

# Leistungsmessungen Benchmarking

Henri Kordt, 15.06.2020

# Gliederung

- Einstieg:
  - Motivation zur Leistungsmessung
  - wie Benchmarks
  - Benchmark auswertung
- Statistische Einführung:
  - Statistische Größen und Werte
- Torsten Hoeflers paper
  - 12 Regeln

# Motivation zur Leistungsmessung

- Was kann mein Produkt?
- Welches Produkt ist schneller, besser, effizienter?
- Wo kann ich verbessern?
- HPC

# Wie Benchmarken?

- Was/Wie soll gemessen werden?
  - FLOPS
  - LINPACK
  - Ausführungszeiten/Zugriffszeiten
- I/O bound
  - IOR
- CPU bound
- HPCG
- Wie viele Tests?
- Ergebnisse auswerten

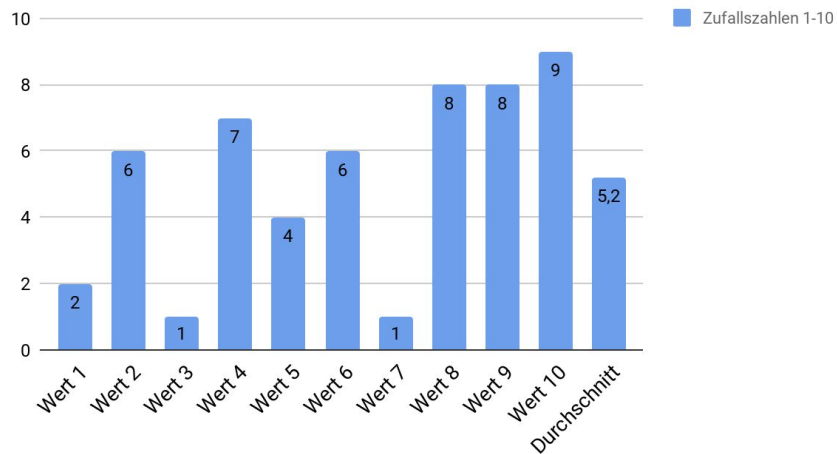
# Benchmark auswertung

- Arithmetisches Mittel
- Median
- Varianz & Standardabweichung
- Konfidenzintervalle.

