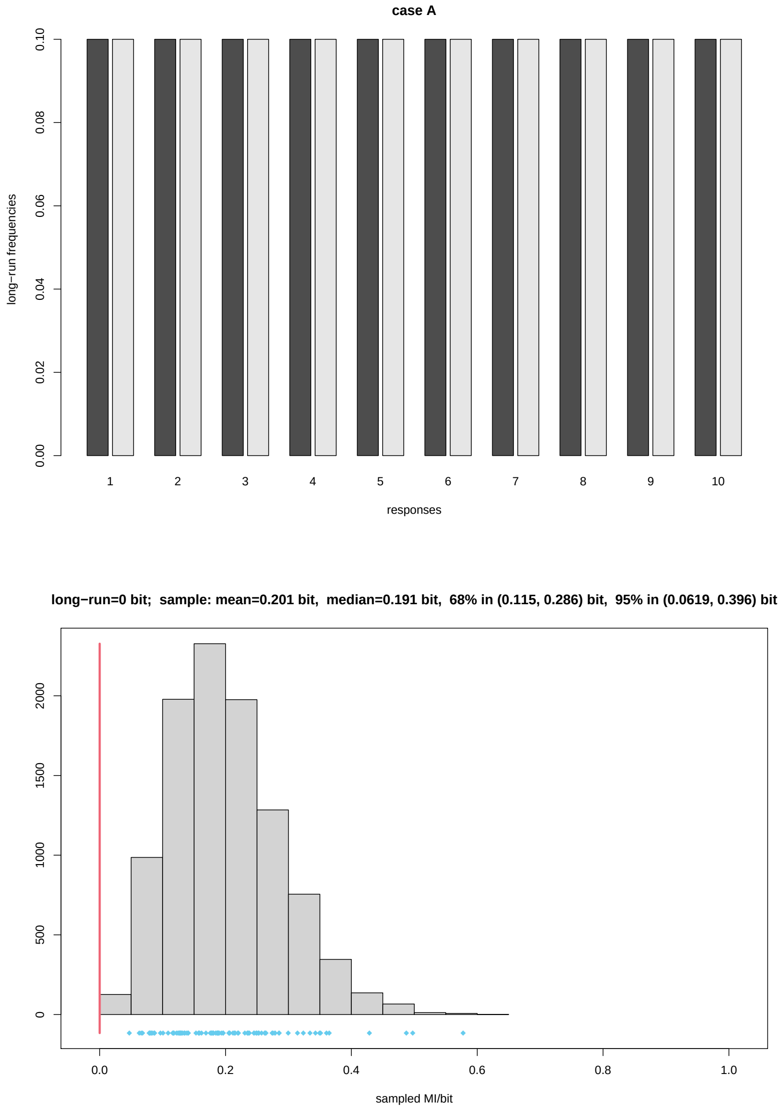
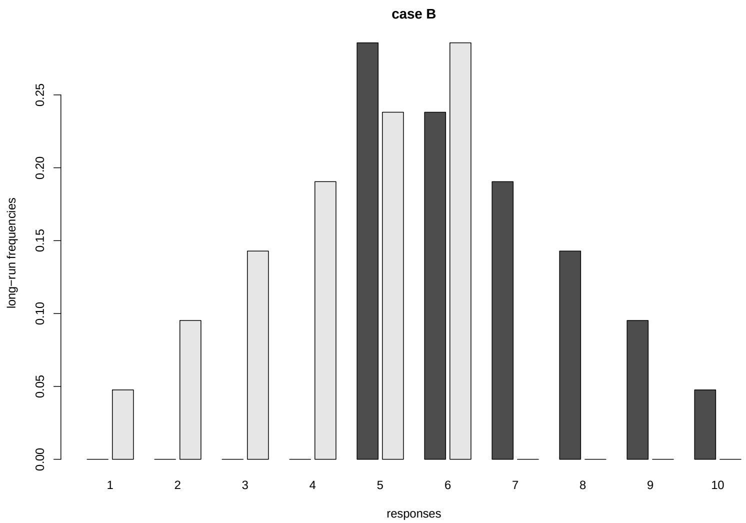


