Assessing the Performance of Uncertainty Calculations by Simulation

Blair Hall

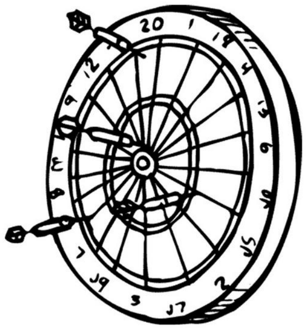
Measurement Standards Laboratory
of New Zealand

Industrial Research Ltd.,
PO Box 31-310, Lower Hutt,
New Zealand.

b.hall@irl.cri.nz

Introduction

There is no universally ‘correct method’ of evaluating uncertainty



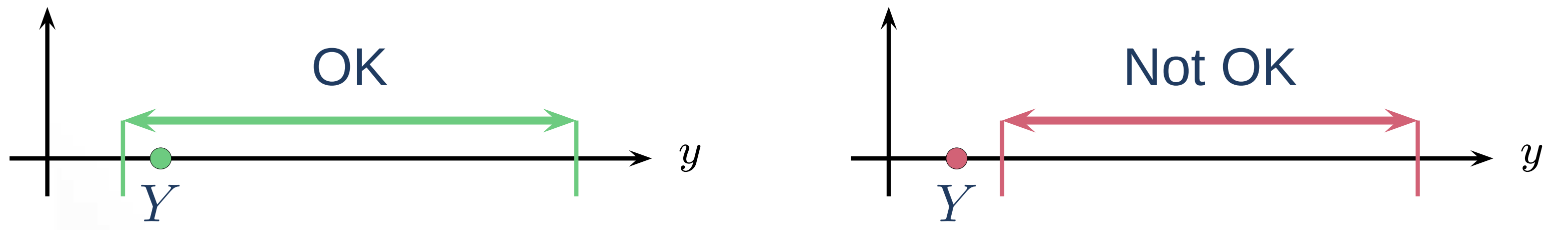
Correctness means achieving (or bettering) a nominal rate of ‘success’ over repeated measurements (level of confidence)

Success is understood to be when the measurand is inside the stated uncertainty, but in real measurements the measurand is unknown

Testing uncertainty calculations

- Devise a simulation model that produces independent data sets
- Test uncertainty calculations as follows:

1. Choose a fixed value for the measurand Y
2. Simulate N (a large number of) measurements
3. For each measurement, evaluate an uncertainty interval
4. Test whether Y is contained
5. Report the number of successes



A simple reflectometer simulation

- Fixed measurand

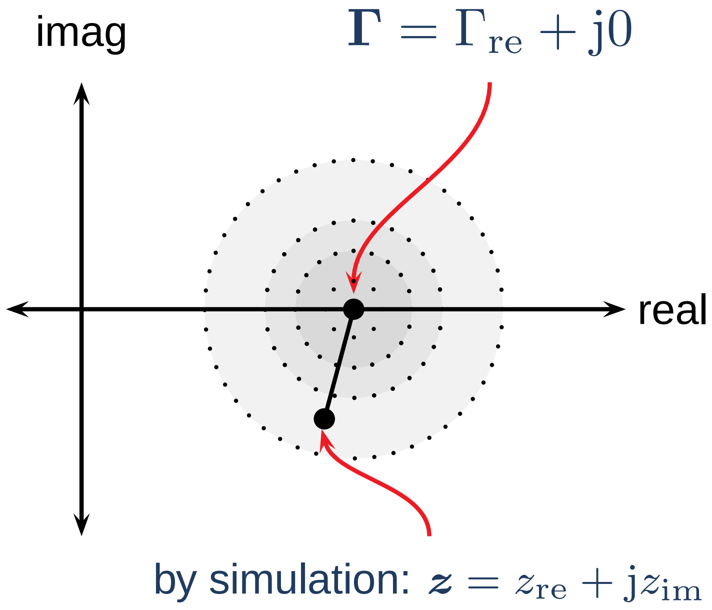
$$\mathbf{\Gamma} = \Gamma_{\mathrm{re}} + \mathrm{j}0$$

- Random errors

$$\varepsilon_{\mathrm{re}} \sim \mathsf{n}(0, u) \,,$$
$$\varepsilon_{\mathrm{im}} \sim \mathsf{n}(0, u)$$

