Homogeneity tests of weighted samples in ROOT

Jakub Trusina

FNSPE CTU in Prague, Department of Mathematics

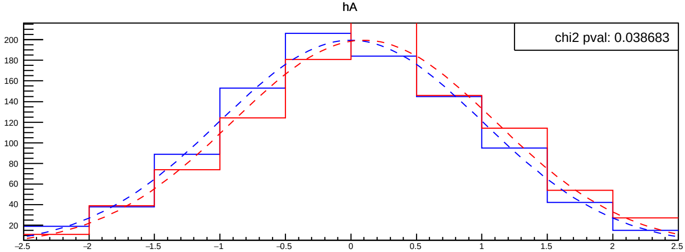
24 June 2019
ATLAS CZ+SK Workshop

Introduction

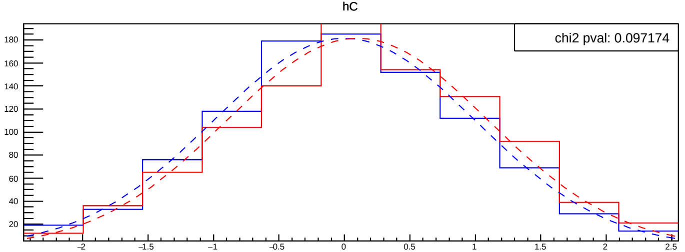
- Homogeneity testing
  - MC vs. data
  - Instability of ROOT's tests
- Weighted samples
  - $(\boldsymbol{X}, \boldsymbol{W}) = ((X_1, W_1), ..., (X_n, W_n))^T$
  - $(\boldsymbol{X}, \boldsymbol{W}) \to \widetilde{\boldsymbol{X}} \sim F^{(n)}$
  - weighted observations vs. weighted histogram

