Confidence Intervals for ratios of two random quantities

Francisco Matorras
IFCA
Instituto de Física de Cantabria (Santander, Spain)

A simple problem?

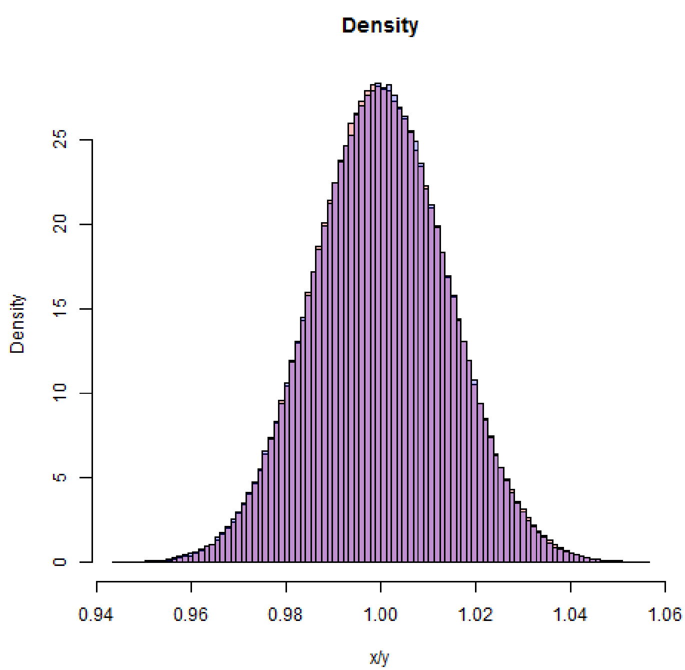
- I will speak about the (apparently) trivial problem:
  - Given two random variables x and y (will assume they are independent), how can I set the confidence interval of $z=x/y$?
- Some of the many cases where you have found this situation:
  - x and y are measurements with a given error, what is the error on x/y?
    - Note this usually means normal distribution and 68% CI ($\mu\pm\sigma$)
  - In a given experiment (or bin in a histogram) I observe some counts, how much off I am from a model whose prediction is obtained by MC simulation? $N_{data}/N_{mc} \sim 1$?
  - Comparing rates of two different decay modes
- Shouldn't just use "error propagation"?
  - Calculate the new σ with the usual formula and define $x/y \pm n\sigma$ intervals

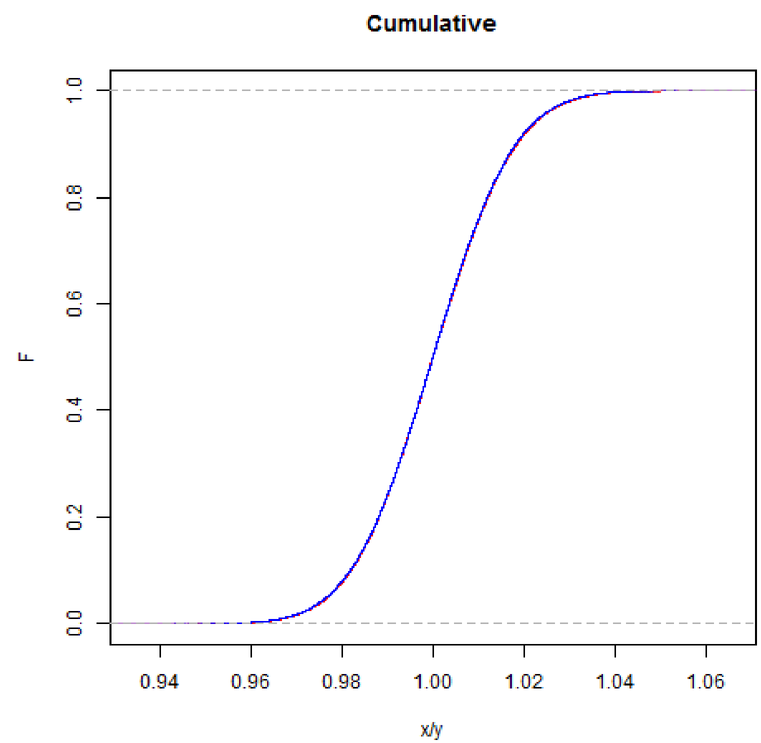
Not so simple…

- We often forget that to use this formula we need that:
  - x and y follow a Gaussian law
  - The expression, x/y here, can be approximately linearized
    - Note x/y is clearly non-linear if we get close to 0
    - σ_x and especially σ_y need to be small w.r.t. x and y means
- How small?
  - Depends on what you want to get! Not the same to draw an error bar on a plot than to extract a 5-σ CI…
- Do not panic ☺!, in most cases both assumptions are approximately fulfilled, but please be aware that:
  - We rarely have purely gaussian distributions
  - Our CI will fail if there is sizeable chance that the denominator gets close to 0
  - If you need precise results error prop. will probably not be enough
- The talk follows for the cases where you have to go beyond

Some examples: ratio of gaussians

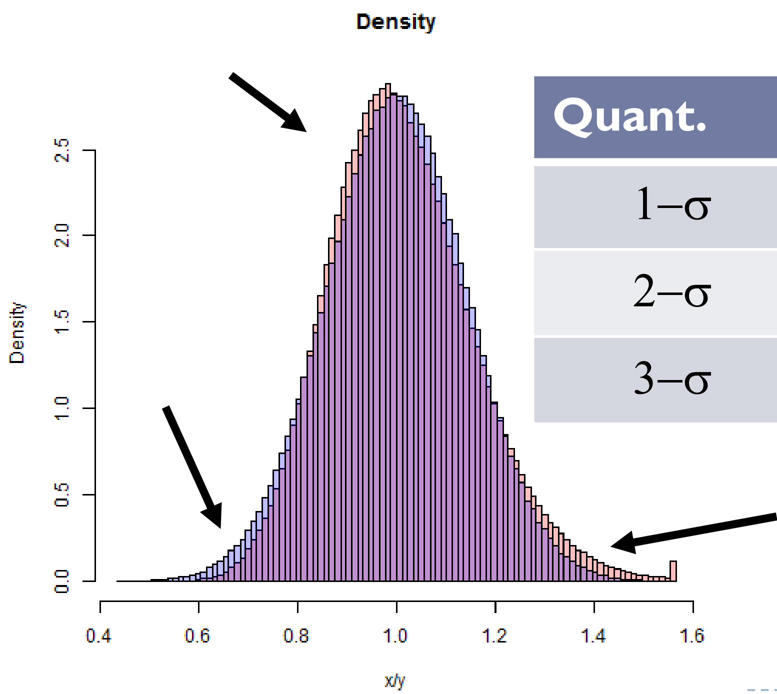
- 10^6 toys x, y are $\mathbb{N}(1,0.01)$
- Red/orange actual ratio
- Blue error prop approx.
- $\sigma/\mu \sim 1\%$ IS small, hardly a difference



