DATA PRESENTATION

Gaia Maselli
maselli@di.uniroma1.it

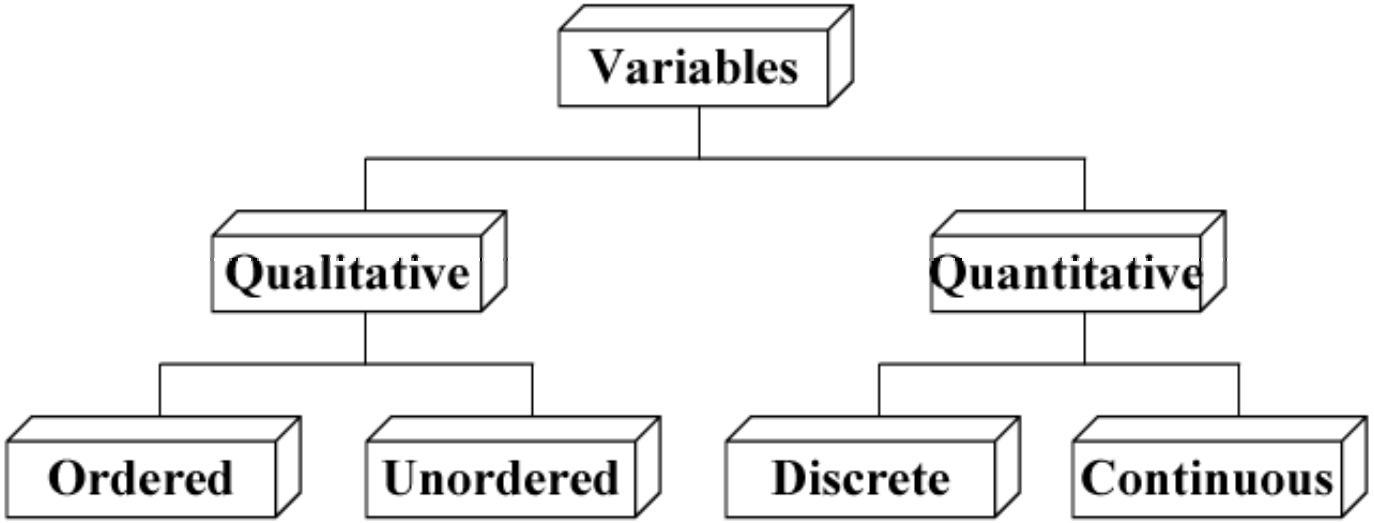
Overview

- The art of data presentation
  - Guidelines
  - Common mistakes

Data presentation

- One of the important steps in every performance evaluation study is the presentation of final results
- The eventual aim of every performance analysis is to help in decision making
- An analysis whose results cannot be understood by the decision makers is as good as one that was never performed
- The analysis has to be presented as clearly and simply as possible
- Graphic charts are commonly used in presenting performance results
  - A picture is worth a thousand words
  - A graphic chart saves reader' time and present information concisely
  - Figures allow to quickly grasp the main points of the study and read the text only for details

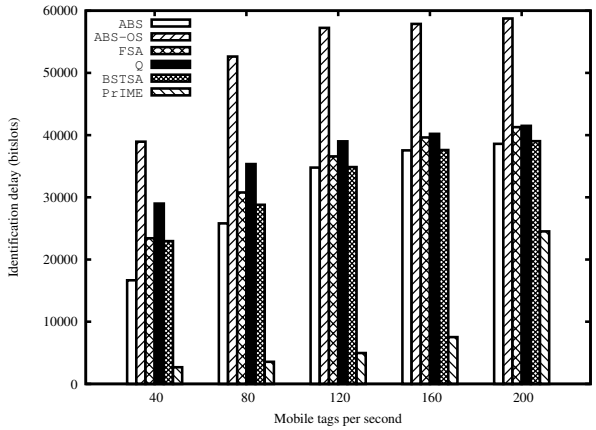
Types of variables



- Qualitative: have states, levels, or categories
- Quantitative: levels can be expressed numerically
  - Discrete: all values of the variable can be counted or put into a one-to-one correspondence with some subset or all the set of positive integers
  - Continuous: can take uncountably infinite values

Discrete vs. continuous

- The type of graphic chart to be used depends upon the type of variable
- If x is a discrete variable ➔ column or bar chart
- If x is a continuous variable ➔ line chart