Problems of TH1::Chi2Test – Binning effect

- nbins = 10
- min = -2.5
- max = 2.5



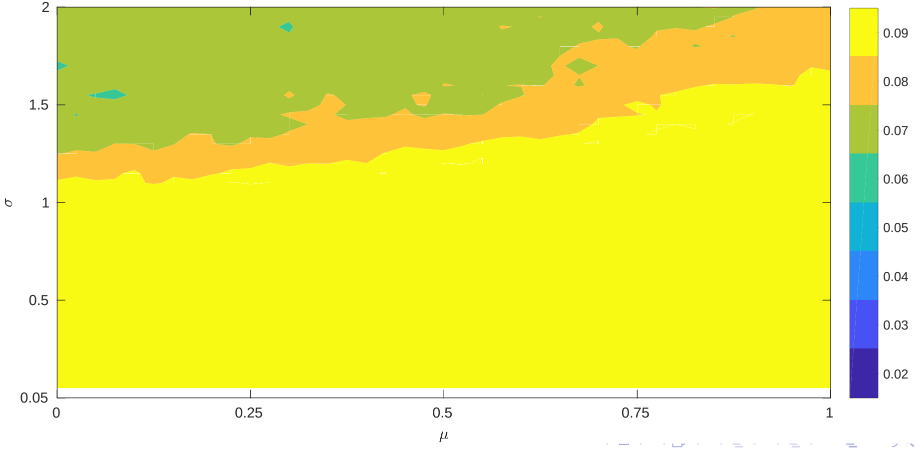
- nbins = 11
- min = -2.45
- max = 2.55



Problems of TH1::Chi2Test – Effect of weights

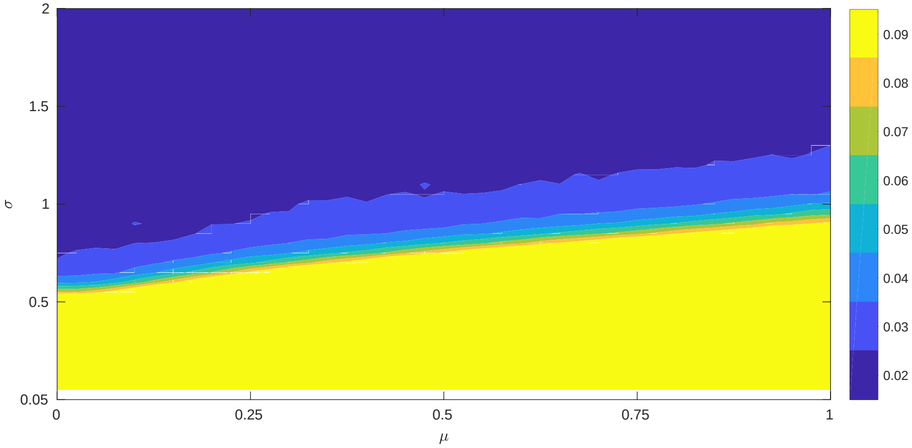
- $(\boldsymbol{X}, \boldsymbol{W})$ vs. $(\boldsymbol{Y}, \boldsymbol{1})$
- $X_i \sim \mathcal{N}(\mu, \sigma^2)$, $Y_i \sim \mathcal{N}(0, 1)$
- $W_i = \frac{f_Y(X_i)}{f_X(X_i)} \implies H_0$ is true

- R … ratio of rejection
- $R \approx \alpha = 0.05$ (significance level)



Problems of TH1::KolmogorovTest

- KS test is originally derived for continuous data samples
- computed *p*-value is higher than the true one



Possible solution: tests based on empirical distribution function

- EDF contains full information about the sample
- we need to modify some constants and functions
- invariant to scaling of weights
- is stable even if the number of events are different
- three generalized tests: KS, CvM and AD

Unweighted	Weighted
n	$n_e = \left(\sum_{i=1}^{n} W_i\right)^2 / \sum_{i=1}^{n} W_i^2$
$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{(-\infty, x_i]}(x)$	$F_n^{\boldsymbol{W}}(x) = \frac{1}{W_{\bullet}} \sum_{i=1}^{n} W_i \mathbf{1}_{(-\infty, x_i]}(x)$

