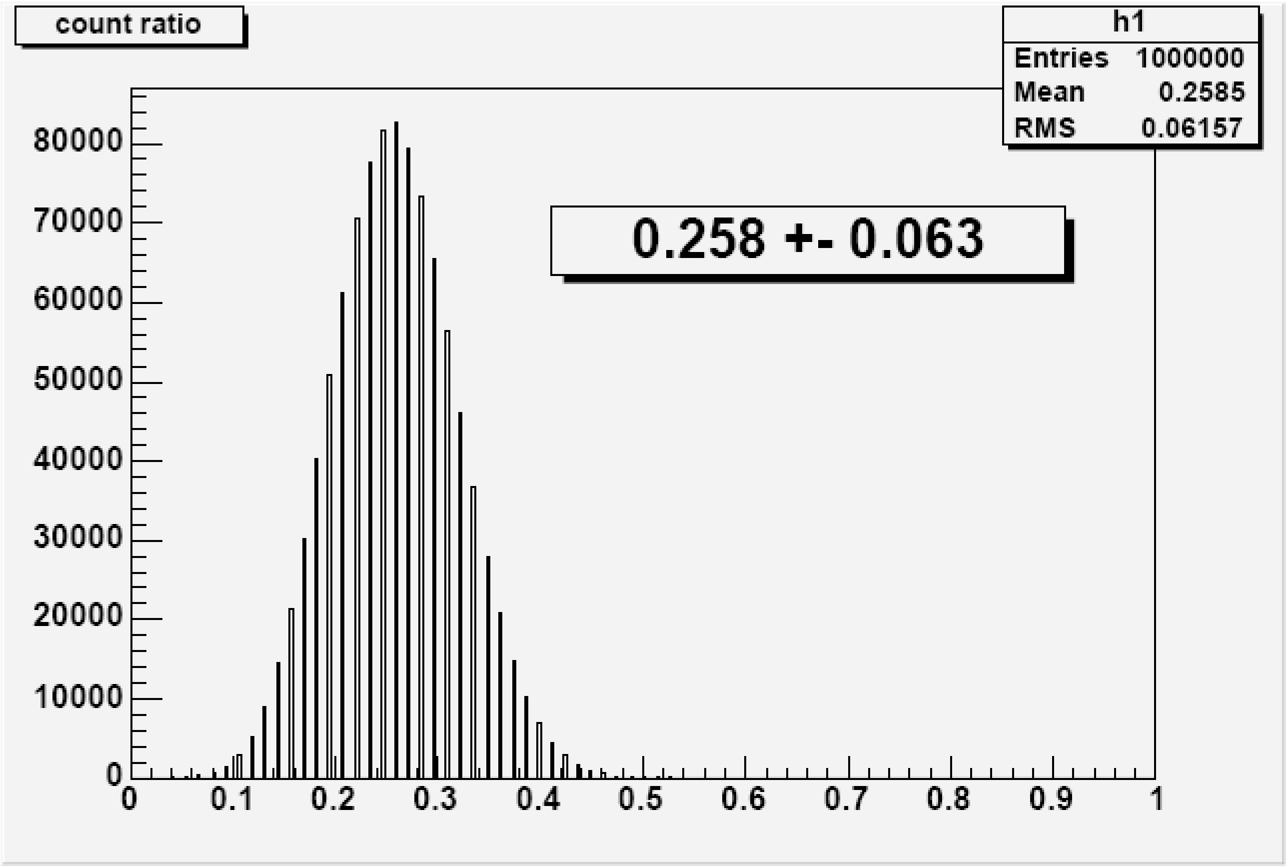
# Bootstrap Methods

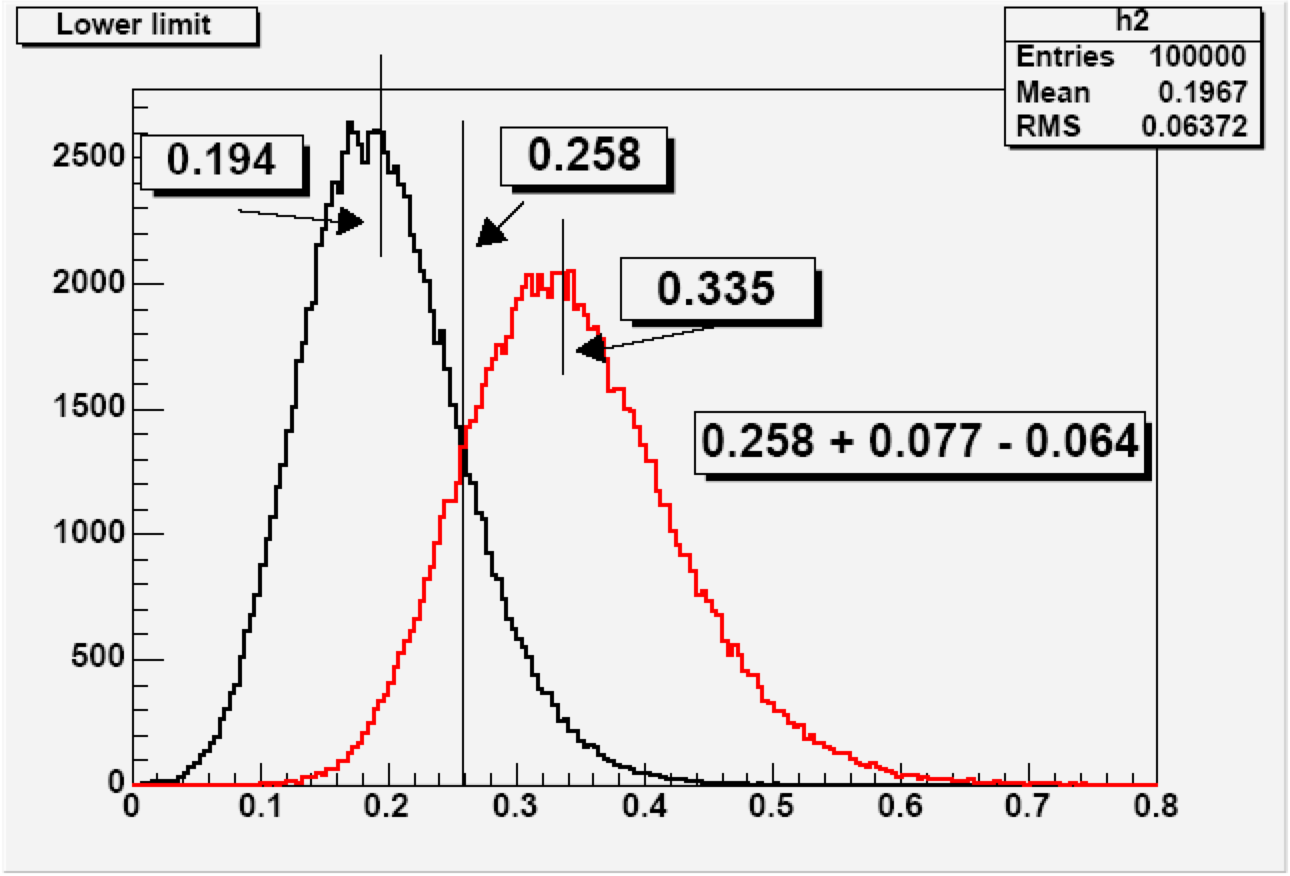
# The bootstrap method for confidence levels