# Arithmetisches Mittel

$$\bar{x} = \frac{x_1 + x_2 + x_3 \dots + x_n}{n}$$

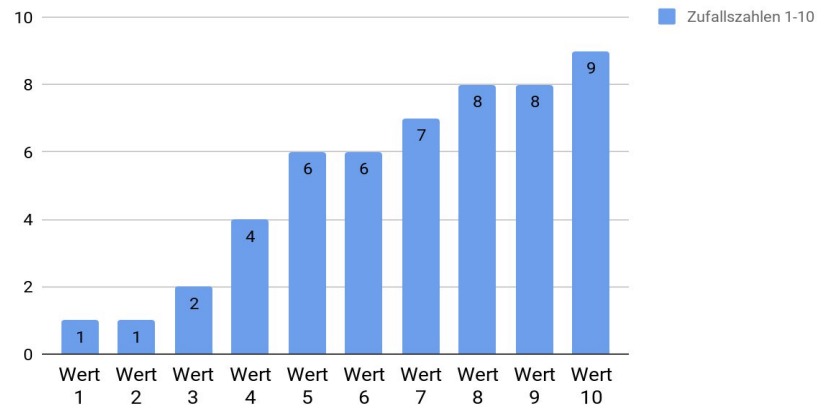
Arithmetisches Mittel Bsp.1



# Median

$$[x_1, \dots, x_{n+1}, \dots, x_{2 \cdot n+1}]$$

Median



# Varianz & Standardabweichung

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}$$

$$s = \sqrt{s^2}$$

# Konfidenzintervall

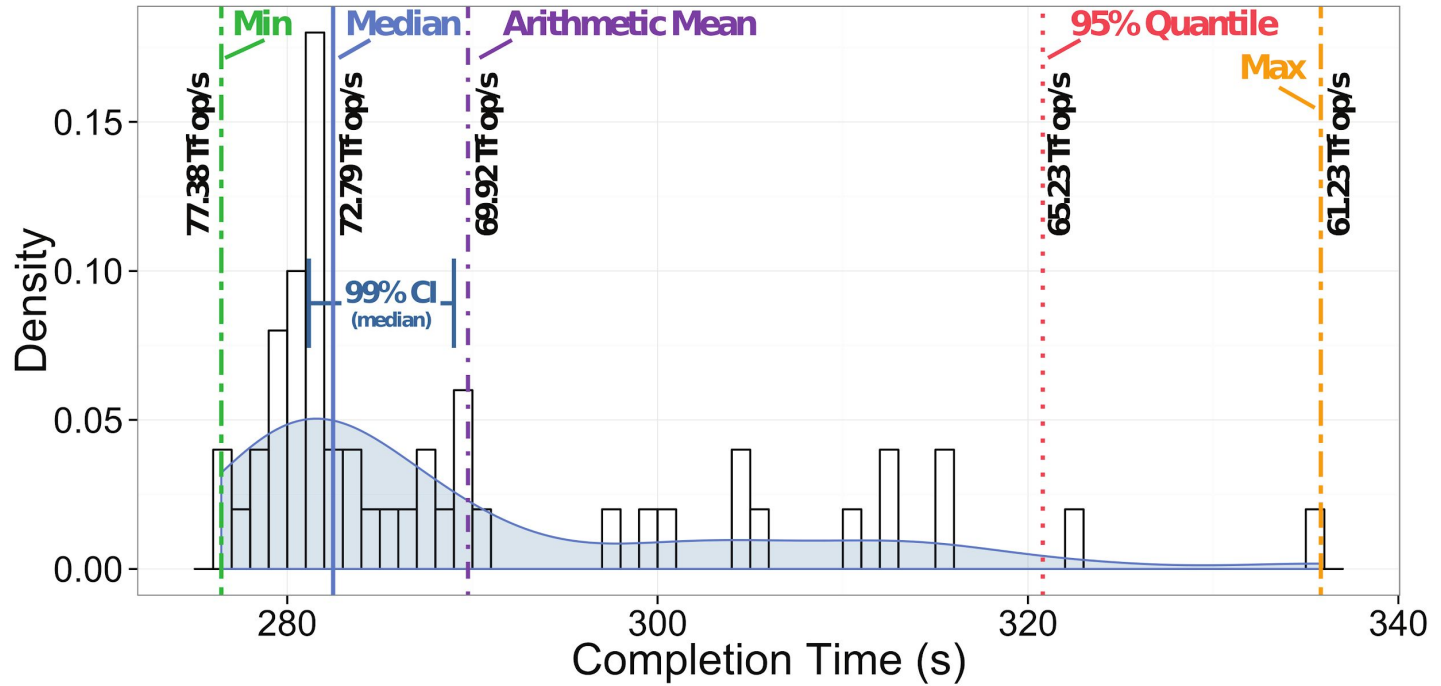
- Normalverteilung

$$\bar{x} \pm z_{\left(1-\frac{\alpha}{2}\right)} \frac{s}{\sqrt{n}}$$

- Central limit Theorem
- Bootstrapping



Grafikquelle 1:



# Scientific Benchmarking of Parallel Computing Systems

Twelve ways to tell the masses when reporting performance results

- Fehler bei der Datenauswertung
- 12 Regeln
- Torsten Hoefler ETH Zürich

## Rule 1:

When publishing parallel speedup, report if the base case is a single parallel process or best serial execution, as well as the absolute execution performance of the base case

- Zu jedem Verhältnis auch absolute Werte angeben!

Rule 2: Specify the reason for only reporting subsets of standard benchmarks or applications or not using all system resources.

- Gib immer die gesamten Ergebnisse an
- Begründe warum nur bestimmte Bereiche gebenchmarkt werden

Rule 3: Use the arithmetic mean only for summarizing costs. Use the harmonic mean for summarizing rates.