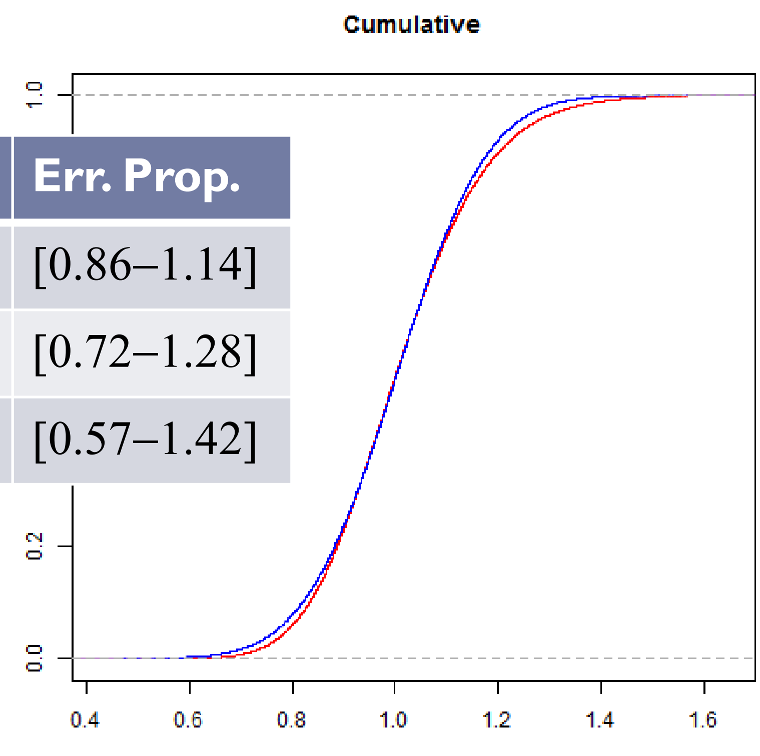


Some examples : ratio of gaussians

- 10^6 toys x, y are $\mathbb{N}(1,0.1)$
- $\sigma/\mu \sim 10\%$ starts to show differences
- Ratio becomes slightly asymmetric
- Central part similar but longer tail to the right

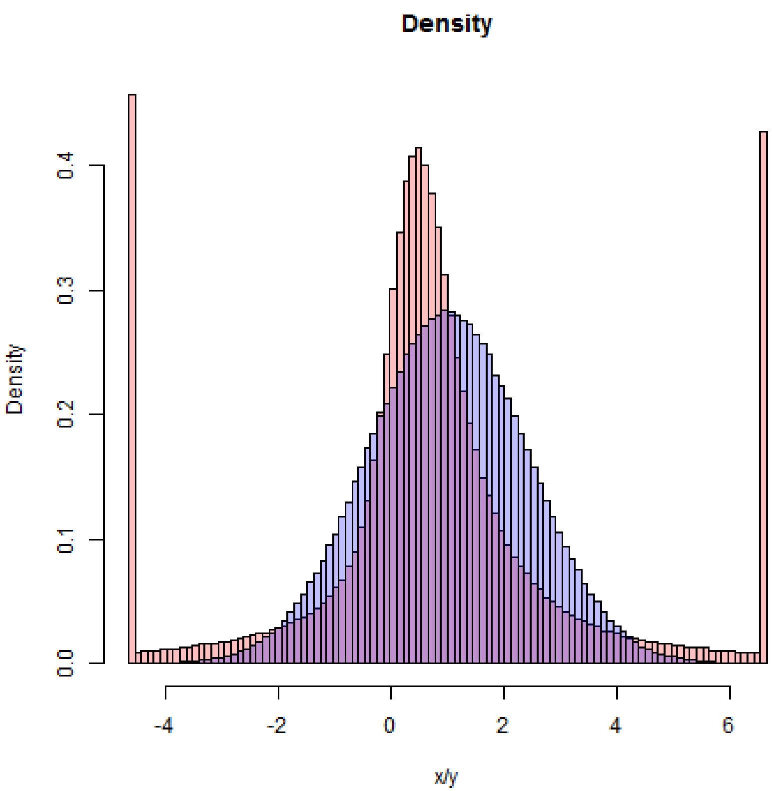


Quant.	Exact	Err. Prop.
1–σ	[0.87–1.15]	[0.86–1.14]
2–σ	[0.75–1.33]	[0.72–1.28]
3–σ	[0.64–1.55]	[0.57–1.42]



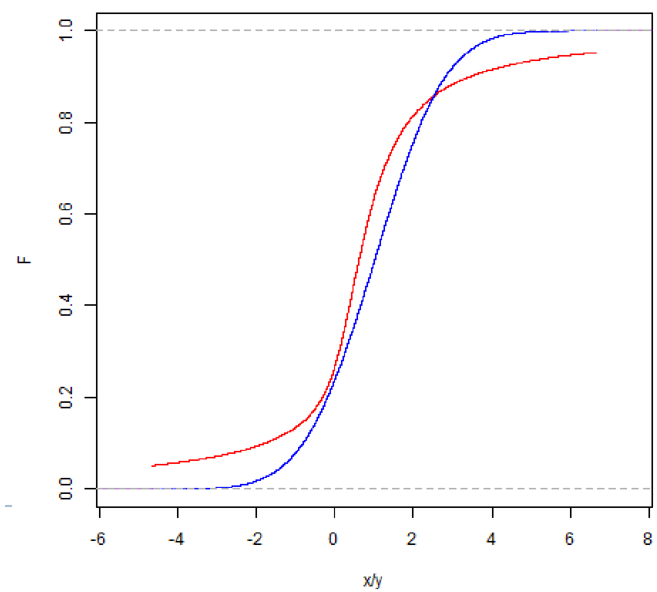
Some examples : ratio of gaussians

- 10^6 toys x, y are $\mathbb{N}(1,1)$
- $\sigma/\mu\sim100\%$ error prop. not valid at all!



