

MATH 203: PRINCIPLES OF STATISTICS 1

FALL 2024

LARGE SAMPLE TESTING CONSISTENCY

For a two sample problem, to test

$$H_0 : \mu_1 = \mu_2$$

$$H_a : \mu_1 \neq \mu_2$$

we compare the observed value of test statistic

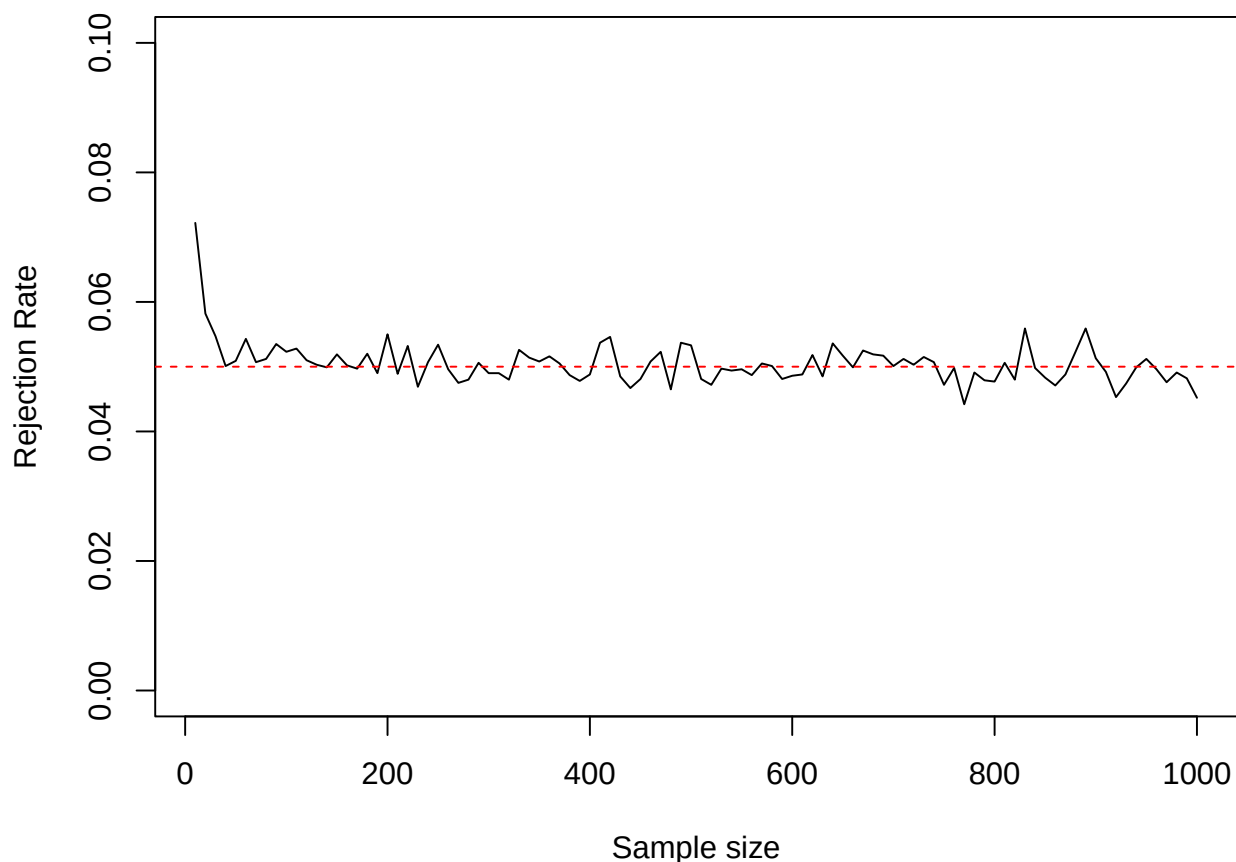
$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}}}$$