With fixed efficiencies we have the binomial/gaussian distribution



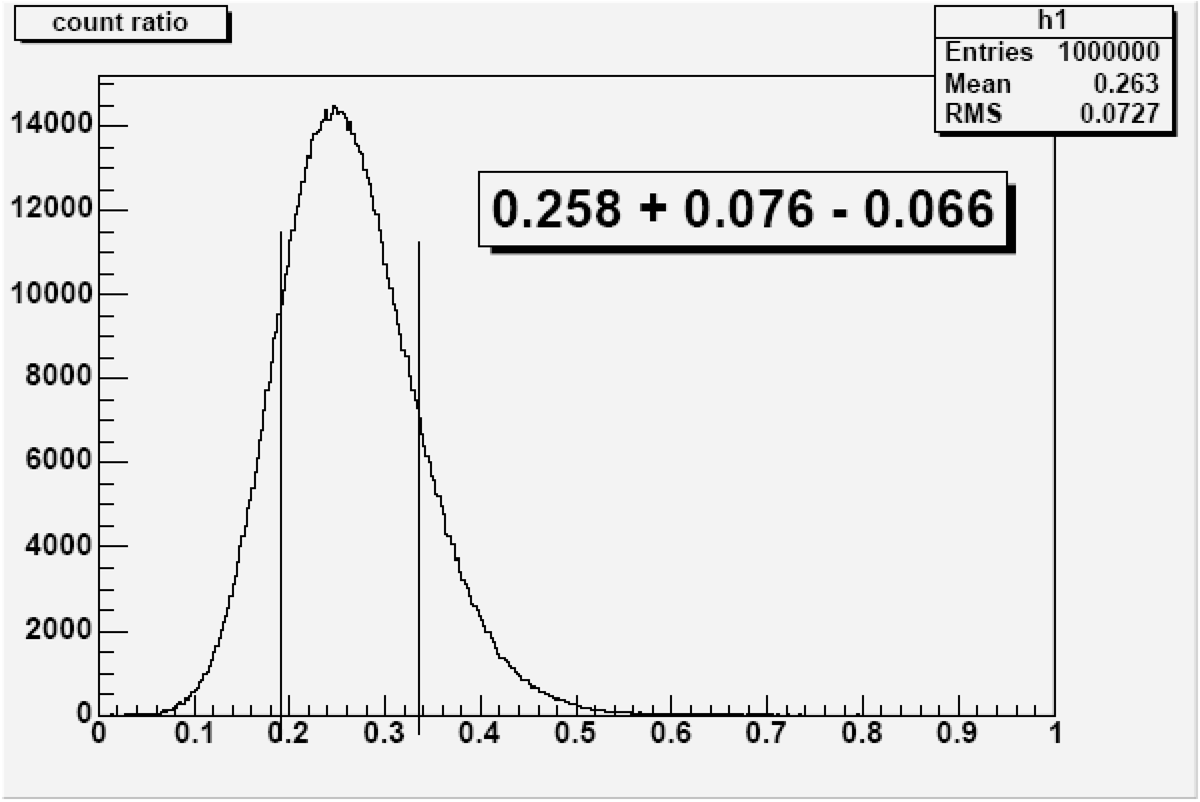
# The grid method for confidence levels

For each value of $p = N_s/N_t$ a sample of $100\,000$ events is generated sampling randomly the $\varepsilon$ and $b$ efficiencies.



