Statistical Tests

primitive → complicated → easy

Prof. Dr. Uwe Saint-Mont

HOCHSCHULE NORDHAUSEN
University of Applied Sciences

Frankfurt, 19. September 2018

Basic Setup

Two tiers

General: $X \sim P_H$ (hypothetical law, hypothesis)

Specific: x (observation)

Fisher (1935): Every experiment may be said to exist only to give the facts a chance of disproving the null hypothesis.

Evaluating Observations

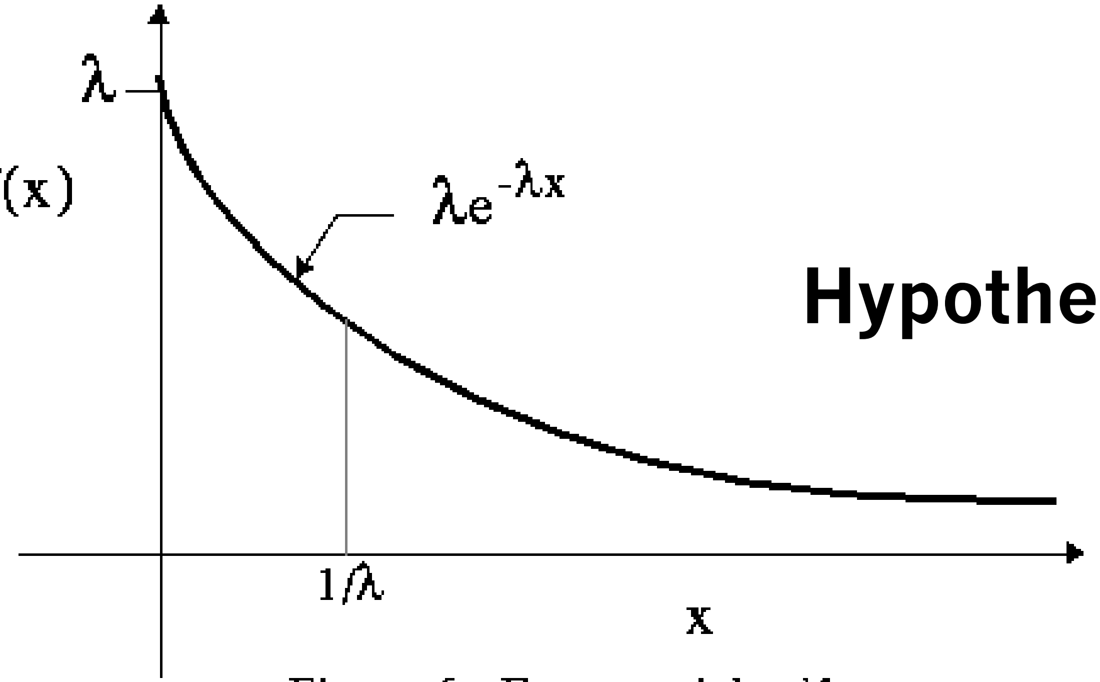