- Simulated observations

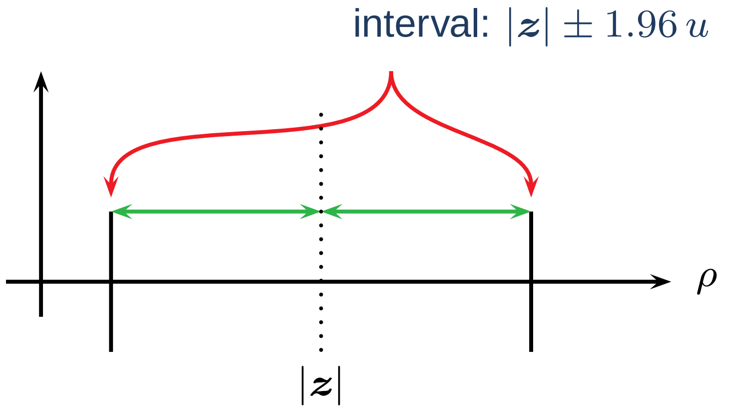
$$z_{\mathrm{re}} = \Gamma_{\mathrm{re}} + \varepsilon_{\mathrm{re}} \,,$$
$$z_{\mathrm{im}} = \Gamma_{\mathrm{im}} + \varepsilon_{\mathrm{im}}$$



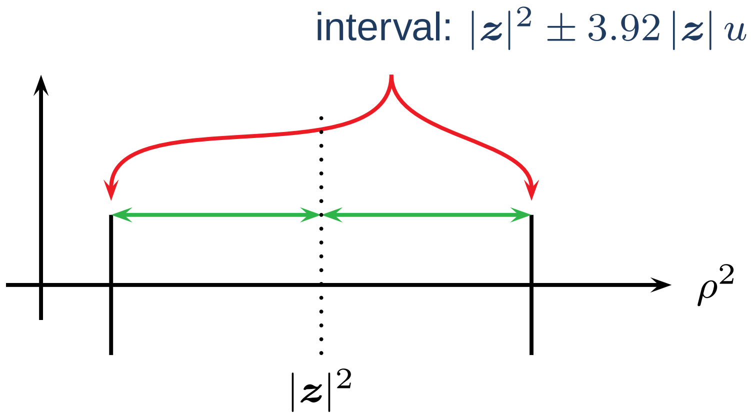
The GUM method for $|\Gamma|$ and $|\Gamma|^2$

Data from a simulated experiment $\boldsymbol{z} = z_{\text{re}} + \text{j}\, z_{\text{im}}$

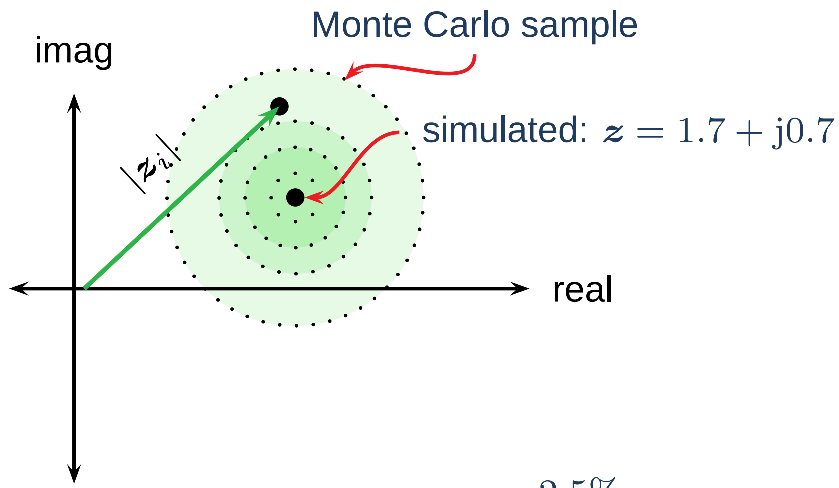
- $|\boldsymbol{\Gamma}| \approx |\boldsymbol{z}| = \sqrt{z_{\text{re}}^2 + z_{\text{im}}^2}$
- interval centered on $|\boldsymbol{z}|$
- $U = 1.96\, u$
- U does not depend on $\boldsymbol{z}$

