Statistical Hypothesis Testing, Statistical Decision Making, & Bayes' Rule

Null Hypothesis Significance Testing

- Model the distribution of a test statistic $T(x)$ under the assumption that the null hypothesis is true.
- Find a critical value of $T_{crit}(x)$ such that more extreme values will have low probability (e.g., p value less than 0.05).
- Compare observed $|T_{obs}(x)|$ to critical $|T_{crit}(x)|$, or compare p_{obs} to p_{crit}.
- If $|T_{obs}(x)| > |T_{crit}(x)|$ or $p_{obs} < p_{crit}$, reject the null hypothesis. Otherwise, fail to reject the null.
- Optional: Model the distribution of $T(x)$ under a specific alternative hypothesis, and use this model and $|T_{crit}(x)|$ to calculate the probability of a statistically significant result assuming the alternative is true (this is called power).

NHST: Test Statistics

- Model the distribution of a test statistic *T(x)* under the assumption that the null hypothesis is true.
  - The test statistic will depend on study design and properties of the measurements of independent and dependent variables.
  - e.g., if you have a nominal independent variable with two levels and a continuous dependent variable (and various other assumptions hold), you would likely calculate a *t* statistic.
  - e.g., if you have a nominal independent variable with more than two levels and a continuous dependent variable (and various other assumptions hold), you would likely calculate an *F* statistic.

NHST: Critical Values, Type I Error

- Find a critical value of $T_{crit}(x)$ such that more extreme values will have low probability (e.g., p value less than 0.05).

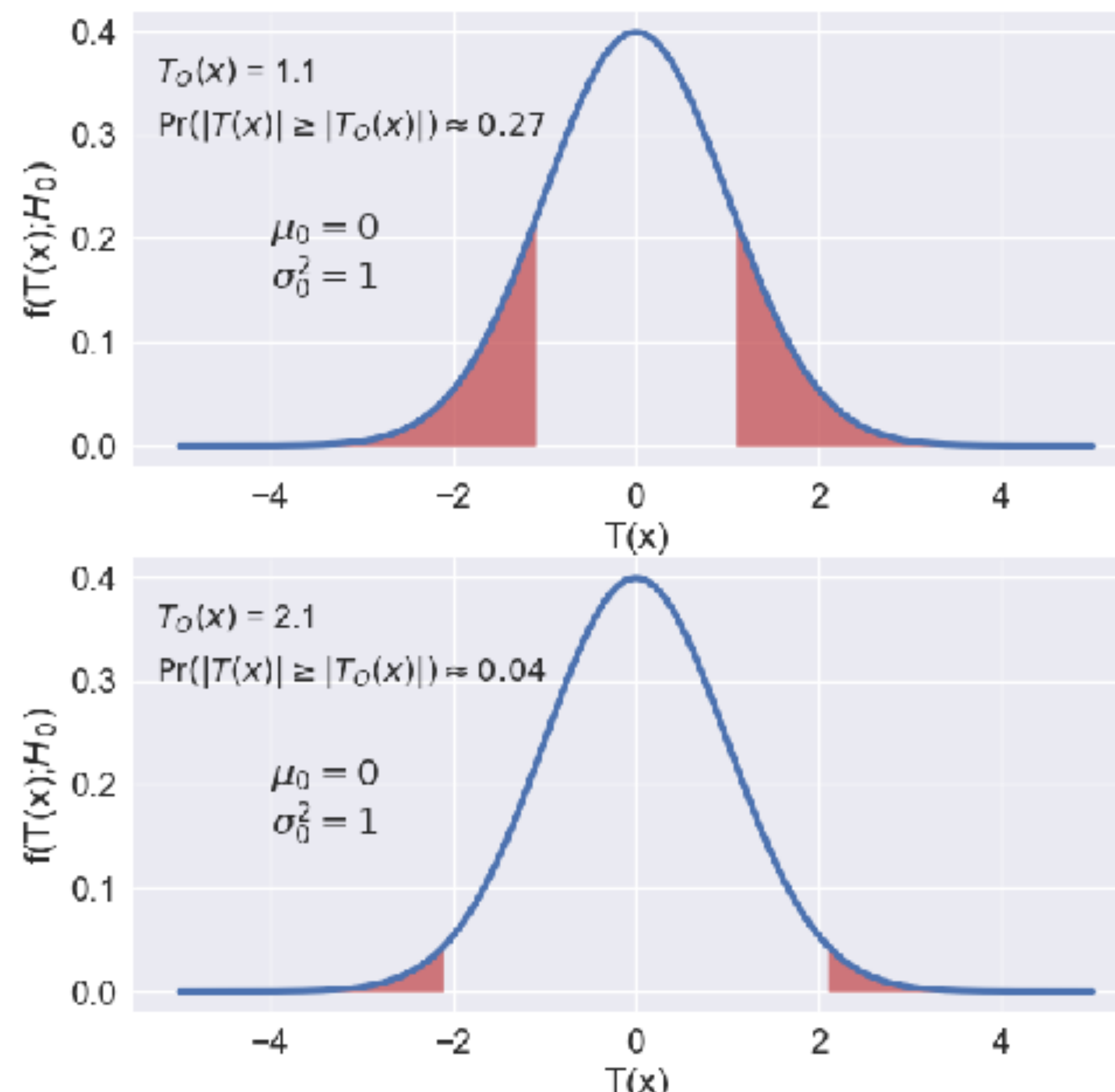
- Compare observed $|T_{obs}(x)|$ to critical $|T_{crit}(x)|$, or compare p_{obs} to p_{crit}.