Figure 6. Exponential *pdf*

Hypothesis H

Evaluating Observations

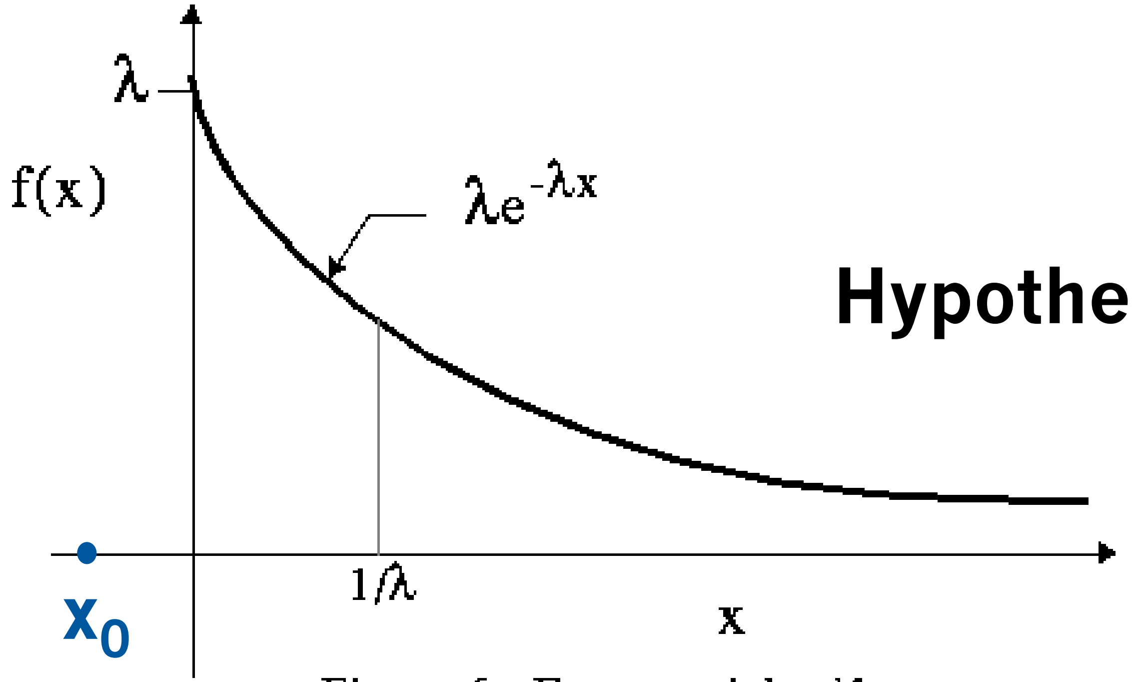


Figure 6. Exponential *pdf*

Hypothesis H

Impossible observation x_0

⇨ hypothesis must be wrong

Evaluating Observations

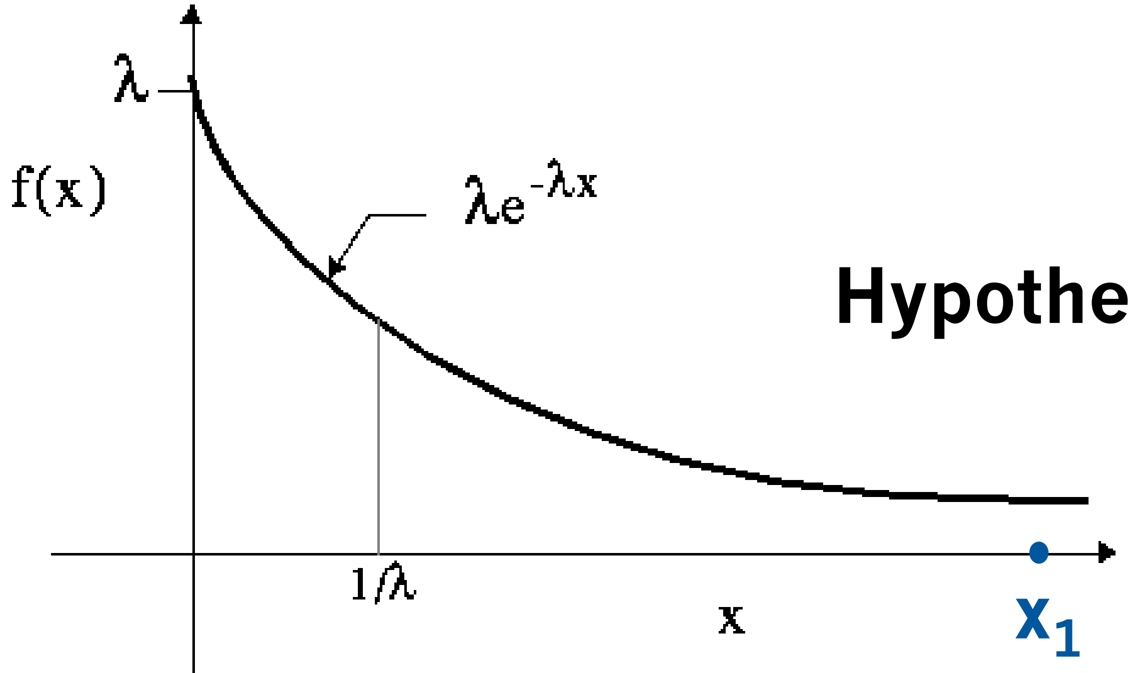


Figure 6. Exponential *pdf*

Hypothesis H

Possible observation x_1 ⇨ hypothesis may be wrong ⇨ rejection of H only "plausible"

Evaluating Observations

Two kinds of implausible observations:

