Kyle Cranmer
Draft 2, April 1, 2011

HistFactory Likelihood

1 The Likelihood Template

The parametrized probability density function constructed by the HistFactory is of a concrete form, but sufficiently flexible to describe many analyses based on template histograms. In general, the HistFactory produces probability density functions of the form

$$\mathcal{P}(n_m, a_i|\mu, \alpha_i) = \prod_{m\in\text{bins}} \text{Pois}(n_m|\nu_m) \cdot G(L_0|L, \Delta_L) \prod_{i\in\text{Syst}}^{5} N(a_i|\alpha_i, 1) \tag{1}$$

where m is an index over the bins of the template histograms, i is an index over systematic effects, n_m is the observed number of events in bin m, $N(a_i|\alpha_i, 1)$ is the normal distribution for the auxiliary measurement a_i that constrains the nuisance parameter α_i and ν_m is the expected number of events in bin m given by

$$\nu_m = \sum_{j\in\text{Samp}} LF_j \eta_j(\boldsymbol{\alpha})\, \sigma_{jm}(\boldsymbol{\alpha}), \tag{2}$$

where F_j is a product of unconstrained normalization factors for sample j

$$F_j = \prod_{n\in\text{NormFact}_j} f_n \tag{3}$$

that typically include the parameter of interest (eg. $\mu = \sigma/\sigma_{SM}$). The term $\eta_j(\boldsymbol{\alpha})$ parametrizes relative changes in the overall normalization, and $\sigma_{jm}(\boldsymbol{\alpha})$ contains the nominal normalization and parametrizes uncertainties in the shape of the distribution of the discriminating variable. Here j is an index of contributions from different processes with $j = 1$ being the signal process. The nuisance parameters α_i are associated to the source of the systematic effect (e.g. the muon momentum resolution uncertainty), while $\eta_j(\boldsymbol{\alpha})$ and $\sigma_{jm}(\boldsymbol{\alpha})$ represent the effect of that uncertainty. The α_i are scaled so that $\alpha_i = 0$ corresponds to the nominal expectation and $\alpha_i = \pm 1$ correspond to the $\pm 1\sigma$ variations of the source, thus $N(\alpha_i)$ is the standard normal distribution. The effect of these variations about the nominal predictions $\eta_j(0) = 1$ and σ^0_{jm} is quantified by dedicated studies that provide $\eta^{\pm}_{ij}$ and $\sigma^{\pm}_{ijm}$, which are then used to form

$$\eta_j(\boldsymbol{\alpha}) = \prod_{i\in\text{Syst}} I(\alpha_i; \eta^+_{ij},\ \eta^-_{ij}) \tag{4}$$

and

$$\sigma_{jm}(\boldsymbol{\alpha}) = \sigma^0_{jm} \prod_{i\in\text{Syst}} I(\alpha_i; \sigma^+_{ijm}/\sigma^0_{jm},\ \sigma^-_{ijm}/\sigma^0_{jm}) \tag{5}$$

with

$$I(\alpha; I^+, I^-) = \begin{cases} 1 + \alpha(I^+ - 1) & \text{if } \alpha > 0 \\ 1 & \text{if } \alpha = 0 \\ 1 - \alpha(I^- - 1) & \text{if } \alpha < 0 \end{cases} \tag{6}$$

enabling piece-wise linear interpolation in the case of asymmetric response to the source of systematic. In the next version of HistFactory, a quadratic interpolation will be available, and may become the default, as it avoids discontinuities in the second derivative that often confuse MINUIT and HESSE.