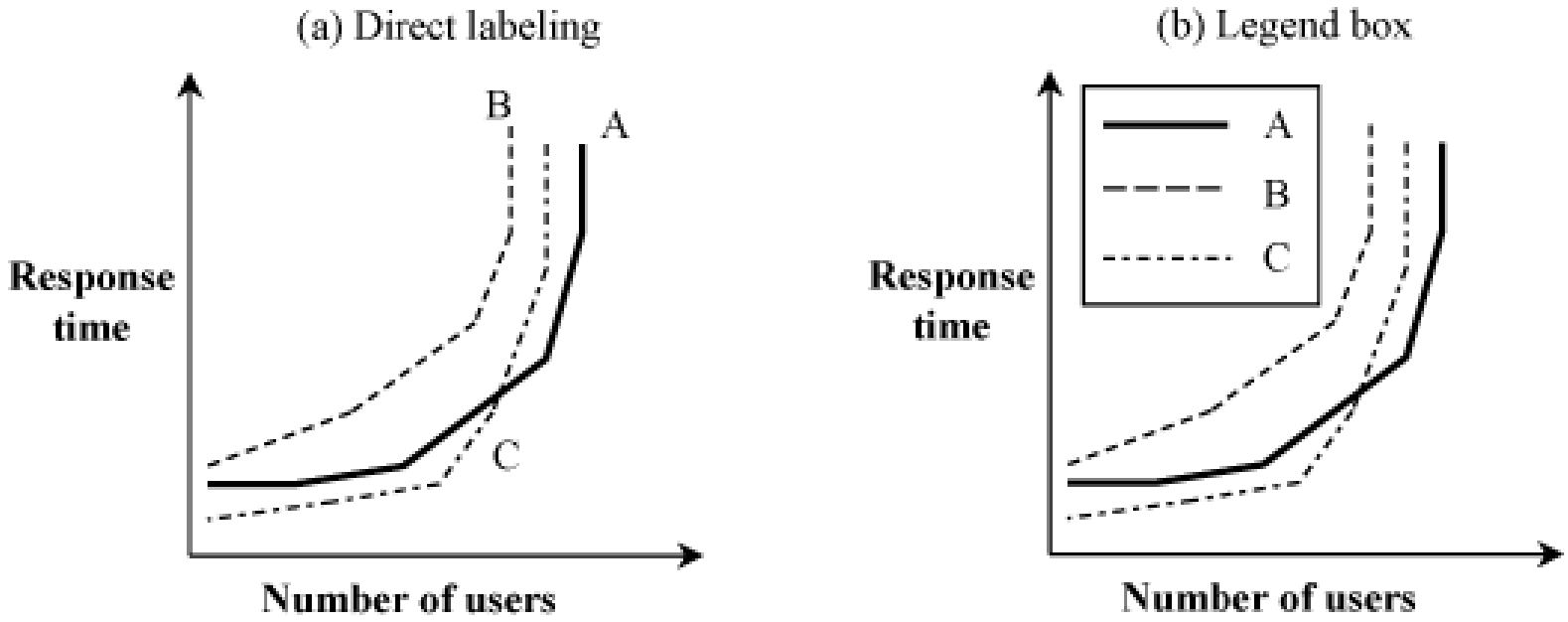
Command-line driven graphing utility

- http://www.gnuplot.info
- http://gnuplot.sourceforge.net/demo/

Guidelines
(with exceptions depending on the context…)

1. *Require minimum effort from the reader*

   Direct labeling vs. legend box



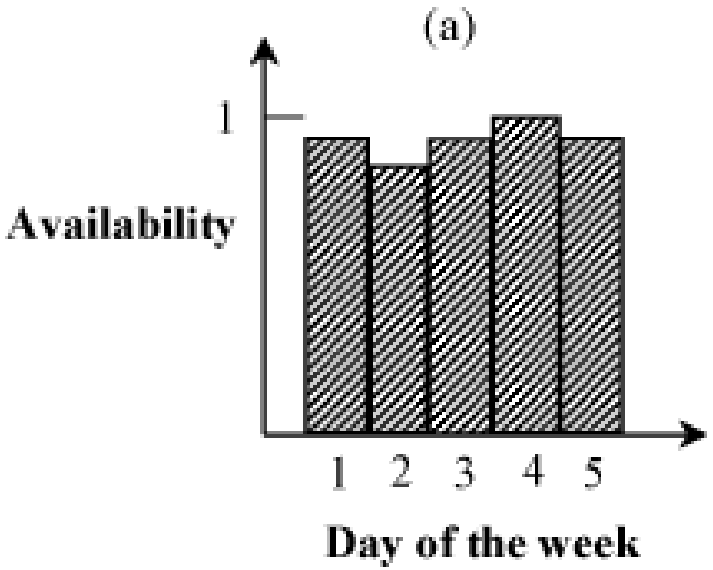
Direct labeling is preferable particularly if the number of curves is large