# The bootstrap method for confidence levels

In this case also the approximate bootstrap method gives the same result.
This method is called Parametric Bootstrap



# The bootstrap method for BR

- A=15 Reaction A events in a N=200 event sample
- B=30 Reaction B events in a N=200 event sample

Standard error propagation for 95% CL

$$\sigma(A/B) = 1.96 \times \frac{A}{B}\sqrt{\frac{\sigma^2(A)}{A^2} + \frac{\sigma^2(B)}{B^2}} = 0.15 \ \rightarrow \ [0.21, 079]\ , \ \ \mathrm{CL} = 95\%$$

$$\sigma(A) = \sqrt{A\left(1 - \frac{A}{N}\right)}$$

Bootstrap methods:

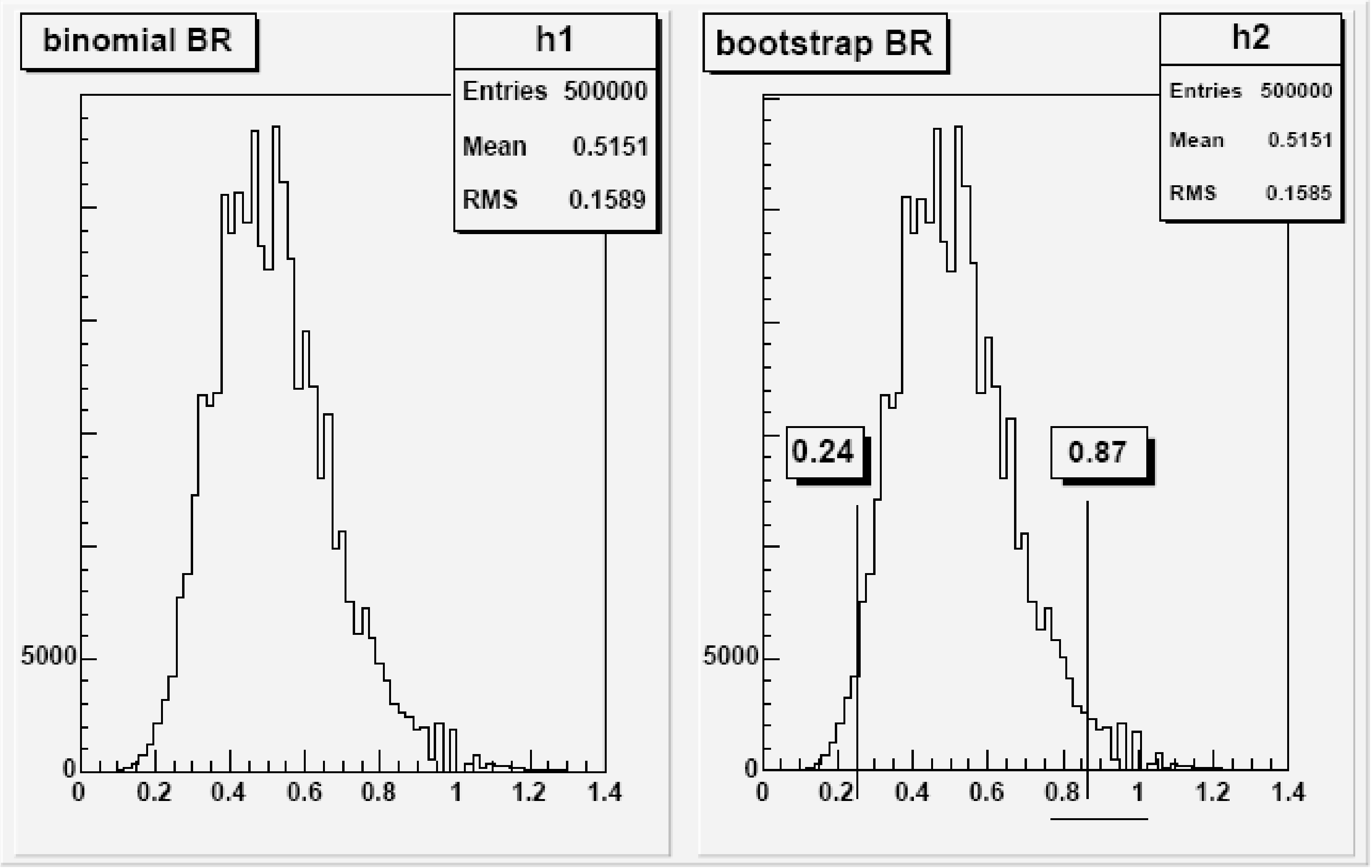
- parametric bootstrap: the events $A$ and $B$ are MC-sampled from two binomial distributions with $N = 200$ and $p_1 = A/N$ and $p_2 = B/N$;
- non parametric bootstrap: the events $A$ and $B$ are sampled with replacement from two experimental samples with $N = 200$ and $A$ or $B$ events $= 1$, the others $= 0$

Obviously, in this case the two methods give the same result:

$$[A/B] \in [0.24, 0, 86]\ , \quad CL = 95\%$$

Are the published BR really all RELIABLE??

# The bootstrap method for BR



Consider a sample $X$ containing $N$ objects. We need an estimate of $\theta$ as $\hat{\theta}(X)$.

