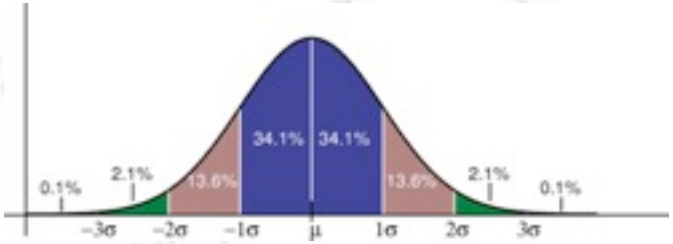
# Applied Statistics

## Hypothesis Testing

Troels C. Petersen (NBI)



*"Statistics is merely a quantization of common sense"*

# Taking decisions

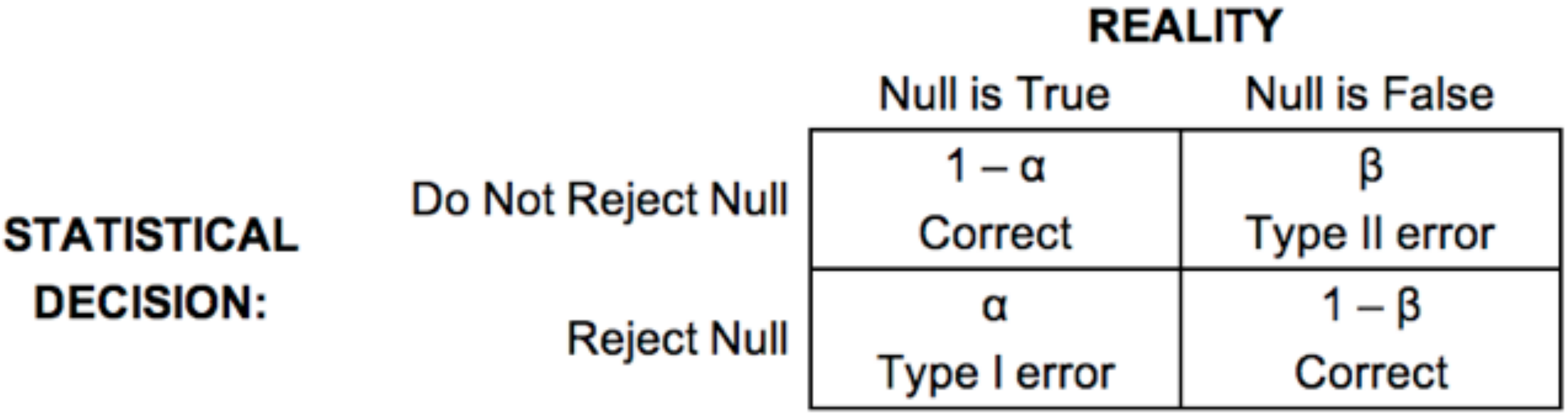
You are asked to take a decision or give judgement - it is yes-or-no.

**Given data - how to do that best?**

That is the basic question in hypothesis testing.

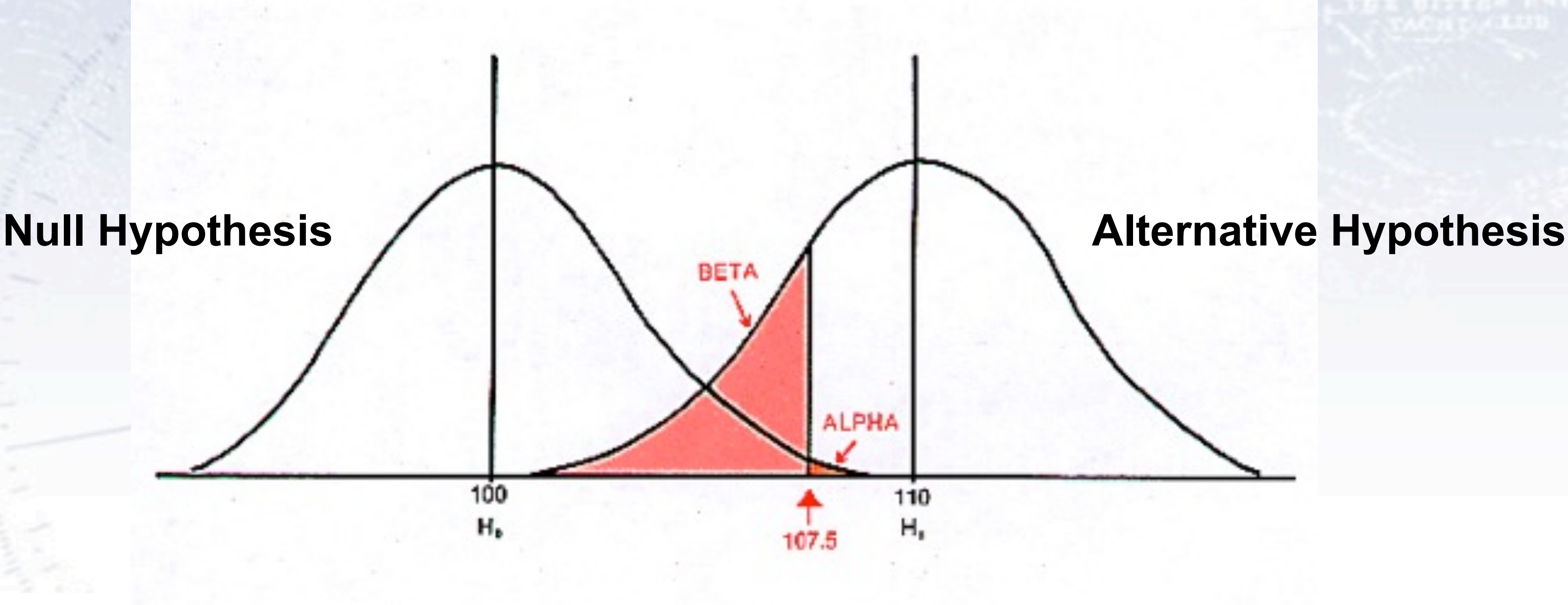
Trouble is, you may take the wrong decision, and there are TWO errors:

- The hypothesis is **true**, but you **reject** it (Type I).
- The hypothesis is **wrong**, but you **accept** it (Type II).