2 The HistFactory XML Schema

Note, when using the `HistFactory` the production modes l and backgrounds j correspond to a single XML `Sample` element. The `HistoName` attribute inside each sample element specifies the histogram with the σ^0_{ijm}. The index $j='J'$ is set by the `Name` attribute of the `Sample` element (eg. `<Sample Name='J'>`). Between the open `<Sample>` and close `</Sample>` one can add

- An `OverallSys` element where the `Name='I'` attribute identifies which α_I is the source of the systematic and implies that the Gaussian constraint $N(a_i|\alpha_I, 1)$ is present. The `High` attribute corresponds to η^+_{IJ}, eg when the source of the systematic is at $+1\sigma$ and $\alpha_I = 1$. Similarly, the `Low` attribute corresponds to η^-_{IJ}, eg when the source of the systematic is at -1σ and $\alpha_I = -1$. The nominal value is $\eta^0_{IJ} = 1$ for the ovreall systematics. The distinction between the sign of the source α and the effect η allows one to have anti-correlated systematics. The `HistFactory` is able to deal with asymmetric uncertainties as well, by using a piece-wise linear interpolation for the $\alpha_I > 0$ and $\alpha_I < 0$ regions.

- A `NormFactor` element is used to introduce an overall constant factor into the expected number of events. In the example below, the term $\mu = \sigma/\sigma_{SM}$ corresponds to the line `<NormFactor Name='SigXsecOverSM'>`. In this case, the histograms were normalized to unity, so additional `NormFactor` elements were used to give the overall cross-sections σ_J.

- A `HistoSys` element is used to introduce shape systematics and the `HistoNameHigh` and `HistoNameLow` attributes have the variational histograms σ^+_{ijm} and σ^-_{ijm} corresponding to $\alpha_i = +1$ and $\alpha = -1$, respectively.

Below is an example XML file for the electron channel.

```
<!DOCTYPE Channel  SYSTEM 'HistFactorySchema.dtd'>
  <Channel Name="channelEle" InputFile="./data/central_Ele_5jet_inc_35invpb.root" HistoName="" >
    <!--<Data HistoName="data" HistoPath="" />-->
    <Sample Name="bbAtautau120" HistoPath="" NormalizeByTheory="True" HistoName="bbAtautau120All">
      <OverallSys Name="JES" High="1.05" Low="0.95"/>
      <OverallSys Name="EVTEFF" High="1.122" Low="0.878"/>
      <OverallSys Name="bbAtautau" High="1.15" Low="0.85"/>
      <NormFactor Name="NEle_bbAtautau120" Val=".83202" Low=".83202" High=".83202" Const="True" />
      <NormFactor Name="SigXsecOverSM" Val="0" Low="-10." High="30." Const="True" />
    </Sample>
    <Sample Name="Atautau120" HistoPath="" NormalizeByTheory="True" HistoName="Atautau120All">
      <OverallSys Name="JES" High="1.05" Low="0.95"/>
      <OverallSys Name="EVTEFF" High="1.122" Low="0.878"/>
      <OverallSys Name="Atautau" High="1.15" Low="0.85"/>
      <NormFactor Name="NEle_Atautau120" Val=".24224" Low=".24224" High=".24224" Const="True" />
      <NormFactor Name="SigXsecOverSM" Val="0" Low="-10." High="30." Const="True" />
    </Sample>
    <Sample Name="bbAtautau130" HistoPath="" NormalizeByTheory="True" HistoName="bbAtautau130All">
      <OverallSys Name="JES" High="1.05" Low="0.95"/>
      <OverallSys Name="EVTEFF" High="1.122" Low="0.878"/>
      <OverallSys Name="bbAtautau" High="1.15" Low="0.85"/>
      <NormFactor Name="NEle_bbAtautau130" Val=".01767" Low=".01767" High=".01767" Const="True" />
      <NormFactor Name="SigXsecOverSM" Val="0" Low="-10." High="30." Const="True" />
    </Sample>
    <Sample Name="Atautau130" HistoPath="" NormalizeByTheory="True" HistoName="Atautau130All">
      <OverallSys Name="JES" High="1.05" Low="0.95"/>
      <OverallSys Name="EVTEFF" High="1.122" Low="0.878"/>
      <OverallSys Name="Atautau" High="1.15" Low="0.85"/>
      <NormFactor Name="NEle_Atautau130" Val=".02441" Low=".02441" High=".02441" Const="True" />
      <NormFactor Name="SigXsecOverSM" Val="0" Low="-10." High="30." Const="True" />
    </Sample>
    <Sample Name="Ztautau" HistoPath="" NormalizeByTheory="True" HistoName="ZtautauAll">
      <OverallSys Name="JES" High="1.05" Low="0.95"/>
      <OverallSys Name="EVTEFF" High="1.122" Low="0.878"/>
      <OverallSys Name="Alpgen" High="1.131" Low="0.869"/>
      <OverallSys Name="Ztautau" High="1.15" Low="0.85"/>
      <NormFactor Name="NEle_Ztautau" Val="1.26818" Low="1.26818" High="1.26818" Const="True" />
    </Sample>
   <Sample Name="AddOn" HistoPath="" NormalizeByTheory="False" HistoName="AddOnAll">
      <OverallSys Name="AddOn" High="1.173" Low=".827"/>
      <NormFactor Name="NEle_AddOn" Val=".88267" Low=".88267" High=".88267" Const="True" />
    </Sample>
```

```
    <Sample Name="SameSign" HistoPath="" NormalizeByTheory="False" HistoName="SameSignAll">
      <OverallSys Name="SameSign" High="1.06828" Low=".93172"/>
      <NormFactor Name="NEle_SameSign" Val="4.00568" Low="4.00568" High="4.00568" Const="True" />
    </Sample>
    <Sample Name="Others" HistoPath="" NormalizeByTheory="True" HistoName="OthersAll">
      <OverallSys Name="JES" High="1.05" Low="0.95"/>
      <OverallSys Name="EVTEFF" High="1.122" Low="0.878"/>
      <OverallSys Name="QFAC" High="1.03" Low="0.97"/>
      <OverallSys Name="Alpgen" High="1.131" Low="0.869"/>
      <OverallSys Name="Others" High="1.15" Low="0.85"/>
      <NormFactor Name="NEle_Others" Val=".17949" Low=".17949" High=".17949" Const="True" />
    </Sample>
  </Channel>
```