Harmonisches Mittel:

$$\frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}} = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

Rates:

- Einheit pro Einheit
- FLOPS

Rule 4: Avoid summarizing ratios; summarize the costs or rates that the ratios base on instead. Only if these are not available use the geometric mean for summarizing ratios

Geometrisches Mittel:

$$\sqrt[n]{x_1 \cdot \dots \cdot x_n}$$

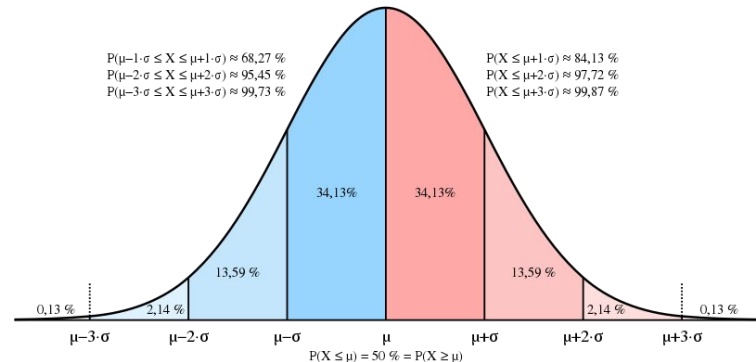
**Rule 5: Report if the measurement values are deterministic. For nondeterministic data, report confidence intervals of the measurement.**

- Leistungsmessungen nicht deterministisch
- nutze statistische Größen um Genauigkeit aufzuzeigen

# Rule 6: Do not assume normality of collected data (e.g., based on the number of samples) without diagnostic checking.

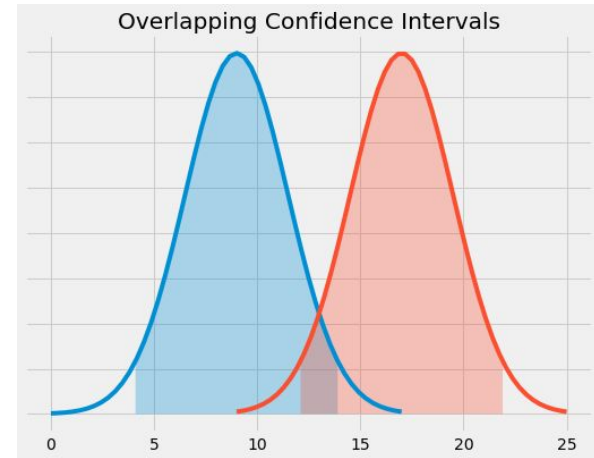
- Nicht unbegründet eine Normalverteilung der Testergebnisse annehmen

Grafikquelle 2:





Rule 7: Compare nondeterministic data in a statistically sound way, e.g., using non-overlapping confidence intervals or ANOVA.



Grafikquelle 3:

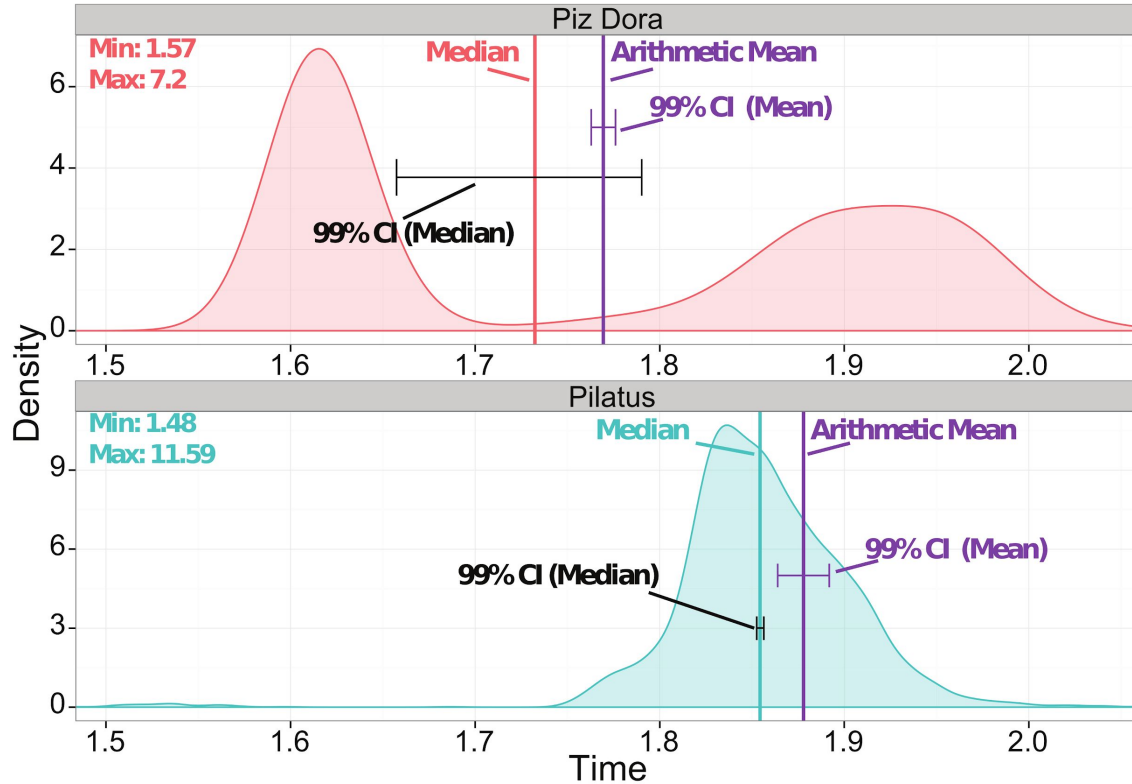
# Anova & T-tests

- Tests um festzustellen, ob sich Stichproben Signifikant unterscheiden
  1. Varianz aller Stichproben zusammen berechnen
  2. Varianz der Stichproben durchschnitt berechnen
  3. Varianz der Durchschnitte durch Varianz der Stichproben teilen
  4. vergleichen mit Testniveau abhängig von Freiheitsgraden und gewünschter Genauigkeit

$$\frac{v_{\bar{x}}}{v_x}$$

Rule 8: Carefully investigate if measures of central tendency such as mean or median are useful to report. Some problems, such as worst-case latency, may require other percentiles.

Grafikquelle: 4



Rule 9: Document all varying factors and their levels as well as the complete experimental setup (e.g., software, hardware, techniques) to facilitate reproducibility and provide interpretability.

