Compressor

JI and SC

December 10, 2014

1 Estimators

Lets call $g(x, Q)$ the prior set of 1000 replicas and $f(x, Q)$ the set of compressed replicas with $N_{\rm rep}$ (replica r denoted by $f^r(x, Q)$), which is minimized by the GA and $r(x, Q)$ the set of random replicas. The number of random trials of random sets is denoted by $N_{\rm rand}$. The current GA minimization sums over all estimators presented below. The normalization factors are extracted as the upper 68% cl. band obtained from the random set.

1.1 Central value

$$ERF_{CV} = \frac{1}{N_{\rm CV}} \sum_{i=-n_f}^{n_f} \sum_{j=1}^{N_x} \left(\frac{f_i^{\rm CV}(x_j, Q) - g_i^{\rm CV}(x_j, Q)}{g_i^{\rm CV}(x_j, Q)} \right)^2 \tag{1}$$

where

$$f_i^{\rm CV}(x_j, Q) = \frac{1}{N_{\rm rep}} \sum_{r=1}^{N_{\rm rep}} f_i^r(x_j, Q) \tag{2}$$

and

$$N_{\rm CV} = \frac{1}{N_{\rm rand}} \sum_{d=1}^{N_{\rm rand}} \sum_{i=-n_f}^{n_f} \sum_{j=1}^{N_x} \left(\frac{r_i^{d,\rm CV}(x_j, Q) - g_i^{\rm CV}(x_j, Q)}{g_i^{\rm CV}(x_j, Q)} \right)^2 \tag{3}$$

1.2 Standard deviation

$$ERF_{STD} = \frac{1}{N_{\rm STD}} \sum_{i=-n_f}^{n_f} \sum_{j=1}^{N_x} \left(\frac{f_i^{\rm STD}(x_j, Q) - g_i^{\rm STD}(x_j, Q)}{g_i^{\rm STD}(x_j, Q)} \right)^2 \tag{4}$$

where

$$f_i^{\rm STD}(x_j, Q) = \sqrt{\frac{1}{N_{\rm rep}} \sum_{r=1}^{N_{\rm rep}} f_i^r(x_j, Q)^2 - f_i^{\rm CV}(x_j, Q)^2} \tag{5}$$

and

$$N_{\mathrm{STD}} = \frac{1}{N_{\mathrm{rand}}} \sum_{d=1}^{N_{\mathrm{rand}}} \sum_{i=-n_f}^{n_f} \sum_{j=1}^{N_x} \left(\frac{r_i^{d,\mathrm{STD}}(x_j,Q) - g_i^{\mathrm{STD}}(x_j,Q)}{g_i^{\mathrm{STD}}(x_j,Q)} \right)^2 \tag{6}$$

1.3 Skewness

$$ERF_{SKE} = \frac{1}{N_{\mathrm{SKE}}} \sum_{i=-n_f}^{n_f} \sum_{j=1}^{N_x} \left(\frac{f_i^{\mathrm{SKE}}(x_j,Q) - g_i^{\mathrm{SKE}}(x_j,Q)}{g_i^{\mathrm{SKE}}(x_j,Q)} \right)^2 \tag{7}$$

where

$$f_i^{\mathrm{SKE}}(x_j,Q) = \frac{1}{N_{\mathrm{rep}}} \sum_{r=1}^{N_{\mathrm{rep}}} \left(f_i^r(x_j,Q) - f_i^{\mathrm{CV}}(x_j,Q) \right)^3 / \left(f_i^{\mathrm{STD}}(x_j,Q) \right)^3 \tag{8}$$

and

$$N_{\mathrm{SKE}} = \frac{1}{N_{\mathrm{rand}}} \sum_{d=1}^{N_{\mathrm{rand}}} \sum_{i=-n_f}^{n_f} \sum_{j=1}^{N_x} \left(\frac{r_i^{d,\mathrm{SKE}}(x_j,Q) - g_i^{\mathrm{SKE}}(x_j,Q)}{g_i^{\mathrm{SKE}}(x_j,Q)} \right)^2 \tag{9}$$