- The critical probability used to determine $T_{crit}(x)$ is denoted α.

- α determines your long-run false alarm (FA, or Type I error) rate.

- Type I error: probability of rejecting the null when it's true.

- In one approach to NHST, you only consider H_0 (the null).



NHST: Logical Asymmetry

- If $|T_{obs}(x)| > |T_{crit}(x)|$ or $p_{obs} < p_{crit}$, reject the null hypothesis. Otherwise, fail to reject the null.

- It is very important to keep this logical asymmetry in mind!

  - If you observe an extreme test statistic, the rules of NHST allow you to reject the null hypothesis. Strictly speaking, this is a fallacy:

    - **Valid**: 1) If P, then Q. 2) Not Q. 3) Therefore, not P.
    - **Not Valid**: 1) If P, then probably Q. 2) Probably not Q. 3) Therefore, (probably) not P.
    - e.g., If H_0 is true, then Intervention X will have no effect. Intervention X (seems to have) had an effect ($M_A > M_B$). Therefore, H_0 is (probably) not true.

  - If you observe a not-so-extreme test statistic, the rules of NHST only allow you to *fail to reject* the null.

NHST: Type II Error, Power

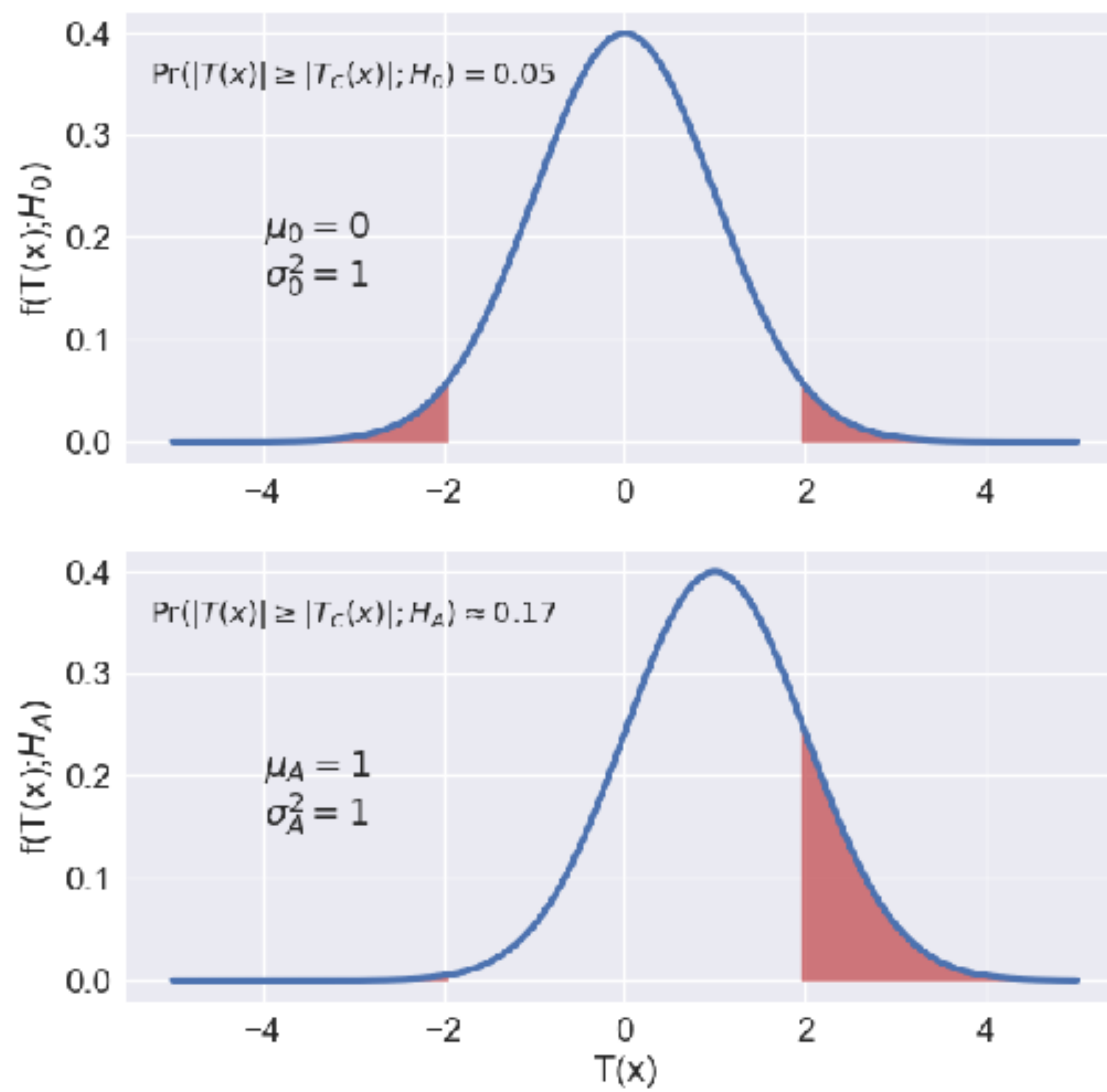
- Optional: Model the distribution of $T(x)$ under a specific alternative hypothesis H_A, and use this model and $|T_{crit}(x)|$ to calculate the probability of statistical significance assuming H_A is true (i.e., power).

- Type II Error: Failing to reject the null when it's false.