Tests based on EDF

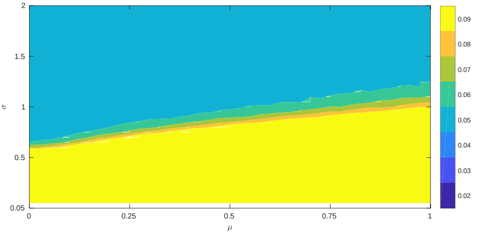


Figure: KS test

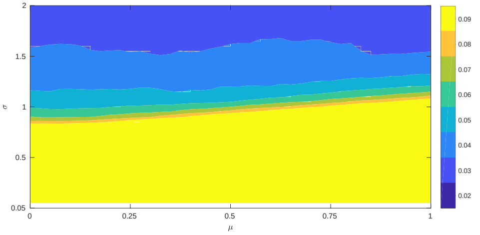


Figure: AD test

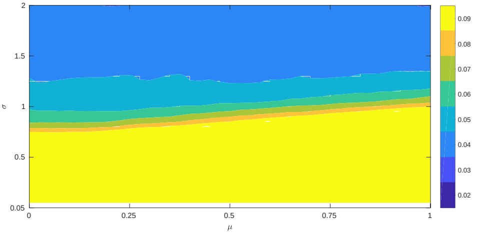


Figure: CvM test

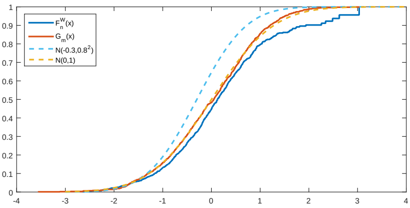


Figure: problematic configuration

Ratios of rejected tests - weights *i.i.d.*

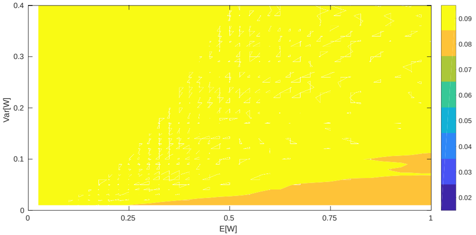


Figure: TH1::Chi2Test

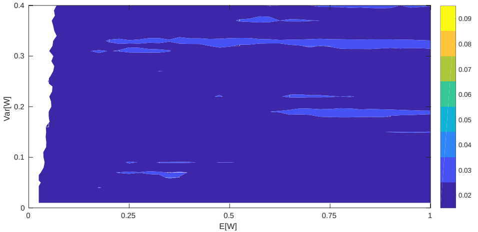


Figure: TH1::KolmogorovTest

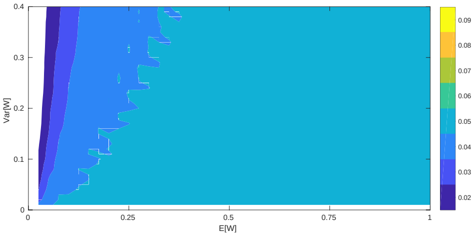


Figure: KS test

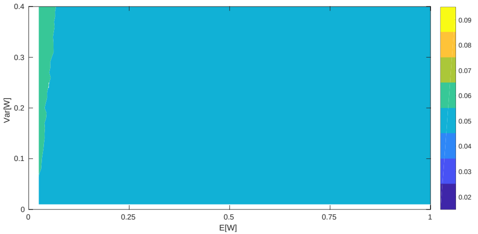


Figure: AD test

Power of test

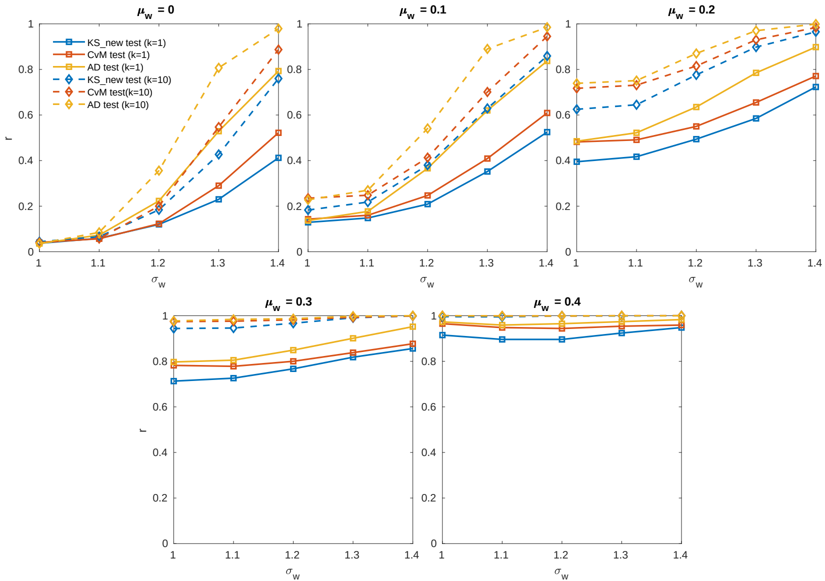


Figure: Various μ_w, σ_w, k and fixed $\mu_s = 0.2$, $\sigma_s = 1.3$, $nevents = 200$.