No model of the $X$ distribution is known or considered

Statisticians have elaborated the following (non parametric) methods:

**The non parametric Sampling methods**

- **Jackknife** (Quenouille, 1949):
  $N$ samples are generated leaving out one element at a time;

- **Subsampling**:
  S resamples of dimension $N_B$ are created by repeatedly sampling without replacement from the ex imental sample. Obviously one has $N$

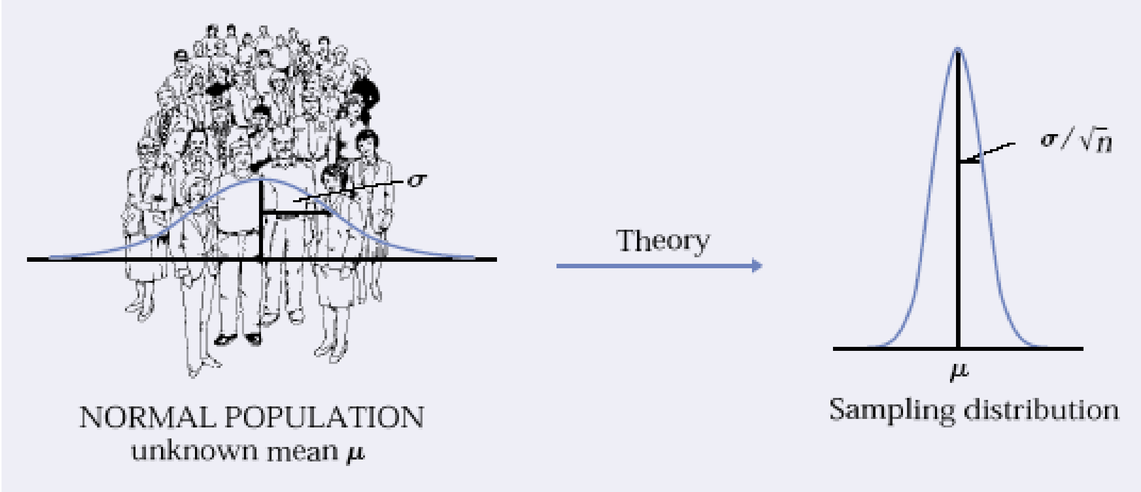
- **Bootstrap** (Efron 1979):
  S resamples of dimension $N_B$ are created by repeatedly sampling with replacement from the experimental sample. Usually $N_B = N$ is set.

**The best one !!!**

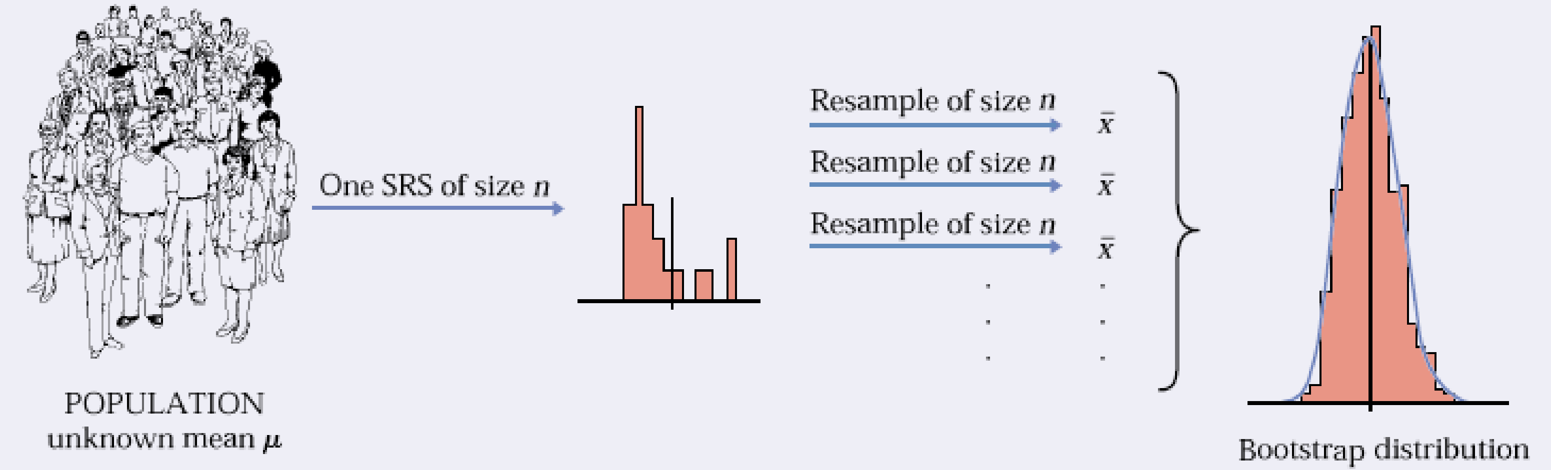
- **Permutation**:
  used in the test between two hypotheses, by resampling in a way that moves observations between the two groups, under the assumption that the null hypothesis is true

These methods, familiar among statisticians, are practically not (yet) used by physicists (only 3 papers with non parametric Bootstrap!) ← Up to 2006

# Non parametric Bootstrap




(b)




(c)

- **Center:** A statistic is biased as an estimate of the parameter if its sampling distribution is not centered at the true value of the parameter. We can check bias by seeing whether the bootstrap distribution of the statistic is centered at the value of the statistic for the original sample.

  bias

  More precisely, the **bias** of a statistic is the difference between the mean of its sampling distribution and the true value of the parameter. The **bootstrap estimate of bias** is the difference between the mean of the bootstrap distribution and the value of the statistic in the original sample.

  bootstrap estimate of bias

- **Spread:** The bootstrap standard error of a statistic is the standard deviation of its bootstrap distribution. The bootstrap standard error estimates the standard deviation of the sampling distribution of the statistic.

# The non parametric BOOTSTRAP

Consider a sample $X$ containing $N$ objects. We need an estimate of $\theta$ as

$$\hat{\theta}(X)$$

Using the Bootstrap sample, we obtain the estimator

$$\hat{\theta}^* = \hat{\theta}(X^*)$$

The Bootstrap samples have expectation values $\hat{\theta}^*$ that differ from the true one $\theta$ (bias), but ...
the Bootstrap approximates the distribution of

$$\hat{\theta} - \theta$$

with the distribution of

$$\hat{\theta}^* - \hat{\theta}$$

obtained by resampling.