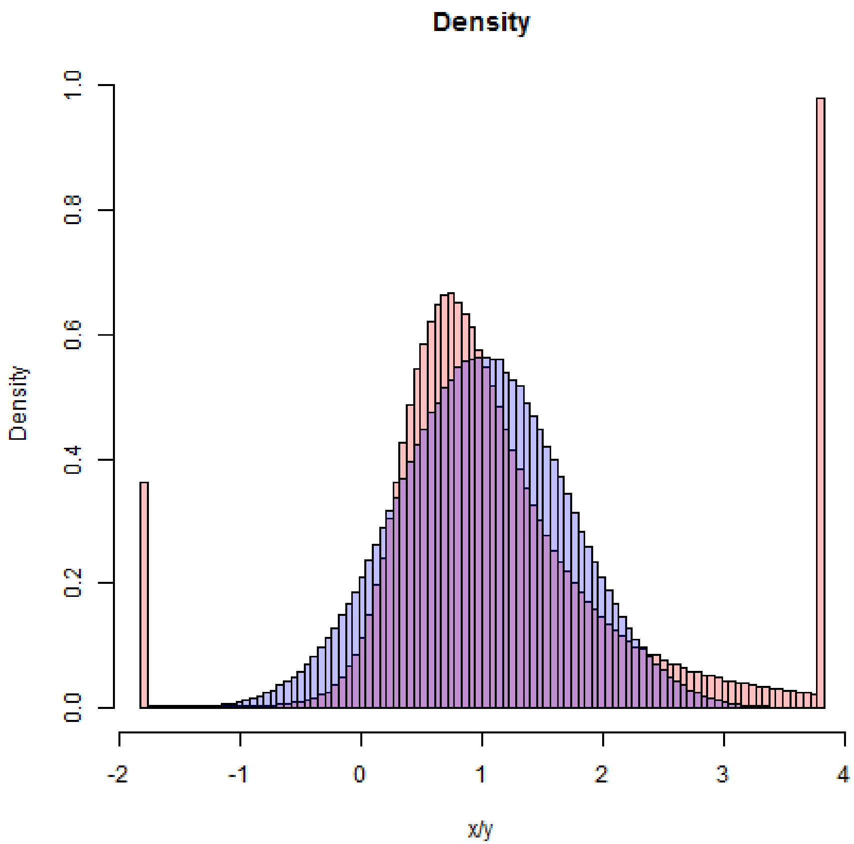
- Narrower central part and very long tails

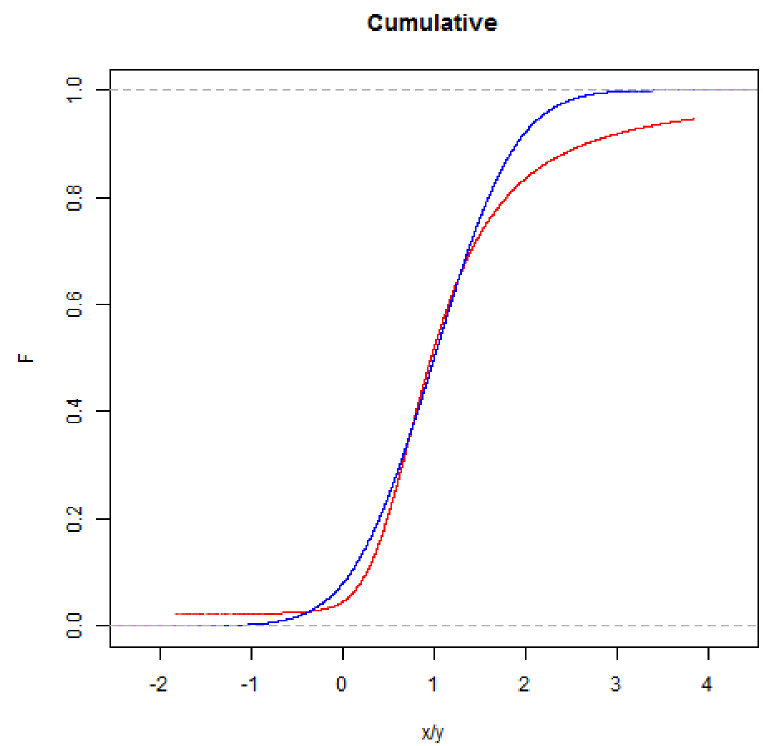
Quant.	Exact	Err. Prop.
1−σ	[−0.7,2.3]	[−0.4,2.4]
2−σ	[−12,13]	[−1.8,3.8]
3−σ	[−210,207]	[−3.8,5.2]



Some examples : ratio of gaussians

- 10^6 toys x, y N(1,0.5)