Power of test

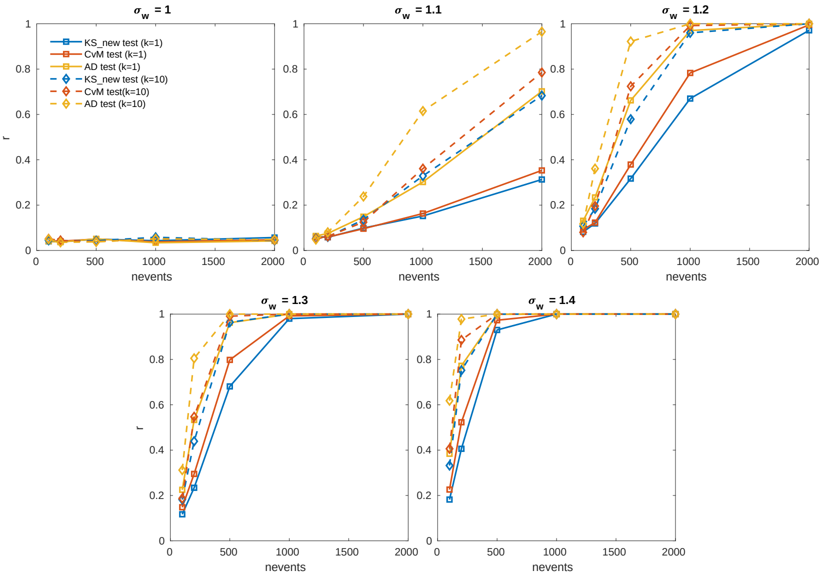


Figure: Various σ_w, *nevents*, k and fixed $\mu_s = 0.1$, $\sigma_s = 1.2$, $\mu_w = 0$.

Power of test

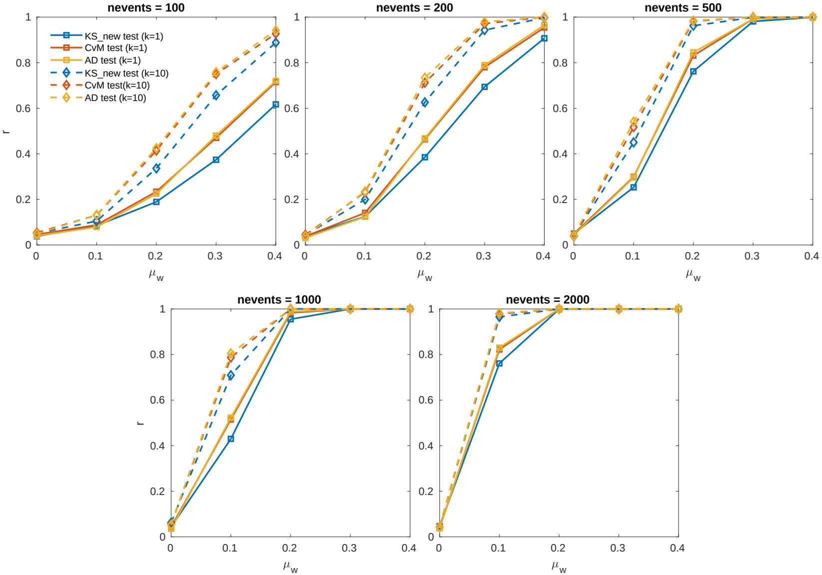


Figure: Various *nevents*, μ_w, k and fixed $\mu_s = 0$, $\sigma_w = 1.4$, $\sigma_w = 0$.

Conclusion

- Weighted homogeneity tests for binned data are generally unstable
- Tests based on EDF compute *p*-values correctly
  - ROOT code: `http://gams.fjfi.cvut.cz/homtests`
  - Implementation in ROOT::Math::GoFTest soon

Backup – Weighted samples

$$
\begin{aligned}
(\boldsymbol{X}, \boldsymbol{W}) &= ((X_1, \ldots, X_n)^T, (W_1, \ldots, W_n)^T) \\
(\boldsymbol{Y}, \boldsymbol{V}) &= ((Y_1, \ldots, Y_m)^T, (V_1, \ldots, V_m)^T)
\end{aligned}
$$

Effective entries size

$$n_e = \left(\sum_{i=1}^{n} W_i \right)^2 / \sum_{i=1}^{n} W_i^2$$

Weighted empirical distribution function (WEDF)

$$F_n^{\boldsymbol{W}}(x) = \frac{1}{W_\bullet} \sum_{i=1}^{n} W_i \mathbf{1}_{(-\infty, X_i]}(x)$$

WEDF of mixed sample

$$H_{n_e, m_e}^{\boldsymbol{W}, \boldsymbol{V}}(x) = \frac{n_e F_n^{\boldsymbol{W}}(x) + m_e F_m^{\boldsymbol{V}}(x)}{n_e + m_e}$$

1. Weights – *i.i.d.*

$$W_i \sim \mathrm{P}^W$$

2. Weights – ratio of PDFs

$$W_i = \frac{g(X_i)}{f(X_i)}$$

3. Weights – CDF

$$W_i = G(X_i) - G(X_{i-1})$$

Backup – Properties of WEDF

Weights – *i.i.d.*

$F_n^{\boldsymbol{W}}(x)$ is an unbiased and strongly consistent estimate of $F_X(x)$.

