Chi-square goodness of fit tests for weighted histograms. Review and improvements

N.D. Gagunashvili

University of Akureyri,
Borgir, v/Nordurslóð, IS-600 Akureyri, Iceland

E-mail: nikolai@unak.is

ABSTRACT: Weighted histograms are used for the estimation of probability density functions. Computer simulation is the main domain of application of this type of histogram. A review of chi-square goodness of fit tests for weighted histograms is presented in this paper. Improvements are proposed to these tests that have size more close to its nominal value. Numerical examples are presented in this paper for evaluation of tests and to demonstrate various applications of tests.

KEYWORDS: Analysis and statistical methods; Data processing methods; Simulation methods and programs

ArXiv ePrint: 1307.1582

1Tel.: +354-4608505; fax: +354-4608998. Present address: Max-Planck-Institut für Kernphysik, P.O. Box 103980, 69029 Heidelberg, Germany.
1 Introduction

A histogram with $m$ bins for a given probability density function (PDF) $p(x)$ is used to estimate the probabilities

$$p_i = \int_{S_i} p(x)dx, \quad i = 1, \ldots, m$$

(1.1)

that a random event belongs to bin $i$. Integration in (1.1) is done over the bin $S_i$.

A histogram can be obtained as a result of a random experiment with PDF $p(x)$. Let us denote the number of random events belonging to the $i$th bin of the histogram as $n_i$. The total number of events $n$ in the histogram is equal to

$$n = \sum_{i=1}^{m} n_i.$$  

(1.2)

The quantity

$$\hat{p}_i = n_i/n$$

(1.3)

is an estimator of probability $p_i$ with expectation value

$$\mathbb{E}[\hat{p}_i] = p_i.$$  

(1.4)
The distribution of the number of events for bins of the histogram is the multinomial distribution \[ \text{PDF} = \frac{n!}{n_1!n_2!...n_m!} p_1^{n_1} \cdots p_m^{n_m}, \quad \sum_{i=1}^{m} p_i = 1. \] (1.5)

A weighted histogram or a histogram of weighted events is used again for estimating the probabilities \( p_i \) (1.1), see ref. [2]. It is obtained as a result of a random experiment with probability density function \( g(x) \) that generally does not coincide with PDF \( p(x) \). The sum of weights of events for bin \( i \) is defined as:

\[ W_i = \sum_{k=1}^{n_i} w_i(k), \] (1.6)

where \( n_i \) is the number of events at bin \( i \) and \( w_i(k) \) is the weight of the \( k \)th event in the \( i \)th bin. The statistic

\[ \hat{p}_i = W_i / n \] (1.7)

is used to estimate \( p_i \), where \( n = \sum_{i=1}^{m} n_i \) is the total number of events for the histogram with \( m \) bins. Weights of events are chosen in such a way that the estimate (1.7) is unbiased,

\[ E[\hat{p}_i] = p_i. \] (1.8)

The usual histogram is a weighted histogram with weights of events equal to 1.

The two examples of weighted histograms are considered below:

### 1.1 Example 1

To define a weighted histogram let us write the probability \( p_i \) (1.1) for a given PDF \( p(x) \) in the form

\[ p_i = \int_{S_i} p(x) dx = \int_{S_i} w(x) g(x) dx, \] (1.9)

where

\[ w(x) = p(x) / g(x) \] (1.10)

is the weight function and \( g(x) \) is some other probability density function. The function \( g(x) \) must be \( > 0 \) for points \( x \), where \( p(x) \neq 0 \). The weight \( w(x) = 0 \) if \( p(x) = 0 \), see ref. [3].

The weighted histogram is obtained from a random experiment with a probability density function \( g(x) \), and the weights of the events are calculated according to (1.10).

### 1.2 Example 2

The probability density function \( p_{\text{rec}}(x) \) of a reconstructed characteristic \( x \) of an event obtained from a detector with finite resolution and limited acceptance can be represented as

\[ p_{\text{rec}}(x) \propto \int_{\Omega} p_{\text{tr}}(x') A(x') R(x|x') dx', \] (1.11)

where \( p_{\text{tr}}(x') \) is the true PDF, \( A(x') \) is the acceptance of the setup, i.e. the probability of recording an event with a characteristic \( x' \), and \( R(x|x') \) is the experimental resolution, i.e. the probability of obtaining \( x \) instead of \( x' \) after the reconstruction of the event. The integration in (1.11) is carried
out over the domain \( \Omega' \) of the variable \( x' \). Total probability that an event will not be registered is equal to
\[
\mathcal{P} = \int_{\Omega'} p_{\text{tr}}(x') \left( 1 - A(x') \right) dx'.
\] (1.12)

The sum of probabilities
\[
\int_{\Omega} \int_{\Omega'} p_{\text{tr}}(x') A(x') R(x|x') dx' dx + \int_{\Omega'} p_{\text{tr}}(x') \left( 1 - A(x') \right) dx' = 1
\] (1.13)
because
\[
\int_{\Omega} \int_{\Omega'} p_{\text{tr}}(x') A(x') R(x|x') dx' dx = \int_{\Omega'} p_{\text{tr}}(x') A(x'), dx',
\] (1.14)
where \( \Omega \) domain of the variable \( x \).

A histogram of the PDF \( p_{\text{rec}}(x) \) can be obtained as a result of a random experiment (simulation) that has three steps [3]:

1. A random value \( x' \) is chosen according to a PDF \( p_{\text{tr}}(x') \).
2. We go back to step 1 again with probability \( 1 - A(x') \), and to step 3 with probability \( A(x') \).
3. A random value \( x \) is chosen according to the PDF \( R(x|x') \).

The quantity \( \hat{p}_i = n_i / n \), where \( n_i \) is the number of events belonging to the \( i \)th bin for a histogram with total number of events \( n \) in random experiment (at step 1), is an estimator of \( p_i \),
\[
p_i = \int_{S_i} \int_{\Omega'} p_{\text{tr}}(x') A(x') R(x|x') dx' dx, \quad i = 1, \ldots, m,
\] (1.15)
with the expectation value of the estimator
\[
E [\hat{p}_i] = p_i.
\] (1.16)

The quantity \( \hat{\mathcal{P}} = \bar{n} / n \), where \( \bar{n} \) is the number of events that were lost, is an estimator of \( \mathcal{P} \) (1.12) with the expectation value of the estimator
\[
E [\hat{\mathcal{P}}] = \mathcal{P}.
\] (1.17)

Notice that
\[
\sum_{i=1}^{m} p_i + \mathcal{P} = 1 \quad \text{and} \quad \sum_{i=1}^{m} n_i + \bar{n} = n.
\] (1.18)

In experimental particle and nuclear physics, step 3 is the most time-consuming step of the Monte Carlo simulation. This step is related to the simulation of the process of transport of particles through a medium and the rather complex registration apparatus.

To use the results of the simulation with some PDF \( g_{\text{tr}}(x') \) for calculating a weighted histogram of events with a true PDF \( p_{\text{tr}}(x') \), we write the equation for \( p_i \) in the form
\[
p_i = \int_{S_i} \int_{\Omega'} w(x') g_{\text{tr}}(x') A(x') R(x|x') dx' dx,
\] (1.19)
where
\[
w(x') = p_{\text{tr}}(x') / g_{\text{tr}}(x')
\] (1.20)
is the weight function.
The weighted histogram for the PDF $p_{\text{rec}}(x)$ can be obtained using events with reconstructed characteristic $x$ and weights calculated according to (1.20).

In this way, we avoid step 3 of the simulation procedure, which is important in cases where one needs to calculate Monte Carlo reconstructed histograms for many different true PDFs.

The probability that an event will not be registered can be represented as

$$\mathcal{P} = \int_{\Omega} w(x') g_{\text{tr}}(x')(1 - A(x')) \, dx',$$

and is estimated the same way using events with weights calculated according formula (1.20).