$P_H(x_1)$ small, i.e., "an exceptionally rare chance has occurred" (Fisher 1956)

x_1 is an outlier, e.g., rather large for the hypothesis under consideration

Fisher‘s posterior test

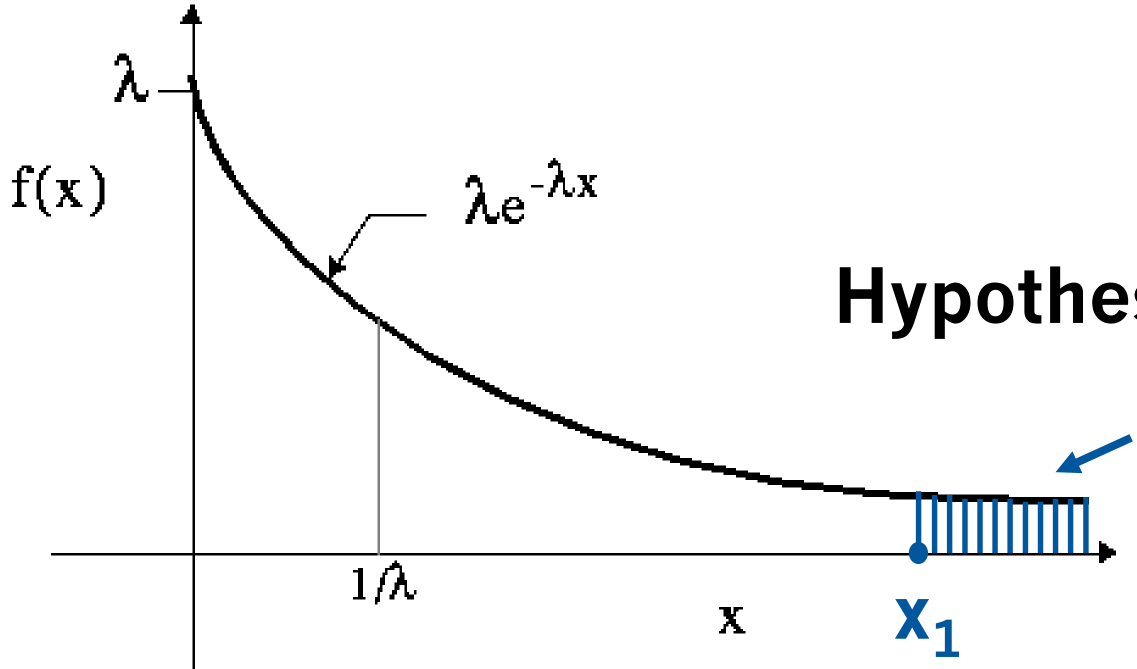


Figure 6. Exponential *pdf*

Hypothesis H

Given observation x_1 ⇨ compute $p = P_H(X \geq x_1)$

$$p = \int_{x_1}^{\infty} f(x)dx$$

Crucial Objection

Mathematically, the observed value x_1 is treated like all unobserved values $x > x_1$

Thus: An hypothesis that may be true is rejected because it has failed to predict observable results that have not occurred. This seems a remarkable procedure. (Jeffreys 1939)