Figure 1 Case A



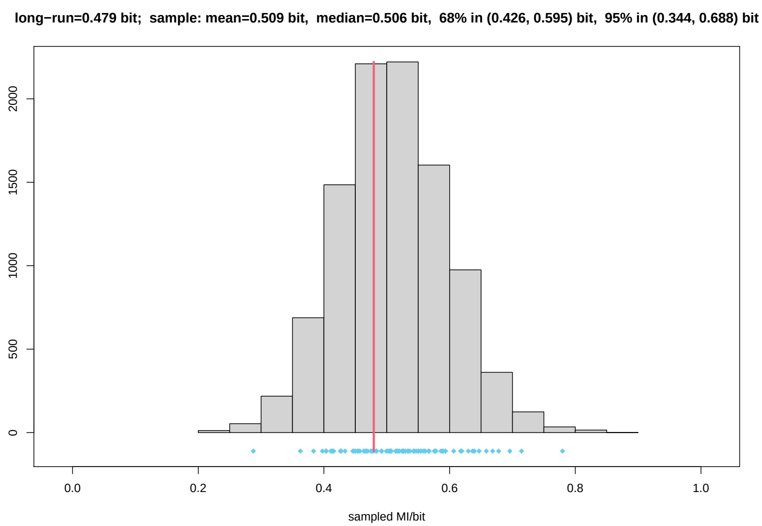


Figure 2 Case B

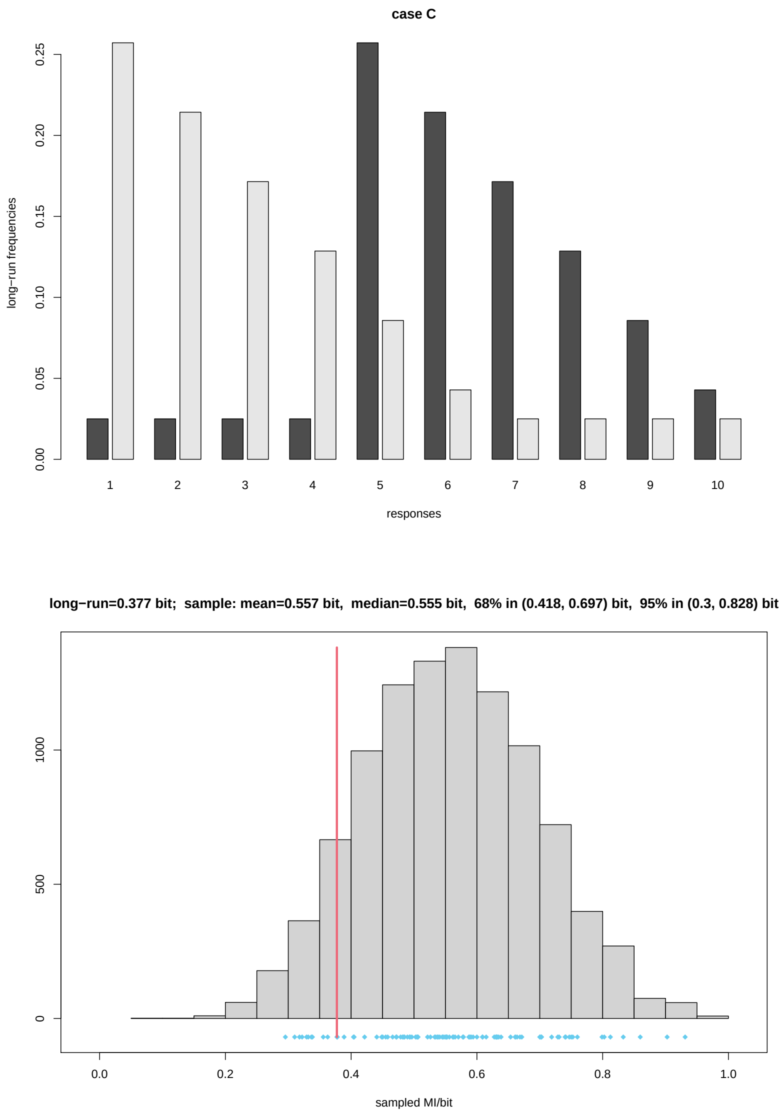


Figure 3 Case C

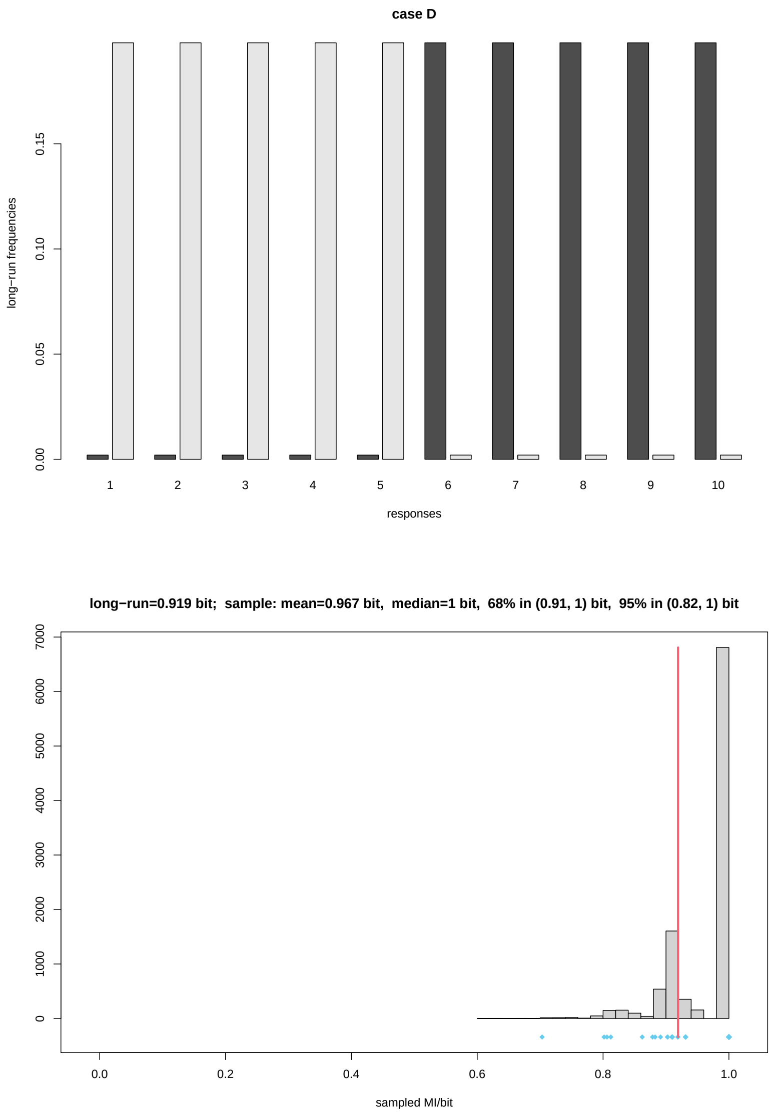


Figure 4 Case D

blocks; this is the result of the discreteness of the possible values of the frequencies from the sample: they must be multiples of $1/20$). The long-run mutual information is within the inner 68% quantile.

What preliminary conclusions can we draw from the examples above?

i. The possible values of the sample mutual information can, in some cases, cover a very broad range, even more than 50% of the maximal possible range of the mutual information, $[0, 1]$ bit.

ii. The distribution of these possible values is in some cases very skewed.

iii.

2 Caveats

The long-run frequencies of the two kinds of stimulus are also very important. Even if the conditional responses seem to yield low mutual info when stimuli occur equally often, they may yield a sufficient amount of mutual info when the stimuli don't occur equally often. In fact, this is partly a *decision* problem, which can't be judged only by checking frequencies or probabilities.