2 Goodness of fit tests

The problem of goodness of fit is to test the hypothesis

$$H_0: \quad p_1 = p_{10}, \ldots, p_{m-1} = p_{m-1,0} \quad \text{vs.} \quad H_a: \quad p_i \neq p_{i0} \quad \text{for some } i,$$

where $p_{i0}$ are specified probabilities, and $\sum_{i=1}^{m} p_{i0} = 1$. The test is used in a data analysis for comparing theoretical frequencies $np_{i0}$ with observed frequencies $n_i$. This classical problem remains of current practical interest. The test statistic for a histogram with unweighted entries

$$X^2 = \sum_{i=1}^{m} \frac{(n_i - np_{i0})^2}{np_{i0}},$$

was suggested by Pearson [4]. Pearson showed that the statistic (2.2) has approximately a $\chi^2_{m-1}$ distribution if the hypothesis $H_0$ is true.

2.1 The contemporary proof of Pearson’s result

The expectation values of the observed frequency $n_i$, if hypothesis $H_0$ is valid, equal to:

$$E[n_i] = np_{i0}, \quad i = 1, \ldots, m$$

and its covariance matrix $\Gamma$ has elements:

$$\gamma_{ij} = \begin{cases} np_{i0}(1 - p_{i0}) & \text{for } i = j \\ -np_{i0}p_{j0} & \text{for } i \neq j \end{cases}$$

Notice that the covariance matrix $\Gamma$ is singular [5].

Let us now introduce the multivariate statistic

$$(n - np_0)^t \Gamma_k^{-1}(n - np_0),$$

where $n = (n_1, \ldots, n_{k-1}, n_{k+1}, \ldots, n_m)^t$, $p_0 = (p_{10}, \ldots, p_{k-1,0}, p_{k+1,0}, \ldots, p_{m0})^t$ and $\Gamma_k = (\gamma_{ij})_{(m-1) \times (m-1)}$ is the covariance matrix for a histogram without bin $k$. The matrix $\Gamma_k$ has the form

$$\Gamma_k = n \text{ diag}(p_{10}, \ldots, p_{k-1,0}, p_{k+1,0}, \ldots, p_{m0}) - np_0p_0^t.$$
The special form of this matrix permits one to find analytically $\Gamma_k^{-1}$ [7]:

$$
\Gamma_k^{-1} = \frac{1}{n} \text{diag} \left( \frac{1}{p_{10}}, \ldots, \frac{1}{p_{k-1,0}}, \frac{1}{p_{k+1,0}}, \ldots, \frac{1}{p_{m0}} \right) + \frac{1}{np_{k,0}} \Theta,
$$

(2.6)

where $\Theta$ is $(m-1) \times (m-1)$ matrix with all elements unity. Finally the result of the calculation of expression (2.4) gives us the $X^2$ test statistic (2.2). Notice that the result will be the same for any choice of bin number $k$.

Asymptotically the vector $\mathbf{n}$ has a normal distribution $\mathcal{N}(np_0, \Gamma_k^{1/2})$, see ref. [5], and therefore the test statistic (2.2) has $\chi^2_{m-1}$ distribution if hypothesis $H_0$ is true

$$
X^2 \sim \chi^2_{m-1}.
$$

(2.7)

2.2 Generalization of the Pearson’s chi-square test for weighted histograms

The total sum of weights of events in $i$th bin $W_i$, $i = 1, \ldots, m$, as proposed in ref. [2], can be considered as a sum of random variables

$$
W_i = \sum_{k=1}^{n_i} w_i(k),
$$

(2.8)

where also the number of events $n_i$ is a random value and the weights $w_i(k), k = 1, \ldots, n_i$ are independent random variables with the same probability distribution function. The distribution of the number of events for bins of the histogram is the multinomial distribution and the probability of the random vector $(n_1, \ldots, n_m)$ is

$$
P(n_1, \ldots, n_m) = \frac{n!}{n_1! n_2! \ldots n_m!} g_1^{n_1} \cdots g_m^{n_m}, \quad \sum_{i=1}^m g_i = 1,
$$

(2.9)

where $g_i$ is the probability that a random event belongs to the bin $i$.

Let us denote the expectation values of the weights of events from the $i$th bin as

$$
E[w_i] = \mu_i
$$

(2.10)

and the variances as

$$
\text{Var}[w_i] = \sigma_i^2.
$$

(2.11)

The expectation value of the total sum of weights $W_i, i = 1, \ldots, m$, see ref. [6], is:

$$
E[W_i] = E\left[ \sum_{k=1}^{n_i} w_i(k) \right] = E[w_i] E[n_i] = n \mu_i g_i.
$$

(2.12)

The diagonal elements $\gamma_{ii}$ of the covariance matrix of the vector $(W_1, \ldots, W_m)$, see ref. [6], are equal to

$$
\gamma_{ii} = \sigma_i^2 g_i n + \mu_i^2 g_i (1 - g_i) n = n \alpha_{2i} g_i - n \mu_i^2 g_i,
$$

(2.13)

where

$$
\alpha_{2i} = E[w_i^2].
$$

(2.14)
The non-diagonal elements \( \gamma_{ij}, i \neq j \) are equal to:

\[
\gamma_{ij} = \sum_{k=0}^{n} \sum_{l=0}^{n} E \left[ \sum_{u=1}^{k} \sum_{v=1}^{l} w_i(u) w_j(v) \right] h(k,l) - E[W_i]E[W_j]
\]

\[
= \sum_{k=0}^{n} \sum_{l=0}^{n} E[w_i w_j] h(k,l) k l - \mu_k n g_j \mu_j n g_j
\]

\[
= \mu_i \mu_j (-g_i g_j n + g_i g_j n^2) - \mu_i n g_j \mu_j n g_j
\]

\[
= -n \mu_i \mu_j g_i g_j,
\]

(2.15)

where \( h(k,l) \) is the probability that \( k \) events belong to bin \( i \) and \( l \) events to bin \( j \).

For weighted histograms again the problem of goodness of fit is to test the hypothesis

\[
H_0 : \ u = u_0, \ldots, p_{m-1} = p_{m-1,0} \quad \text{vs.} \quad H_a : \ u_i \neq p_{i0} \quad \text{for some} \quad i,
\]

(2.16)

where \( p_{i0} \) are specified probabilities, and \( \sum_{i=1}^{m} p_{i0} = 1 \). If hypothesis \( H_0 \) is true then

\[
E[W_i] = n \mu_i g_i = n p_{i0}, \quad i = 1, \ldots, m
\]

(2.17)

and

\[
g_i = p_{i0} / \mu_i, \quad i = 1, \ldots, m.
\]

(2.18)

We can substitute \( g_i \) to eqs. (2.13) and (2.15) which gives the covariance matrix \( \Gamma \) with elements:

\[
\gamma_{ij} = \begin{cases} 
np_{i0}(r_i^{-1} - p_{i0}) & \text{for } i = j \\
-n p_{i0} p_{j0} & \text{for } i \neq j,
\end{cases}
\]

(2.19)

where

\[
r_i = \mu_i / \alpha_{2i}
\]

is the ratio of the first moment of the distribution of weights of events \( \mu_i \) to the the second moment \( \alpha_{2i} \) for a particular bin \( i \). Notice that for usual histograms the ratio of moments \( r_i \) is equal to 1 and the covariance matrix coincides with the covariance matrix of the multinomial distribution.