# Limits of non parametric BOOTSTRAP

Drawback: the Bootstrap samples are correlated.
Some important results on this:

- the sharing of the same elements in different samples reduces the variance $s_{\rm res}$ of the (re)samples:

$$s^2_{\rm res} \to (1-\rho)\sigma^2$$

where $\rho = N_B/N$ in subsampling without replacement;

- the sampling with replacement in bootstrap increases the variance of the (re)samples:

$$s^2_{\rm res} \to (1-\rho)\rho_1\sigma^2$$

- in many cases in the bootstrap the positive bias due to the within sample correlation and the negative bias due to the between sample correlation cancel exactly

$$\sqrt{1-\rho}\sqrt{\rho_1} \simeq 1$$

# The non parametric BOOTSTRAP

**When does the Bootstrap work?**
**For the consistency of the method, the reliability must be Bootstrap-checked, through the Bootstrap samples themselves!**
**The important checks are:**

- **check the symmetry of the Bootstrap distribution, that assures the bootstrap property. Find if necessary a transformation $h$ such as**

$$h(\hat{\theta}) - h(\theta) \quad \text{and} \quad h(\hat{\theta}^*) - h(\hat{\theta})$$

  **are pivotal, that is follow the same distribution. Then make the estimate of the $h$ intervals before anti-transforming with $h^{-1}$**

- **make different estimates with different bootstrap samples (with replacement) $N_B \leq N$ and verify that the variances scales as $1/N_B$. This verify the condition**

$$\sqrt{1-\rho}\sqrt{\rho_1} \simeq 1$$

**There exists a wide statistical literature on the subject....**

PHYSICAL REVIEW D VOLUME 39, NUMBER 1 1 JANUARY 1989

# Application of the bootstrap statistical method to the tau-decay-mode problem

Kenneth G. Hayes and Martin L. Perl
*Stanford Linear Accelerator Center, Stanford University, Stanford, California 94309*

Bradley Efron
*Department of Statistics, Stanford University, Stanford, California 94309*
(Received 18 July 1988)

The bootstrap statistical method is applied to the discrepancy in the one-charged-particle decay modes of the tau lepton. This eliminates questions about the correctness of the errors ascribed to the branching-fraction measurements and the use of Gaussian error distributions for systematic errors. The discrepancy is still seen when the results of the bootstrap analysis are combined with other measurements and with deductions from theory. But the bootstrap method assigns less statistical significance to the discrepancy compared to a method using Gaussian error distributions.

At present there is a problem[1,2] in fully understanding the decay modes of the tau lepton to one-charged particle. The average directly measured value[1] of the inclusive, one-charged-particle, branching fraction $B_1$ is (86.6±0.3)%. The same number should be obtained by adding up the branching fractions of the individual one-charged-particle modes. Examples of these individual branching fractions are

$$B_e,\quad \tau^- \rightarrow \nu_\tau + e^- + \bar{\nu}_e \ ,$$

$$B_\mu,\quad \tau^- \rightarrow \nu_\tau + \mu^- + \bar{\nu}_\mu \ ,$$

$$B_\pi,\quad \tau^- \rightarrow \nu_\tau + \pi^- \ ,$$

$$B_\rho,\quad \tau^- \rightarrow \nu_\tau + \rho^- \rightarrow \nu_\tau + \pi^- + \pi^0 \ ,$$

$$B_{\pi 2\pi^0},\quad \tau^- \rightarrow \nu_\tau + \pi^- + 2\pi^0 \ ,$$

$$B_{\pi 3\pi^0},\quad \tau^- \rightarrow \nu_\tau + \pi^- + 3\pi^0 \ .$$

As shown in Table I from Ref. 2, this sum is less than (80.6±1.5)%, 6% less than the directly measured value of $B_1$. This is the $\tau$-decay-mode problem.

# Bootstrap of $B_1$ and $B_i$ data

TABLE II. $B_1$ topological branching fractions in percent. The statistical error is given first, the systematic error second.

| $B_1$ Measurement | Combined error | Energy (GeV) | Experimental group | Reference |
|---|---|---|---|---|
| 84.0 | ±2.0 | 32.0–36.8 | CELLO | H. J. Behrend *et al.*, Phys. Lett. **114B**, 282 (1982) |
| 85.2±2.6±1.3 | ±2.9 | 14.0 | CELLO | H. J. Behrend *et al.*, Z. Phys. C **23**, 103 (1984) |
| 85.1±2.8±1.3 | ±3.1 | 22.0 | CELLO | H. J. Behrend *et al.*, Z. Phys. C **23**, 103 (1984) |
| 87.8±1.3±3.9 | ±4.1 | 34.6 average | PLUTO | Ch. Berger *et al.*, Z. Phys. C **28**, 1 (1985) |
| $84.7\pm1.1^{+1.6}_{-1.3}$ | $^{+1.9}_{-1.7}$ | 13.9–43.1 | TASSO | M. Althoff *et al.*, Z. Phys. C **26**, 521 (1985) |
| 86.7±0.3±0.6 | ±0.7 | 29.0 | MAC | E. Fernandez *et al.*, Phys. Rev. Lett. **54**, 1624 (1985) |
| 86.9±0.2±0.3 | ±0.4 | 29.0 | HRS | C. Akerlof *et al.*, Phys. Rev. Lett. **55**, 570 (1985) |
| 86.1±0.5±0.9 | ±1.0 | 30.0–46.8 | JADE | W. Bartel *et al.*, Phys. Lett. **161B**, 188 (1985) |
| 87.9±0.5±1.2 | ±1.3 | 29.0 | DELCO | W. Ruckstuhl *et al.*, Phys. Rev. Lett. **56**, 2132 (1986) |
| 87.2±0.5±0.8 | ±0.9 | 29.0 | Mark II | W. B. Schmidke *et al.*, Phys. Rev. Lett. **57**, 527 (1986) |
| 84.7±0.8±0.6 | ±1.0 | 29.0 | TPC | H. Aihara *et al.*, Phys. Rev. D **35**, 1553 (1987) |