| | | REALITY | |
|---|---|---|---|
| | | Null is True | Null is False |
| STATISTICAL DECISION: | Do Not Reject Null | 1 – α<br>Correct | β<br>Type II error |
| | Reject Null | α<br>Type I error | 1 – β<br>Correct |

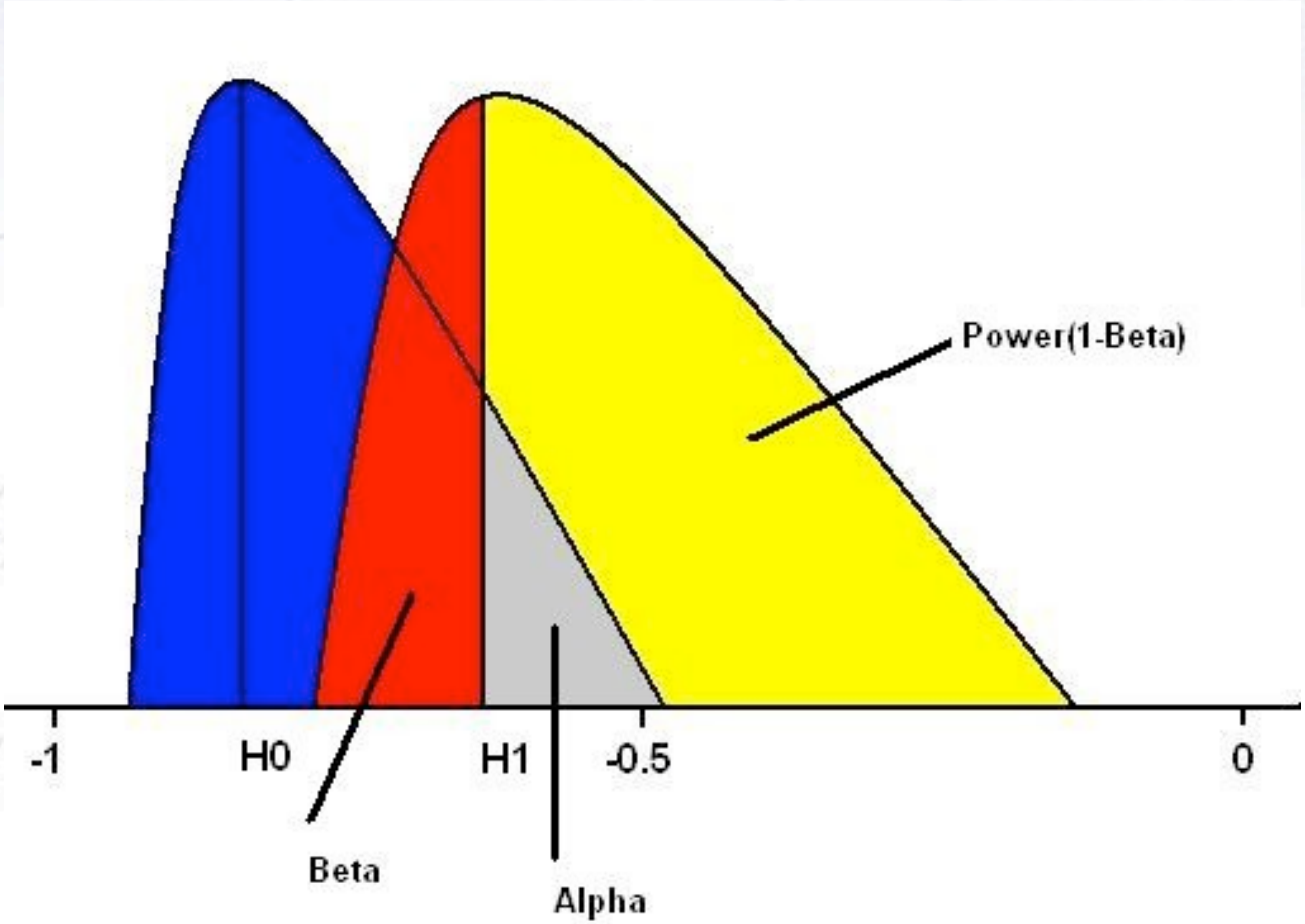
# Taking decisions



| | | REALITY | |
|---|---|---|---|
| | | Null is True | Null is False |
| STATISTICAL DECISION: | Do Not Reject Null | $1 - \alpha$<br>Correct | $\beta$<br>Type II error |
| | Reject Null | $\alpha$<br>Type I error | $1 - \beta$<br>Correct |

# Taking decisions

The purpose of a test is to yield distributions for the Null and Alternative, which are as separated from each other as possible (to minimize α and β).