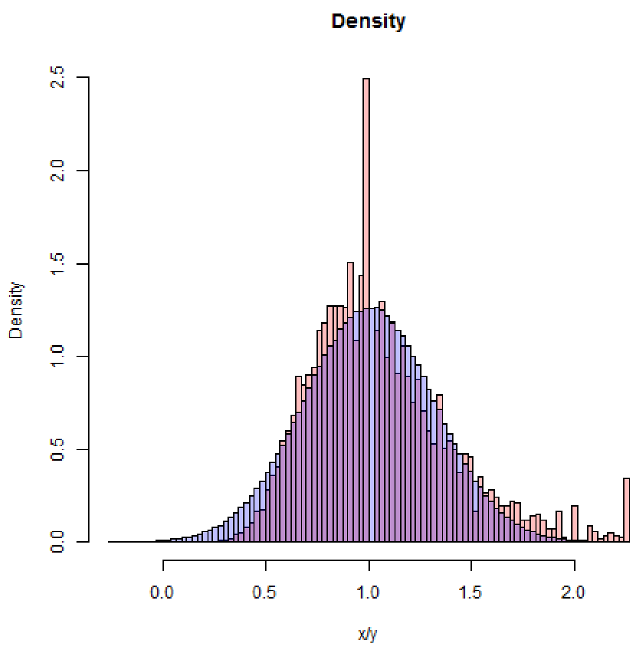


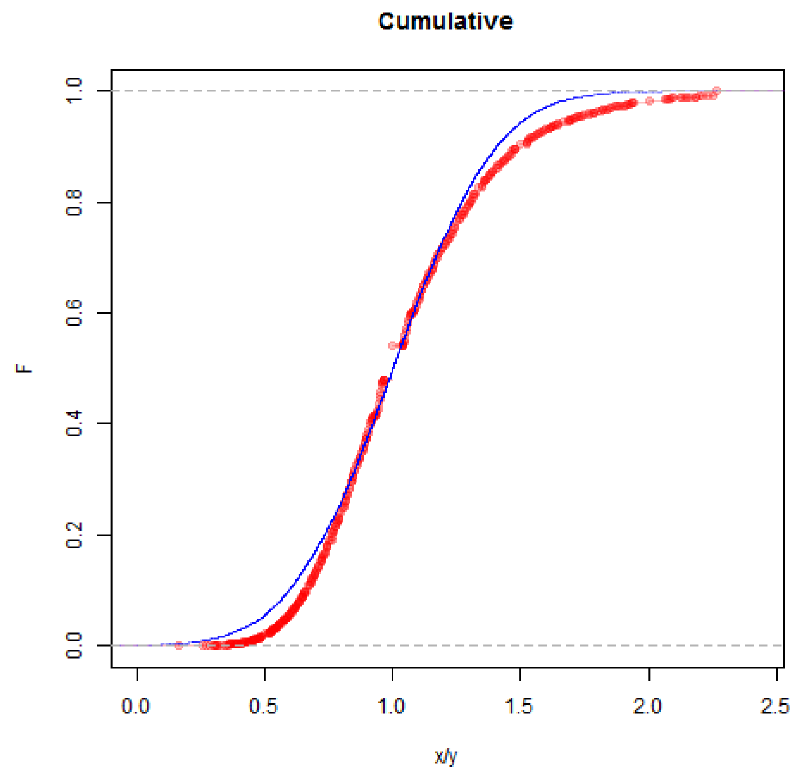
Ratio of Poisson distributions

- Will not detail here how to deal with the ratio of Gaussians
  - but given in the additional material
- Will concentrate on a much more interesting case, the ratio of two Poisson distributions
  - General case when we compare two counting experiments
  - Or when we compare an observation with a MC prediction
  - Comparison of the decay BR of two decays modes

Ratio of two random variables following Poisson law

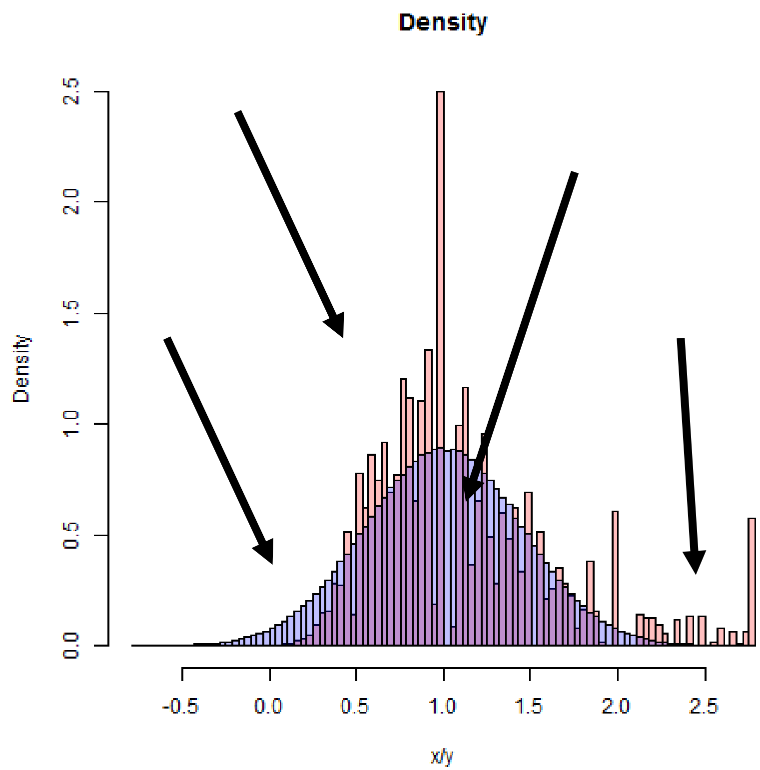
- $\frac{\wp(20)}{\wp(20)}$, ratio of Poisson distribution with mean=20
  - where we usually would believe the gaussian regime is valid!
- Clearly different pdf