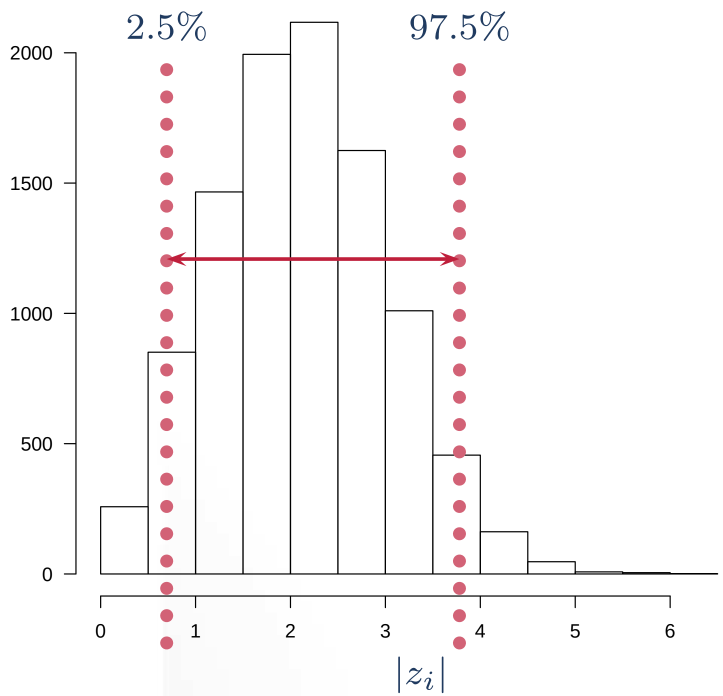


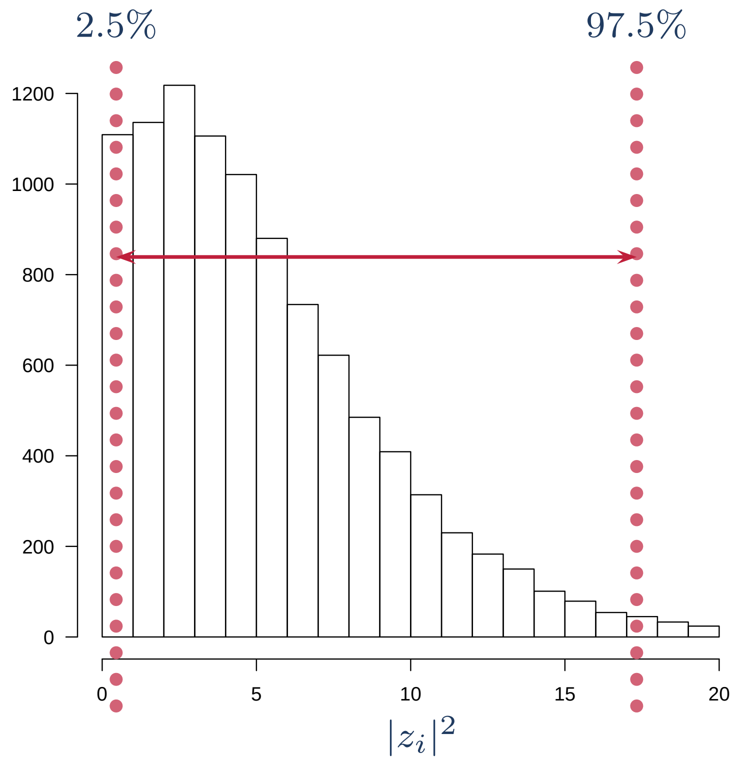
- $|\boldsymbol{\Gamma}|^2 \approx |\boldsymbol{z}|^2 = z_{\text{re}}^2 + z_{\text{im}}^2$
- interval centered on $|\boldsymbol{z}|^2$
- $U = 3.92\, |\boldsymbol{z}|\, u$
- U depends on $\boldsymbol{z}$