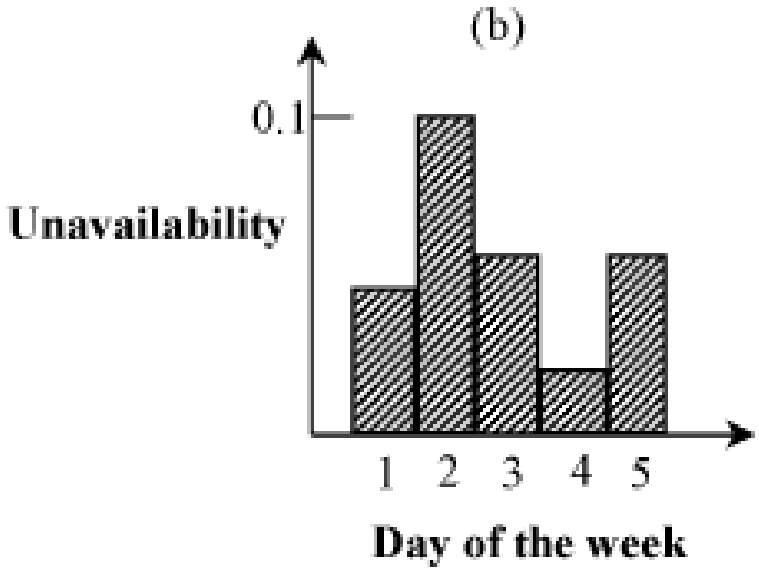
Guidelines (cont)

2. *Maximize information*: there should be sufficient information on the graph to make it self-sufficient
   - Use key words in place of symbols
   - The axes labels should be as informative as possible
   - Include units in the labels
   - Example: “Daily CPU usage (sec.)”
     is better than “CPU usage”
     “CPU time in seconds”

Guidelines (cont)

3. *Minimize ink*: present as much information as possible with as little ink as possible
   - No grid lines unless they are required for to accurately read the values
   - Give more information for the same data is preferable





Guidelines (cont)

4. *Use commonly accepted practices* (present what people expect)

   Most people expect

   - the origin to be (0,0)
   - the independent variable or cause to be plotted along the x-axis
   - The dependent variable or effect to be plotted along the y-axis
   - Scales to be linear
   - Scales to increase left to right and bottom to top