The multivariate statistic is represented as

\[
(W - n p_0)' \Gamma_k^{-1} (W - n p_0),
\]

(2.20)

where \( W = (W_1, \ldots, W_{k-1}, W_{k+1}, \ldots, W_m)' \), \( p_0 = (p_{10}, \ldots, p_{k-1,0}, p_{k+1,0}, \ldots, p_{m0})' \) and \( \Gamma_k = \gamma_{ij} \backslash (m-1) \times (m-1) \) is the covariance matrix for a histogram without bin \( k \). The matrix \( \Gamma_k \) has the form

\[
\Gamma_k = n \ \text{diag} \left( p_{10} / r_1, \ldots, p_{k-1,0} / r_k, p_{k+1,0} / r_{k+1}, \ldots, p_{m0} / r_m \right) - n p_0 p_0'.
\]

(2.21)

The special form of this matrix permits one to find analytically the inverse matrix

\[
\Gamma_k^{-1} = \frac{1}{n} \ \text{diag} \left( r_1 / p_{10}, \ldots, r_k / p_{k-1,0}, r_{k+1} / p_{k+1,0}, \ldots, r_m / p_{m0} \right) + \frac{1}{n (1 - \sum_{i \neq k} r_i p_{i0})} r r',
\]

(2.22)

where \( r = (r_1, \ldots, r_{k-1}, r_{k+1}, \ldots, r_m)' \).
After that, the multivariate statistic can be written as

\[ X_k^2 = \sum_{i \neq k} r_i \left( \frac{W_i - np_0}{np_0} \right)^2 + \frac{\left( \sum_{i \neq k} r_i (W_i - np_0) \right)^2}{n \left( 1 - \sum_{i \neq k} r_i p_0 \right)}, \]  

(2.23)

and can also be transformed to form

\[ X_k^2 = \frac{1}{n} \sum_{i \neq k} r_i W_i^2 + \frac{1}{n} \frac{\left( n - \sum_{i \neq k} r_i W_i \right)^2}{1 - \sum_{i \neq k} r_i p_0} - n \]

(2.24)

which is convenient for numerical calculations. Asymptotically the vector \( \mathbf{W} \) has a normal distribution \( \mathcal{N}(n \mathbf{p}_0, \mathbf{\Gamma}_k^{1/2}) \) [8] and therefore the test statistic (2.23) has \( \chi^2_{m-1} \) distribution if hypothesis \( H_0 \) is true

\[ X_k^2 \sim \chi^2_{m-1}. \]  

(2.25)

For usual histograms when \( r_i = 1, i = 1, \ldots, m \) the statistic (2.23) is Pearson’s chi-square statistic (2.2).

The expectation value of statistic (2.23), as shown in ref. [2], is equal to

\[ \mathbb{E}[X_k^2] = m - 1, \]  

(2.26)

as for Pearson’s test [1].

The ratio of moments \( r_i = \mu_i / \sigma_2i \), that is used for the test statistic calculation, is not known in majority of cases. An estimation of \( r_i \) can be used:

\[ \hat{r}_i = \frac{W_i}{W_{2i}}, \]

(2.27)

where \( W_{2i} = \sum_{k=1}^{n_i} w_i^2 (k) \).

Let us now replace \( r_i \) with the estimate \( \hat{r}_i \) and denote the estimator of matrix \( \mathbf{\Gamma}_k \) as \( \hat{\mathbf{\Gamma}}_k \). Then for positive definite matrices \( \hat{\mathbf{\Gamma}}_k, k = 1, \ldots, m \) the test statistic is given as

\[ \hat{X}_k^2 = \sum_{i \neq k} \hat{r}_i \left( \frac{W_i - np_0}{np_0} \right)^2 + \frac{\left( \sum_{i \neq k} \hat{r}_i (W_i - np_0) \right)^2}{n \left( 1 - \sum_{i \neq k} \hat{r}_i p_0 \right)}. \]

(2.28)

Formula (2.28) for usual histograms does not depend on the choice of the excluded bin, but for weighted histograms there can be a dependence. A test statistic that is invariant to the choice of the excluded bin and at the same time is a Pearson’s chi square statistic (2.2) for the unweighted histograms can be represented as the median value for the set of statistics \( \hat{X}_k^2 \) (2.28) with positive definite matrices \( \hat{\mathbf{\Gamma}}_k \)

\[ \hat{X}_{\text{Med}}^2 = \text{Med}\{ \hat{X}_1^2, \hat{X}_2^2, \ldots, \hat{X}_m^2 \}. \]  

(2.29)

Statistic \( \hat{X}_{\text{Med}}^2 \) first time was proposed in ref. [2] and approximately has \( \chi^2_{m-1} \) distribution if hypothesis \( H_0 \) is true

\[ \hat{X}_{\text{Med}}^2 \sim \chi^2_{m-1}. \]  

(2.30)

The usage of \( \hat{X}_{\text{Med}}^2 \) to test the hypothesis \( H_0 \) with a given significance level is equivalent to making a decision by voting. It was noticed that size of test can be slightly greater than nominal value of size of test even for large value of total number of events \( n \).
2.3 New generalizations of Pearson’s chi-square test for weighted histograms

Set of statistics \( \{ \hat{X}^2_1, \hat{X}^2_2, \ldots, \hat{X}^2_m \} \), with positive definite matrices \( \hat{\Gamma}_k \) only, is used for calculating the median statistic \( \hat{X}^2_{\text{Med}} \) (2.29). It can be used for any weighted histograms, including histograms with unweighted entries. One bin is excluded because the full covariance matrix of an unweighted histogram is singular and hence can not be inverted.

Let us consider estimation of a full covariance matrix \( \hat{\Gamma} \) for the weighted histogram with more detail. The symmetric matrix is positive definite if the minimal eigenvalue of the matrix larger then 0. We denote minimal eigenvalue of the matrix \( n^{-1} \hat{\Gamma} \) by \( \lambda_{\text{min}} \) then follow to ref. [10] it can be shown that

\[
\min_i \left\{ \frac{p_{i0}}{\hat{r}_i} \right\} - \sum_{i=1}^{m} \frac{p_{i0}^2}{p_{i0}} \leq \lambda_{\text{min}} \leq \min_i \left\{ \frac{p_{i0}}{\hat{r}_i} \right\}
\]  

(2.31)

and the eigenvalue \( \lambda_{\text{min}} \) is the root of secular equation

\[
1 - \sum_{i=1}^{m} \frac{p_{i0}^2}{p_{i0}/\hat{r}_i - \lambda} = 0.
\]

(2.32)

In case of a histogram with unweighted entries, all \( \hat{r}_i = 1 \) and \( \lambda = 0 \) is zero of equation (2.32). Matrix \( \hat{\Gamma} \) for this case is not positive definite and is singular, but matrix \( \hat{\Gamma}_k \) is positive definite and therefore invertible. Number of events \( n_i \) in bins of usual histogram satisfy to equation \( n_1 + n_2, \ldots, + n_m = n \) that is why the covariance matrix of multinomial distribution is not positive definite and is singular.

Matrix \( \hat{\Gamma} \) for a histogram with weighted entries can be also non-positive definite. There are two reasons why this can be. First of all, the total sums of weights \( W_i \) in bins of a weighted histogram are related with each other, because satisfy the equation \( E \left[ \sum_{i=1}^{m} W_i \right] = n \) and second, due fluctuations of matrix elements.

The test statistic obtained with full matrix \( \hat{\Gamma} \) is unstable and can have large variance especially for the case of low number \( n \) of events in a histogram.

The fact the matrix is not positive definite is equivalent to the fact that the minimal eigenvalue \( \lambda_{\text{min}} \) of the matrix \( \hat{\Gamma} \) is \( \leq 0 \). A case when the minimal eigenvalue is positive but rather small is also not desirable, especially for computer calculations.

Due to the above mentioned reasons it is wise to use the test statistic for a weighted histograms

\[
\hat{X}^2 = \hat{X}^2_k = \sum_{i \neq k} \frac{(W_i - np_{i0})^2}{np_{i0}} + \frac{(\sum_{i \neq k} \hat{r}_i (W_i - np_{i0}))^2}{n \left(1 - \sum_{i \neq k} \hat{r}_i p_{i0}\right)},
\]

(2.33)

for \( k \) where

\[
\frac{p_{i0}}{\hat{r}_i} = \min_i \left\{ \frac{p_{i0}}{\hat{r}_i} \right\}.
\]

(2.34)

A secular equation for the new minimal eigenvalue can be solved numerically, by bisection method, to check whether a matrix \( \hat{\Gamma}_k \) is positive definite or not. Numerical experiments show that it is very rare that the matrix \( \hat{\Gamma}_k \) is not positive definite and it happens only for histograms with a small number \( n \) of events in a histogram. If hypothesis \( H_0 \) is valid, statistic \( \hat{X}^2 \) asymptotically has distribution

\[
\hat{X}^2 \sim \chi^2_{m-1}.
\]

(2.35)
It is plausible that power of the new test is not lower than power of tests with statistic $\hat{X}^2_{\text{Med}}$ and with other statistics $\{\hat{X}^2_i, i \neq k\}$. The distribution of the statistic $\hat{X}^2$ is closer to $\chi^2_{m-1}$ then distribution of median statistic $\hat{X}^2_{\text{Med}}$. Also the statistic $\hat{X}^2$ is easier to calculate than the statistic $\hat{X}^2_{\text{Med}}$.