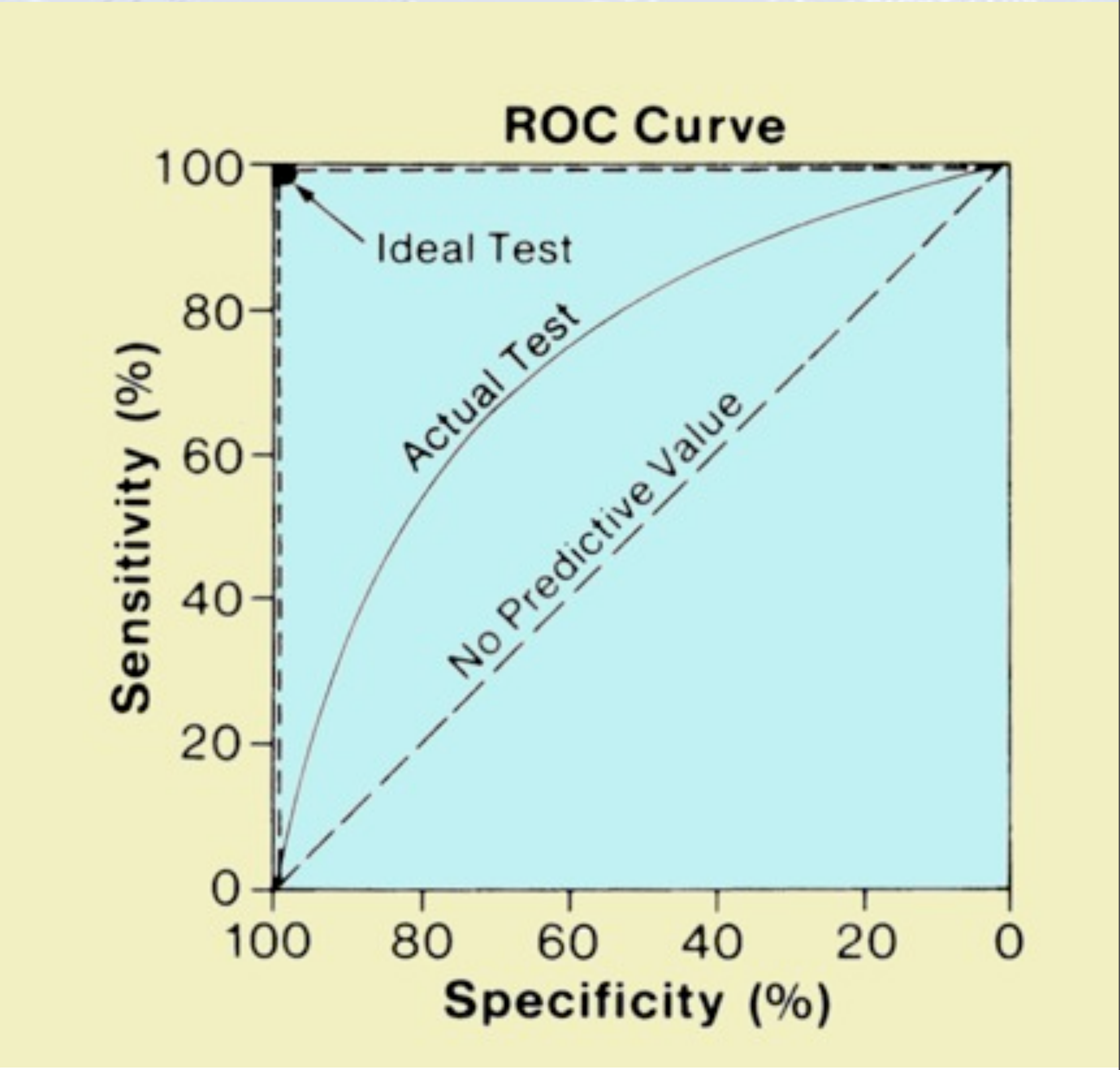
# ROC-curves

The **Receiver Operating Characteristic** or just ROC-curve is a graphical plot of the sensitivity, or true positive rate, vs. false positive rate.

It is calculated as the integral of the two hypothesis distributions, and is used to evaluate the power of a test.

Often, it requires a testing data set to actually see how well a test is performing.

It can also detect overtraining!



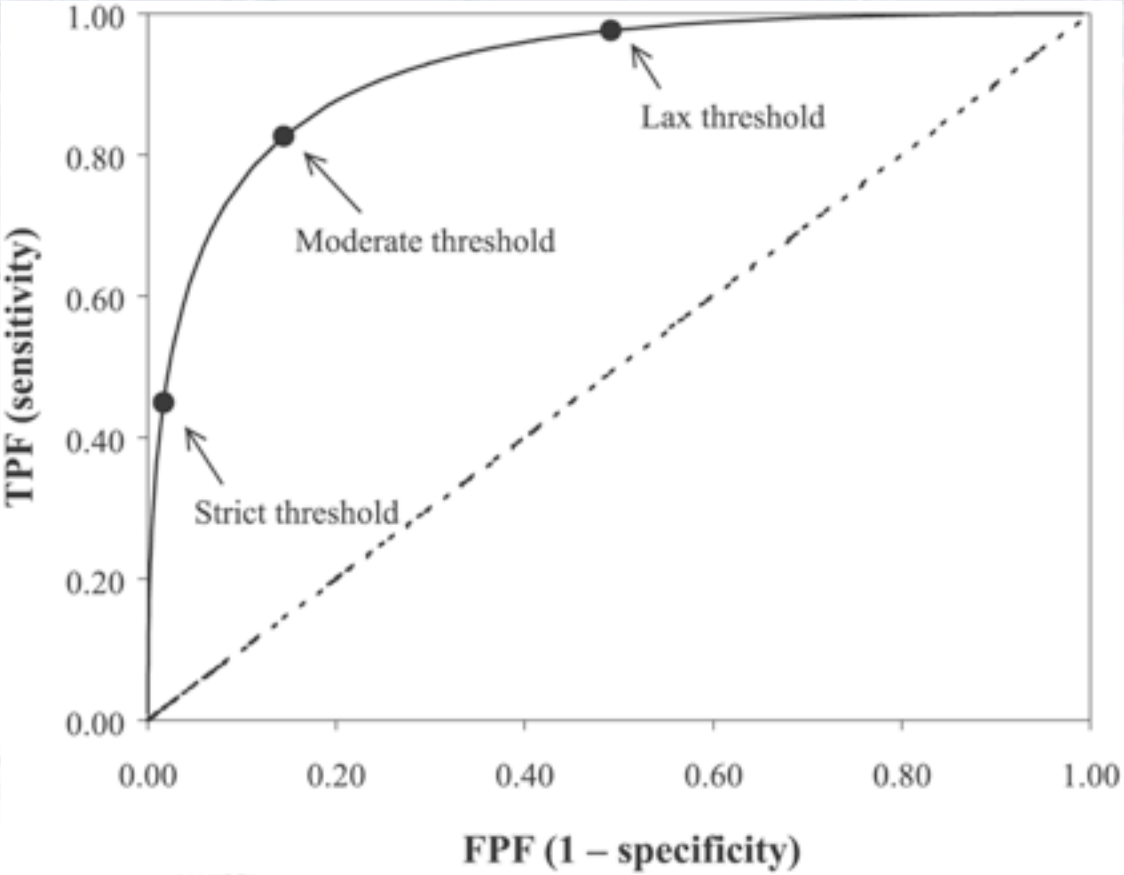
# ROC-curves

The **Receiver Operating Characteristic** or just ROC-curve is a graphical plot of the sensitivity, or true positive rate, vs. false positive rate.

It is calculated as the integral of the two hypothesis distributions, and is used to evaluate the power of a test.