- Dokumentiere so viel wie sinnvoll/möglich

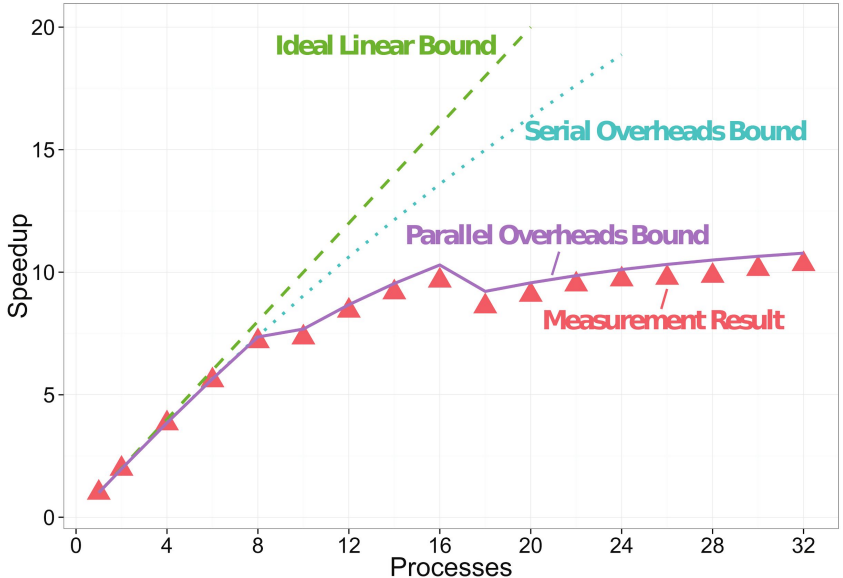
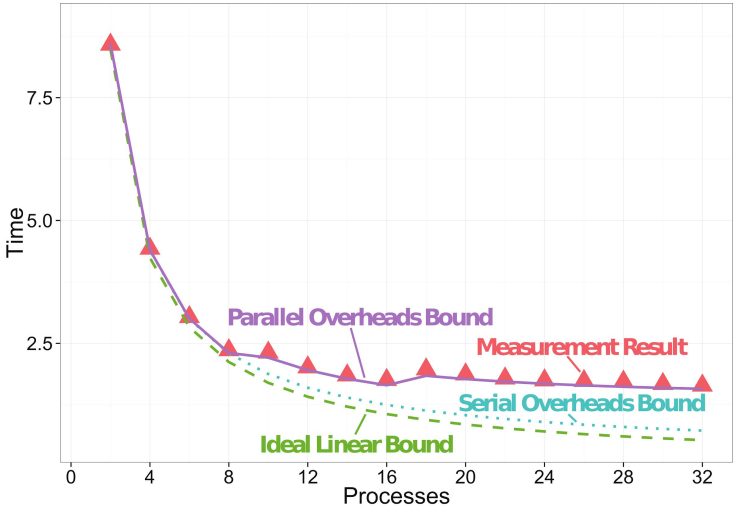
Rule 10: For parallel time measurements, report all measurement, (optional) synchronization, and summarization techniques.

- Wie wurden Messungen durchgeführt
- Wie wurden Messungen Synchronisiert

Rule 11: If possible, show upper performance bounds to facilitate interpretability of the measured results.

- Zeige optimale Graphen zur erleichterten Einordnung

Grafikquelle: 5





Rule 12: Plot as much information as needed to interpret the experimental results. Only connect measurements by lines if they indicate trends and the interpolation is valid

- Stelle so viele Informationen wie möglich graphisch dar

# Grafikquellen:

- Grafik 2: <https://glossar.hs-augsburg.de/Normalverteilung>
- Grafik 3: <https://medium.com/@kangeugine/overlapping-and-difference-confidence-intervals-d163a86b3a00>
- Grafik 1, 4, 5: <https://hlor.inf.ethz.ch/publications/img/hoefler-scientific-benchmarking.pdf>

# Quellen:

- <https://hlor.inf.ethz.ch/publications/img/hoefler-scientific-benchmarking.pdf>
- <https://www.statistikhq.de/konfidenzintervall.php#:~:text=Mit%20s%20%3D%20Standardabweichung%20und%20z.betr%C3%A4gt%20CE%B1%20gleich%20%2C05>
- <https://www.netlib.org/benchmark/hpl/>
- <https://www.hpcg-benchmark.org/>
- <https://towardsdatascience.com/on-average-youre-using-the-wrong-average-geometric-harmonic-means-in-data-analysis-2a703e21ea0>
- [https://en.wikipedia.org/wiki/HPCG\\_benchmark](https://en.wikipedia.org/wiki/HPCG_benchmark)
- <https://buildmedia.readthedocs.org/media/pdf/ior/latest/ior.pdf>
- [https://beza1e1.tuxen.de/articles/basic\\_benchmark\\_statistics.html](https://beza1e1.tuxen.de/articles/basic_benchmark_statistics.html)
- <https://www.youtube.com/watch?v=Bmi7w-vKSCs>
- <https://de.wikipedia.org/wiki/Bootstrapping-Verfahren>
- [https://de.wikipedia.org/wiki/Zentraler\\_Grenzwertsatz](https://de.wikipedia.org/wiki/Zentraler_Grenzwertsatz)
- <https://statisticsbyjim.com/basics/central-limit-theorem/#:~:text=The%20central%20limit%20theorem%20in.variable's%20distribution%20in%20the%20population>