

1. skewed data; but our axes stretch to cover the large values that make up a relatively small proportion of the data. This is also means that our eye focuses on those values too.
Does not change the ranking, but change the distance
Good situation: ratio is a better match for our "perceptual" distance

3. ???

4. density $h_k = \text{\#observations in bin } k / w * n$

5. Pdf: $f(x) = P(X=x)$

Cdf: $F(x) = P(X \leq x)$

CLT

Then the **central** limit theorem says that if the original distribution has mean μ_{true} and variance τ_{true}^2 , then the distribution of \bar{X} for a sample of size n will be approximately

$$N(\mu_{true}, \frac{\tau_{true}^2}{n})$$

LLN: $\bar{X} \rightarrow \text{mean}$ when n gets large

6. Density hist is an estimation of pdf Kernel density function: In statistics, kernel density estimation (KDE) is a non-parametric way to estimate the probability density function of a random variable

Does $\hat{p}(x)$ form a proper density, i.e. is the area under its curve equal 1? We can answer this question by integrating $\hat{p}(x)$

$$\begin{aligned} \int_{-\infty}^{\infty} \hat{p}(x) dx &= \int_{-\infty}^{\infty} \frac{1}{n} \sum_{i=1}^n f(x, x_i) dx \\ &= \frac{1}{n} \sum_{i=1}^n \int_{-\infty}^{\infty} f(x, x_i) dx \end{aligned}$$

So if $\int_{-\infty}^{\infty} f(x, x_i) dx = 1$ for any x_i , we will have,

$$\int_{-\infty}^{\infty} \hat{p}(x) dx = \frac{1}{n} \sum_{i=1}^n 1 = 1.$$

}

Is this the case? Well, considering $f(x, x_i)$ as a function of x with a fixed x_i value, it is equal to $1/w$ when x is within $w/2$ of x_i , and zero otherwise (i.e. the same function as before, but now centered at x_i):

7. p-value : the probability (under the null distribution) of randomly getting a statistic X under the null hypothesis *as extreme as or more extreme* than the statistic we observed in our data

Power: $1 - \text{type-II}$. (set smaller significance level to increase power)

type-I error: reject a true H_0

type-II error: fail to reject a false H_0

8. Assumption: SRS

Consequence of violating assumption: p-value could be wrong

duality between Hypothesis tests and Confidence Intervals: CI gives a range of “acceptable” values of the population parameter, which consists of the values that would not be rejected by a hypothesis test.

$$|Z| = \frac{|\bar{X} - \bar{Y}|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}},$$

significance level: same as type I error

now we have

$$|T| = \frac{|\bar{X} - \bar{Y}|}{\sqrt{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}}}.$$

10.

Assumption: Good estimates of \hat{F} , \hat{G}

A big assumption of the bootstrap is that our sample distribution \hat{F} , \hat{G} is a good estimate of F and G . We've already seen that will not necessarily be the case. Here are some examples of why that might fail:

- Sample size n_1/n_2 is too small
- The data is not a SRS

Assumption: Data generation process

Another assumption is that our method of generating our data X_i^* , and Y_i^* matches the way X_i and Y_i were generated from F , G . In particular, in the bootstrap procedure above, we are assuming that X_i and Y_i are i.i.d from F and G (i.e. a SRS with replacement).

Assumption: Well-behaved test statistic

We also need that the parameter θ and the estimate $\hat{\theta}$ to be well behaved in certain ways

- $\hat{\theta}$ needs to be an **unbiased** estimate of θ , meaning across many samples, the average of the $\hat{\theta}$ is equal to the true parameter θ ²⁵
- θ is a function of F and G , and we need that the value of θ changes smoothly as we change F and G . In other words if we changed from F to F' , then θ would change to θ' ; we want if our new F' is very “close” to F , then our new θ' would be very close to θ . This is a pretty mathematical requirement, and requires a precise definition of “close” for two distributions that is not too important for this class to understand.

11. Steps: we need to randomly draw a permutation. We'll do this many times (i.e. many calls to the function `sample`), and this will create a SRS of permutations. Once we have a SRS of permutations, we can calculate the test statistic for each permutation, and get an estimate of the true null distribution.

Assumptions: randomness; randomly assigned labels model

Why permutation: much less assumptions than its alternatives

12. if each of these tests has level 0.05, then even if none of the pairs are truly different from each other, I might expect on average around 2 to be rejected at level 0.05 just because of variation in sampling

Specifically, say we will quantify our notion of `stricter` to require "of all the tests I ran, there's only a 5% chance of a type I error". Let's make this a precise statement. Suppose that of the K tests we are considering, there are $V \leq K$ tests that are type I errors, i.e. the null is true but we rejected. We will define our cumulative error rate across the set of K tests as

$$P(V \geq 1)$$

So we if we can guarantee that our testing procedure for the set of K tests has $P(V \geq 1) \leq \gamma$ we have controlled the **family-wise error rate** to level γ .