The SUP method for $|\Gamma|$ and $|\Gamma|^2$







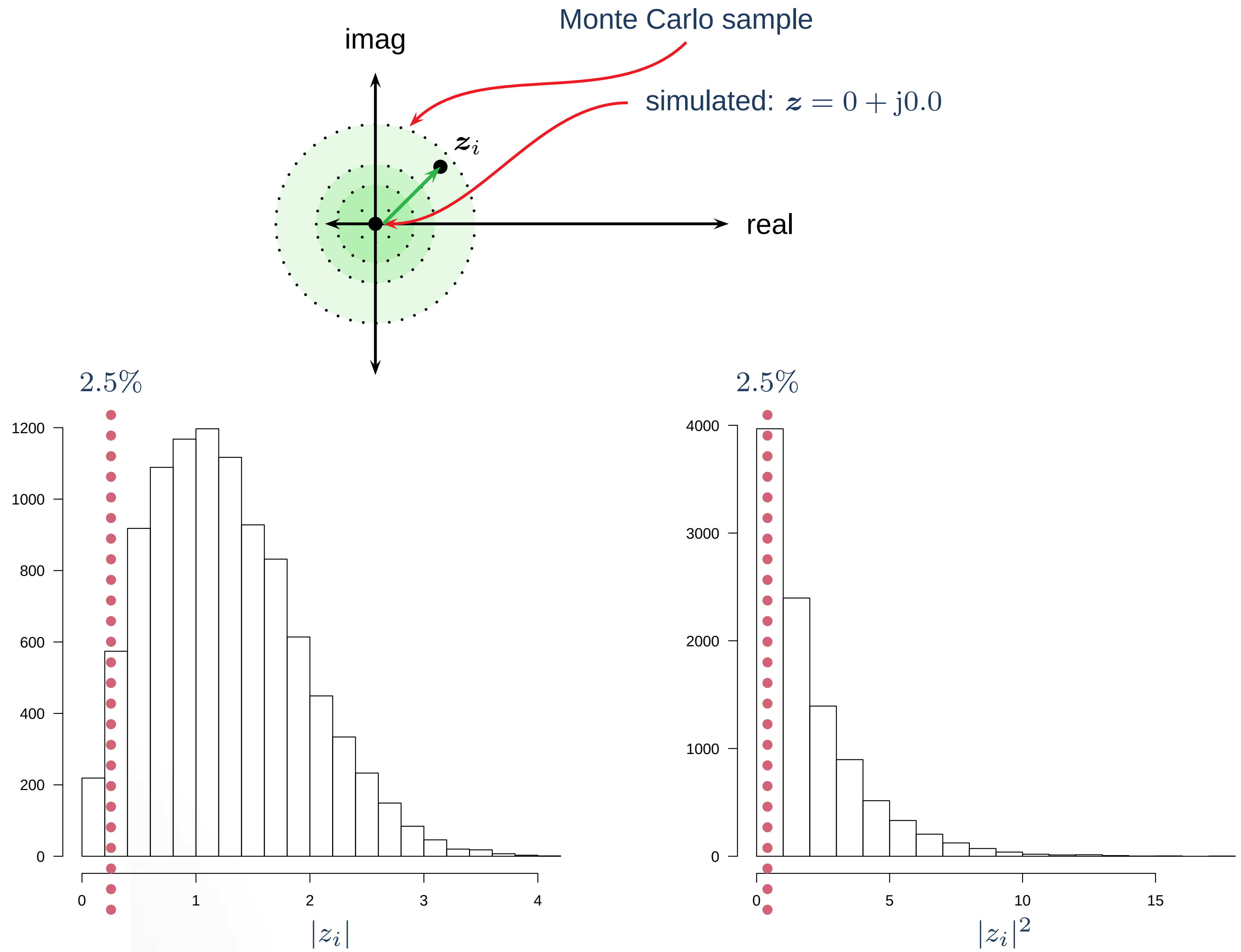
Performance results

Successes in 1000 simulated experiments:

$\mathbf{\Gamma}/u$	$\lvert\mathbf{\Gamma}\rvert$		$\lvert\mathbf{\Gamma}\rvert^2$	
	GUM	SUP	GUM	SUP
0.1	884	0	1000	0
0.2	880	0	999	0
0.5	920	735	999	765
1.0	955	908	984	913
2.0	956	934	943	959
5.0	948	945	942	950
10.0	943	943	948	950

- Monte Carlo (MC) sample size of $L = 10^4$
- Standard uncertainty in these results ≈ 7
- The MC samples generated for $|\mathbf{\Gamma}|$ and $|\mathbf{\Gamma}|^2$ were independent
- Strange things are happening above the line at $\mathbf{\Gamma}/u = 1$