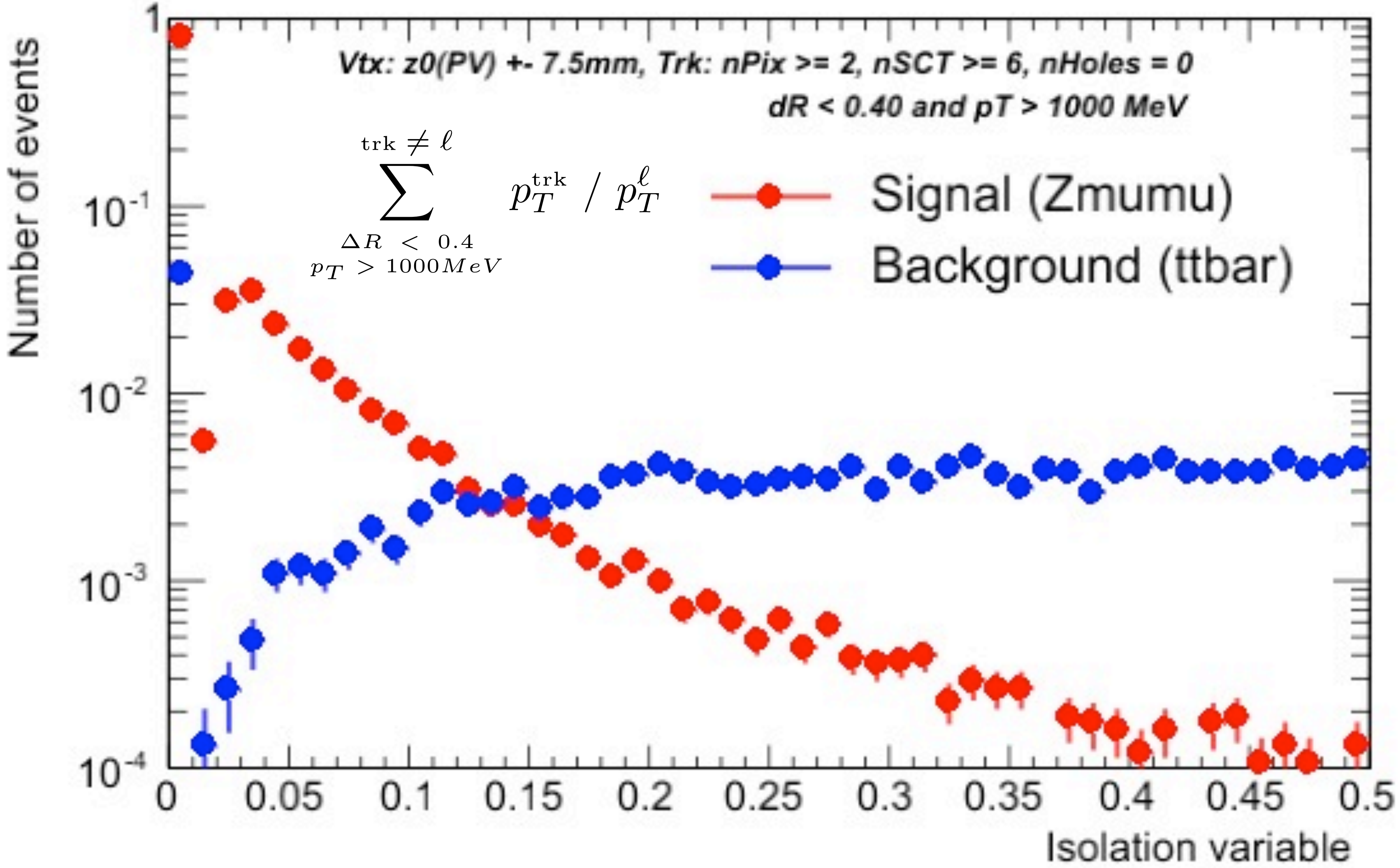
Often, it requires a testing data set to actually see how well a test is performing.

It can also detect overtraining!



