Animation in Statistics: Dynamic Graphics for Statistical Models and Practical Applications

Yihui XIE

School of Statistics, Renmin University of China

Dec 12, 2007
Contents

1 Why Employ Animations?
   - Interesting
   - Intuitional

2 How to Make Animations?
   - R Graphical Devices
   - HTML & JavaScript

3 Where to Use Animations?
   - Iterative Algorithms
   - Random Experiments

4 What’s More?
   - Statistical Theories
   - Practical Applications
Interest is the Best Teacher
Nobody likes dull stuff...

Buffon’s Needle

- Buffon’s Needle is one of the oldest problems in the field of geometrical probability; It was first stated in 1777.
- It involves dropping a needle on a lined sheet of paper and determining the probability of the needle crossing one of the lines on the page.
- The remarkable result is that the probability is directly related to the value of $\pi$.
- From frequency to probability.
Interest is the Best Teacher (cont’d)

Simulation of Buffon’s Needle

\[ y = \frac{L}{2} \sin(\phi) \]

Dropping times

\[ \frac{\pi}{\lambda} \]

\[ \frac{\pi}{\lambda} \]
Two Pictures are Worth 2000 Words

You don’t like those abstract theories, do you?
Produce Animations inside R

- The R package grDevices has offered a variety of graphics devices.
- We may just create animations in the Windows graphics devices (Windows) or X Window System graphics devices (Linux) or MacOS X Quartz devices (MacOS X).
- The most critical elements are loops and `Sys.sleep()`.
- It’s convenient to produce single image files – there are several choices such as PNG, JPEG, BMP, PDF, PS, \TeX/\LaTeX and WMF, etc.
Produce Animations outside R

- It’s convenient to produce single image files using R – there are several choices such as PNG, JPEG, BMP, PDF, PS, \TeX/\LaTeX and WMF, etc.
- We may use JavaScript to animate these image frames – quite naive way.

\[ | / — — \backslash | \]
\[ . . \circ \circ O \]
- Many examples in this talk have been implemented in the R package ‘animation’; you may download it from CRAN.
- The package ‘animation’ has offered both ways of creating animations.
- See \texttt{buffon.needle()}, \texttt{flip.coin()}, \texttt{kmeans.ani()}, \texttt{boot.iid()}, \texttt{knn.ani()}, ...
K-Means Cluster Algorithm
Flipping Coins

Where to Use Animations?

Random Experiments

School of Statistics

Yihui XIE

Dec 12, 2007 9 / 16
Probability Theory: Central Limit Theorem

ECDF of $x$

Density Estimation of $x$

Density Estimation of $\bar{x}$
Mathematical Statistics: Confidence Interval

CI: $[ar{x} - z_{\alpha/2}\sigma/\sqrt{n}, \bar{x} + z_{\alpha/2}\sigma/\sqrt{n}]$

Coverage rate: 92% (average: 92.07%)
Machine Learning: Bootstrapping

Demonstration of bootstrapping for i.i.d data

Distribution of bootstrap estimates

What's More? Statistical Theories

School of Statistics
Yihui XIE
Dec 12, 2007
Machine Learning: $k$-Nearest Neighbor Algorithm

Demonstration for kNN Classification

- $X_1$
- $X_2$

○ first class
△ second class

training set
test set
Simulation of A Chemical Experiment
Social-Economic Changes Over Time

What's More?

Practical Applications

Yihui XIE

Dec 12, 2007
Thanks!

A weird name card...

Contact Me

xieyihui(at)gmail.com

http://www.yihui.name