Compressor

JI and SC

November 27, 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.

1.1 Central value

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

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

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

1.2 Standard deviation

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

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

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

1.3 Skewness

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

where

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

and

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

1.4 Kurtosis

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

where

$$f_i^{\text{KUR}}(x_j, Q) = \frac{1}{N_{\text{rep}}} \sum_{r=1}^{N_{\text{rep}}} \left(f_i^r(x_j, Q) - f_i^{\text{CV}}(x_j, Q)\right)^4 / \left(f_i^{\text{STD}}(x_j, Q)\right)^4 - 3 \tag{11}$$

and

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

1.5 Kolmogorov

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

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:

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

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.

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

1.6 Correlations

We define the correlation matrix C_{ij}, where i, j are the flavors indexes [-3,3] as:

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

where

$$\langle i \rangle = \frac{1}{N_{\text{rep}}} \sum_{n=1}^{N_{\text{rep}}} \sum_{k=1}^{N_x} f_i^n(x_k, Q) \tag{17}$$

$$\langle ij \rangle = \frac{1}{N_{\text{rep}}} \sum_{n=1}^{N_{\text{rep}}} \sum_{k=1}^{N_x} f_i^n(x_k, Q) \cdot f_j^n(x_k, Q) \tag{18}$$

$$\sigma_i = \sqrt{\frac{1}{N_{\text{rep}}} \sum_{n=1}^{N_{\text{rep}}} \left(\left(\sum_{k=1}^{N_x} f_i^n(x_k, Q) \right) - \langle i \rangle \right)^2} \tag{19}$$

The ERF is then applied to the eigenvalues:

$$ERF == \frac{1}{N_{CORR}} \sum_{i=1}^{n_f} (Eigval_{comp} - Eigval_{\text{prior}})^2 \tag{20}$$