3 Goodness of fit tests for weighted histograms with deviations from main model

Here, different deviations from the main model of weighted histograms will be considered as well as goodness of fit tests for those cases.

3.1 Goodness of fit test for weighted histogram with unknown normalization

In practice one is often faced with the case that all weights of events are defined up to an unknown normalization constant $C$ see ref. [2]. It happens because in some cases of computer simulation is rather difficult give analytical formula for the PDF, but the PDF up to multiplicative constant is possible, that is enough for the generation of events according to the PDF, for example, by very popular Neumann’s method [11]. For the goodness of fit test it means that if hypothesis $H_0$ is valid

$$E [W_i] \cdot C = np_{0i}, \quad i = 1, \ldots, m$$

(3.1)

with unknown constant $C$. Then the test statistic (2.24) can be written as

$$c\hat{X}^2_k = \sum_{i \neq k} \hat{r}_i \frac{(W_i - np_{0i}/C)^2}{np_{0i}/C} + \frac{\left(\sum_{i \neq k} \hat{r}_i (W_i - np_{0i}/C)\right)^2}{n(1 - C^{-1}\sum_{i \neq k} \hat{r}_i p_{0i})}.$$ 

(3.2)

An estimator for the constant $C$ can be found by minimizing eq. (3.2).

$$\hat{C}_k = \sum_{i \neq k} \hat{r}_i p_{0i} + \sqrt{\frac{\sum_{i \neq k} \hat{r}_i p_{0i}}{\sum_{i \neq k} \hat{r}_i W_i^2 / p_{0i}} \left(n - \sum_{i \neq k} \hat{r}_i W_i\right)},$$

(3.3)

where $\hat{C}_k$ is an estimator of $C$. Substituting (3.3) to (3.2), we get the test statistic

$$c\hat{X}^2_k = \sum_{i \neq k} \hat{r}_i \frac{(W_i - np_{0i}/\hat{C}_k)^2}{np_{0i}/\hat{C}_k} + \frac{\left(\sum_{i \neq k} \hat{r}_i (W_i - np_{0i}/\hat{C}_k)\right)^2}{n(1 - \hat{C}_k^{-1}\sum_{i \neq k} \hat{r}_i p_{0i})}.$$ 

(3.4)

The statistic (3.4) has a $\chi^2_{m-2}$ distribution if hypothesis $H_0$ is valid.

Formula (3.4) can be also transformed to

$$c\hat{X}^2_k = \frac{s_k^2}{n} + 2s_k,$$

(3.5)

where

$$s_k = \sqrt{\sum_{i \neq k} \hat{r}_i p_{0i} \sum_{i \neq k} \hat{r}_i W_i^2 / p_{0i} - \sum_{i \neq k} \hat{r}_i W_i}$$

(3.6)

which is convenient for calculations, see [2]. Median statistics can be used for the same reason as in section 2.2

$$c\hat{X}^2_{\text{Med}} = \text{Med}\{c\hat{X}^2_1, c\hat{X}^2_2, \ldots, c\hat{X}^2_m\}$$

(3.7)

and has approximately $\chi^2_{m-2}$ distribution if hypothesis $H_0$ valid, see ref. [2]

$$c\hat{X}^2_{\text{Med}} \sim \chi^2_{m-2}.$$ 

(3.8)
3.2 New goodness of fit test for weighted histogram with unknown normalization

The new estimator of constant $C$ is

$$\hat{C} = \sum_{i \neq k} \hat{r}_i p_{i0} + \sqrt{\frac{\sum_{i \neq k} \hat{r}_i p_{i0}}{\sum_{i \neq k} \hat{r}_i W_i^2 / p_{i0}}} \left( n - \sum_{i \neq k} \hat{r}_i W_i \right),$$

(3.9)

for $k$ where

$$\frac{p_{i0}}{\hat{r}_k} = \min_i \left\{ \frac{p_{i0}}{\hat{r}_i} \right\}. \quad (3.10)$$

And the test statistic can be written as

$$c\hat{X}^2 = \sum_{i \neq k} \frac{(W_i - n p_{i0} / \hat{C})^2}{n p_{i0} / \hat{C}} + \frac{(\sum_{i=1}^{m} \hat{r}_i (W_i - n p_{i0} / \hat{C}))^2}{n (1 - \hat{C} - \sum_{i=1}^{m} \hat{r}_i p_{i0})}. \quad (3.11)$$

Statistic $c\hat{X}^2$ asymptotically has $\chi^2_{m-2}$ distribution if hypothesis $H_0$ is valid

$$c\hat{X}^2 \sim \chi^2_{m-2}. \quad (3.12)$$

3.3 Goodness of fit test for weighted Poisson histograms

Poisson histogram [12] can be defined as histogram with multi-Poisson distributions of a number of events for bins

$$P(n_1, \ldots, n_m) = \prod_{i=1}^{m} e^{-n_{i0}} (n_{i0} p_i)^{n_i} / n_i!, \quad (3.13)$$

where $n_{0}$ is a free parameter. The discrete probability distribution function (probability mass function) of a Poisson histogram can be represented as a product of two probability functions: a Poisson probability mass function for a number of events $n$ with parameter $n_{0}$ and a multinomial probability mass function of the number of events for bins of the histogram, with total number of events equal to $n$, see ref. [1]

$$P(n_1, \ldots, n_m) = e^{-n_{0}} (n_{0})^n / n! \times \frac{n!}{n_1! n_2! \ldots n_m!} p_1^{n_1} \ldots p_m^{n_m}. \quad (3.14)$$

A Poisson histogram can be obtained as a result of two random experiments, namely, where the first experiment with Poisson probability mass function gives us the total number of events in histogram $n$, and then a histogram is obtained as a result of a random experiment with PDF $p(x)$ and the total number of events is equal to $n$.

As in the case of multinomial histograms, also for Poisson histograms there is the problem of goodness of fit test with the hypothesis:

$$H_0: \quad p_1 = p_{10}, \ldots, p_{m-1} = p_{m-1,0} \quad \text{vs.} \quad H_a: \quad p_i \neq p_{i0} \quad \text{for some } i, \quad (3.15)$$

where $p_{i0}$ are specified probabilities, and $\sum_{i=1}^{m} p_{i0} = 1$. If $n_{0}$ is known, then the statistic, see ref. [13]:

$$X^2_{\text{pois}} = \sum_{i=1}^{m} \frac{(n_i - n_{0} p_{i0})^2}{n_{0} p_{i0}}, \quad (3.16)$$
can be used and has asymptotically a $\chi^2$ distribution if the hypothesis $H_0$ is valid

$$X^2_{\text{pois}} \sim \chi^2_m.$$  \hfill (3.17)

The hypothesis $H_0$ becomes complex if parameter $n_0$ is unknown for the Poisson histogram. This is an opposite situation to the case of a multinomial histogram, where the hypothesis is simple.

In [13] there are proposed statistics for goodness of fit test for a weighted Poisson histogram with known parameter $n_0$

$$X^2_{\text{corr0}} = \sum_{i=1}^{m} \frac{(W_i - n_0 p_{i0})^2}{W_2 p_{i0}/W_i},$$  \hfill (3.18)

and for the case the $n_0$ is not known:

$$X^2_{\text{corr}} = \sum_{i=1}^{m} \frac{(W_i - \hat{n}_0 p_{i0})^2}{W_{2i} \hat{n}_0 p_{i0}/W_i},$$  \hfill (3.19)

with estimation of $n_0$ obtained by minimization of equation (3.18)

$$\hat{n}_0 = \left[ \frac{\sum_{i=1}^{m} W_i^2/(W_{2i} p_{i0})}{\sum_{i=1}^{m} W_i p_{i0}/W_{2i}} \right]^{1/2}.$$  \hfill (3.20)

The distribution of statistic $X^2_{\text{corr0}}$ in case hypothesis $H_0$ is valid

$$X^2_{\text{corr0}} \sim \chi^2_m,$$  \hfill (3.21)

and for the statistic $X^2_{\text{corr}}$ is

$$X^2_{\text{corr}} \sim \chi^2_{m-1}.$$  \hfill (3.22)

according ref. [13].