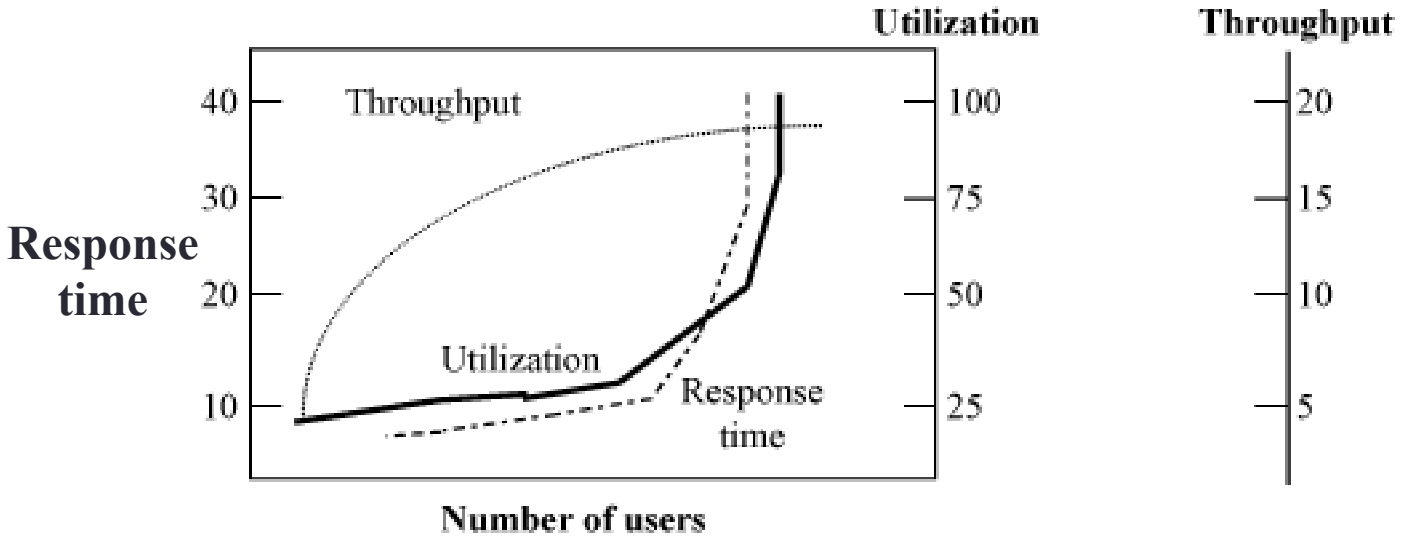
5. *Avoid ambiguity*
   - Do not plot multiple variables in the same chart

N.B. see checklist in box 10.1 (Jain's book) The checklist is arranged so that a "yes" answer to each question, in general, leads to a better graph.

In practice it is necessary to make several trials before arriving at the final graph