Weights – ratio of PDFs

$F_n^{\boldsymbol{W}}(x)$ is asymptotically unbiased and strongly consistent est. of $F_Y(x)$.

Weights – CDF

$F_n^{\boldsymbol{W}}(x)$ is strongly consistent estimate of $F_Y(x)$.

Generalized Glivenko-Cantelli theorem

Let $\boldsymbol{X} = (X_1, ..., X_n)^T$ be *i.i.d.* random vector from F. Let $\boldsymbol{W} = (W_1, ..., W_n)^T$ be its weights. Let $F_n^{\boldsymbol{W}}$ be their WEDF which value $F_n^{\boldsymbol{W}}(x)$ is for all $x \in \mathbb{R}$ strongly consistent estimate of value $G(x)$. Then $\sup_{x \in \mathbb{R}} |F_n^{\boldsymbol{W}}(x) - G(x)| \xrightarrow{\text{s.j.}} 0$.

Backup –Homogeneity tests of weighted samples

Kolmogorov-Smirnov test

$$T_{n,m}^{\boldsymbol{W},\boldsymbol{V}} = \sqrt{\frac{n_e m_e}{n_e+m_e}} \sup_{x\in\mathbb{R}} \left|F_n^{\boldsymbol{W}}(x) - F_m^{\boldsymbol{V}}(x)\right|$$

$$K(\lambda) = 1 - 2\sum_{k=1}^{+\infty}(-1)^{k+1}e^{-2k^2\lambda^2}$$

Cramér-von Mises test

$$T_{n,m}^{\boldsymbol{W},\boldsymbol{V}} = \frac{n_e m_e}{n_e+m_e}\int_{\mathbb{R}}\left(F_n^{\boldsymbol{W}}(x) - F_m^{\boldsymbol{V}}(x)\right)^2 dH_{n_e,m_e}^{\boldsymbol{W},\boldsymbol{V}}$$

$$L(z) = \frac{1}{\pi\sqrt{z}}\sum_{k=0}^{+\infty}(-1)^k\binom{-\frac{1}{2}}{k}\sqrt{1+4k}\ \exp\left(-\frac{(1+4k)^2}{16z}\right)K_{\frac{1}{4}}\left(\frac{(1+4k)^2}{16z}\right)$$

Anderson-Darling test

$$T_{n,m}^{\boldsymbol{W},\boldsymbol{V}} = \frac{n_e m_e}{n_e+m_e}\int_{0<H_{n_e,m_e}^{\boldsymbol{W},\boldsymbol{V}}(x)<1}\frac{\left(F_n^{\boldsymbol{W}}(x)-F_m^{\boldsymbol{V}}(x)\right)^2}{H_{n_e,m_e}^{\boldsymbol{W},\boldsymbol{V}}(x)\left(1-H_{n_e,m_e}^{\boldsymbol{W},\boldsymbol{V}}(x)\right)}dH_{n_e,m_e}^{\boldsymbol{W},\boldsymbol{V}}$$

$$M(z) = \frac{\sqrt{2\pi}}{z}\sum_{k=0}^{+\infty}\binom{-\frac{1}{2}}{k}(1+4k)\exp\left(-\frac{(1+4k)^2\pi^2}{8z}\right)\int_0^{+\infty}\exp\left(\frac{z}{8(w^2+1)} - \frac{(1+4k)^2\pi^2 w^2}{8z}\right)dw$$