One can convert this Gaussian constraints into a Poisson/Gamma systematic by adding lines like

```
<ConstraintTerm Type="Gamma" RelativeUncertainty="0.1">JES</ConstraintTerm>
```

to the `Measurement` element. For example:

```
  <Measurement Name="AllSYS" Lumi="35.2" LumiRelErr="0.11" BinLow="0" BinHigh="20" Mode="comb"  ExportOnly="True">
    <POI>SigXsecOverSM</POI>
    <ParamSetting Const="True">NEle_AddOn,NEle_Atautau120,NEle_Atautau130,NEle_Others,
      NEle_SameSign,NEle_Ztautau,NEle_bbAtautau120,NEle_bbAtautau130,NMuo_AddOn,
      NMuo_Atautau120,NMuo_Atautau130,NMuo_Others,NMuo_SameSign,NMuo_Ztautau,NMuo_bbAtautau120,
      NMuo_bbAtautau130
    </ParamSetting>
    <ConstraintTerm Type="Gamma" RelativeUncertainty="0.1">JES</ConstraintTerm>
    <!--<ConstraintTerm Type="LogNormal" RelativeUncertainty="0.1">JES</ConstraintTerm>-->
  </Measurement>
```

3 Usage of the HistFactory

ROOT installation

Download, install, and setup ROOT v5.28 or greater. It is recommended to use one of the patch releases of v5.28 as the ”standard form” described below was not available before the patch releases.

```
cd $ROOTSYS
source bin/thisroot.sh
```

This will setup your `MANPATH` environment variable so that you can use the command line help.