Generally, the power of the tests for Poisson histograms will be slightly lower than for multinomial histograms with the number of events $n = n_0$ which is explained by the fact that the total number of events for Poisson histograms fluctuates.

The choice of the type of the histogram depends on what type of a physical experiment is produced. If the number of events $n$ is constant, then it is a multinomial histogram; if the number of events $n$ is a random value that has Poisson distribution, then it is a Poisson histogram.

A weighted histogram very often is the result of modeling and the number of simulated events is known exactly, and therefore the choice of a multinomial histogram is reasonable. It is also reasonable to use tests developed for the multinomial histograms in the case, if the number of events $n$ is random value but with unknown distribution [14].

4 Restriction for goodness of fit tests applications

For the histograms with unweighted entries, the use of Pearson’s chi-square test (2.2) is inappropriate if any expected frequency $np_{i0}$ is below 1 or if the expected frequency is less than 5 in more than 20% of bins [15].

Restrictions for weighted histograms, due to fluctuation of the estimation of ratio of moments $\hat{r}_i$, can be made stronger. Namely, the use of new chi-square tests (2.33) and (3.11) is inappropriate if any expected frequency $E[n_i]$ is less than 5.

Following ref. [16] a disturbance is regarded as unimportant when the nominal size of the test is 5% and the size of the test lies between 4% and 6% for a goodness of fit tests.
5 Numerical evaluation of the tests’ power and sizes

The main parameters which characterize the effectiveness of a test are size and power. The nominal significance level was taken to be equal to 5% for calculating of size of tests in presented numerical examples. Hypothesis $H_0$ is rejected if test statistic $\hat{X}^2$ is larger than some threshold. Threshold $k_{0.05}$ for a given nominal size of test 5% can be defined from the equation

$$0.05 = P(\chi^2 > k_{0.05}) = \int_{k_{0.05}}^{+\infty} \frac{x^{l/2-1}e^{-x/2}}{2^{l/2}\Gamma(l/2)}dx,$$  \hfill (5.1)

where $l = m - 1$.

Let us define the test size $\alpha$ for a given nominal test size 5% as the probability

$$\alpha = P(\hat{X}^2 > k_{0.05}|H_0).$$  \hfill (5.2)

This is the probability that hypothesis $H_0$ will be rejected if the distribution of weights $W_i$ for bins of the histogram satisfies hypothesis $H_0$. Deviation of the test size from the nominal test size is an important test characteristic.

A second important test characteristic is the power. Let us define the test power as

$$P(\hat{X}^2 > k_{0.05}|H_a).$$  \hfill (5.3)

This is the probability that hypothesis $H_0$ will be rejected if the distribution of weights $W_i$ for bins of the histogram does not satisfy hypothesis $H_0$.

Notice that the power calculated by formula (5.3) can give misleading result in case of comparing of different tests. To overcome this problem here we define the power of test $\pi$ as

$$\pi = P(\hat{X}^2 > K_{0.05}|H_a)$$  \hfill (5.4)

with the threshold $K_{0.05}$ calculated by Monte-Carlo method from equation

$$0.05 = P(\hat{X}^2 > K_{0.05}|H_0).$$  \hfill (5.5)

All definitions proposed above for statistics $\hat{X}^2$ can be used for other test statistics with appropriate number of degree of freedom $l$ in the formula (5.1).

The size and power of tests depend on the number of events and the binning that was discussed for usual histograms in ref. [1]. The power for weighted histograms also depends on the choice of PDF $g(x)$ (subsection 1.1) or $g_{tr}$ (subsection 1.2) and can be even higher than for histograms with unweighted entries as well as lower. Below we demonstrate two examples of an application of the previously discussed tests. The size and power of the tests are calculated for a different total number of events in the histograms. In numerical examples were demonstrated applications of:

- Pearson’s goodness of fit test [4], see subsection 2.1 and first paragraph of section 2. The test statistic is $X^2$ (2.2).
- Goodness of fit test for weighted histograms with normalized weights [2], see subsection 2.2. The test statistic is $\hat{X}^2_{Med}$ (2.29).
• Goodness of fit test for weighted histograms with unnormalized weights \[2\], see subsection 3.1. The test statistic is \(c \tilde{X}_\text{Med}^2 (3.7)\).

• New goodness of fit test for weighted histograms with normalized weights, see subsection 2.3. The test statistic is \(\tilde{X}^2 (2.33)\).

• New goodness of fit test for weighted histograms with unnormalized weights, see subsection 3.2. The test statistic is \(c \tilde{X}_\text{Med}^2 (3.11)\).

• Goodness of fit test for Poisson histograms with unweighted entries and known parameter \(n_0 [13]\), see subsection 3.3. The test statistic is \(X^2_{\text{pois}} (3.16)\).

• Goodness of fit test for weighted Poisson histograms with known parameter \(n_0 [13]\), see subsection 3.3. The test statistic is \(X^2_{\text{corr}0} (3.18)\).

• Goodness of fit test for weighted Poisson histograms with unknown parameter \(n_0 [13]\), see subsection 3.3. The test statistic is \(X^2_{\text{corr}} (3.19)\).

The published program, see ref. [17], was used for the calculation of the test statistics with minor modification needed for the new tests.

5.1 Numerical example 1

A simulation study was done for the example from ref. [2]. Weighted histograms described in subsection 1.1, are used here. The PDF for hypothesis \(H_0\) is:

\[ p_0(x) \propto \frac{2}{(x-10)^2 + 1} + \frac{1.15}{(x-14)^2 + 1} \]  

(5.6)

against alternative \(H_a\):

\[ p(x) \propto \frac{2}{(x-10)^2 + 1} + \frac{1}{(x-14)^2 + 1} \]  

(5.7)

represented by the weighted histogram. Both PDF’s are defined on the interval \([4, 16]\). A calculation was done for three cases of a PDF, used for the event generation, see figure 1

\[ g_1(x) = p(x) \]  

(5.8)

\[ g_2(x) = 1/12 \]  

(5.9)

\[ g_3(x) \propto \frac{2}{(x-9)^2 + 1} + \frac{2}{(x-15)^2 + 1} \]  

(5.10)

Distribution (5.8) gives an unweighted histogram. Distribution (5.9) is a uniform distribution on the interval \([4, 16]\). Distribution (5.10) has the same type of parameterizations as eq. (5.6), but with different values of the parameters. Histograms with 20 bins and equidistant binning were used. At figure 2 presented probabilities \(p_i, i = 1, \ldots, 20\) for the PDF \(p(x)\) and \(p_0, i = 1, \ldots, 20\) for the PDF \(p_0(x)\). Size and power of tests with statistics \(X^2 (2.33), c \tilde{X}^2 (3.11), \tilde{X}^2_{\text{Med}} (2.29)\) and \(c \tilde{X}^2_{\text{Med}} (3.7)\) were calculated for weighted histograms with weights of events equal to \(p(x)/g_2(x)\) and \(p(x)/g_3(x)\). Statistics \(\tilde{X}^2 (2.33)\) and \(\tilde{X}^2_{\text{Med}} (2.29)\) coincide with Pearson’s statistic \(X^2 (2.2)\) and were used for histograms with unweighted entries. The results of calculations for 100000 runs are presented in table 1.
Conclusion and interpretation of results presented in table 1.

- The size of new tests $\hat{X}^2 (2.33)$ (rows 2, 6) and $\hat{c}X^2 (3.11)$ (row 3, 7) are generally closer to nominal value 5% than median tests $\hat{X}_{\text{Med}}^2 (2.29)$ (rows 4, 8) and $\hat{c}X_{\text{Med}}^2 (3.7)$ (rows 5, 9) when the application of the test satisfies restrictions formulated in section 4.