- β = probability of Type II Error

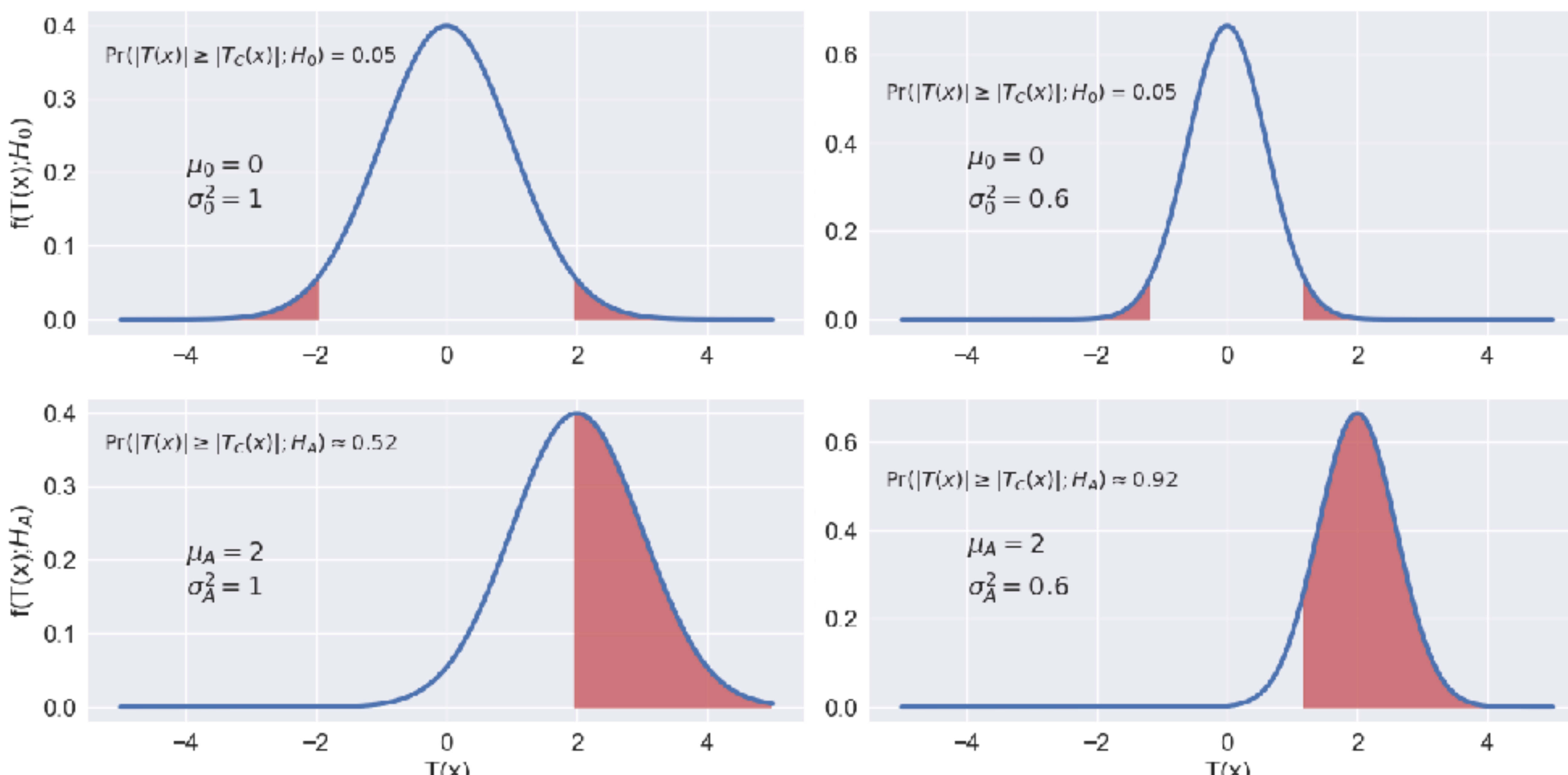
- 1-β = power

- You have to have a specific alternative hypothesis to calculate power.



NHST: More Power

- A number of factors influence statistical power, including, but not limited to, better measurement, more powerful manipulations/interventions (i.e., larger effect size), larger samples.



Statistical Decision Making

	Lack of Statistical Significance	Statistical Significance
H_0 True	True Negative	False Positive (Type I Error)
H_0 False	False Negative (Type II Error)	True Positive

- Accuracy: True Negative + True Positive / Total
- Prevalence: H_0 False/Total

- Hit Rate: True Positive/H_0 False
- FA Rate: False Positive/H_0 True
- Sensitivity = Hit Rate
- Specificity: 1 - FA Rate

- Positive Predictive Value: True Positive / Stat. Sig.
- False Discovery Rate: False Positive / Stat. Sig.

Bayes' Rule

- NHST is concerned with $\Pr(|T_{obs}(x)| > |T_{crit}(x)| ; H_0)$, i.e., α, and with $\Pr(|T_{obs}(x)| > |T_{crit}(x)| ; H_A)$, i.e., $1-\beta$.
  - H_0 and H_A are considered non-random properties of the universe
  - A (very) common misinterpretation of p values is that they mean $\Pr(H_0 \mid |T_{obs}(x)| > |T_{crit}(x)|)$ rather than $\Pr(|T_{obs}(x)| > |T_{crit}(x)| ; H_0)$

- That is, people give p values a Bayesian interpretation.

- Here's Bayes' rule:

  - $\Pr(A|B) = \Pr(B|A)\Pr(A) / [\Pr(B|A)\Pr(A) + \Pr(B|{\sim}A)\Pr({\sim}A)]$

  - Pr(A|B): probability of A given B. Pr(A): prior probability of A

  - In NHST terms, a p value gives us just Pr(B|A), but people often misinterpret it as Pr(A|B).

