

Statistical Software Repositories for HEP

Ilya Narsky, Caltech

Workshops

- Michigan State U, 2004
 - http://user.pa.msu.edu/linnemann/public/workshop/workshop_summary.pdf
- Fermilab, August 2005
 - organized by Jim Linnemann and Mark Fischler
 - <http://whcdf03.fnal.gov/PHYSTATworkshop>
- Results summarized by Jim Linnemann at Phystat 2005
 - <http://www.physics.ox.ac.uk/phystat05/Talks/LinnemanOxford.pdf>
- Status and plans

Statistical Resources on the Web

- http://www.pa.msu.edu/people/linnemann/stat_resources.html
 - lots of links; probably the most complete list maintained by a physicist
- No central repository for statistics software in HEP
- Repositories maintained by other communities:
 - StatLib at CMU <http://lib.stat.cmu.edu/>
 - StatCodes at Penn State <http://www.astro.psu.edu/statcodes/>
 - and more!!!

A repository for HEP

- Stage I – “a very open archival repository”
 - anyone can upload
 - “submission effort should be less than or comparable to submitting a paper to arxiv.org”
- Stage II – a supervised repository
 - distinguish between archival entries and actively maintained packages
 - user feedback or quality assessment by a designated group of people
 - more restrictive standards for submissions to improve efficiency and portability (naming conventions, package structure, allowed base libraries, writing code for missing functionality etc)

Actions

- Set up a temp repository
 - <http://whcdf03.fnal.gov/PHYSTATworkshop/ContributedSoftware>
 - If you'd like to submit anything, send email to Marc Paterno paterno@fnal.gov with a brief description and tar file
- Propose Fermilab as the sponsor for a permanent repository
 - in the process of discussion with Fermilab management (manpower, security etc)
- Solicit support from collaborations

What we could do at BaBar

- On behalf of BaBar, SWG could send our comments/recommendations
 - What kind of repository would we be interested in? (How much supervision is needed? How restrictive do submission requirements need to be?)
 - linnemann@pa.msu.edu; mf@fnal.gov
- And even if we have none, a letter of support would be useful!

R, the new can-do-it-all tool

- A high-level interpreter
- Free
- Easy to install (from what I hear)
- Choice of many statisticians and practitioners from other fields
- <http://www.r-project.org/>
- Integration of R and Root
 - Adam Lyon (Fermilab) wrote a Root Tree reader for R
 - Rene Brun expressed interest

R, a good thing or a bad thing?

- Good
 - implements many useful things
 - can make it easier to promote well-established methods that are not yet recognized in HEP
- Bad
 - also implements many things that HEP will never need (just look at the list of R modules)
 - unified high-level interface comes with a price – computational inefficiency (I would really love to see some benchmarks!)
 - marriage of two monsters (R and Root) does not necessarily make a genius child
 - harder to integrate in C++ environment than C++ code
 - how hard is it to add new methods?

Summary

- What kind of statistical repository would you like? Make your opinion known!
- Better learn R or the future may find you statistically clueless