Neyman&Pearson‘s prior test

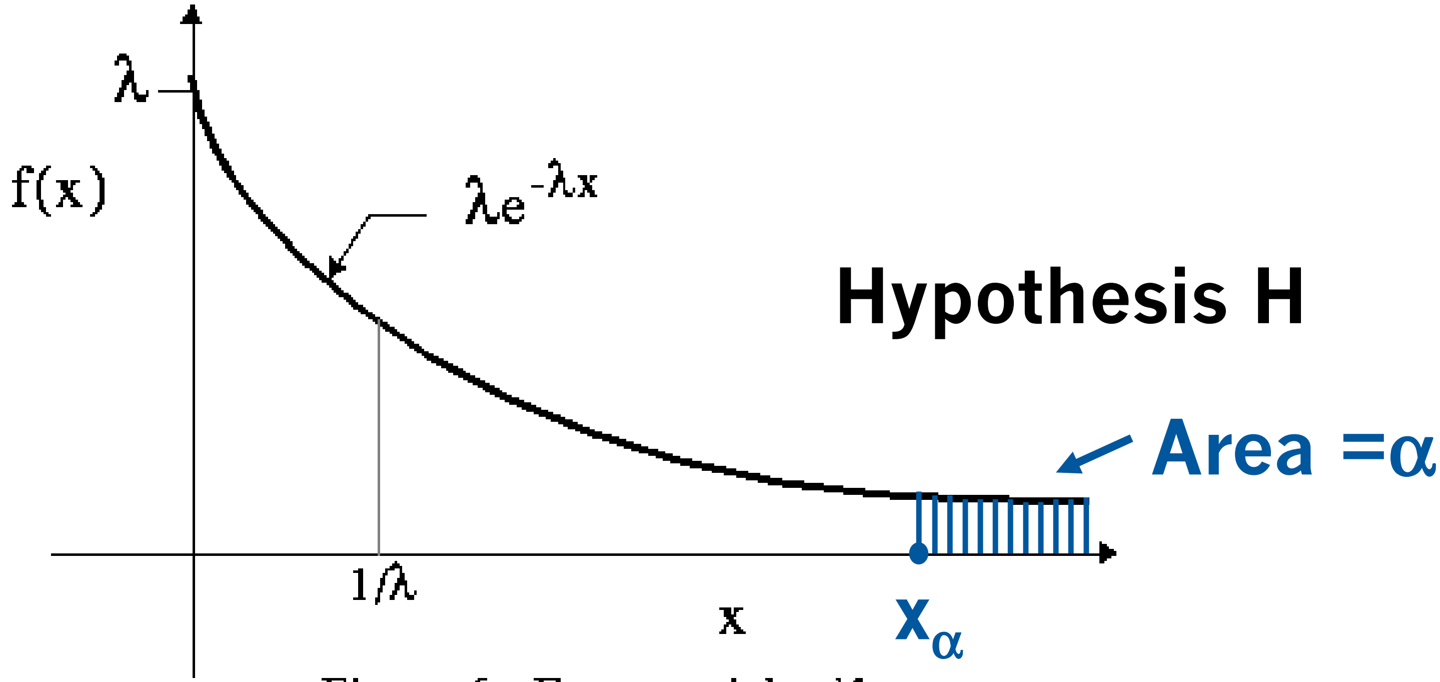


Figure 6. Exponential *pdf*

Given H, fix α such that $\alpha = \int_{x_\alpha}^{\infty} f(x)\,dx$

Reject H if $x > x_\alpha$ and do not reject H if $x \leq x_\alpha$

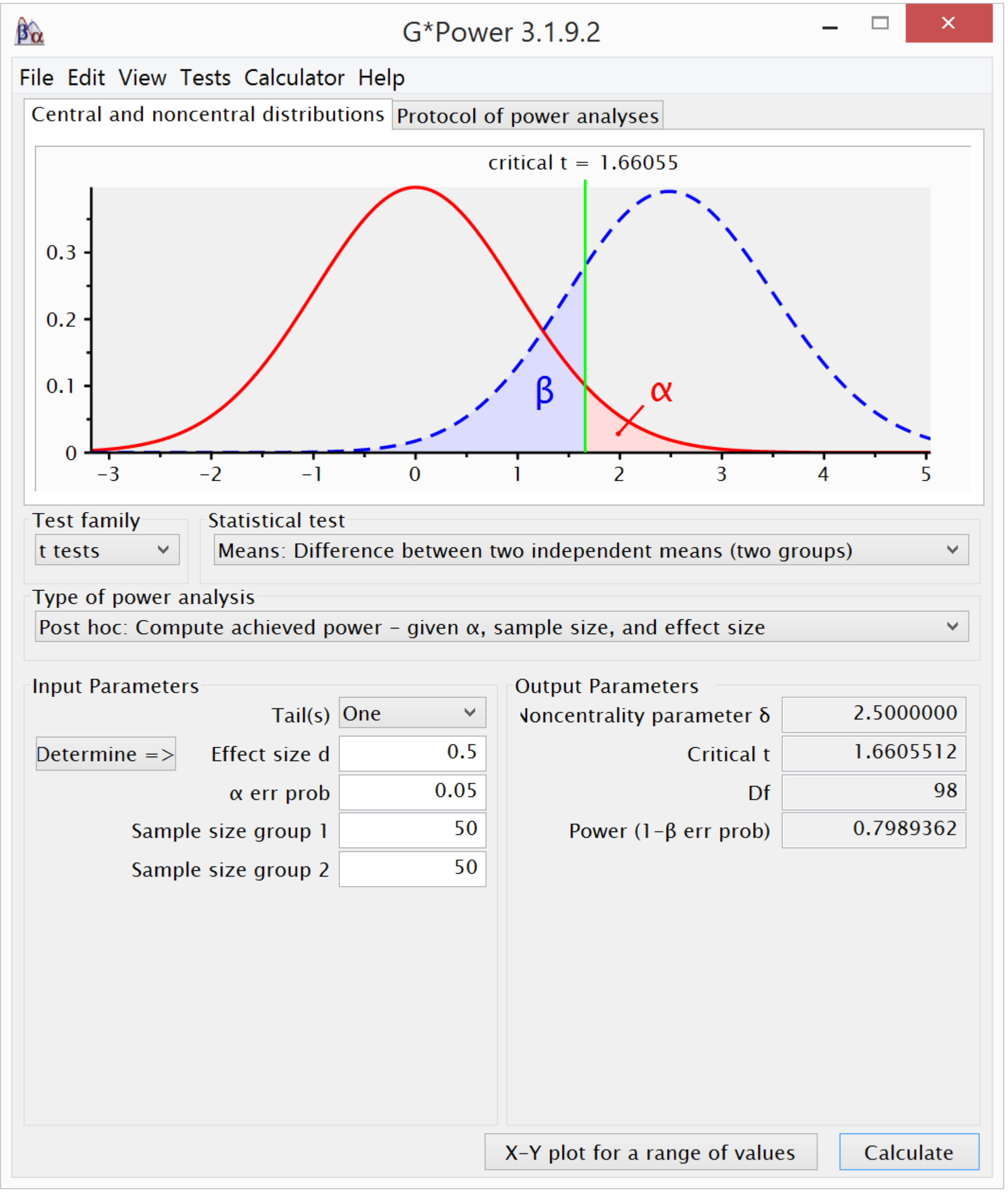
Objections

- Strict prior view ⇨ Dichotomous decision

- Information in the data is neglected

- Although the approaches are very different conceptionally, they are almost identical mathematically ⇨ Confusion

N&P's standard refined test



Given two hypotheses and the prior view, one has to deal with probabilities of error (α, β), the effect size, and sample sizes

Objections

- **Although complicated, the result is just an "optimal" dichotomous decision**

Objections

- **Although complicated, the result is just an “optimal” dichotomous decision**

- **Invites “cherry picking”, some kind of “ritual” (Gigerenzer), and further confusion: “alphabet soup” (Hubbard)**

Objections

- **Although complicated, the result is just an "optimal" dichotomous decision**

- **Invites "cherry picking", some kind of "ritual" (Gigerenzer), and further confusion: "alphabet soup" (Hubbard)**

- **Many further "side effects", in particular the idea that the smaller the sample the better. (Saint-Mont 2018)**

Main Objection

- Why use integrals as “proxies”?
- Given a hypothesis H and an alternative K, there is a
- Straightforward comparison:

$P_H(x)$ vs. $P_K(x)$

Likelihood: Results

- **Likelihood-ratio test (optimal):**

  decide in favour of H if $P_H(x) > P_K(x)$

  decide in favour of K if $P_K(x) > P_H(x)$

Likelihood: Results

- **Likelihood-ratio test (optimal):**

  decide in favour of H if $P_H(x) > P_K(x)$

  decide in favour of K if $P_K(x) > P_H(x)$