**Trimmed mean 50%**
**Correlation between measurements**
**Weighted resampling Int($1/\sigma^2$) times**
**The error on measurements is not considered**

Scope of the analysis: to test wether errors only or the data itself are unreliable

TABLE V. (a) Independent measurement of the $\tau^- \to e^- \bar{\nu}_e \nu_T$ and $\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau$ branching fractions $B_e$ and $B_\mu$ in percent. The statistical error is given first, the systematic error second. (b) Constrained or correlated measurements of the $\tau^- \to e^- \bar{\nu}_e \nu_\tau$ and $\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau$ branching fractions $B_e$ and $B_\mu$ in percent. The statistical error is given first, the systematic error second.

| $B_e$ Measurement | $B_e$ Combined error | $B_\mu$ Measurement | $B_\mu$ Combined error | Energy (GeV) | Experimental group | Reference |
|---|---|---|---|---|---|---|
| | | | | | (a) | |
| | | 17.5±2.7±3.0 | ±4.0 | 3.8–7.8 | Mark I | M. L. Perl *et al.*, Phys. Lett. **70B**, 487 (977) |
| | | 22 | $^{+10}_{-7}$ | 4.8 | | M. Cavalli-Storza *et al.*, Lett. Nuovo Cimento **20**, 337 (1977) |
| | | 15 | ±3.0 | 3.6–5.0 | PLUTO | J. Burmester *et al.*, Phys. Lett. **68B**, 297 (1977) |
| 16.0 | ±1.3 | | | 3.1–7.4 | DELCO | W. Bacino *et al.*, Phys. Rev. Lett. **41**, 13 (1978) |
| | | 22 | $^{+7}_{-8}$ | 6.4–7.4 | Iron Ball | J. G. Smith *et al.*, Phys. Rev. D **18**, 1 (1978) |
| | | 21±5±3 | ±6 | 3.6–7.4 | DELCO | W. Bacino *et al.*, Phys. Rev. Lett. **42**, 6 (1979) |
| 19 | ±9.0 | 35 | ±14 | 12–31.6 | TASSO | R. Brandelik *et al.*, Phys. Lett. **92B**, 199 (1980) |
| | | 17.8±2.0±1.8 | ±2.7 | 9.4–31.6 | PLUTO | Ch. Berger *et al.*, Phys. Lett. **99B**, 489 (1981) |
| 18.3±2.4±1.9 | ±3.1 | 17.6±2.6±2.1 | ±3.3 | 34.0 | CELLO | H. J. Behrend *et al.*, Phys. Lett. **127B**, 270 (1983) |
| $20.4\pm3.0^{+1.4}_{-0.9}$ | $^{+3.3}_{-3.1}$ | $12.9\pm1.7^{+0.7}_{-0.5}$ | ±1.8 | 13.9–43.1 | TASSO | M. Althoff *et al.*, Z. Phys. C **26**, 521 (1985) |
| 13.0±1.9±2.9 | ±3.5 | 19.4±1.6±1.7 | ±2.3 | 34.6 average | PLUTO | Ch. Berger *et al.*, Z. Phys. C **28**, 1 (1985) |
| | | 17.4±0.6±0.8 | ±1.0 | 14.0–46.8 | Mark J | B. Adeva *et al.*, Phys. Lett. B **179**, 177 (1986) |
| 17.0±0.7±0.9 | ±1.1 | 18.8±0.8±0.7 | ±1.1 | 34.6 average | JADE | W. Bartel *et al.*, Phys. Lett. B **182**, 216 (1986) |
| 19.1±0.8±1.1 | ±1.4 | 18.3±0.9±0.8 | ±1.2 | 29.0 | Mark II | P. R. Burchat *et al.*, Phys. Rev. D **35**, 27 (1987) |
| | | | | | (b) | |
| 18.9±1.0±2.8 | ±3.0 | 18.3±1.0±2.8 | ±3.0 | 3.8–7.8 | Mark I | M. L. Perl *et al.*, Phys. Lett. **70B**, 487 (1977) |
| 22.7 | ±5.5 | 22.1 | ±5.5 | 4.1–7.4 | Lead-Glass Wall | A. Barbaro-Galtiero *et al.*, Phys. Rev. Lett. **39**, 1058 (1977) |
| 18.5±2.8±1.4 | ±3.1 | 18.0±2.8±1.4 | ±3.1 | 3.9–5.2 | DASP | R. Brandelik *et al.*, Phys. Lett. **73B**, 109 (1978) |
| 17.6±0.6±1.0 | ±1.3 | 17.1±0.6±1.0 | ±1.3 | 3.5–6.7 | Mark II | C. A. Blocker *et al.*, Phys. Lett. **109B**, 119 (1982) |
| 18.2±0.7±0.5 | ±0.9 | 18.0±1.0±0.6 | ±1.2 | 3.8 | Mark III | R. M. Baltrusaitis *et al.*, Phys. Rev. Lett. **55**, 1842 (1985) |
| 17.4±0.8±0.5 | ±0.9 | 17.7±0.8±0.5 | ±0.9 | 29.0 | MAC | W. W. Ash *et al.*, Phys. Rev. Lett. **55**, 2118 (1985) |
| 18.4±1.2±1.0 | ±1.6 | 17.7±1.2±0.7 | ±1.4 | 29.0 | TPC | H. Aihara *et al.*, Phys. Rev. D **35**, 1553 (1987) |