prepareHistFactory

```
man prepareHistFactory
prepareHistFactory
```

The command line executable `prepareHistFactory [dir_name]` is a simple script that prepares a working area (and creates the directory dir_name if specified). Within the directory dir_name, it creates a results/, data/, and config/ directory relative to the given path. It also copies the HistFactorySchema.dtd and example XML files into the config/ directory. Additionally, it copies a root file into the data/ directory for use with the examples. Once this is done, one is ready to run the example hist2workspace input.xml or edit the XML files for a new project.

hist2workspace

```
man hist2workspace
hist2workspace config/example.xml
```

The command line executable `hist2workspace [option] [input xml]` is a utility to create RooFit/RooStats workspace from histograms

OPTIONS:

- `-standard_form` default model (from v5.28.00a and beyond), which creates an extended PDF that interpolates between RooHistFuncs. This is much faster for models with many bins and uses significantly less memory.
- `-number_counting_form` this was the original model in 5.28 (without patches). It uses a Poisson for each bin of the histogram. This can become slow and memory intensive when there are many bins.

4 Usage with RooStats tools

Once one runs `hist2workspace` on an XML file there will be output root and eps files in the results directory. The files are named

```
results/[Prefix]_[Channel]_[Measurement]_model.root
```

where Prefix is specified in the `<Combination>` element in the top-level XML file, for example:

```
<Combination OutputFilePrefix="./results/example" Mode="comb" >
```

Measurement is specified in each of the `<Measurement>` elements in the top-level XML file

```
<Measurement Name="AllSYS" ...>
```

and Channel is "combined" for the combined model, but a model file is exported for each individual channel as well using the name taken from the `<Channel>` element of the corresponding channel's XML file

```
<Channel Name="channelEle" ...>
```

These root files have inside a RooWorkspace which contains a RooDataSet and a ModelConfig that can be used with standard RooStats tools (see for example `$ROOTSYS/tutorials/RooStats/Standard*Demo.C`

```
$ hist2workspace config/example.xml
$ root.exe results/example_combined_GaussExample_model.root
root [0]
Attaching file results/example_combined_GaussExample_model.root as _file0...
root [1] .ls
TFile** results/example_combined_GaussExample_model.root
 TFile* results/example_combined_GaussExample_model.root
  KEY: RooWorkspace combined;1 combined
  KEY: TProcessID ProcessID0;1 1222429a-5b98-11e0-9717-0701a8c0beef

root [2] combined->Print()

RooWorkspace(combined) combined contents

variables
---------
...

p.d.f.s
-------
....

functions
--------
....

datasets
--------
RooDataSet::asimovData(channelCat,obs_channel1)
RooDataSet::obsData(channelCat,obs_channel1)

named sets
----------
...

generic objects
---------------
RooStats::ModelConfig::ModelConfig



root [3] using namespace RooStats
root [4] ModelConfig* mc = (ModelConfig*) combined->obj("ModelConfig")
root [5] mc->Print()

=== Using the following for ModelConfig ===
Observables:             RooArgSet:: = (obs_channel1,weightVar,channelCat)
Parameters of Interest:  RooArgSet:: = (SigXsecOverSM)
Nuisance Parameters:     RooArgSet:: = (alpha_syst2,alpha_syst3)
Global Observables:      RooArgSet:: = (nominalLumi,nom_alpha_syst1,nom_alpha_syst2,nom_alpha_syst3)
PDF:                     RooSimultaneous::simPdf[ indexCat=channelCat channel1=model_channel1 ] = 260.156
```