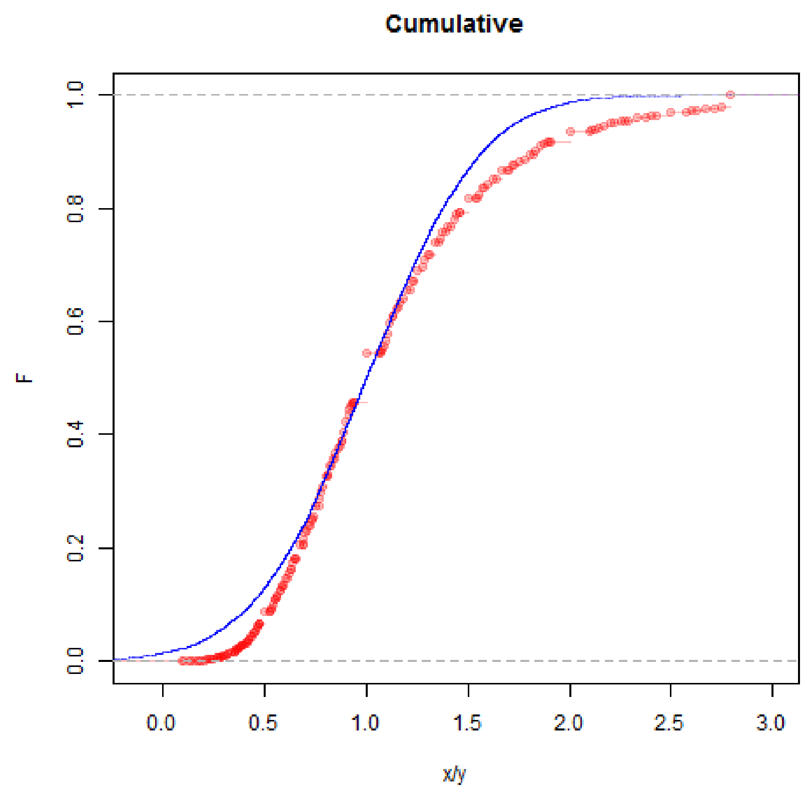




Ratio of two random variables following Poisson law

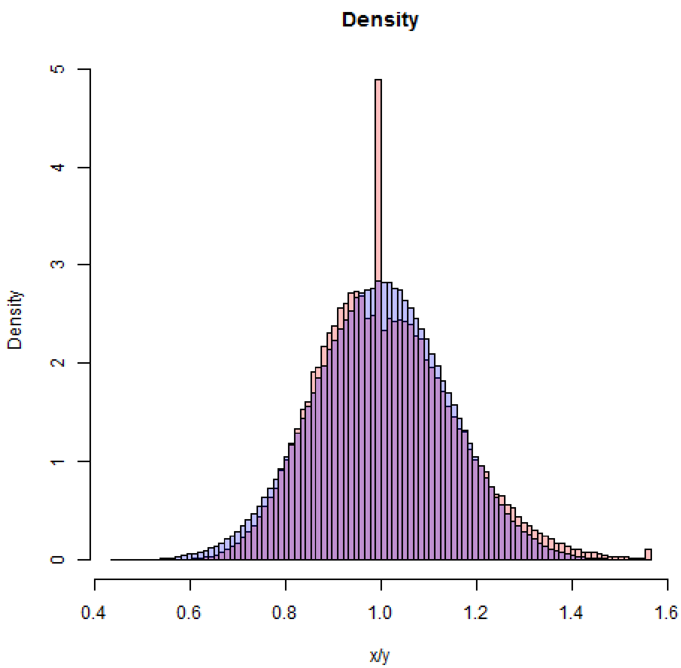
- $\frac{\wp(10)}{\wp(10)}$, ratio of Poisson distribution with mean=10

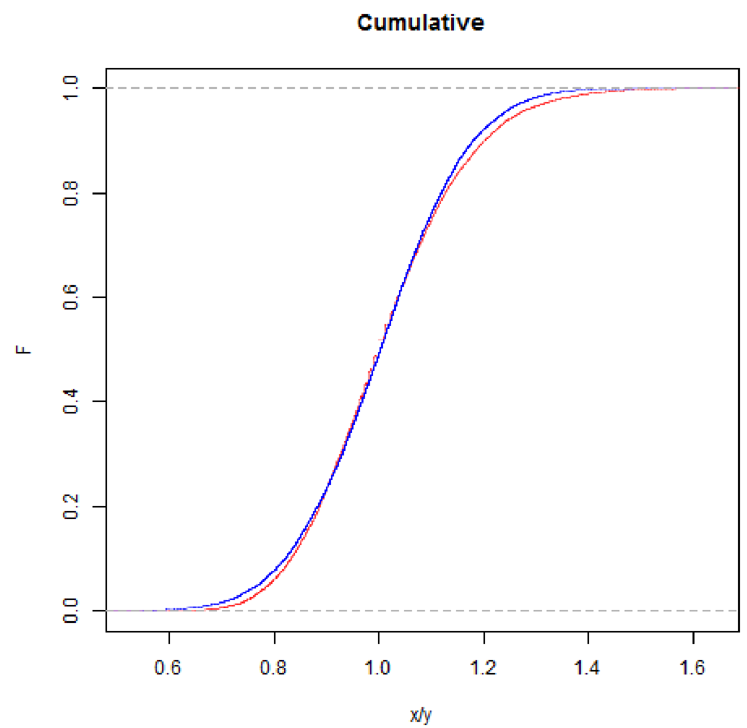




Some examples

- $\frac{\wp(100)}{\wp(100)}$, even for such a large number of counts we get a difference (mostly because of the integer counts)





Ratio of Poissonians

- Despite looking more complex, this problem has an analytical solution
- The CI ratio of two independent Poissonian observations n and m is equivalent to that of a binomial of m/(n+m)
  - Observed by J. Przyborowski and H. Wilenski, "Homogeneity of Results in Testing Samples from Poisson Series," Biometrika 31 (1940)
  - Introduced in HEP by F. James and M. Roos, "Errors on Ratios of Small Numbers of Events," Nuclear Physics B172 (1980).
  - Thoroughly tested by Robert D. Cousins, Kathryn E. Hymes, Jordan Tucker
- The binomial has an analytical solution, so-called exact CI, the Clopper-Pearson (C-P) interval. C.J. Clopper and E.S. Pearson, "The Use of Confidence or Fiducial Limits illustrated in the Case of the Binomial," Biometrika 26 (1934) 404.
  - It is exact in the sense that guarantees a minimum coverage
  - Since binomial (and Poisson) distribution are discrete it often **overcovers**
- C-P available in most packages or can be easily derived from Fisher- F

Example of overcoverage

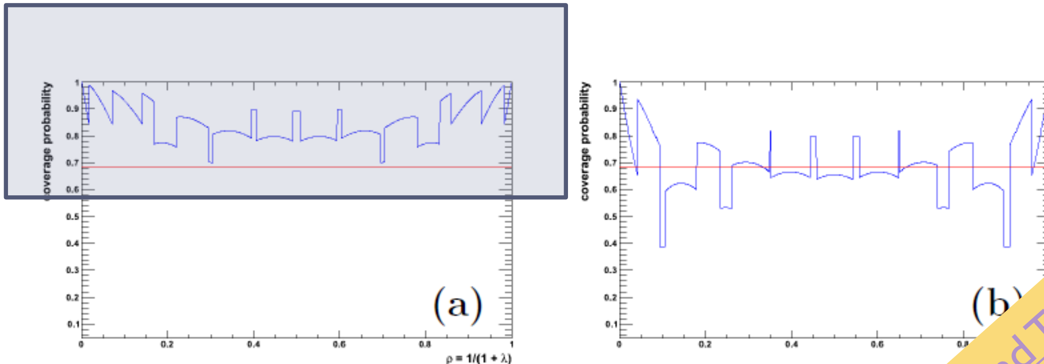


Fig. 1. (a) Coverage of 68.27% C.L. Clopper-Pearson intervals, [illegible] coverage of intervals calculated using a Bayesian method with Jeffreys [illegible] containing 68.27% posterior probability, both as a function of ρ, for f[illegible] $= 10$. (a) and (b) are horizontal slices of Figs. 3a and b, respectively.