The SUP method close to the origin

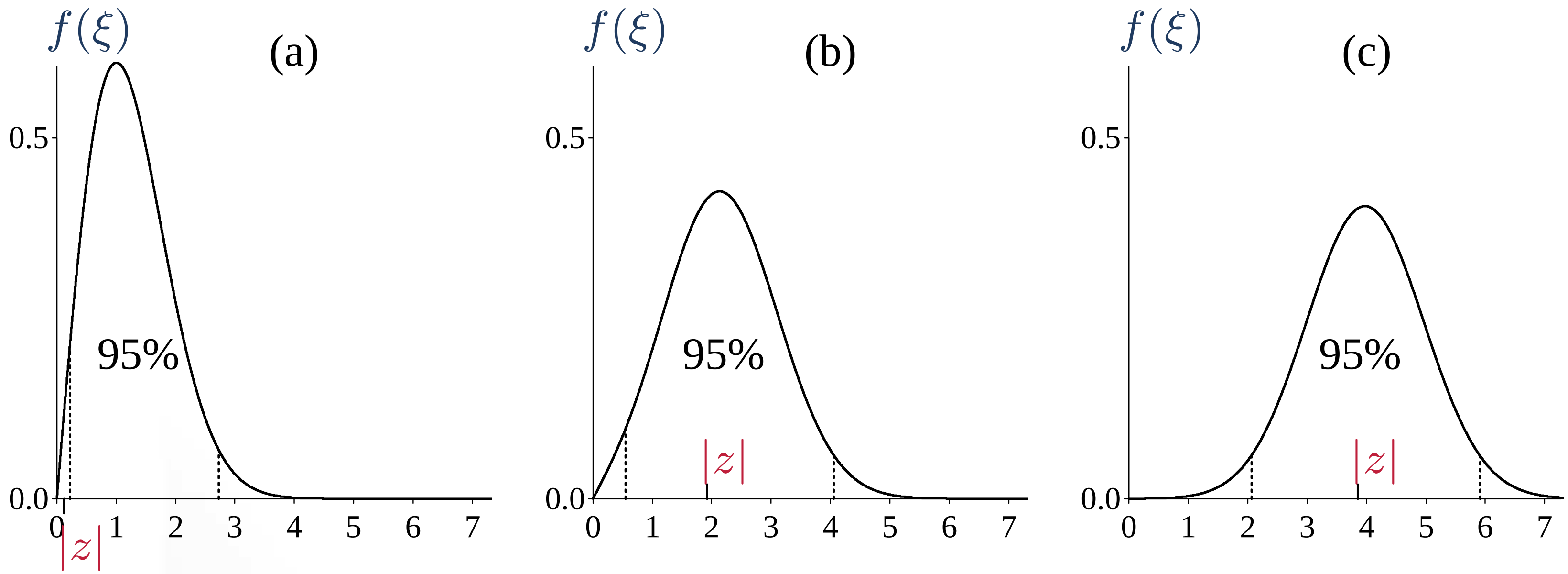


SUP distributions for $|\Gamma|$

The shape of the MC sample will depend on the observed value $|z|$ even when the measurand does not change

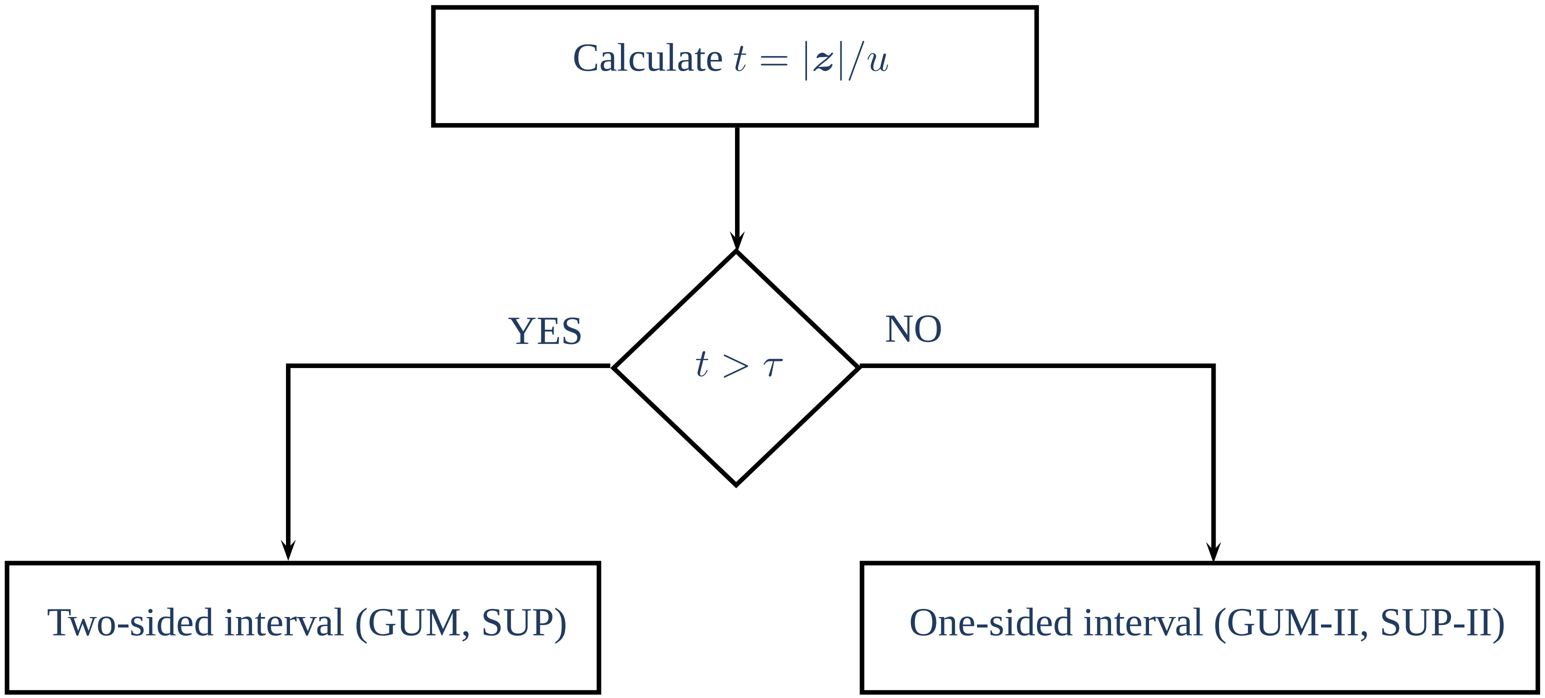
Here are three possible MC distributions for a measurement of $|\mathbf{\Gamma}|$:

(a) $|z| = 0.12$, **(b)** $|z| = 1.92$ and **(c)** $|z| = 3.85$



Can we change a standard method?

Methods can be made adaptive, taking $|\boldsymbol{z}| = 0$ as a lower limit when the data suggest that $\boldsymbol{\Gamma}$ is small



What is the best value for τ?

Choosing τ for measurements of $|\Gamma|$

Simulations can evaluate different choices of τ

Successes out of 1000 for SUP-II

Γ/u	$\tau = 0$	$\tau = 1$	$\tau = 2$	$\tau = 2.5$	$\tau = 3$
0.1	0	369	874	937	993
0.2	0	377	858	947	991
0.5	760	781	830	945	981
1	907	897	910	919	960
2	958	956	941	951	941
5	937	958	957	952	938
10	956	953	948	959	951

Successes out of 1000 for GUM-II

Γ/u	$\tau = 0$	$\tau = 1$	$\tau = 2$	$\tau = 2.5$	$\tau = 3$
0.1	874	875	881	937	993
0.2	899	875	889	947	991
0.5	935	928	935	945	981
1	946	939	953	957	960
2	967	960	956	954	943
5	942	956	956	954	952
10	956	953	948	959	951

Choosing τ for measurements of $|\Gamma|^2$

Successes out of 1000 for SUP-II

$\mathbf{\Gamma}/u$	$\tau = 0$	$\tau = 1$	$\tau = 2$	$\tau = 2.5$	$\tau = 3$
0.1	0	398	865	961	989
0.2	0	397	859	950	989
0.5	759	806	825	936	983
1	911	904	922	886	962
2	932	950	942	951	949
5	951	943	949	964	957
10	960	948	946	955	950

- Note the dip in performance when $\mathbf{\Gamma}/u \approx 1$
- No changes seen when GUM-II is applied to $|\mathbf{\Gamma}|^2$. Success-rates were already very high, suggesting wide intervals

Interval widths are also important

We seek the most accurate results (for a given level of confidence)

Mean widths from 1000 simulations

$\mathbf{\Gamma}/u$	$\|\mathbf{\Gamma}\|$		$\|\mathbf{\Gamma}\|^2$	
	GUM-II	SUP-II	GUM-II	SUP-II
0.1	2.9	3.2	6.3	10.5
0.2	2.9	3.2	6.4	10.6
0.5	2.9	3.2	6.9	11.0
1.0	3.1	3.3	8.4	12.3
2.0	3.6	3.6	15.1	17.9
5.0	3.9	3.9	40.6	41.0
10.0	3.9	3.9	78.7	78.8

- GUM-II is never worse than SUP-II (often better for $|\mathbf{\Gamma}|^2$)
- GUM-II had higher coverage for measurements of $|\mathbf{\Gamma}|^2$

Conclusions

- Performance problems hard to anticipate
- Evaluate scenarios of interest using simulations
- Long-run success rate is a performance measure
- ‘Common-sense’ adaptations can be verified