Appendix A Parametric vs. non-parametric hypothesis testing

The main difference between parametric and non-parametric hypothesis testing is that the former parameterizes probability density functions (PDFs) (Appendix D) and the latter does not. This distinction exists at two levels:

- Experimental data: parametric hypothesis testing assumes that the data are drawn from a population with a known, parameterizable PDF (usually the Gaussian distribution), but non-parametric procedures generally makes no such assumption.
- Test statistic: parametric procedures base inferences on parameterized test statistic PDFs which are analytically derived from the population PDF, but non-parametric procedures generally base inferences on empirically derived test statistic PDFs.

Below we consider these points in detail.

Parametric PDFs

The fundamental PDF upon which most parametric inference is based is the normal (Gaussian) distribution, which is parameterized by the true population mean μ and true population standard deviation σ (Fig.A1):

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{A.1}$$

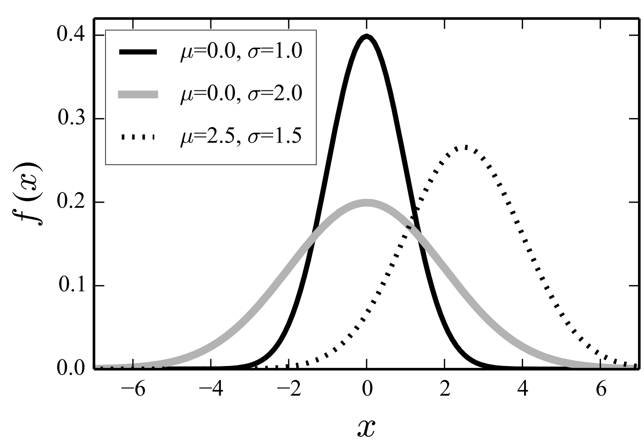

Figure A1: Gaussian probability density functions.

If we assign numerical values to μ and σ, then we can compute arbitrary probabilities using Eqns.D.1 and D.2 (Appendix D). For example, if μ=0 and σ=1, then the survival function (Eqn.D.2) predicts $P(x>0.0)$=0.500 and $P(x>2.0)$=0.023. These probabilities respectively imply that 50% of random values drawn from this distribution are expected to be greater than zero, and only 2.3% are expected to be greater than 2.0. To re-emphasize the meaning of 'parametric', we note that two simple parameters (μ and σ) completely specify the probabilistic behavior of Gaussian data.

The Gaussian PDF (Eqn.A.1) is nevertheless seldom used directly when conducting statistical inference. One reason is that the Gaussian PDF describes a random variable x, which is analogous to the raw data we measure experimentally. Most experiments are less interested in x itself than in averages (one-sample tests), average differences (two-sample tests), and correlations between x and an independent variable (regression tests). To address these empirical pursuits, the parametric approach funnels the Gaussian PDF into a particular experimental design, and generates predictions regarding what Gaussian data would do in that particular setting, over an infinite number of identical experiments.

Another reason the Gaussian PDF is not used directly for statistical inference is that μ and σ are true population parameters, but we rarely know these true values because we rarely have access to the entire population. We instead have to estimate μ and σ using a sample drawn from that population, but those estimates are imperfect, especially if the data are not sampled randomly. Even when the data are sampled randomly, estimates of μ and σ worsen as sample size decreases (Fig.A2), and parametric inference must account for this sample size-dependent behavior. Student solved this problem in 1908 through use of PDF which depends only on sample size:

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} \tag{A.2}$$

Here ν is the degrees of freedom and Γ is the gamma function. The ν parameter specifies the number of values which can vary freely in a particular statistic's computation. For example, in the one-sample t test (Table F2) there are J responses, but not all response values can vary freely. In particular, after one estimates the mean, there are only $(J-1)$ responses which can vary freely to produce the same mean, so the SD estimate is normalized using $(J-1)$ rather than J (Table F2).