- The power of new tests $\hat{X}^2 (2.33)$ (rows 2, 6) are greater than for analogous median tests $\hat{X}_{\text{Med}}^2 (2.29)$ (rows 4, 8). The power of tests $\hat{c}X^2 (3.11)$ (rows 3, 7) are greater than for analogous median tests $\hat{c}X_{\text{Med}}^2 (3.7)$ (rows 5, 9).
Table 1. Numerical example 1. Size ($\alpha$) and power ($\pi$) of different test statistics $X^2$ (2.2), $\hat{X}^2$ (2.33), $\bar{c}X^2$ (3.11), $\hat{X}_{\text{Med}}^2$ (2.29), $\bar{c}\hat{X}_{\text{Med}}^2$ (3.7) obtained for different weighted functions $w(x)$. Italic type marks a size of test with inappropriate number of events in the bins of histograms.

| № | $w(x)$ | $n$ | 200 | 400 | 600 | 800 | 1000 | 3000 | 5000 | 7000 | 9000 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | $\alpha$ | 5.7 | 5.4 | 5.3 | 5.2 | 5.1 | 5.0 | 5.0 | 5.1 | 1 |
| 1 | $\pi$ | 6.0 | 7.1 | 8.2 | 9.8 | 11.2 | 29.9 | 52.7 | 71.6 | 84.9 |
| 2 | $\alpha$ | 5.5 | 5.3 | 5.2 | 5.1 | 5.0 | 5.1 | 5.1 | 4.9 | 1 |
| 2 | $\pi$ | 6.1 | 7.0 | 8.2 | 9.2 | 10.5 | 26.2 | 45.8 | 64.0 | 78.7 |
| 3 | $\alpha$ | 5.0 | 5.1 | 5.0 | 5.0 | 4.9 | 5.0 | 5.0 | 5.2 | 4.9 |
| 3 | $\pi$ | 6.0 | 7.0 | 8.1 | 9.1 | 10.4 | 26.0 | 45.6 | 63.0 | 78.1 |
| 4 | $\alpha$ | 5.4 | 5.4 | 5.3 | 5.2 | 5.1 | 5.3 | 5.2 | 5.3 | 5.0 |
| 4 | $\pi$ | 6.0 | 6.9 | 8.0 | 9.1 | 10.3 | 25.7 | 45.3 | 63.1 | 78.2 |
| 5 | $\alpha$ | 5.6 | 5.8 | 5.7 | 5.7 | 5.7 | 5.7 | 5.7 | 5.8 | 5.5 |
| 5 | $\pi$ | 5.9 | 6.9 | 8.0 | 9.1 | 10.2 | 25.4 | 44.9 | 62.5 | 77.5 |
| 6 | $\alpha$ | 7.3 | 6.6 | 6.1 | 5.8 | 5.6 | 5.2 | 5.2 | 4.9 | 5.0 |
| 6 | $\pi$ | 16.2 | 29.7 | 40.1 | 48.5 | 56.1 | 95.7 | 99.8 | 100.0 | 100.0 |
| 7 | $\alpha$ | 4.7 | 4.9 | 5.0 | 5.0 | 5.1 | 5.1 | 5.0 | 4.9 | 5.0 |
| 7 | $\pi$ | 6.9 | 8.3 | 9.9 | 11.6 | 13.4 | 36.5 | 61.6 | 80.5 | 91.2 |
| 8 | $\alpha$ | 5.5 | 5.3 | 5.4 | 5.3 | 5.5 | 5.4 | 5.2 | 5.3 | 5.2 |
| 8 | $\pi$ | 7.9 | 11.8 | 15.8 | 20.2 | 25.0 | 75.3 | 96.6 | 99.8 | 100.0 |
| 9 | $\alpha$ | 5.4 | 5.4 | 5.7 | 5.6 | 5.8 | 5.7 | 5.6 | 5.7 | 5.5 |
| 9 | $\pi$ | 6.8 | 8.4 | 9.7 | 11.4 | 13.1 | 36.0 | 60.7 | 79.2 | 90.7 |

- The power of all tests calculated for histograms with weights of events equal to $p(x)/g_2(x)$ (rows 2–5) are lower than for histograms with unweighted entries (row 1), but the power of all tests calculated for histograms with weights of events equal to $p(x)/g_3(x)$ (rows 6–9) are greater. The explanation is that in latter case we increase the statistics of events for domains with high deviation of the distribution presented by the histogram from the tested distribution.

Properties of tests in applications to Poisson histograms with the same weighted functions and distributions of events were investigated. In this case, the total number of events $n$ is random and was simulated according Poisson distribution for a given parameter $n_0$. Size and power of tests $X^2_{\text{pois}}$ (3.16), $X^2_{\text{corr0}}$ (3.18) with exactly known parameter $n_0$ and $X^2_{\text{corr}}$ (3.19) developed ad hoc for the Poisson histogram in [13] also was calculated. Results of the calculations are presented in table 2.

Conclusion and interpretation of results presented in table 2.

- The size of all tests are close to nominal value 5%.
- The power of new tests $\hat{X}^2$ (2.33) (rows 5, 9) and $\bar{c}\hat{X}^2$ (3.11) (rows 6, 10) used for Poisson histograms are greater than the power of tests developed ad hoc for the Poisson histograms $X^2_{\text{corr0}}$ (3.18) (rows 3, 7) with the exactly known parameter $n_0$ and $X^2_{\text{corr}}$ (3.19) (rows 4, 8) with the unknown parameter $n_0$ in ref. [13].
- The power of Pearson’s test $X^2$ (2.2) (row 2) used for Poisson histograms is greater than test $X^2_{\text{pois}}$ (3.16) (row 1) with the exactly known parameter $n_0$ proposed in ref. [13].
Table 2. Numerical example 1. Size ($\alpha$) and power ($\pi$) of different test statistics $X^2_{\text{pois}}$ (5.16), $X^2$ (2.2), $X^2_{\text{corr}}$ (3.18), $X^2_{\text{corr}}$ (3.19), $\hat{X}^2$ (2.33), $c\hat{X}^2$ (3.11) in application for Poisson histograms. Italic type marks a size of test with inappropriate number of events in the bins of histograms.