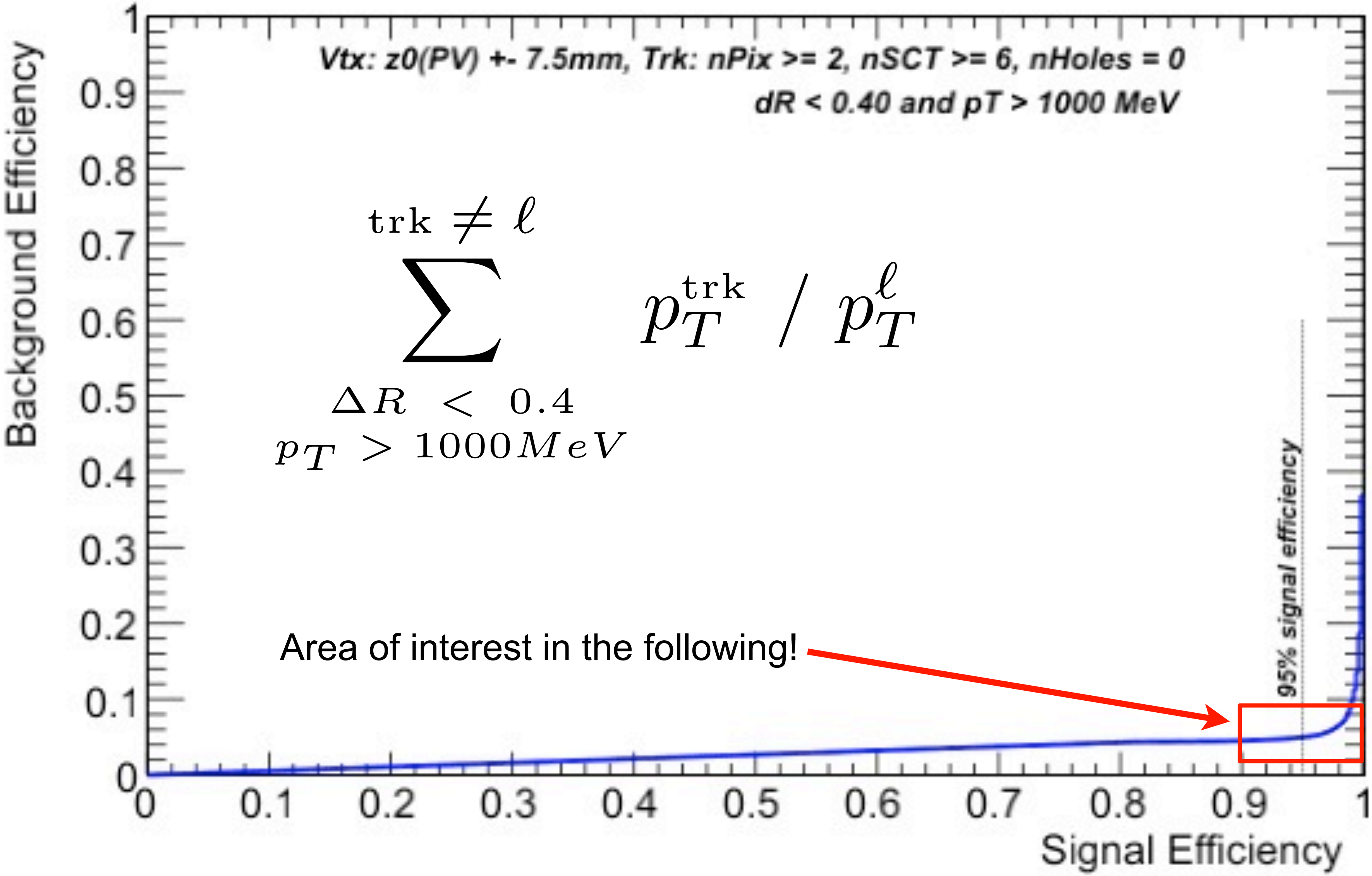
# Example of ROC curves in use

# Basic steps - distributions



$$\sum_{\substack{\Delta R\ <\ 0.4 \\ p_T\ >\ 1000MeV}}^{\mathrm{trk}\ \neq\ \ell} p_T^{\mathrm{trk}}\ /\ p_T^{\ell}$$