Equation A.2 is the analytical result obtained when Gaussian data (Eqn.A.1) are funneled into t statistic equations (Table F2). In other words, Gaussian data behave in a sample-size dependent manner (Fig.A3) when the sample is smaller than the population size.

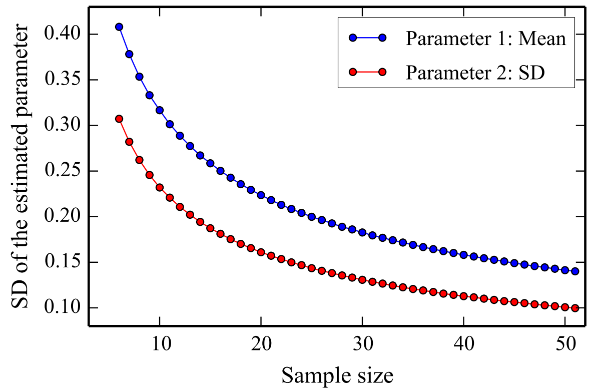

Figure A2: Variability of population parameter estimates as a function of sample size. The true mean and SD were 0 and 1, respectively. These results were constructed by simulating 10^6 samples of each sample size, computing each sample's mean and SD, then computing the SD of each parameter across all 10^6 samples.

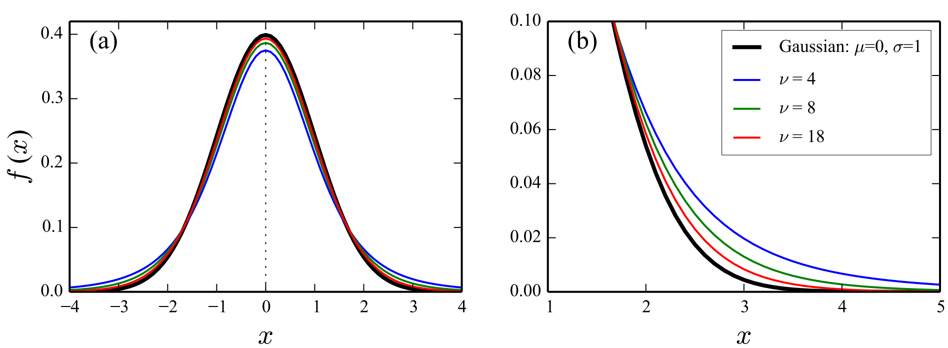

Figure A3: Comparison of various t PDFs with the standard normal PDF (μ=0, σ=1). The PDFs in panels (a) and (b) are identical, but panel (b) zooms in on one part of the PDF for clarity.

The t PDF approaches the standard normal PDF (μ=0, σ=1) as ν increases (Fig.A3b). Equivalently and conversely, large t values become increasingly likely as sample size decreases. Although the effect of ν may appear small in Fig.A3, consider the following numerical results: $P(x>3.0)$=0.020 when ν=4, but $P(x>3.0)$=0.00384 when ν=18. This implies that Gaussian data are approximately five times more likely to produce t values larger than 3.0 for ν=4 vs. ν=18.

Last, let us consider a full numerical example, which we shall repeat with non-parametric analyses below. Imagine that an experiment yields Group A and Group B responses of {1.14, 1.21, 1.25, 1.43, 1.57} and {1.37, 1.52, 1.61, 1.74, 1.54}, respectively. A two-sample independent t test (ν=8) yields t=2.378. From Eqns.D.2 and A.2 we may conclude that Gaussian data are expected to produce a t value this large with a probability of p=0.022 over many random samplings.

To summarize, the t statistic's PDF (Eqn.A.2) is completely specified by one parameter: ν, and that PDF is derived from the Gaussian PDF (Eqn.A.1), which is also parametric. More generally, parametric procedures use a small number of parameters to specify both the PDF from which experimental data are assumed to have been randomly drawn, and the test statistic PDF upon which statistical inference is based.

