Combined Neyman-Pearson Chi-square: An Improved Approximation to the Poisson-likelihood Chi-square

Xiangpan Ji (xji@bnl.gov)

July 30, 2019
DPF 2019, Northeastern University
Introduction: $\chi^2_{\text{Poisson}}$

- Poisson distribution: popular for modeling the number of events occur in a measurement

- Poisson-likelihood chi-square: most commonly used in HEP for parameter($\theta$) estimation

$\chi^2_{\text{Poisson}} = -2 \ln \lambda(\theta) = 2 \sum_{i=1}^{n} \left( \mu_i(\theta) - M_i + M_i \ln \frac{M_i}{\mu_i(\theta)} \right)$

- minimizing $\chi^2 \rightarrow$ best estimation of $\theta$
- $\lambda(\theta)$ the likelihood ratio

$M$: number of events
$\mu$: expected number of events

Poisson distribution

<table>
<thead>
<tr>
<th>$\mu$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(M, $\mu$)</td>
<td>0.5</td>
<td>0.4</td>
<td>0.3</td>
<td>0.2</td>
<td>0.1</td>
<td>0</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
</tr>
</tbody>
</table>

M: number of events

Xiangpan Ji, BNL
A Simple Example

A set of $n$ independent counting experiments to measure a common expected value $\mu$

$\chi^2_{\text{Poisson}}$ Bias is zero!

$n = 100$ and $\mu = 15$
Additional two common chi-squares

- At large statistics, the Poisson distribution can be approximated as Normal (Gaussian) distribution with mean $\mu_i$ and standard deviation $\sigma_i^2 = \mu_i$ (or $M_i$)

- Then, two common chi-squares*: Pearson’s and Neyman’s chi-squares. Popular because of their close connection to the covariance matrix format (talk later)

$$\chi^2_{\text{Pearson}} = \sum_i \frac{(\mu_i(\theta) - M_i)^2}{\mu_i(\theta)}$$

$$\chi^2_{\text{Neyman}} = \sum_i \frac{(\mu_i(\theta) - M_i)^2}{M_i}$$

* *Nucl. Instrum. Meth. 221 (1984) 437*
Bias: $\chi^2_\text{Neyman}$ and $\chi^2_\text{Pearson}$

Well known: estimator from $\chi^2_\text{Neyman}$ or $\chi^2_\text{Pearson}$ leads to bias especially when the large statistics condition is not met

$n = 100$ and $\mu = 15$

* F. James, Statistical methods in experimental physics. 2006
Biases trace back to $\chi^2$ definition

\[
\chi^2_{\text{Poisson}} = 2 \sum_{i=1}^{n} \left( \mu - M_i + M_i \ln \frac{M_i}{\mu} \right)
\]

\[
\chi^2_{\text{Neyman}} = \sum_{i} \frac{(\mu - M_i)^2}{M_i}
\]

\[
\chi^2_{\text{Pearson}} = \sum_{i} \frac{(\mu - M_i)^2}{\mu}
\]

Minimizing $\chi^2$

Obtain the estimator $\hat{\mu}$

$\hat{\mu}_{\text{Neyman}} = \frac{n}{\sum_{i=1}^{n} \frac{1}{M_i}}$

$\hat{\mu}_{\text{Pearson}} = \sqrt{\frac{\sum_{i=1}^{n} M_i^2}{n}}$

$\hat{\mu}_{\text{Neyman}} \leq \hat{\mu}_{\text{Poisson}} \leq \hat{\mu}_{\text{Pearson}}$

(equal only when all $M_i$ are same)

\[
\chi^2_{\text{Poisson}} - \chi^2_{\text{Neyman}} \approx - \sum_{i} \frac{2 (\mu - M_i)^3}{3 M_i^2} + O \left( \frac{(\mu - M_i)^4}{M_i^3} \right)
\]

\[
\chi^2_{\text{Poisson}} - \chi^2_{\text{Pearson}} \approx \sum_{i} \frac{1 (\mu - M_i)^3}{3 M_i^2} - O \left( \frac{(\mu - M_i)^4}{M_i^3} \right)
\]

$\sim x^{-2}$
Combined Neyman-Pearson Chi-square

\[ \chi^2_{\text{CNP}} = \frac{1}{3} (\chi^2_{\text{Neyman}} + 2\chi^2_{\text{Pearson}}) = \sum_{i=1}^{n} \frac{(\mu - M_i)^2}{3 \left( \frac{1}{M_i} + 2 \frac{1}{\mu} \right)} \]

- Approximately equal to \( \chi^2_{\text{Poisson}} \) up to \( O\left( \frac{(\mu-M_i)^4}{M_i^3} \right) \)
- The estimator from \( \chi^2_{\text{CNP}} \) in the previous example is

\[ \hat{\mu}_{\text{CNP}} = 3 \sqrt[3]{\frac{\sum_{i=1}^{n} M_i^2}{\sum_{i=1}^{n} \frac{1}{M_i}}} = 3 \sqrt[3]{\hat{\mu}^2_{\text{Pearson}} \cdot \hat{\mu}_{\text{Neyman}}} \]

the geometric mean of two \( \hat{\mu}_{\text{Pearson}} \) and one \( \hat{\mu}_{\text{Neyman}} \)
The Simple Example

\[ n = 100 \text{ and } \mu = 15 \]
A Realistic Example: Backgrounds + Systematics uncertainties