with the Rejection Region

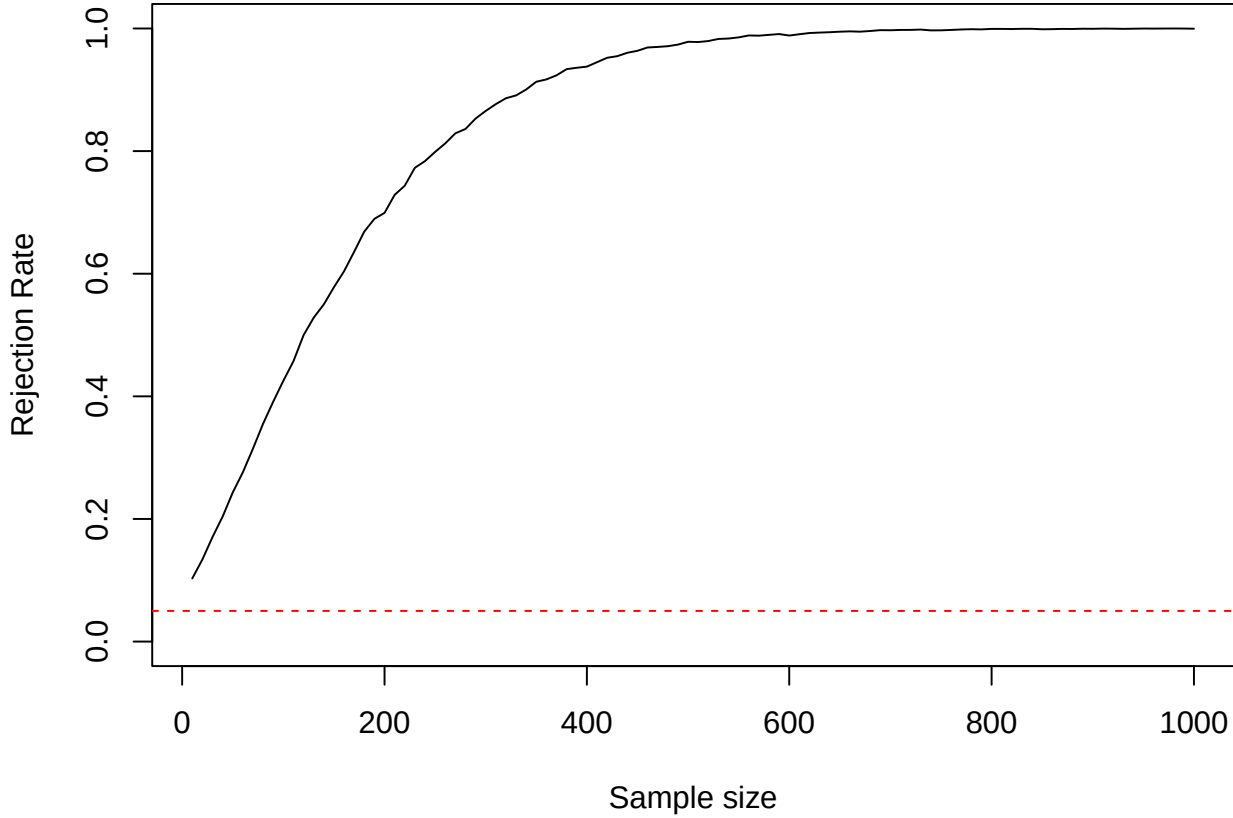
$$\{|T| \geq z_{\alpha/2}\} = \{T \leq -z_{\alpha/2}\} \cup \{T \geq z_{\alpha/2}\}$$

If H_0 is **true**, we see in the plot of the average rejection rate over $N = 10000$ replicate data sets that as the sample size grows, the probability with which we reject H_0 converges to α . This is by design, as α is the probability of a Type 1 error. With $\alpha = 0.05$, the Rejection Region is

$$\{|T| \geq 1.96\}$$



However, if H_0 is not true, we will eventually reject the null hypothesis in favour of the alternative. In the simulation, $\mu_1 = 0$ and $\mu_2 = 0.25$, and $\sigma_1 = \sigma_2 = 1$, with the data Normally distributed. For simplicity, the two sample sizes are equal, with $m = n$ taking values between 10 and 1000.



In this problem, we have in reality that

$$\bar{X} - \bar{Y} \sim Normal\left(\mu_1 - \mu_2, \left(\frac{1}{m} + \frac{1}{n}\right)\right) = Normal\left(0.25, \frac{2}{n}\right).$$

as $m = n$. Thus, using the z-transform approach, we have

$$Z \sim \frac{(\bar{X} - \bar{Y}) - 0.25}{\sqrt{2/n}} = \sqrt{\frac{n}{2}}((\bar{X} - \bar{Y}) - 0.25) \sim Normal(0, 1).$$

However, the null hypothesis assumes (at least approximately) that

$$\bar{X} - \bar{Y} \sim Normal\left(\mu_1 - \mu_2, \left(\frac{1}{m} + \frac{1}{n}\right)\right) = Normal\left(0, \frac{2}{n}\right).$$

so that

$$T = \frac{(\bar{X} - \bar{Y})}{\sqrt{1/m + 1/n}} = \sqrt{\frac{n}{2}}(\bar{X} - \bar{Y}) \sim Normal(0, 1)$$

so we have the relation

$$Z = T - 0.25\sqrt{\frac{n}{2}}$$

which shows that Z becomes arbitrarily large and negative as n grows, and so eventually it falls into the Rejection Region.