Non-parametric PDFs

Non-parametric PDFs are identical to parametric PDFs in the sense that they describe the behavior of randomly sampled data. The main difference is that non-parametric PDFs generally make no assumptions regarding the distribution from which data are drawn, and instead build PDFs empirically, directly from experimental data. If the underlying data are in fact Gaussian distributed, then non-parametric PDFs converge to parametric PDFs (Fig.A4) and non-parametric results converge to parametric results (Appendix E). If experimental data deviate from Gaussian behavior then parametric approaches based on the Gaussian PDF (like the t PDF) are generally not valid.

To emphasize these points it is sufficient to describe one non-parametric approach to PDF construction. Below we describe a simple two-sample permutation procedure similar to the one used in the main manuscript, but somewhat different from the one-sample procedure described in Appendix E . Returning to the numerical example above, the two-sample permutation approach starts by labeling the original data as follows:

Label	A	A	A	A	A	B	B	B	B	B
Value	1.14	1.21	1.25	1.43	1.57	1.37	1.52	1.61	1.74	1.54

As we saw before, this particular labeling (AAAAA–BBBBB) yields t=2.378. To build the permutation PDF, we simply permute these ten labels and recompute the t statistic for each permutation. For example, labels of BAAAA–ABBBB and BBAAA–AABBB yield t=1.208 and t=0.154, respectively. Repeating for many or all label permutations builds a permutation PDF (Fig.A4). In this example there are ten labels, but once we choose positions for the five A labels, the positions of the five B labels are decided. There are thus $\binom{10}{5} = 10!/(5!5!) = 252$ unique permutations. Assembling all or a large number of permutation t values forms a permutation PDF (or empirical PDF), from which probability values can be computed as follows:

$$P(t \geq u) = \frac{\text{Number of permutation values greater than or equal to } u}{\text{Number of permutations}} \tag{A.3}$$

Since this example has 252 permutations, the minimum possible p value is $1/252 = 0.004$. Of those 252 permutations, this example yields a total of eight which satisfy $t \geq u$, including a maximum t value of 4.804 for a labeling of: AAAAB–ABBBB. Thus the p value is 8/252=0.0318, which is similar to the parametric p value of 0.022. This indirectly suggests that the parametric approach's assumption of normality is a reasonable one.

Which p value is correct, the parametric or non-parametric one? Both are correct, but their meanings are different. The interpretation of the parametric p value is as follows: if there were truly no difference between Groups A and B and if the population data are Gaussian distributed then a t value as large as the observed value would be expected in 2.2% of an infinite number of identical experiments. The interpretation of the non-parametric p value is: if there were truly no difference between Groups A and B and the group labels were assigned randomly to the data then only 3.18% of relabelings would yield a t value as large as the observed value. The primary difference between the two approaches is thus that the parametric p value assumes that the population distribution is Gaussian but the non-parametric p value does not.

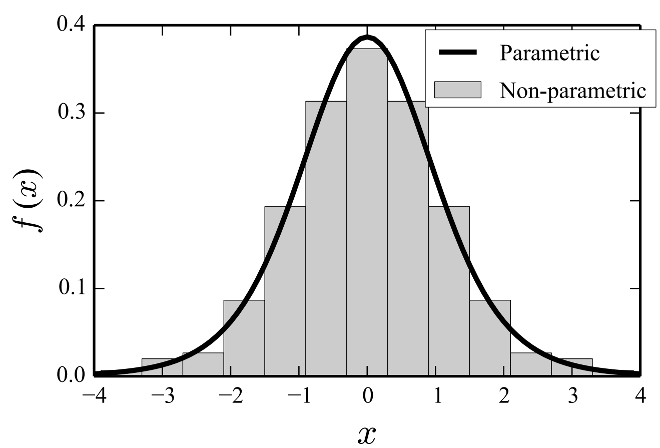

Figure A4: Comparison of parametric and non-parametric PDFs for the two-sample t test example described in the text. Here ν=8 completely parameterizes the parametric PDF. The non-parametric PDF is a histogram of the t values computed from all 252 permutations.