Bayes' Rule, NHST, & False Discovery

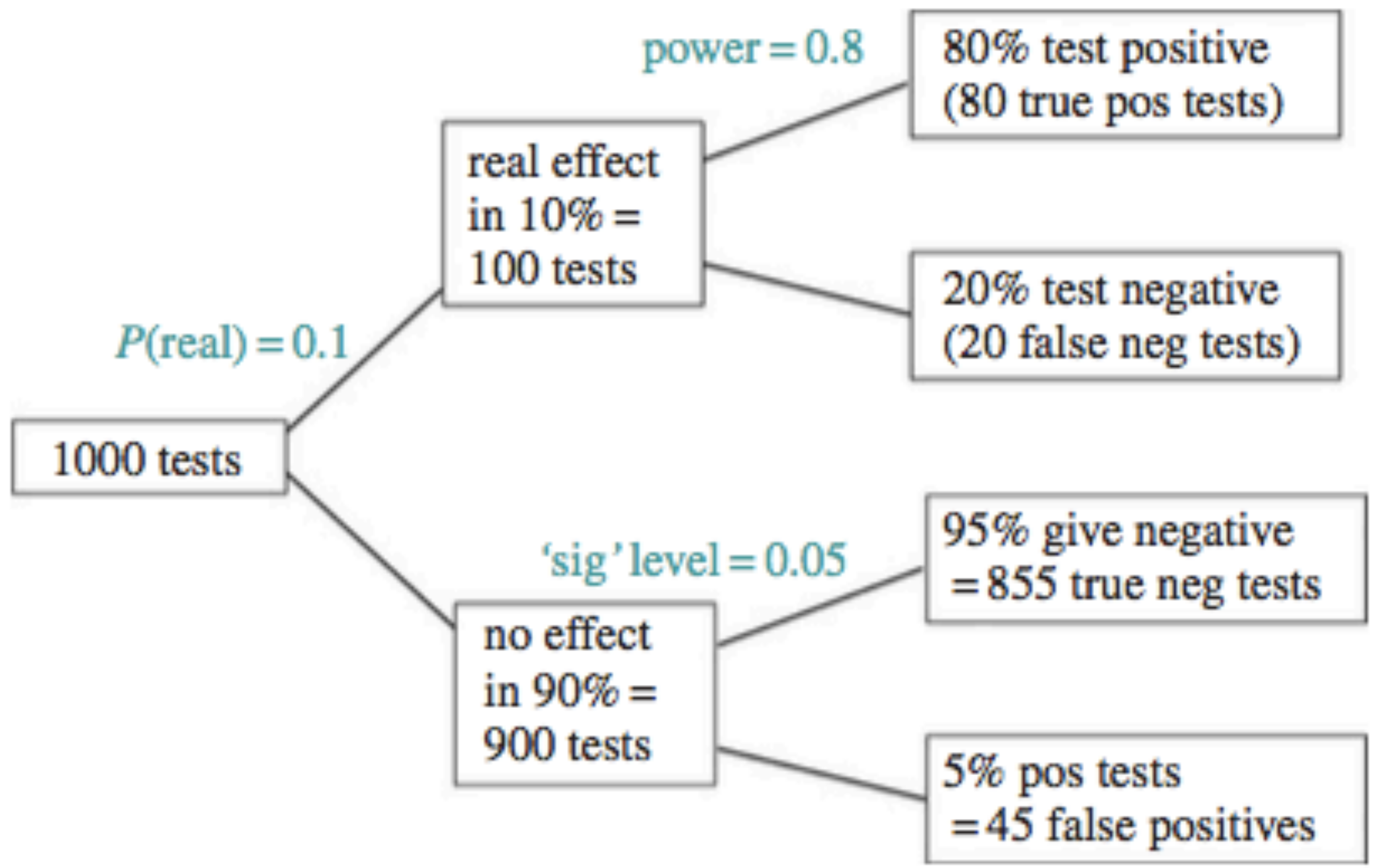


Figure 2. Tree diagram to illustrate the false discovery rate in significance tests. This example considers 1000 tests, in which the prevalence of real effects is 10%. The lower limb shows that with the conventional significance level, $p = 0.05$, there will be 45 false positives. The upper limb shows that there will be 80 true positive tests. The false discovery rate is therefore $45/(45 + 80) = 36\%$, far bigger than 5%.

- This figure is from Colquhoun (2014), the fourth paper we'll be reading. Note that the false discovery rate also gives the (posterior) probability of the null being true given that you've observed a statistically significant result, i.e.,
  - $Pr(H_0|p < 0.05) = Pr(p < 0.05|H_0)Pr(H_0)/[\ Pr(p < 0.05|H_0)Pr(H_0) + Pr(p < 0.05|{\sim}H_0)Pr({\sim}H_0)\ ]$