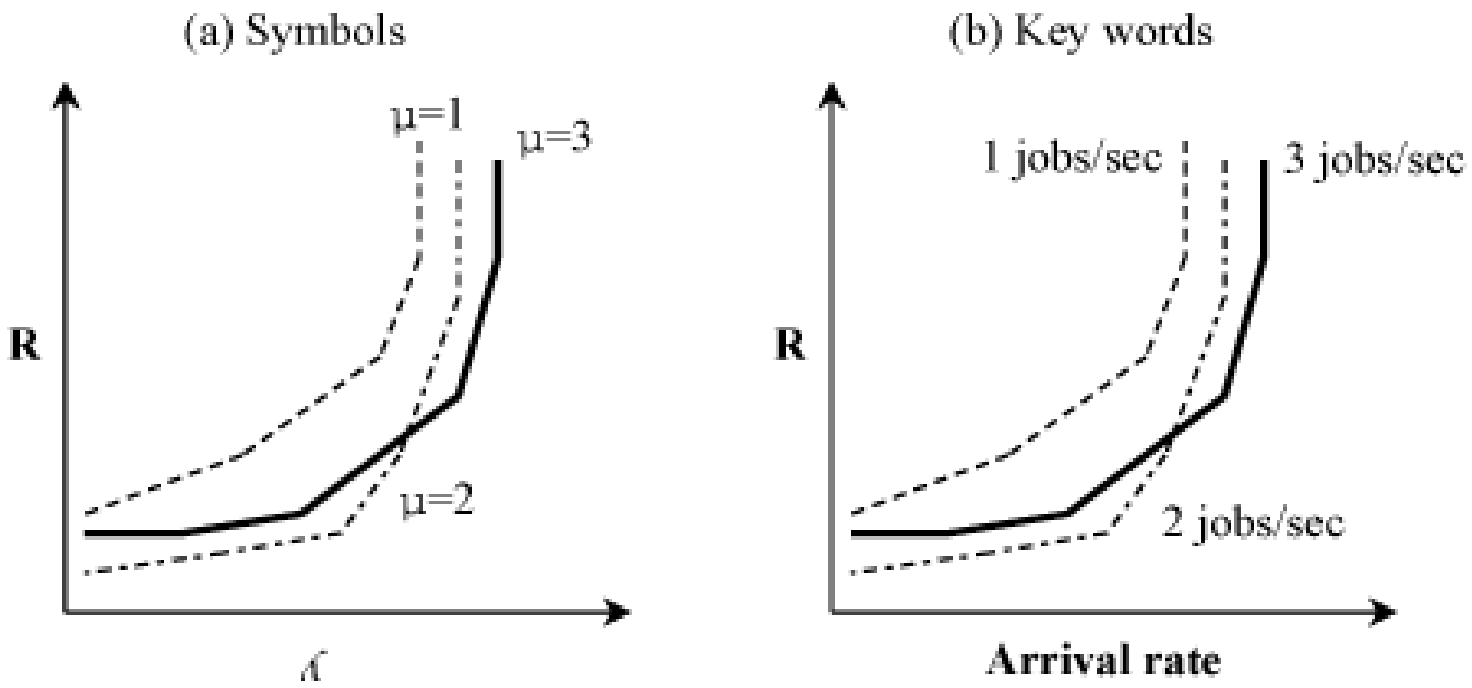
Common mistakes in preparing charts

- *Presenting too many alternatives on a single chart*
  Max 5 to 7 messages => Max 6 curves in a line charts, no more than 10 bars in a bar chart, max 8 components in a pie chart
- *Presenting many y variables on a single chart*



Common mistakes in preparing charts (cont)

- *Using symbols in place of text*



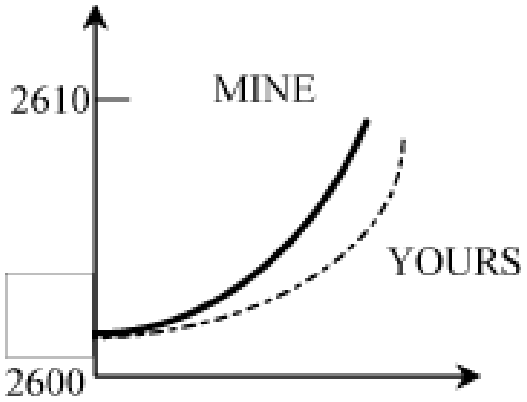
Common mistakes in preparing charts (cont)

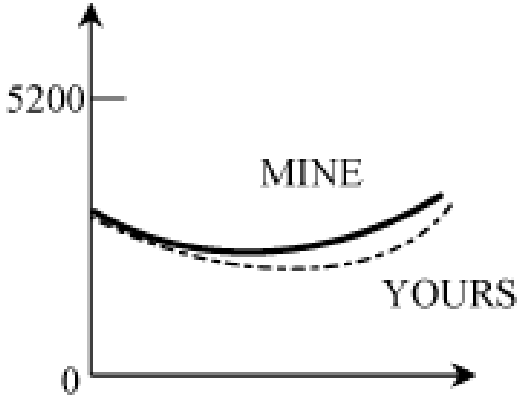
- *Placing extraneous information on the chart*: grid lines, granularity of the grid lines (grid lines on a line chart should be there only if the reader is expected to read the values precisely)
- *Selecting scale ranges improperly*: automatic selection by programs may not be appropriate
- *Using a line chart in place of column chart*: line ➔ Continuity

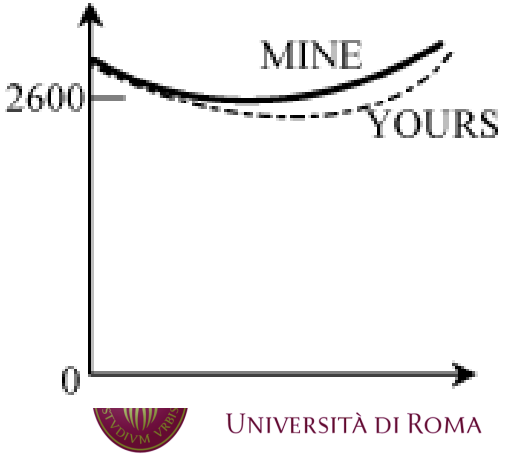
the lines joining successive points on a line chart signify the fact that the intermediate values can be approximately interpolated