- **Bayes test: prior probs. of the hypos. Q_H, Q_K**

  decide in favour of H if $Q_H \, P_H(x) > Q_K \, P_K(x)$

  decide in favour of K if $Q_K \, P_K(x) > Q_H \, P_H(x)$

Testing: Look at the Distance

- **Straightforward link to information theory:**
  KL-divergence $D(K||H) = \int_{-\infty}^{\infty} P_K(x) log \frac{P_K(x)}{P_H(x)} dx$

- **Hoeffding's universal test (1965)**
  $P(X^n)$ empirical distribution of iid sample
  decide in favour of H if $D(P(X^n)||H) < c_n$

Elegant General Theory

- Likelihood function if there are many hyps.: decide in favour of the hypothesis H with the highest probability (maximum likelihood)

- Close connection with general criteria: Maximum Entropy, Minimum Message Length, Minimum Description Length, AIC, BIC,...

References

Fisher, R.A. (1935). The Design of Experiments.

Fisher, R.A. (1956). Statistical Methods and Scientific Inference.

Hoeffding, W. (1965). Asymptotically optimal tests for multinomial distributions. Ann. Math. Stat. 36, 369-408.

Jeffreys, H. (1939). Theory of Probability.

Pearson, E. S. (1955). Statistical Concepts and their Relation to Reality. J. Royal Stat. Society, Ser. B 17(2), 204-207.

Saint-Mont, U. (2018). Where Fisher, Neyman and Pearson went astray: On the logic (plus some history and philosophy) of statistical tests. Adv. Soc. Sci. Research J. 5(8), 672-691.

Small Samples seem to be optimal

- **E.S. Pearson (1955):**

The appropriate test is one which, while involving through the choice of its significance level α only a very small risk of discarding my working hypothesis H prematurely will enable me to demonstrate with assurance $1-\beta$ (**but without any unnecessary amount of experimentation**) the reality of the influences which I suspect may be present.