from Cousins, Hymes, and Tucker

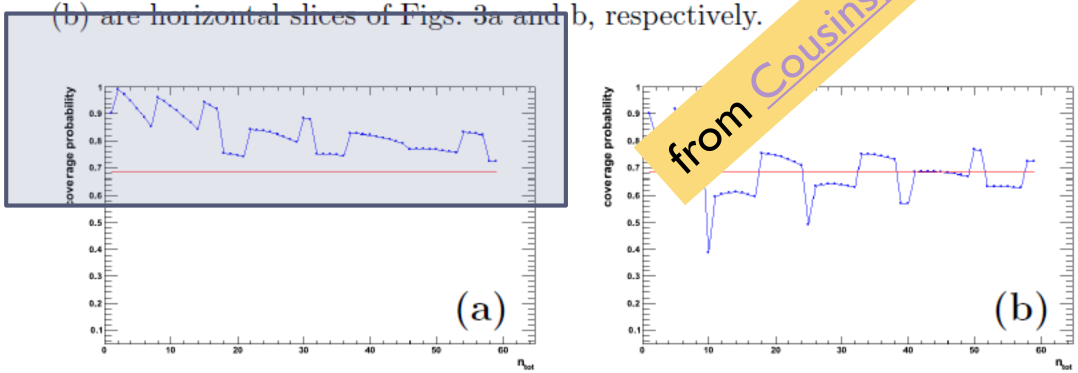


Fig. 2. (a) Coverage of 68.27% C.L. Clopper-Pearson intervals, and (b) coverage of intervals calculated using a Bayesian method with Jeffreys prior and containing 68.27% posterior probability, as a function of n_{tot}, for fixed $\rho = 0.1$. (a) and (b) are vertical slices of Figs. 3a and b, respectively.

Ratio of Poisson Distributions

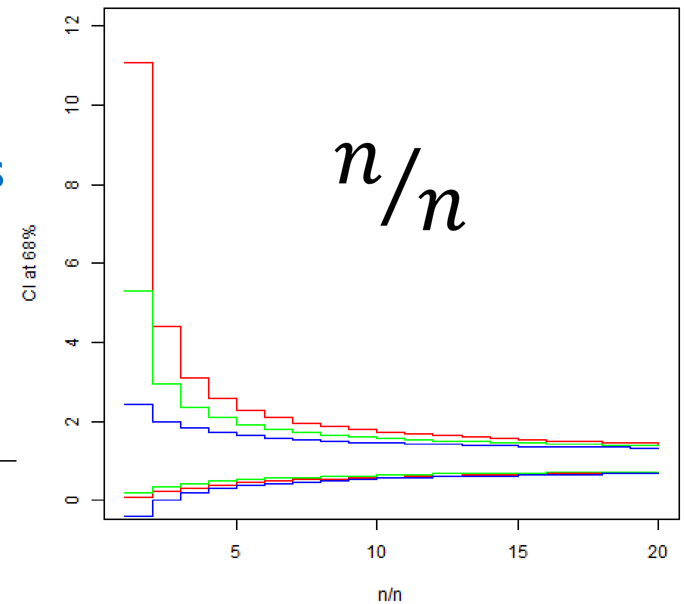
- We tend to believe that overcoverage is always conservative, but that's not always true
  - For example, we calculate a ratio and want to see if it is compatible with 1. Overcoverage means an error bar larger than desired and hence could show compatibility when it is not true.
- Alternative methods to C-P are reviewed by Cousins, Hymes, and Tucker
  - Robert D. Cousins, Kathryn E. Hymes, Jordan Tucker, Frequentist Evaluation of Intervals Estimated for a Binomial Parameter and for the Ratio of Poisson Means, Nucl. Instr A. 612-2 (2010)

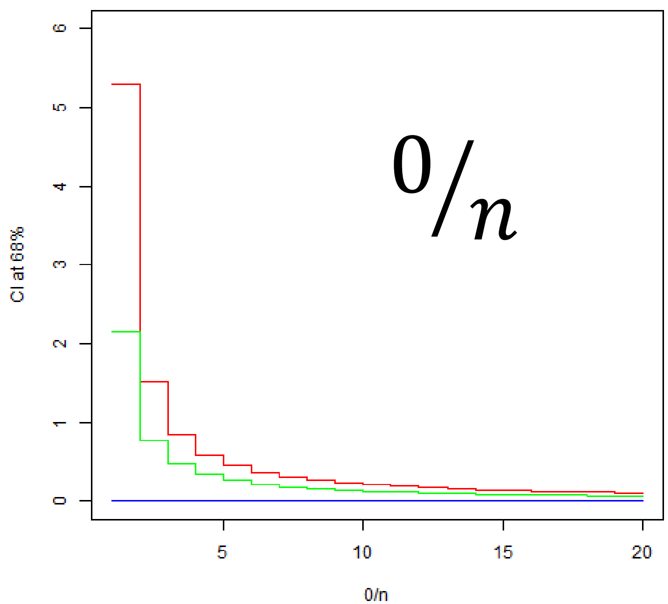
Ratio of Poisson distributions

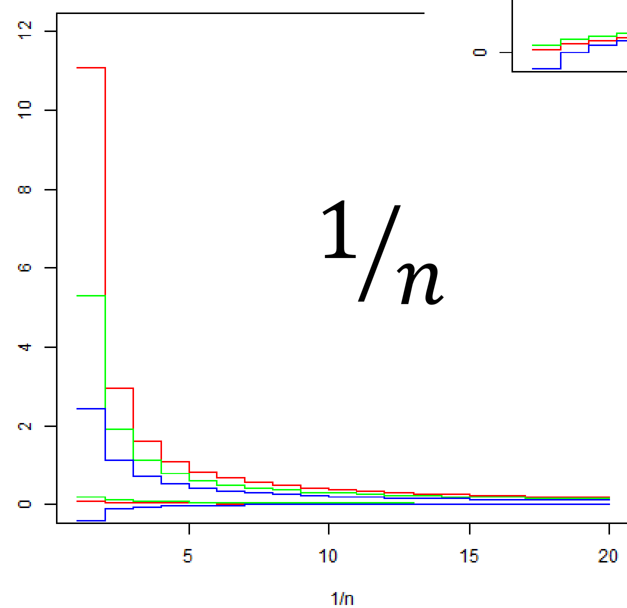
- Good performance obtained if Clopper-Pearson replaced by the so-called Lancaster mid-P CI
  - "an intermediate value of the tail probability" to overcome discreteness
  - Sort of an "average"
  - Proposed by Lancaster
    - H.O. Lancaster, Significance Tests in Discrete Distributions," J. Amer. Stat. Assoc. 56 (1961) 223.
- Works as expected for the ratio of two Poisson (keeps good coverage properties with much narrower intervals)
- Jordan Tucker developed a method for Root, now implemented by Lorenzo Moneta

Examples of 68% CI

- Intervals estimated for a given observation for C-P, mid-P or error propagation
- All methods tend to give the same intervals at large values
- Err prop gives often unphysical intervals
- C-P produces huge intervals at small values





