Background-only hypothesis test with different methods for calculating significance

Adi Ashkenazi (CERN)
Renaud Bruneliere (Freiburg)
Riccardo Bianchi (Freiburg)
Outline

• Methods for calculating significance
• Results
• Implementation example
• Conclusions
Methods

• Poisson distribution from data in signal region

$$\frac{e^{-\mu} \mu^n}{n!}$$
Methods

- Poisson distribution from data in signal region
  \[ e^{-\mu} \frac{\mu^n}{n!} \]

- Due to uncertainty in \( \mu = \mu_s + \mu_b \):
  multiply the Poisson with a Poisson/Gaussian from extra measurement
  \[ e^{-\mu_s + \mu_b} \left( \mu_s + \mu_b \right)^n \times \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(\mu_b - \langle\mu_b\rangle)^2}{2\sigma^2}} \]
  \[ e^{-\mu_s + \mu_b} \left( \mu_s + \mu_b \right)^n \times e^{-\tau\mu_b} \left( \tau\mu_b \right)^{n_{\text{off}}} \]

\( \mu_s \) is signal, \( \mu_b \) is background = nuisance parameter
## Methods

- Poisson distribution from data in signal region
  \[ \frac{e^{-\mu} \mu^n}{n!} \]

- Due to uncertainty in \( \mu \):
  multiply the Poisson with a Poisson/Gaussian from extra measurement

\[
e^{-\mu_s+\mu_b} \left( \mu_s + \mu_b \right)^n \times \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(\mu_b-\langle \mu_b \rangle)^2}{2\sigma^2}} \times e^{\mu_s+\mu_b} \left( \mu_s + \mu_b \right)^n \times e^{-\tau\mu_b} \left( \tau\mu_b \right)^{n_{\text{off}}} = \frac{e^{-\mu} \mu^n}{n!} \times e^{-\tau\mu_b} \left( \tau\mu_b \right)^{n_{\text{off}}}
\]

\( \mu_s \) is signal, \( \mu_b \) is background = nuisance parameter

- In order to get rid of nuisance parameters, can either integrate or minimize over them
- Statistical significance for background only hypothesis = pvalue = sum from nosbs to infinity
Methods

• $Z_{Bi}$  Binomial : exact classical solution
Methods

• $Z_{Bi}$  Binomial : exact classical solution

• Hybrid recipe with Integration:
  
  $Z_{\Gamma}$
  
  $Z_{N}$  
  - Using Cumulative Distribution Function
  - Using CreateIntegral
Problem with c.d.f

Before

A RooPlot of "n"

Projection of integral of p.d.f

0 1 2 3 4 5 6 7 8 9 10
n
Problem with c.d.f

Before

After
Problem with c.d.f

Before

After

• c.d.f for as if continuous poisson distribution = $\int f(x)dx$
• pvalue(n) = 1 – c.d.f(n)
• what is in RooStats now
Problem with c.d.f

Before

After

• c.d.f for discrete poisson distribution
  $= \Sigma f(n)$
• pvalue(n) $= 1 - c.d.f(n - 1)$
Methods

- $Z_{Bi}$: Binomial: exact frequentist solution

- Hybrid recipe with Integration:
  - $Z_{\Gamma}$
  - $Z_{N}$
    - Using CDF
    - Using CreateIntegral

- Likelihood profile = minimization:
  - $Z_{PL}^{Poisson}$
  - $Z_{PL}^{Gaussian}$
    - Using ProfileLikelihoodCalculator & MINUIT
# Results

## Significance Comparison

<table>
<thead>
<tr>
<th>$n_{\text{obs}}$</th>
<th>4.0</th>
<th>6.0</th>
<th>9.0</th>
<th>17.0</th>
<th>50.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_0$</td>
<td>1.0</td>
<td>1.3</td>
<td>3.8</td>
<td>3.8</td>
<td>27.5</td>
</tr>
<tr>
<td>$\mu_\sigma$</td>
<td>0.477</td>
<td>0.3</td>
<td>0.9</td>
<td>0.6</td>
<td>3.71</td>
</tr>
<tr>
<td>$Z_0$</td>
<td>1.66</td>
<td>1.665</td>
<td>2.63</td>
<td>2.631</td>
<td>4.46</td>
</tr>
<tr>
<td>$Z_T$</td>
<td>1.66</td>
<td>1.666</td>
<td>2.63</td>
<td>2.688</td>
<td>4.46</td>
</tr>
<tr>
<td>$Z_X$</td>
<td>1.88</td>
<td>1.873</td>
<td>2.71</td>
<td>2.713</td>
<td>4.55</td>
</tr>
<tr>
<td>$Z_{\text{PGaussian}}$</td>
<td>2.00</td>
<td>2.013</td>
<td>2.83</td>
<td>2.831</td>
<td>4.62</td>
</tr>
<tr>
<td>$Z_{\text{PPoisson}}$</td>
<td>1.95</td>
<td>1.948</td>
<td>2.82</td>
<td>2.816</td>
<td>4.57</td>
</tr>
</tbody>
</table>

## Significance Comparison

<table>
<thead>
<tr>
<th>$n_{\text{obs}}$</th>
<th>67.0</th>
<th>200.0</th>
<th>523.0</th>
<th>498428.0</th>
<th>2119449.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_0$</td>
<td>30.0</td>
<td>100.0</td>
<td>388.6</td>
<td>493434.0</td>
<td>2109732.0</td>
</tr>
<tr>
<td>$\mu_\sigma$</td>
<td>7.75</td>
<td>31.6</td>
<td>8.1</td>
<td>702.4</td>
<td>433.8</td>
</tr>
<tr>
<td>$Z_0$</td>
<td>2.89</td>
<td>2.893</td>
<td>2.20</td>
<td>2.203</td>
<td>5.93</td>
</tr>
<tr>
<td>$Z_T$</td>
<td>2.89</td>
<td>3.087</td>
<td>2.20</td>
<td>2.203</td>
<td>5.93</td>
</tr>
<tr>
<td>$Z_X$</td>
<td>3.44</td>
<td>3.429</td>
<td>2.90</td>
<td>2.900</td>
<td>5.93</td>
</tr>
<tr>
<td>$Z_{\text{PGaussian}}$</td>
<td>3.45</td>
<td>3.444</td>
<td>2.90</td>
<td>2.898</td>
<td>5.93</td>
</tr>
<tr>
<td>$Z_{\text{PPoisson}}$</td>
<td>3.04</td>
<td>3.042</td>
<td>2.38</td>
<td>2.384</td>
<td>5.95</td>
</tr>
</tbody>
</table>
Results

Very good agreement between Cousins et al. and our RooStats implementation.
Conclusion

• Tested and implemented all methods described by Cousins et al. in Roostats

• Good agreement between their results and ours

• Simple code which could be used as short tutorial for Roostats and comparison between groups