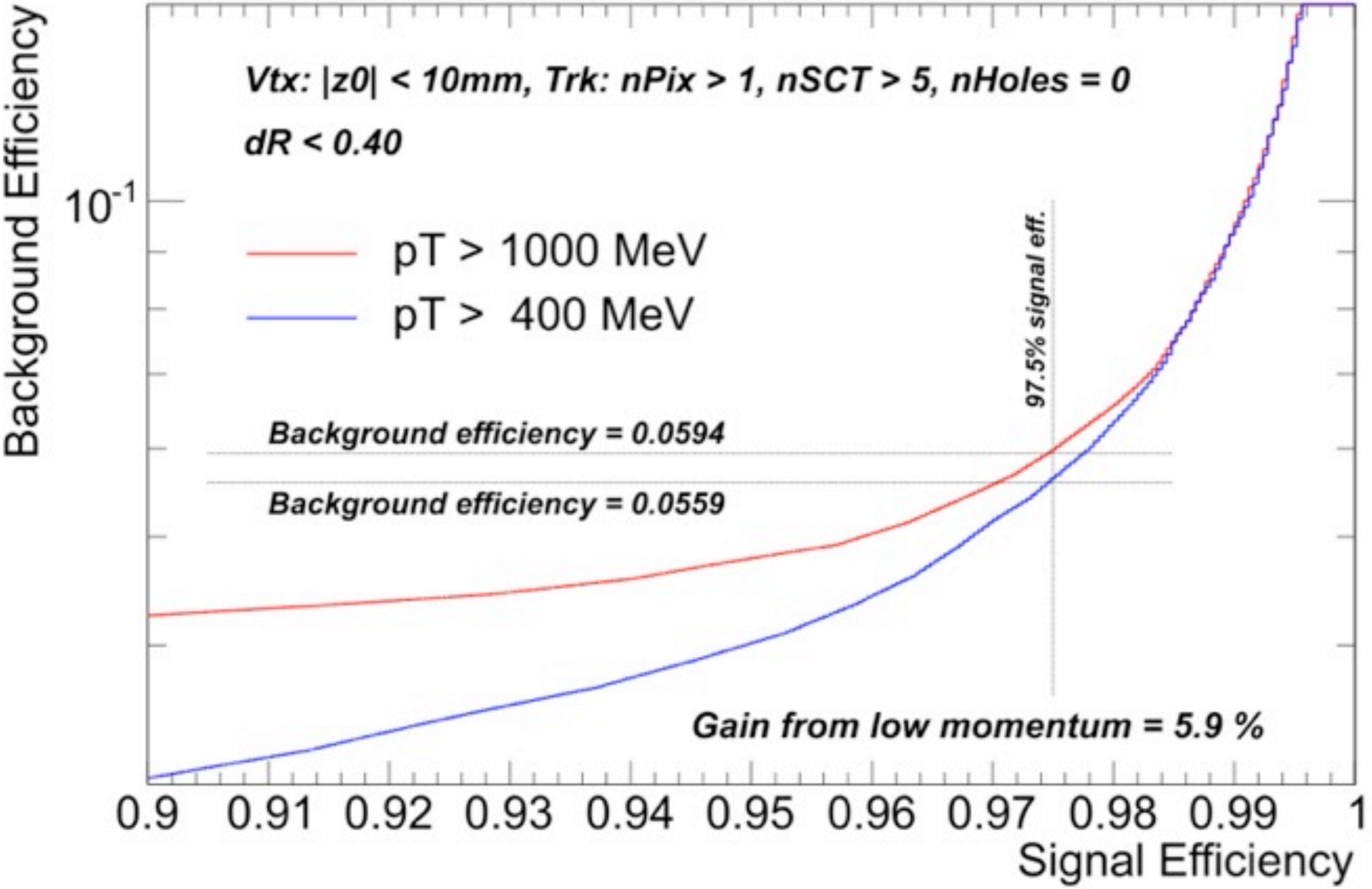
# Basic steps - ROC curves



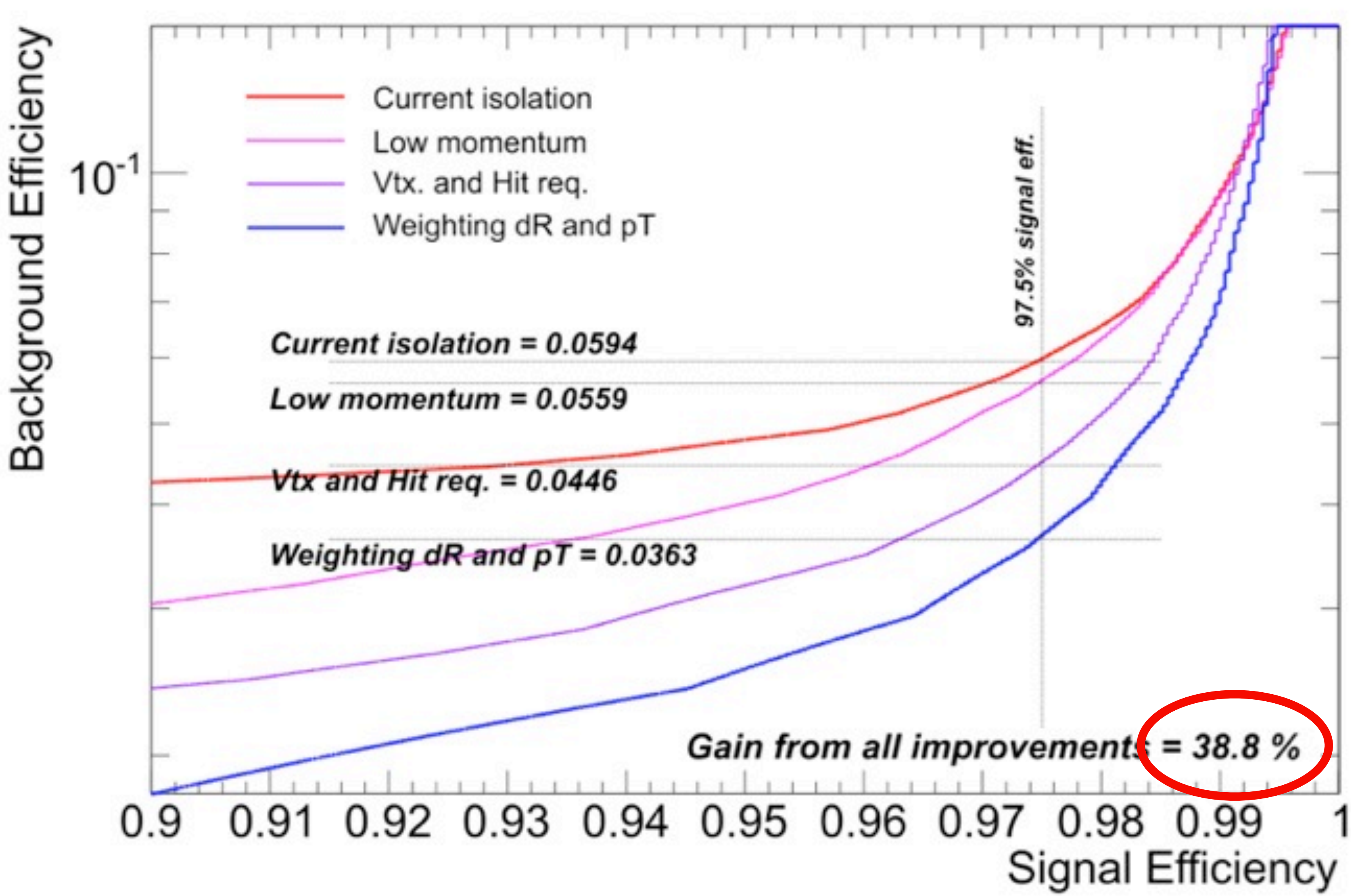
$$\sum^{\text{trk}\neq\ell}_{\substack{\Delta R \; < \; 0.4 \\ p_T \; > \; 1000 MeV}} \quad p_T^{\text{trk}} \; / \; p_T^{\ell}$$

Area of interest in the following!

# Using tracks down to 400 MeV



# Overall improvement



# Typical statistical tests

# ChiSquare test

A good example of a test is the ChiSquare test:

$$\chi^2(\bar{\theta}) = \sum_{i=1}^{N} \frac{(y_i - \lambda(x_i;\bar{\theta}))^2}{\sigma_i^2}$$

Calculating the probability from Chi2 and Ndof, this turns out to be a very good test.

**If the p-value is small, the hypothesis is unlikely...**

However, there are other (and more powerful) tests.

# Neyman-Pearson lemma

Consider a **likelihood ratio** between the null and the alternative model:

$$D = -2 \ln \left( \frac{\text{likelihood for null model}}{\text{likelihood for alternative model}} \right)$$

The Neyman-Pearson lemma (loosely) states, that this is the most powerful test there is.

In reality, the problem is that it is not always easy to write up a likelihood for complex situations!

# Wald-Wolfowitz runs test

A different test to the Chi2 (and in fact a bit orthogonal!) is the Wald-Wolfowitz runs test.

It measures the number of “runs”, defined as sequences of same outcome (only two types).

Example:

++++−−−+++−−++++++−−−−

If random, the mean and variance is known:

$$\mu = \frac{2\ N_+\ N_-}{N} + 1$$

$$\sigma^2 = \frac{2\ N_+\ N_-\ (2\ N_+\ N_- - N)}{N^2\ (N-1)} = \frac{(\mu-1)(\mu-2)}{N-1}$$