# Results

TABLE VI. Means and standard deviations (SD) for $B_1$, $B_\rho$, $B_\pi$, $B_e$, and $B_\mu$. Both quantities are in percent.

| Branching fraction | Bootstrap (method A) Mean, 25% trimmed | SD | Analysis method Bootstrap with weighted measurements (method C) Mean, 25% trimmed | SD | Normal-error method from Ref. 1 Mean, not trimmed | Formal error |
|---|---|---|---|---|---|---|
| $B_1$ | 85.8 | 0.63 | 86.9 | 0.36 | 86.6 | 0.28 |
| $B_\rho$ | 22.5 | 0.35 | 22.5 | 0.19 | 22.5 | 0.85 |
| $B_\pi$ | 10.2 | 0.55 | 10.8 | 0.45 | 10.8 | 0.60 |
| $B_e$ | 18.3 | 0.38 | 17.8 | 0.30 | 17.6 | 0.44 |
| $B_\mu$ | 18.2 | 0.56 | 17.8 | 0.21 | 17.7 | 0.41 |

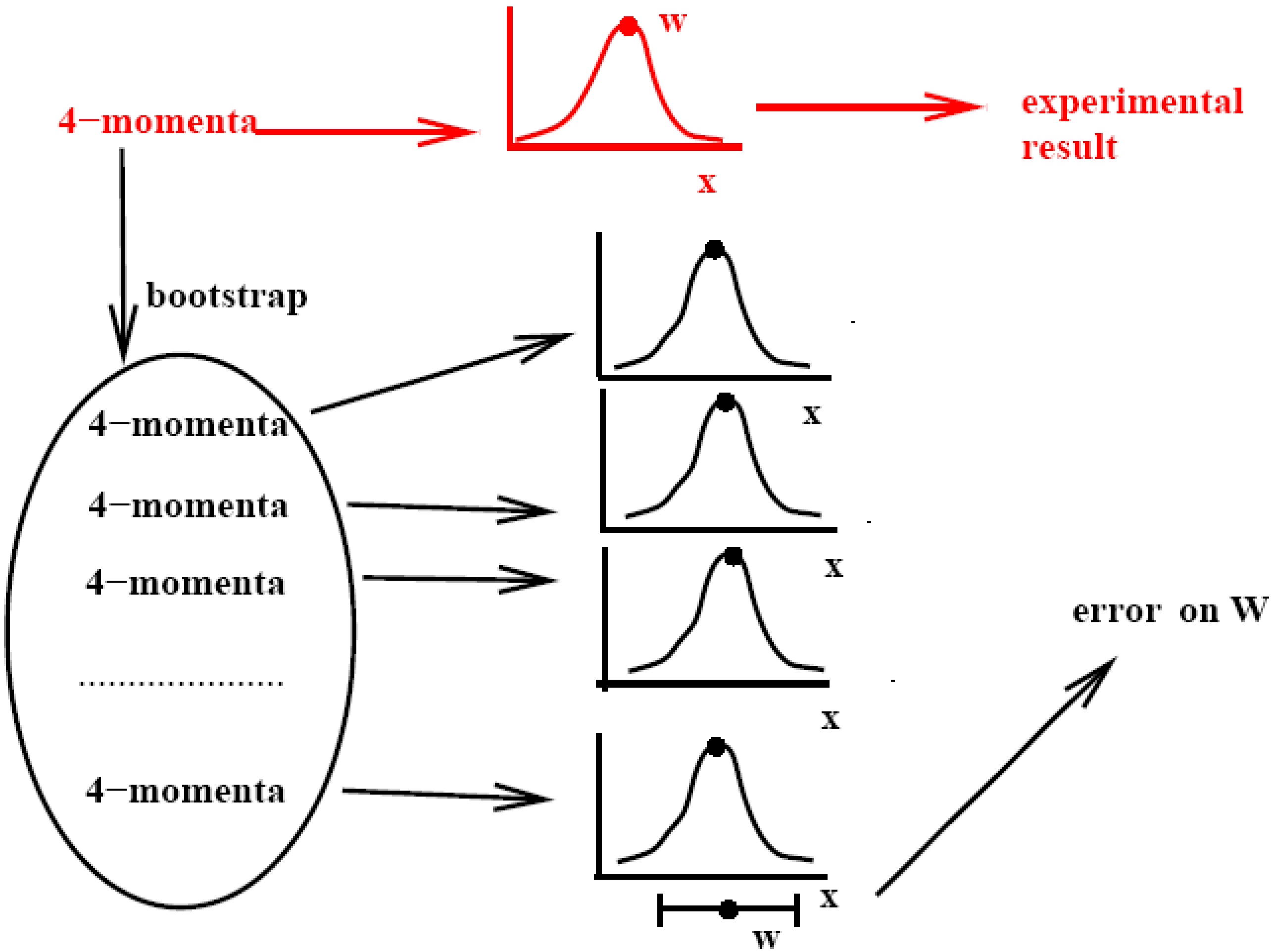
Bootstrap: $$\frac{85.8-81.3}{\sqrt{1.3^2+0.6^2}}=3.1$$

Standard analysis: $$\frac{86.6-80.6}{\sqrt{1.5^2+0.3^2}}=3.9$$

**Some data are unreliable**

# The non parametric BOOTSTRAP

A possible use of the Bootstrap in Nuclear physics



## BOOTSTRAP FOR COMPARING TWO POPULATIONS

Given independent SRSs of sizes $n$ and $m$ from two populations:

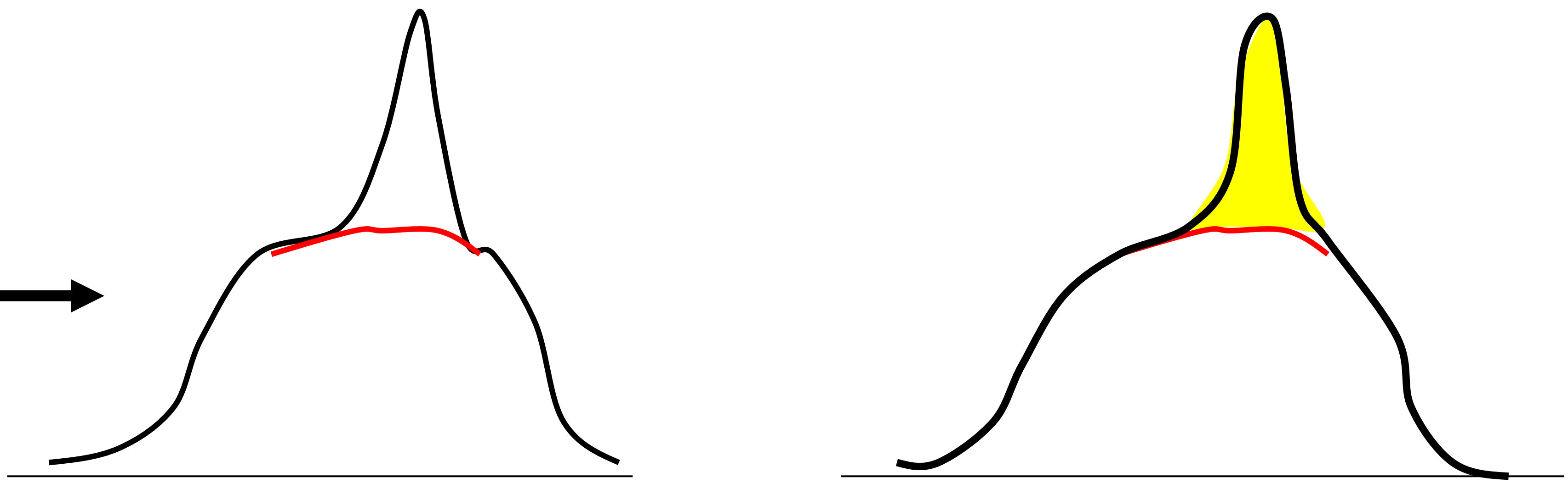
1. Draw a resample of size $n$ with replacement from the first sample and a separate resample of size $m$ from the second sample. Compute a statistic that compares the two groups, such as the difference between the two sample means.

2. Repeat this resampling process hundreds of times.

3. Construct the bootstrap distribution of the statistic. Inspect its shape, bias, and bootstrap standard error in the usual way.

**Useful when the two samples are signal and background....**

# The dual Bootstrap



Fix the background on one sample and calculated the peak signal with another sample to avoid biases !!

Repeat on bootstrap samples (dual bootstrap)

# Standard analysis in nuclear physics experiments

- the 4-momenta are reconstructed and the analysis is performed
- errors are calculated following the standard (gaussian) theory
- a MC toy model is invented and the analysis procedure is checked on this model
- at this point the procedure could be further checked on bootstrapped data!

# A word on the permutation tests

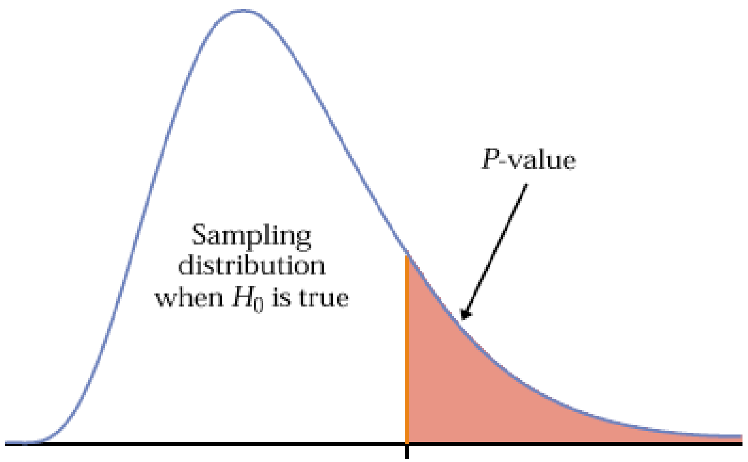



FIGURE 14.19 The *P*-value of a statistical test is found from the sampling distribution the statistic would have if the null hypothesis were true. It is the probability of a result at least as extreme as the value we actually observed.

## GENERAL PROCEDURE FOR PERMUTATION TESTS

To carry out a permutation test based on a statistic that measures the size of an effect of interest:

**1.** Compute the statistic for the original data.

**2.** Choose permutation resamples from the data without replacement in a way that is consistent with the null hypothesis of the test and with the study design. Construct the permutation distribution of the statistic from its values in a large number of resamples.

**3.** Find the *P*-value by locating the original statistic on the permutation distribution.

# Conclusions

- **Poissonian Counting**: most of the tests do not consider the error on background and overestimate the signal. Often true (mean) values and measured values are improperly confused.

- **Binomial counting**: a general theory there exists and should be applied.

- The **errors** should be calculated by MC methods and the procedure checked with MC toy models

- **Nonparametric Bootstrap** methods should be used also by physicists