Pictorial games

- *Using nonzero origin to emphasize the difference*

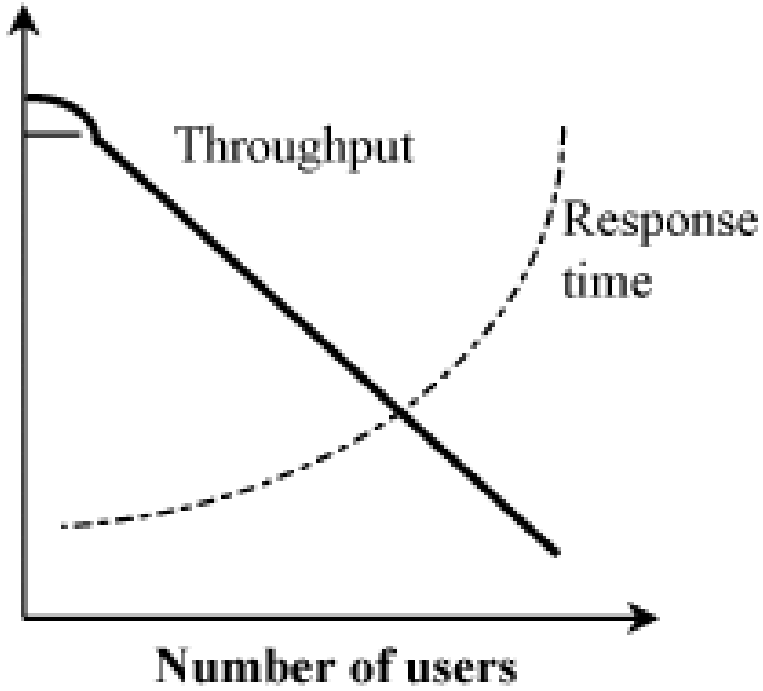






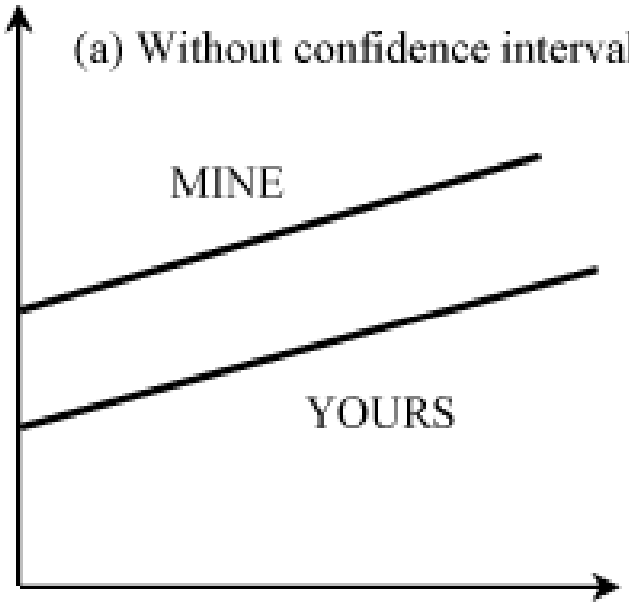
Pictorial games

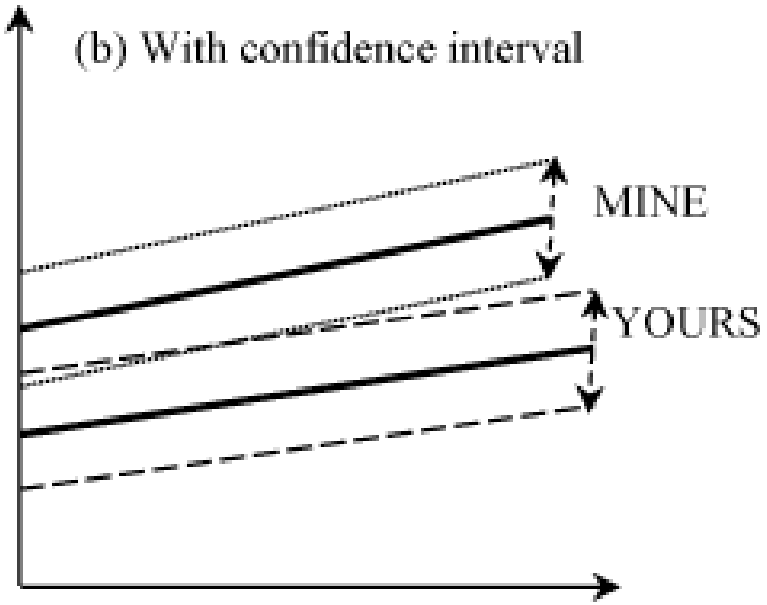
- *Using double-whammy graph for dramatization*: two curves on the same graph can have twice as much impact as one



Pictorial games

- *Plotting random quantities without showing confidence intervals*





- Overlapping confidence intervals are generally enough to deduce that the two random quantities are statistically indifferent