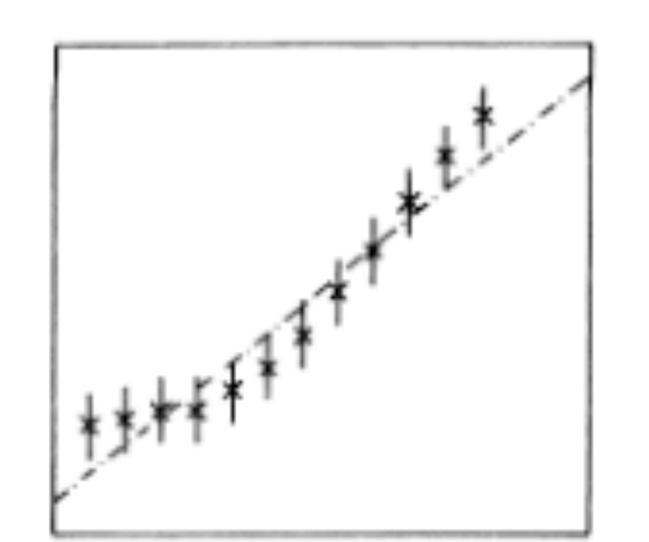


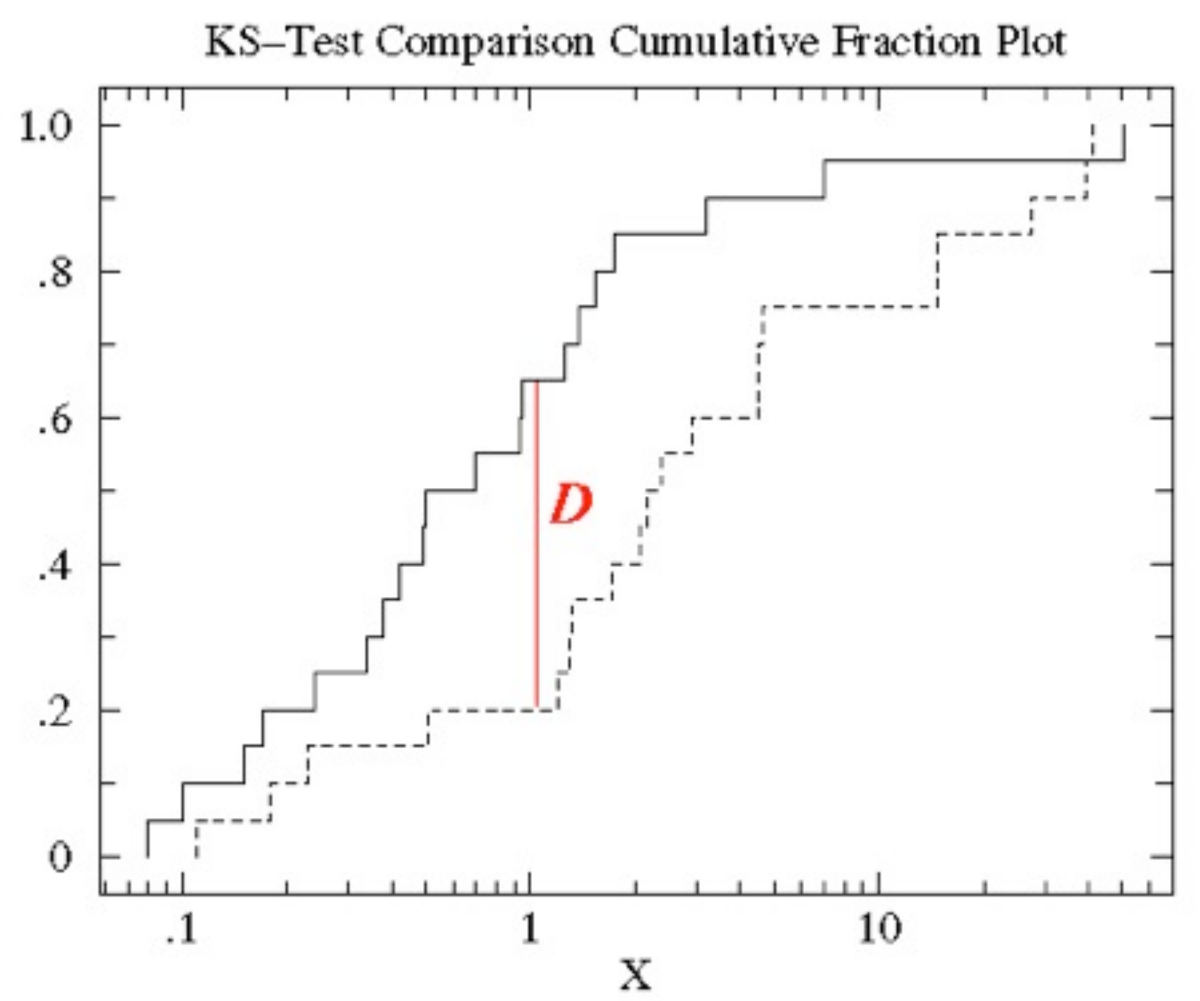
Fig. 8.3. A straight line through twelve data points.

$N = 12$, $N_+ = 6$, $N_- = 6$
$\mu = 7$, $\sigma = 1.76$
$(7-3)/1.65 = 2.4\ \sigma$ (~1%)

Note: The WW runs test requires N > 10-15 for the output to be Gaussian!