| № | $n_0$ | 200 | 400 | 600 | 800 | 1000 | 3000 | 5000 | 7000 | 9000 | w(x) |
|---|------|-----|-----|-----|-----|------|------|------|------|------|------|
| 1 | $X^2_{\text{pois}}$ | $\alpha$ | 6.0 | 5.6 | 5.2 | 5.2 | 5.1 | 5.1 | 5.1 | 5.1 | 5.1 | 1 |
|   | $\pi$  | 5.9 | 7.0 | 8.3 | 9.6 | 11.1 | 29.2 | 50.9 | 70.0 | 83.8 |
| 2 | $X^2$ | $\alpha$ | 5.6 | 5.5 | 5.1 | 5.2 | 5.1 | 5.1 | 5.1 | 5.0 | 5.0 |
|   | $\pi$  | 6.0 | 7.0 | 8.4 | 9.8 | 11.1 | 30.0 | 52.2 | 71.2 | 85.0 |
| 3 | $X^2_{\text{corr}}$ | $\alpha$ | 5.4 | 5.3 | 5.2 | 5.1 | 5.0 | 5.1 | 5.1 | 5.0 | 5.0 |
|   | $\pi$  | 6.0 | 6.7 | 7.8 | 8.8 | 10.0 | 25.0 | 43.9 | 61.5 | 76.2 |
| 4 | $X^2_{\text{corr}}$ | $\alpha$ | 3.9 | 4.4 | 4.6 | 4.7 | 4.7 | 4.5 | 4.5 | 5.0 | 5.0 |
|   | $\pi$  | 6.0 | 7.0 | 8.0 | 9.0 | 10.3 | 25.5 | 40.0 | 61.2 | 77.4 |
| 5 | $X^2$ | $\alpha$ | 5.5 | 5.2 | 5.1 | 5.0 | 5.1 | 5.0 | 5.1 | 5.0 | 5.0 |
|   | $\pi$  | 6.1 | 7.1 | 8.1 | 9.2 | 10.6 | 26.3 | 46.0 | 64.1 | 78.5 |
| 6 | $c\hat{X}^2$ | $\alpha$ | 5.1 | 5.0 | 5.0 | 5.0 | 5.0 | 5.0 | 5.0 | 5.0 | 5.0 |
|   | $\pi$  | 6.0 | 7.0 | 8.1 | 9.2 | 10.5 | 26.0 | 45.5 | 63.4 | 77.6 |
| 7 | $X^2_{\text{corr}}$ | $\alpha$ | 5.1 | 5.0 | 5.2 | 5.0 | 5.2 | 5.1 | 5.1 | 4.9 | 4.9 |
|   | $\pi$  | 6.3 | 7.5 | 8.8 | 10.7 | 12.3 | 35.3 | 60.5 | 79.6 | 91.3 |
| 8 | $X^2_{\text{corr}}$ | $\alpha$ | 3.5 | 4.1 | 4.5 | 4.7 | 4.8 | 4.5 | 4.5 | 5.0 | 5.0 |
|   | $\pi$  | 7.0 | 8.4 | 9.7 | 11.6 | 13.4 | 36.0 | 60.4 | 79.2 | 90.8 |
| 9 | $X^2$ | $\alpha$ | 7.2 | 6.5 | 6.0 | 5.6 | 5.6 | 5.3 | 5.1 | 5.1 | 4.9 |
|   | $\pi$  | 16.4 | 30.1 | 40.1 | 48.8 | 56.1 | 95.7 | 99.8 | 100.0 | 100.0 |
| 10| $c\hat{X}^2$ | $\alpha$ | 4.6 | 4.9 | 4.9 | 5.0 | 5.1 | 5.0 | 5.0 | 5.0 | 5.0 |
|   | $\pi$  | 7.0 | 8.5 | 10.0 | 11.7 | 13.5 | 37.0 | 61.7 | 80.0 | 91.2 |

5.2 Numerical example 2

A simulation study was done for the example described in ref. [18] and also in ref. [19]. Weighted histograms described in subsection 1.2 are used here.

The PDF $p_0(x)$ for the hypothesis $H_0$ is taken according to formula (1.11) with:

$$p_0(x') = 0.4(x' - 0.5) + 1; \quad x' \in [0, 1]$$

(5.11)

$$A(x') = 1 - (x' - 0.5)^2$$

(5.12)

$$R(x|x') = \frac{1}{\sigma \sqrt{2\pi}} \exp \left[ -\frac{(x-x')^2}{2\sigma^2} \right], \quad \text{with } \sigma = 0.3.$$  

(5.13)

For the alternative $H_a$, $p(x)$ is taken with the same acceptance and resolution function according to formula (1.11) with:

$$p_\alpha(x') = 0.6666(x' - 0.5) + 1; \quad x' \in [0, 1]$$

(5.14)

that is presented by the weighted histogram.

A calculation was done for two cases of PDFs used for event generation, see figure 3.

$$h_1(x') = 0.6666(x' - 0.5) + 1; \quad x' \in [0, 1]$$

(5.15)

and

$$h_2(x') = -0.6666(x' - 0.5) + 1; \quad x' \in [0, 1].$$

(5.16)
Figure 3. Probability density functions $h_1(x') = p_{tr}(x')$, $h_2(x')$ and $p_{0tr}(x')$ (dashed line).

Figure 4. Probabilities $p_i, i = 1, \ldots, 20$ for the PDF $p(x)$ (solid line) and $p_{0i}, i = 1, \ldots, 20$ for the PDF $p_0(x)$ (dashed line).

In the first case, a weighted histogram is the histogram with weights of events equal to 1 (histogram with unweighted entries) and, in the second case, weights of events equal to $h_1(x')/h_2(x')$. The results of this calculation for 100000 runs are presented in tables 3. We use a histogram with 20 bins on interval $[-0.3, 1.3]$. Figure 4 presented probabilities $p_i, i = 1, \ldots, 20$ for the PDF $p(x)$ and $p_{0i}, i = 1, \ldots, 20$ for the PDF $p_0(x)$. Here, we add two bins for events with $x \leq -0.3$ and $x > 1.3$ as well as one bin for events that were not registered due to limited acceptance. Total number of bins $m$ is used in test equal to 23. The results of calculations of the sizes and power of tests for 100000 runs are presented in table 3.
Table 3. Numerical example 2. Sizes ($\alpha$) and powers ($\pi$) of different test statistics $X^2$ (2.2), $\hat{X}^2$ (2.33), $c\hat{X}^2$ (3.11), $\hat{X}^2_{\text{Med}}$ (2.29), $c\hat{X}^2_{\text{Med}}$ (3.7) obtained for different weighted functions $w(x)$. Italic type marks a size of test with inappropriate number of events in the bins of histograms.

| №  | $n$ | 200  | 400  | 600  | 800  | 1000 | 3000 | 5000 | 7000 | 9000 | $w(x)$ |
|----|-----|------|------|------|------|------|------|------|------|------|--------|
| 1  | $X^2$ | $\alpha$ | 5.1  | 5.1  | 5.1  | 5.0  | 5.1  | 5.1  | 5.0  | 5.0  | 1      |
|    |      | $\pi$  | 5.6  | 6.6  | 7.5  | 8.8  | 9.8  | 25.9 | 45.7 | 64.9 | 79.4  |
| 2  | $\hat{X}^2$ | $\alpha$ | 7.0  | 6.2  | 5.8  | 5.6  | 5.5  | 5.1  | 5.0  | 4.9  | 4.9  |
|    |      | $\pi$  | 8.4  | 9.4  | 10.9 | 12.8 | 14.6 | 40.9 | 67.1 | 85.3 | 94.5  |
| 3  | $c\hat{X}^2$ | $\alpha$ | 5.6  | 5.6  | 5.5  | 5.4  | 5.3  | 5.1  | 5.0  | 5.0  | 4.9  |
|    |      | $\pi$  | 6.4  | 7.4  | 8.4  | 9.9  | 11.0 | 28.0 | 47.9 | 66.4 | 80.5  |
| 4  | $\hat{X}^2_{\text{Med}}$ | $\alpha$ | 10.9 | 7.4  | 6.6  | 6.1  | 6.1  | 5.7  | 5.6  | 5.6  | 5.6  |
|    |      | $\pi$  | 9.1  | 10.1 | 11.5 | 13.9 | 15.8 | 43.7 | 70.9 | 87.8 | 95.8  |
| 5  | $c\hat{X}^2_{\text{Med}}$ | $\alpha$ | 7.8  | 6.6  | 6.3  | 5.9  | 5.9  | 5.7  | 5.7  | 5.7  | 5.6  |
|    |      | $\pi$  | 6.1  | 7.2  | 8.4  | 9.7  | 10.9 | 27.4 | 46.9 | 65.0 | 79.2  |

Conclusion and interpretation of results presented in table 3.

- The size of new tests $\hat{X}^2$ (2.33) and $c\hat{X}^2$ (3.11) (row 2, 3) is more close to the nominal value 5% then the size of median tests $\hat{X}^2_{\text{Med}}$ (2.29) and $c\hat{X}^2_{\text{Med}}$ (3.7) (rows 4, 5).

- The power of new tests $\hat{X}^2$ (2.33) and $c\hat{X}^2$ (3.11) (rows 2, 3) is roughly the same compared with analogous median tests $\hat{X}^2_{\text{Med}}$ (2.29) and $c\hat{X}^2_{\text{Med}}$ (3.7) (rows 4, 5).

- All tests demonstrate greater power than Pearson’s test $X^2$ (2.2) (row 1) used for the histogram with unweighted entries.

The property of tests in application for Poisson histograms is investigated with the same weighted functions and distributions of events. In this case, the number of events $n$ in a histogram was simulated according Poisson distribution with given parameter $n_0$. The size and power of tests developed for the Poisson histogram in [13] was also calculated. Results of calculations are presented in table 4.

Conclusion and interpretation of results presented by table 4.

- The size of all tests are close to nominal value 5%.