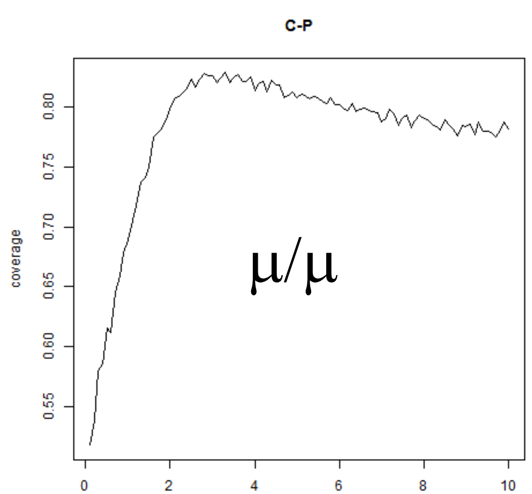


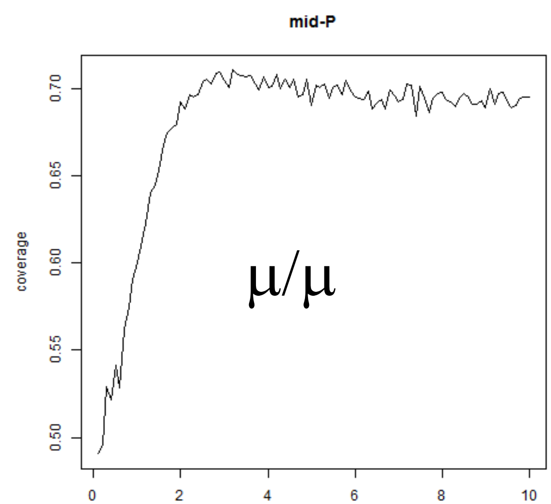
Coverage for 68% CI

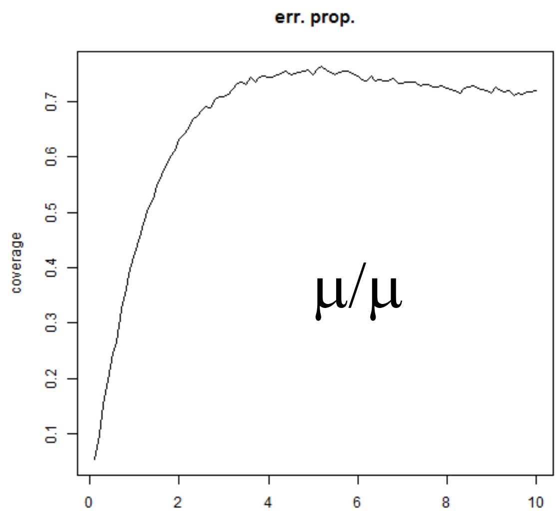
- Generate random numbers according to Poisson law for numerator and denominator with given $\mu 1$, $\mu 2$
- Compute the CI according to the three methods
- Count how often the true ratio is within the CI
- Note: 0/0 ignored, n/0 counted as fail ???
  - This leads to undercoverage when denominator mean is small
  - Not sure if it makes sense when one compares data and MC, will we define a confidence interval if we have data in a bin where the MC predicts exactly 0?
  - Note 0/0 does not appear in binomial case

Coverage for 68% CI

- Scanning for μ1=μ2
- C-P slightly undercovers for small values (see note on previous slide) and overcovers above 1
- Mid-P is very close for essentially all values
- Err prop largely undercovers below 2 and then works
- Remember the quite different CI (at small m, mid-p gives the same coverage despite having ½ smaller interval)

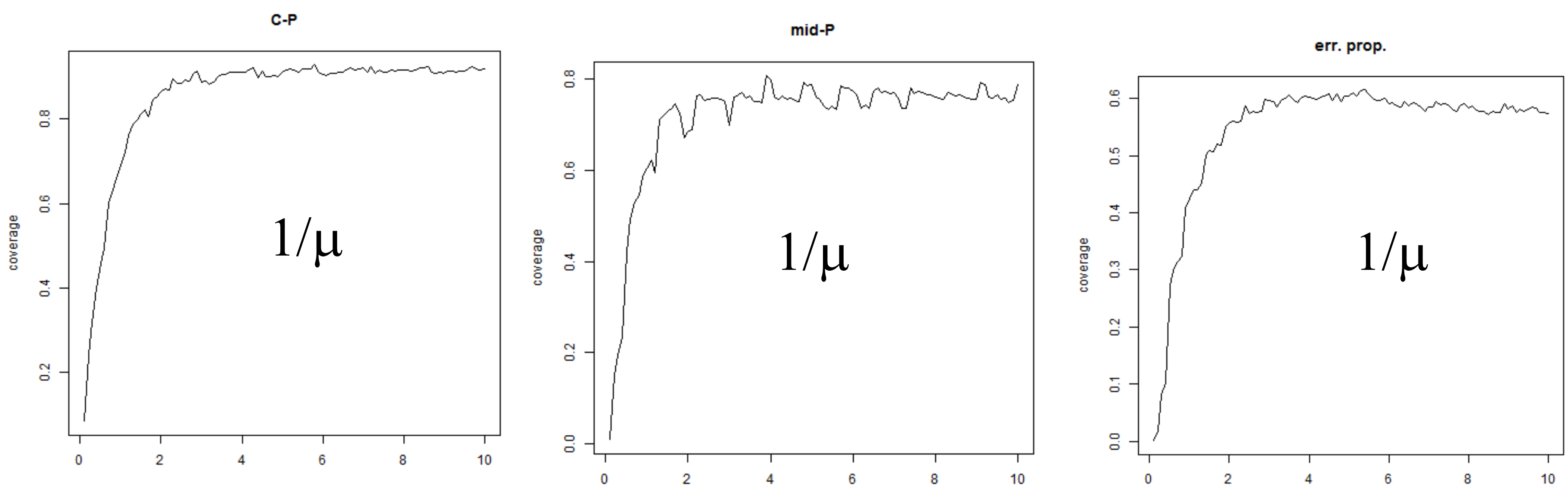






Coverage for 68% CI

- Fixing μ1=1 and scanning for μ2
- Similar behavior, with poor coverage for small values (1/0)
- Above ~2, mid-P produces excellent results



But real life is a bit more complex

- Usually if we compare a measurement (counts) with an expectation we need to use a MC simulation with several contributions, each with a different weight (process A, B, C…)
- The denominator is a linear combination of Poissonians
  - Which is NOT a Poisson distribution (except when you just sum Poisson with equal weights)
- Or even worst, we often have event by event weights
- What to do?
  - Again, gaussian approximation? The denominator often is close to a Gaussian, but a Poisson divided by a Gaussian has no analytical treatment (to my knowledge!)