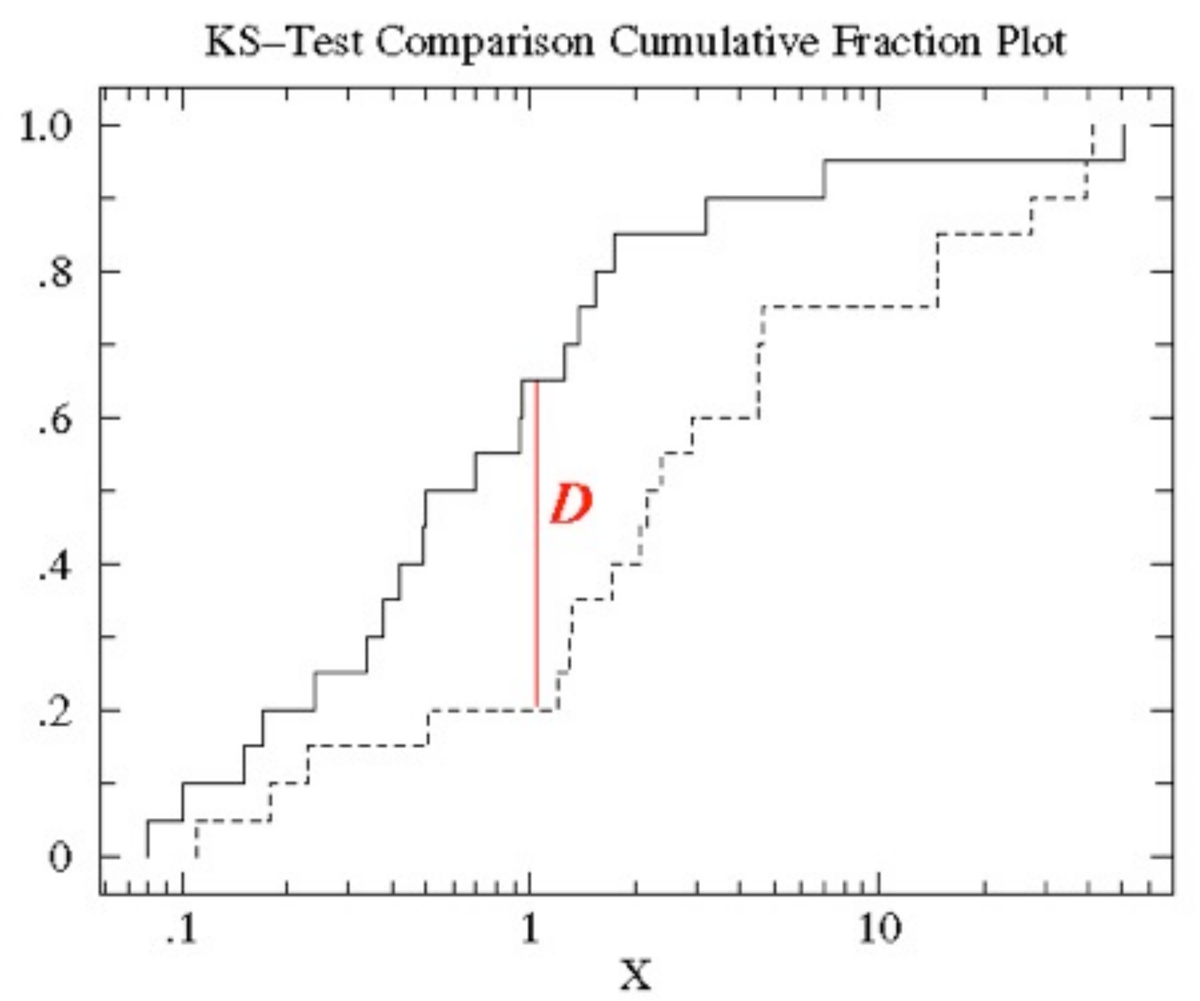
# Kolmogorov test

The Kolmogorov test measures the maximal distance between the integrals of two distributions and gives a probability of being from the same distribution.



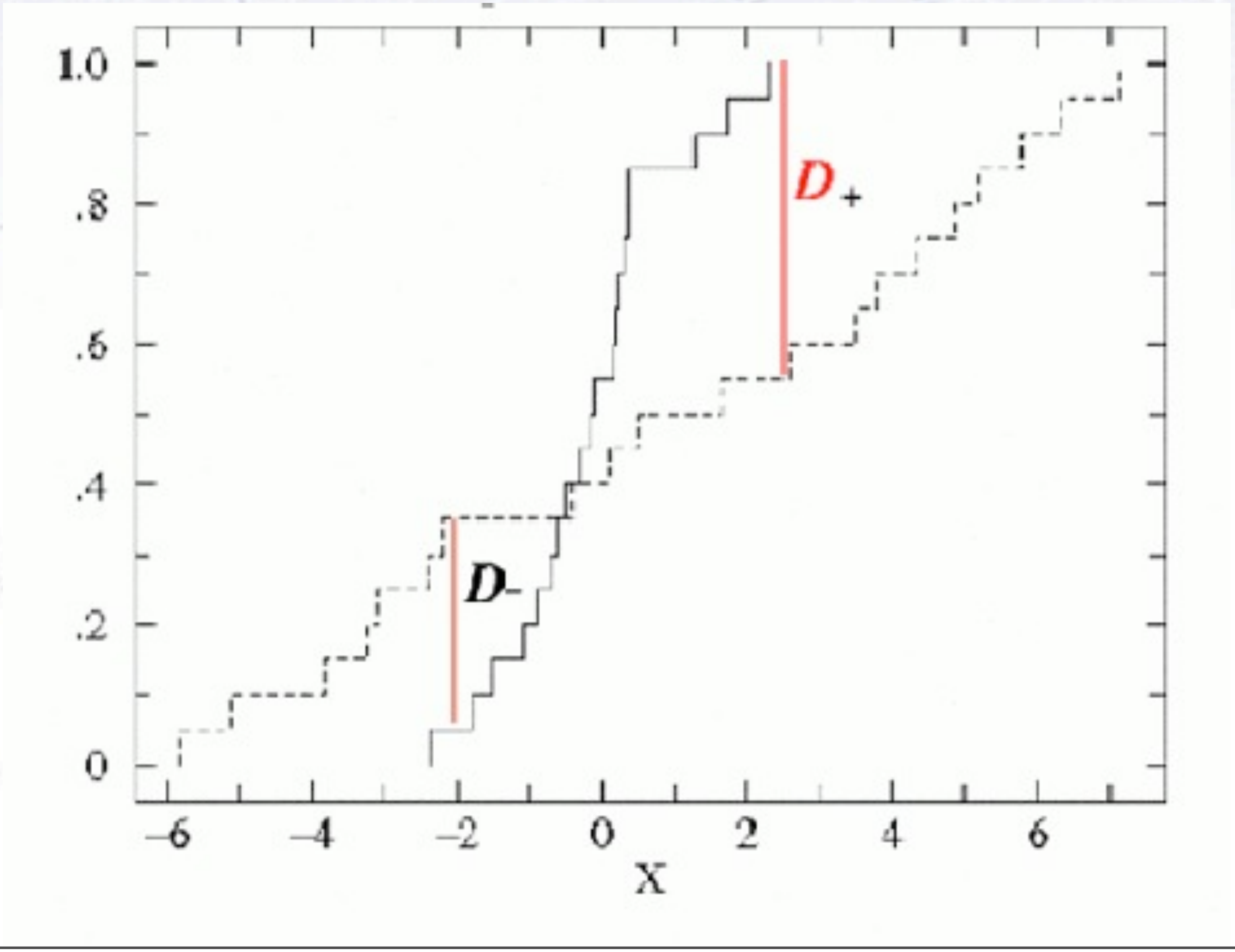
# Kolmogorov-Smirnov test

The Kolmogorov test measures the maximal distance between the integrals of two distributions and gives a probability of being from the same distribution.



# Kuiper test

Is a similar test, but it is more specialized in that it is good to detect SHIFTS in distributions (as it uses the maximal signed distance in integrals).


1.0
.8
.6
.4
.2
0
−6
−4
−2
0
2
4
6
X
$D_+$
$D_-$