- Basically, the power of new tests $\hat{X}^2$ (2.33) and $c\hat{X}^2$ (3.11) (rows 5, 6) in applying for Poisson histograms are greater than the power of tests developed ad hoc for the Poisson histograms $X^2_{\text{corr0}}$ (3.18) with the exactly known parameter $n_0$ and $X^2_{\text{corr}}$ (3.19) (rows 3, 4) with the unknown parameter $n_0$ in ref. [13].

- The power of Pearson’s test $X^2$ (row 2) used for Poisson histograms is greater than power of test $X^2_{\text{pois}}$ (3.16) (row 1) with the exactly known parameter $n_0$.

Generally the numerical example 1 and example 2 demonstrate the superiority of new goodness of fit tests under existing tests for weighted histograms, see ref. [2] and for weighted Poisson histograms, see ref. [13].
Table 4. Numerical example 2. Size ($\alpha$) and power ($\pi$) of different test statistics $X^2_{\text{pois}}$ (3.16), $X^2$, $X^2_{\text{corr}}$ (3.18), $\hat{X}^2$ (2.33), $c\hat{X}^2$ (3.11) in application for Poisson histograms. Italic type marks a size of test with inappropriate number of events in the bins of histograms.

| № | $n_0$ | 200 | 400 | 600 | 800 | 1000 | 3000 | 5000 | 7000 | 9000 | w(x) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | $X^2_{\text{pois}}$ | $\alpha$ | 5.5 | 5.3 | 5.3 | 5.0 | 5.2 | 5.1 | 5.1 | 5.0 | 5.2 |
|   |   | $\pi$ | 5.5 | 6.4 | 7.4 | 8.7 | 9.7 | 25.6 | 45.0 | 64.0 | 77.9 |
| 2 | $X^2$ | $\alpha$ | 5.1 | 5.1 | 5.1 | 4.9 | 5.2 | 5.0 | 5.1 | 5.0 | 5.2 |
|   |   | $\pi$ | 5.5 | 6.5 | 7.5 | 8.9 | 9.8 | 26.3 | 45.9 | 65.0 | 78.8 |
| 3 | $X^2_{\text{corr0}}$ | $\alpha$ | 5.8 | 5.8 | 5.5 | 5.3 | 5.3 | 5.0 | 5.1 | 5.0 | 5.0 |
| 4 | $X^2_{\text{corr}}$ | $\alpha$ | 4.2 | 4.9 | 4.9 | 4.8 | 4.9 | 4.9 | 5.0 | 4.9 | 5.0 |
|   |   | $\pi$ | 6.3 | 7.2 | 8.4 | 9.7 | 11.1 | 27.5 | 46.9 | 65.5 | 79.5 |
| 5 | $\hat{X}^2$ | $\alpha$ | 6.8 | 6.1 | 5.8 | 5.6 | 5.4 | 5.1 | 5.0 | 5.0 | 4.9 |
|   |   | $\pi$ | 8.4 | 9.4 | 11.0 | 12.7 | 14.8 | 41.1 | 67.2 | 85.1 | 94.4 |
| 6 | $c\hat{X}^2$ | $\alpha$ | 5.4 | 5.6 | 5.5 | 5.3 | 5.3 | 5.0 | 4.9 | 5.0 | 4.9 |
|   |   | $\pi$ | 6.5 | 7.4 | 8.5 | 9.8 | 11.0 | 28.0 | 48.2 | 66.4 | 80.4 |

6 Conclusion

A review of goodness of fit tests for weighted histograms was presented. The bin content of a weighted histogram was considered as a random sum of random variables that permits to generalize the classical Pearson’s goodness of fit test for histograms with weighted entries. Improvements of the chi-square tests with better statistical properties were proposed. Evaluation of the size and power of tests was done numerically for different types of weighted histograms with different numbers of events and different weight functions. Generally the size of new tests is closer to nominal value and power is not lower than have existing tests. Except direct application of tests in data analysis, see for example ref. [20], the proposed tests are necessary bases for generalization of test in the case when some parameters must be estimated from the data, see ref. [21], as well as for the generalisation of test for comparing weighted and unweighted histograms or two weighted ones (homogeneity test), see refs. [9, 21, 22]. Parametric fit of data obtained from detectors with finite resolution and limited acceptance is one of important application of methods developed for weighted histograms that can be used for experimental data interpretation, see refs. [19].

Acknowledgments

The author is grateful to Johan Blouw for useful discussions and careful reading of the manuscript and thanks the University of Akureyri and the MPI for Nuclear Physics for support in carrying out the research.

References

[1] M.G. Kendall and A.S. Stuart, The advanced theory of statistics, vol. 2, section 30.4, Griffin Publishing Company, London U.K. (1973).
[2] N.D. Gagunashvili, *Goodness of fit tests for weighted histograms*, Nucl. Instrum. Meth. A 596 (2008) 439.

[3] I.M. Sobol, *Numerical Monte Carlo methods*, chapter 5, Nauka, Moscow U.S.S.R. (1973).

[4] K. Pearson, *On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling*, Philos. Mag. 50 (1900) 157.

[5] M.G. Kendall and A.S. Stuart, *The advanced theory of statistics*, vol. 1, section 15.10, Griffin Publishing Company, London U.K. (1973).

[6] B.V. Gnedenko and V.Yu. Korolev, *Random summation: limit theorems and applications*, CRC Press, Boca Raton U.S.A. (1996).

[7] H.V. Henderson and S.R. Searle, *On deriving the inverse of a sum of matrices*, SIAM Rev. 23 (1981) 53.

[8] H. Robbins, *The asymptotic distribution of the sum of a random number of random variables*, Bull. Amer. Math. Soc. 54 (1948) 1151.

[9] N.D. Gagunashvili, *Chi-square tests for comparing weighted histograms*, Nucl. Instrum. Meth. A 614 (2010) 287 [arXiv:0905.4221].

[10] G.H. Golub, *Some modified matrix eigenvalue problems*, SIAM Rev. 15 (1973) 318.

[11] J. von Neumann, *Various techniques used in connection with random digits*, NBS Appl. Math. Series 12 (1951) 36.

[12] S. Baker and R.D. Cousins, *Clarification of the use of chi-square and likelihood functions in fits to histograms*, Nucl. Instrum. Meth. 221 (1984) 437.

[13] G. Zech, *A goodness-of-fit test for histograms of weighted events*, Nucl. Instrum. Meth. A 691 (2012) 178.

[14] W.T. Eadie, D. Dryard, F.E. James, M. Roos and B. Sadoulet, *Statistical methods in experimental physics*, section 4.1.2, North-Holland Publishing Company, Amsterdam, London (1971).

[15] D.S. Moore and G.P. McCabe, *Introduction to the practice of statistics*, W.H. Freeman Publishing Company, New York U.S.A. (2007).

[16] W.G. Cochran, *The \( \chi^2 \) test of goodness of fit*, Ann. Math. Stat. 23 (1952) 315.

[17] N.D. Gagunashvili, *CHIWEI: a code of goodness of fit tests for weighted and unweighted histograms*, Comput. Phys. Commun. 183 (2012) 418 [arXiv:1104.3733].

[18] G. Bohm and G. Zech, *Introduction to statistics and data analysis for physicists*, Verlag Deutsches Elektronen-Synchrotron (2010).

[19] N.D. Gagunashvili, *Parametric fitting of data obtained from detectors with finite resolution and limited acceptance*, Nucl. Instrum. Meth. A 635 (2011) 86 [arXiv:1011.0662].

[20] D0 collaboration, V.M. Abazov et al., *Dependence of the \( t\bar{t} \) production cross section on the transverse momentum of the top quark*, Phys. Lett. B 693 (2010) 515 [arXiv:1001.1900].

[21] H. Cramer, *Mathematical methods of statistics*, chapter 30, Princeton University Press, Princeton U.S.A. (1999).

[22] N.D. Gagunashvili, *CHICOM: a code of tests for comparing unweighted and weighted histograms and two weighted histograms*, Comput. Phys. Commun. 183 (2012) 193 [arXiv:1105.1288].