To measure the neutrino oscillation parameters, the PROSPECT measured the reactor neutrino spectrum

- ~100 segments, ranging from 7~9 meters to the reactor core
- number of events/segment ~200. Dividing into 16 energy bins gives ~12 events/bin
- Backgrounds and systematics uncertainties in each segment

→ An uncommon statistical regime where we need many bins (100 × 16) but number of events in each bin is not large


DPF 2019

Xiangpan Ji, BNL
- Our toy experiment similar to PROSPECT: 100 detectors $\times$ 16 energy bins, true signals 30/bin, backgrounds 15/bin

- Fit the free normalization parameter $\epsilon$ (the true value is 0 in our toy experiment)

\[
\chi^2_{\text{CNP}} = \sum_{d=1}^{100} \sum_{i=1}^{16} \left( \frac{\mu_d^i(1 + \epsilon + \epsilon_d) + b_d^i - M_d^i}{3/(\frac{1}{M_d^i} + \frac{2}{\mu_d^i(1+\epsilon+\epsilon_d)+b_d^i})} \right)^2 + \sum_{d=1}^{100} \sum_{i=1}^{16} \left( \frac{b_d^i - B_d^i}{3/(\frac{1}{B_d^i} + \frac{2}{b_d^i})} \right)^2 + \sum_{d=1}^{100} \left( \frac{\epsilon_d}{0.02} \right)^2
\]

- CNP’s bias is small
- Backgrounds
- Uncorrelated systematics

Denominator: CNP
Second Advantage
Connection to Covariance Matrix

\[
\chi^2 = \sum_i^n \frac{(\mu_i(\theta) - M_i)^2}{\sigma_i^2} \begin{cases} 
\sigma_i^2 = M_i & \rightarrow \chi^2_{\text{Neyman}} \\
\sigma_i^2 = \mu_i(\theta) & \rightarrow \chi^2_{\text{Pearson}} \\
\sigma_i^2 = 3/(\frac{1}{M_i} + \frac{2}{\mu_i(\theta)}) & \rightarrow \chi^2_{\text{CNP}} = \frac{1}{3} (\chi^2_{\text{Neyman}} + 2\chi^2_{\text{Pearson}})
\end{cases}
\]

- Both $\chi^2_{\text{Neyman}}$ and $\chi^2_{\text{Pearson}}$ are commonly used, partly due to their close connection* to the covariance-matrix formalism

\[
\chi^2_{\text{cov}} = (M - \mu(\theta))^T \cdot V^{-1} \cdot (M - \mu(\theta))
\]

- Reduces the fitting number of nuisance parameters of the $\chi^2$ function
- Leads to a fast minimization of the $\chi^2$ function

* CDF Note 8661 (1999)
At low statistics, $\chi^2_{\text{Neyman}}$ and $\chi^2_{\text{Pearson}}$ have large biases

$\chi^2_{\text{Poisson}}$ difficulty to use together with covariance matrix, can be computationally expensive

$\chi^2_{\text{CNP}}$ approximation to $\chi^2_{\text{Poisson}}$
- Has much smaller bias
- Naturally connects to covariance matrix
A more complicated PROSPECT example:

There are correlations between the energy bins of the background in the same detector:

\[
(b^i_{d})_{\text{on}} = \sum_j R^{ij}_d b^j_d
\]

- \( R^{ij}_d \) is a smearing matrix
  - 0.5 when \( i = j \)
  - 0.25 when \( i = j \pm 1 \)
  - 0 others

Backgrounds when the source on and detector on

Backgrounds when the source off and detector on (\( B^i_d \) measurement values)

\[
\left( \chi^2_{\text{CNP}} \right)_{\text{cov}} = (\mu (1 + \epsilon) + R \cdot b - M)^T \cdot (V^\text{stat}_{\text{CNP}} + V^\text{syst})^{-1} \cdot (\mu (1 + \epsilon) + R \cdot b - M) \\
+ (b - B)^T \cdot (V^\text{bkg}_{\text{CNP}})^{-1} \cdot (b - B),
\]
- Poisson-likelihood chi-square doesn’t have a equivalent covariance matrix format

- In the covariance matrix format, CNP chi-square also has much smaller bias than Neyman’s and Pearson’s chi-squares
We proposed a new test statistics CNP chi-square: an improved approximation to the Poisson-likelihood chi-square

\[ \chi^2_{\text{CNP}} = \frac{1}{3} (\chi^2_{\text{Neyman}} + 2\chi^2_{\text{Pearson}}) = \sum_{i=1}^{n} \frac{(\mu_i(\theta) - M_i)^2}{\frac{1}{M_i} + \frac{2}{\mu_i(\theta)}} \]

Estimator from CNP chi-square has much smaller bias than that from Neyman’s or Pearson’s chi-square

CNP chi-square can use the covariance matrix format