5 Manual entries

```
man prepareHistFactory
PREPAREHISTFACTORY(1)                                                      PREPAREHISTFACTORY(1)

NAME
       prepareHistFactory - create a working directory for the HistFactory tools

SYNOPSIS
       prepareHistFactory [dir_name]

DESCRIPTION
       prepareHistFactory  is  a simple script that prepares a working area (and creates the directory
       dir_name if specified).  Within the directory dir_name, it creates a results/, data/, and  con-
       fig/  directory relative to the given path.  It also copies the HistFactorySchema.dtd and exam-
       ple XML files into the config/ directory.  Additionally, it copies a root file into  the  data/
       directory  for  use  with  the  examples.  Once  this is done, one is ready to run the example
       hist2workspace input.xml or edit the XML files for a new project.

ORIGINAL AUTHORS
              Dominique Tardif
               and Kyle Cranmer

COPYRIGHT
       This library is free software; you can redistribute it and/or modify it under the terms of  the
       GNU  Lesser General Public License as published by the Free Software Foundation; either version
       2.1 of the License, or (at your option) any later version.

       This library is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; with-
       out  even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.  See the
       GNU Lesser General Public License for more details.

       You should have received a copy of the GNU  Lesser  General  Public  License  along  with  this
       library;  if  not,  write  to  the Free Software Foundation, Inc., 51 Franklin St, Fifth Floor,
       Boston, MA  02110-1301  USA

                                             DEC. 2010                         PREPAREHISTFACTORY(1)
```

NAME
    hist2workspace - utility to create RooFit/RooStats workspace from histograms

SYNOPSIS
    hist2workspace [option] input.xml

DESCRIPTION
    hist2workspace is a utility to create RooFit/RooStats workspace from histograms

OPTIONS
    -standard_form default model, which creates an extended PDF that interpolates between RooHistFuncs. This is much faster for models with many bins and uses significantly less memory.
    -number_counting_form this was the original model in 5.28 (without patches). It uses a Poisson for each bin of the histogram. This can become slow and memory intensive when there are many bins.

Prepare working area
    The ROOT release ships with a script prepareHistFactory in the $ROOTSYS/bin directory that prepares a working area. It creates a results/, data/, and config/ directory. It also copies the HistFactorySchema.dtd and example XML files into the config/ directory. Additionally, it copies a root file into the data/ directory for use with the examples.

HistFactorySchema.dtd
    This file is located in $ROOTSYS/etc/ specifies the XML schema. It is typically placed in the config/ directory of a working area together with the top-level XML file and the individual channel XML files. The user should not modify this file.
    The HistFactorySchema.dtd is commented to specify exactly the meaning of the various options.

Top-Level XML File
    (see for example $ROOTSYS/tutorials/histfactory/example.xml) This file is edited by the user.
    It specifies
    - A top level 'Combination' that is composed of:
        - several 'Channels', which are described in separate XML files.
        - several 'Measurements' (corresponding to a full fit of the model) each of which specifies
            - a name for this measurement to be used in tables and files
            - what is the luminosity associated to the measurement in picobarns
            - which bins of the histogram should be used
            - what is the relative uncertainty on the luminosity
            - what is (are) the parameter(s) of interest that will be measured
            - which parameters should be fixed/floating (eg. nuisance parameters)
            - which type of constriants are desired - Gaussian by default - Gamma, LogNormal, and Uniform are also supported
            - if the tool should export the model only and skip the default fit

Channel XML Files
    (see for example $ROOTSYS/tutorials/histfactory/example_channel.xml) This file is edited by the user. It specifies for each channel
    - observed data
        - if absent the tool will use the expectation, which is useful for expected sensitivity
    - several 'Samples' (eg. signal, bkg1, bkg2, ...), each of which has:
        - a name
        - if the sample is normalized by theory (eg N = L*sigma) or not (eg. data driven)
        - a nominal expectation histogram
        - a named 'Normalization Factor' (which can be fixed or allowed to float in a fit)
        - several 'Overall Systematics' in normalization with:
            - a name
            - +/- 1 sigma variations (eg. 1.05 and 0.95 for a 5% uncertainty)
        - several 'Histogram Systematics' in shape with:
            - a name (which can be shared with the OverallSyst if correlated)
            - +/- 1 sigma variational histograms

ORIGINAL AUTHORS
    Kyle Cranmer , Akira Shibata , and Dominique Tardif

COPYRIGHT