1.4 Kurtosis

$$ERF_{KUR} = \frac{1}{N_{\mathrm{KUR}}} \sum_{i=-n_f}^{n_f} \sum_{j=1}^{N_x} \left(\frac{f_i^{KUR}(x_j,Q) - g_i^{\mathrm{KUR}}(x_j,Q)}{g_i^{\mathrm{KUR}}(x_j,Q)} \right)^2 \tag{10}$$

where

$$f_i^{\mathrm{KUR}}(x_j,Q) = \frac{1}{N_{\mathrm{rep}}} \sum_{r=1}^{N_{\mathrm{rep}}} \left(f_i^r(x_j,Q) - f_i^{\mathrm{CV}}(x_j,Q) \right)^4 / \left(f_i^{\mathrm{STD}}(x_j,Q) \right)^4 \tag{11}$$

and

$$N_{\mathrm{KUR}} = \frac{1}{N_{\mathrm{rand}}} \sum_{d=1}^{N_{\mathrm{rand}}} \sum_{i=-n_f}^{n_f} \sum_{j=1}^{N_x} \left(\frac{r_i^{d,\mathrm{KUR}}(x_j,Q) - g_i^{\mathrm{KUR}}(x_j,Q)}{g_i^{\mathrm{KUR}}(x_j,Q)} \right)^2 \tag{12}$$

1.5 Kolmogorov

$$ERF_{KOL} = \frac{1}{N_{KOL}} \sum_{i=-n_f}^{n_f} \sum_{j=1}^{N_x} \sum_{k=1}^{6} \left(\frac{F_i^k(x_j,Q) - G_i^k(x_j,Q)}{G_i^k(x_j,Q)} \right)^2 \tag{13}$$

where F and Q contains the number of replicas contained in the k regions where the Kolmogorov test is applied, there are 6 regions defined as multiples of standard deviations:

$$[-\infty, -2f_i^{\mathrm{STD}}(x_j,Q), -f_i^{\mathrm{STD}}(x_j,Q), 0, f_i^{\mathrm{STD}}(x_j,Q), 2f_i^{\mathrm{STD}}(x_j,Q), +\infty] \tag{14}$$

We count the number of replicas which fall in each region and then we normalize by the total number of replicas of the respective set.

$$N_{KOL} = \frac{1}{N_{\rm rand}} \sum_{d=1}^{N_{\rm rand}} \sum_{i=-n_f}^{n_f} \sum_{j=1}^{N_x} \sum_{k=1}^{6} \left(\frac{R_i^k(x_j,Q) - G_i^k(x_j,Q)}{G_i^k(x_j,Q)} \right)^2 \tag{15}$$

1.6 Correlation

We define the correlation matrix C for any PDF set as:

$$C_{ij} = \frac{\langle ij \rangle - \langle i \rangle \langle j \rangle}{\sigma_i \cdot \sigma_j} \tag{16}$$

where for each flavor [-3,3] we define 3 points in x so:

$$\langle i \rangle = \frac{1}{N_{\rm rep}} \sum_{r=1}^{N_{\rm rep}} f_i^r(x_i,Q) \tag{17}$$

$$\langle ij \rangle = \frac{1}{N_{\rm rep}} \sum_{r=1}^{N_{\rm rep}} f_i^r(x_i,Q) f_j^r(x_j,Q) \tag{18}$$

$$\sigma_i = \sqrt{\frac{1}{N_{\rm rep}} \sum_{r=1}^{N_{\rm rep}} \left(f_i^r(x_i,Q) - \langle i \rangle\right)^2} \tag{19}$$

We verify the prior trace:

$$g = {\rm Tr}\left(P \cdot P^{-1}\right) = 21 \tag{20}$$

and store P^{-1}, then we compute the estimator:

$$f = {\rm Tr}\left(C \cdot P^{-1}\right) \tag{21}$$

for each compressed set. Finally, for the compression we minimize the quantity:

$$ERF_{Cor} = \frac{1}{N_{Cor}} \left(\frac{f-g}{g} \right)^2 \tag{22}$$

where N_{Cor} is computed from the random sets.