Kirchhoff-Insitut für Physik
Physikalisches Insitut

Winter semester 2013-14
KIP CIP Pool (1.401)

Exercises for Statistical Methods in Particle Physics

http://www.physi.uni-heidelberg.de/~nberger/teaching/ws13/statistics/statistics.php

Dr. Niklaus Berger (nberger@physi.uni-heidelberg.de)
Dr. Oleg Brandt (obrandt@kip.uni-heidelberg.de)

Exercise 11: Trees and Fishers

13. January 2014
Hand-in solutions by 14:00, 19. January 2014

Please send your solutions to obrandt@kip.uni-heidelberg.de by 19.1.2014 14:00, punctually. Make sure that you use *SMIPP:Exercise11* as subject line. If plots are requested, please include print statements to produce pdf files in your code, and provide the plots separately. Please add comments to your source code explaining the steps. Test macros and programs before sending them off...

1 Trees, leaves, and simple cuts

On the course website, you find a ROOT file, named `ex_11.root`. It contains an *n-tuple*, i.e. a vector of numbers (which can be of different type) for each measured event. For each event, our n-tuple contains two measured quantities x and y, as well as a variable *type*, which is 0 for the signal Monte Carlo simulations, 1 for the background Monte Carlo simulations, and 2 for the "data".
Explore this file in interactive ROOT. You easily access files by creating a `TBrowser` object on the command line. You can also draw the distributions for each variable by clicking on it in the `TBrowser`. If you want to draw several variables on the same plot or apply cuts, you can also use a tree viewer, which can be launched by right-clicking on the tree and selecting "StartViewer". In the viewer, you can drag-and-drop variables onto the x, y, z axes or also into the cut string (the one with the scissors), the syntax to apply cuts is identical to the examples below. You can also draw distributions using the command line, giving the name of the tree (in this case `searchTree`) and using the `Draw` command, e.g.

```
$ searchTree->Draw("x");
```

You can plot correlations by specifying two variables:

```
$ searchTree->Draw("x:y");
```

You can apply cuts and, more generally, weights in the second argument (for ROOT, a cut is a special case of a binary weight):

```
$ searchTree->Draw("x:y","type == 2");
```

and you can specify drawing option(s) in the third argument:

```
$ searchTree->Draw("x:y","type==2","BOX");
```

Explore the given file.
If you want to read the file in a script or program, do something like the following:

```
double x,y ;
int type ;

TFile* f = new TFile("ex_11.root","READ") ;
TTree* tree = (TTree*)(f->Get("searchTree")) ;

tree->SetBranchAddress("x",&x) ;
tree->SetBranchAddress("y",&y) ;
tree->SetBranchAddress("type",&type) ;

for( int event=0, event_max=tree->GetEntries() ; event!=event_max ; ++event ) {
  tree->GetEntry(event) ;

  // < Do stuff with x,y,type...>
}
```

Try to separate signal and background via a cuts in x. Similarly, do the same in y. For a signal efficiency of 50%, what purity do you obtain for your two cuts, respectively? Now apply both cuts simultaneously, which efficiency do you obtain now? Is that expected? And which purity? (ex_11_1.txt)
For the above scenarios, please provide the plots ex_11_1_x.pdf and ex_11_1_y.pdf which show the distribution of signal and background compared to data in x and y, respectively, for a signal efficiency of 50%. Please indicate the cuts with TLines. Similarly, show ex_11_1_xy.pdf in the $x \times y$ plane with the above cuts (for clarity, you can drop the data distribution from the 2d plot). For MC-data comparisons, assume that the integrated luminosity of the signal and background MC samples corresponds exactly to that of data.

2 The Fisher discriminant: the two-dimensional case

Construct the Fisher discriminant in x and y, using the MC events. For an efficiency of 50%, what purity do you obtain?
Analogously to Problem 11.1, please provide the plot ex_11_2_xy.pdf where you show the signal and background MC together with the direction of the Fisher discriminant, as well as ex_11_2_f.pdf which shows the Fisher discriminant for signal and background MC alongside the data.

Hints:
Similarly to Problem 10, you can use TMatrixD in ROOT for the matrix inversion and other linear algebra operations.
A good description of the Fisher discriminant method can be found in Ref. [1] or in the original publication [2].

Literatur

[1] http://research.cs.tamu.edu/prism/lectures/pr/pr_l10.pdf.

[2] R.A. Fisher, Annals Eugen., **7**, pp. 179-188 (1936), available from: http://hdl.handle.net/2440/15227.