Approaches explored

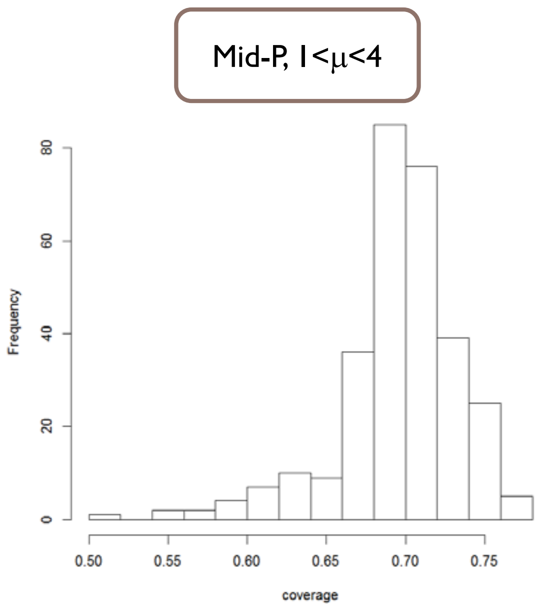
- Everything Gaussian and error propagation
  - As expected works as soon as the number of events are large, fails otherwise
- Profile maximum likelihoods
  - Complex to handle and do not give excellent results
- Search for an equivalent (approximate) problem with ratio of Poisson and apply C-P or (better) mid-P
  - $\frac{\wp(\mu_0)}{\mathbb{N}(\mu,\sigma)} \rightarrow \frac{\wp(\mu_0)}{K\wp(\mu_1)}$ with K and μ1 chosen to preserve μ and σ
  - A simple system of equations: $\mu = K\,\mu_1$, $\sigma^2 = K^2\mu_1$
  - Coverage studies with pseudoexperiments shows good behavior for a wide range of parameters

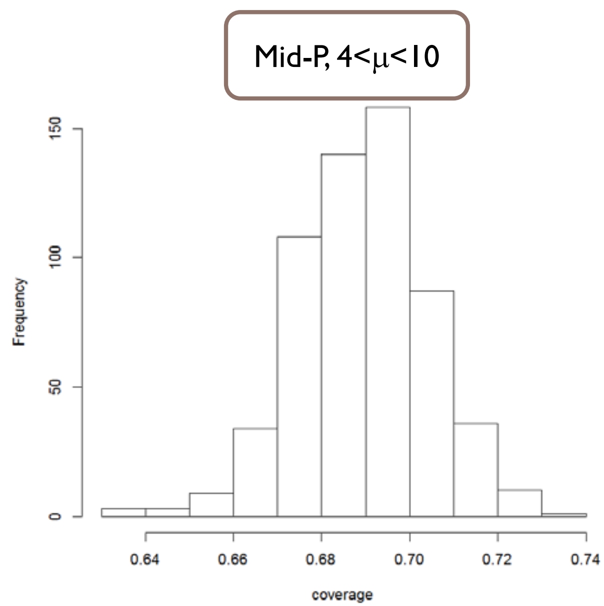
Checking the models

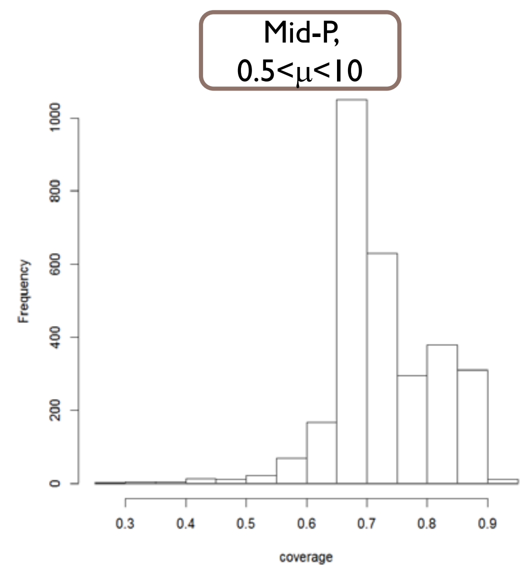
- Compare the coverage for each of the models in different regions of the parameter space
- Random scan in regions of interest
  - Generate random models, drawing from $\wp(\mu_0)$ for the numerator and a linear combination of up to four random numbers following $\wp(\mu_i)$
    - Obtain randomly the different means and linear coefficients
    - For each, do a coverage calculation with pseudoexperiments
  - Three examples shown here, scanning the means of numerator and denominator:
    - between 1 and 4, small
    - between 4 and 10, almost gaussian
    - Between 0.5 and 10
  - Not shown, but for larger numbers perfect coverage

Coverage for general data/MC ratio

- With this approach and mid-P, excellent (realistic) coverage when the means are above 1, even if small, for mid-P
- Still might work for some purposes (like plots) down to even smaller values







Root implementation

- If you want to use this approaches for a better representation of error bars when doing a data over MC ratio plot, can use the Root implementation (by Lorenzo Moneta,) since v6.06
  - TGraphAsymmErrors::Divide(h1,h2,"pois") for exact C-P
  - TGraphAsymmErrors::Divide(h1,h2,"pois midp") for mid-p

Summary

- The definition of the confidence level for x/y is a bit more complex than what we usually think
- In most cases error propagation is fine
  - but please spend a couple of minutes thinking to what extent it is appropriate for your problem
- For ratios of Poisson, you can use analytic functions available in R and Root for binomial
  - If you want to ensure a minimum coverage use C-P
  - Otherwise I recommend Lancaster mid-P
- For ratio data plots wrt MC estimations with several components or weights a reasonably good approximation available in Root

Additional Material

Ratio of Gaussians

- Based on a private communication by Jochen Ott
- You can define an adequate confidence interval for r=x/y from the roots of the inequality:

$$(\hat{y}^2 - n^2\sigma_y^2)r^2 - 2(\hat{x}\hat{y} - n^2\sigma_{xy})r + (\hat{x}^2 - n^2\sigma_x^2) < 0$$

- If you want 68% CI, n=1
- σ_{xy}=0 if there is no correlation
- If the roots are $r_1 < r_2$ one can set the confidence intervals:
  - [r1,r2] if |y|>nσ
  - (-∞, r1] ∪ [r2, ∞] if |y|<nσ
  - Note the unusual "error bar" in this case, caused by the probability that the denominator goes close to 0