Proposed list of corrections for NIST SP 800-90B 6.3 Estimators

Gen'ya SAKURAI

2022/12/29

Chapter 1

Proposed list of corrections for NIST SP 800-90B 6.3 Estimators

1.1 Introduction

This list of corrections for NIST SP 800-90B [1] 6.3 Estimators has been drafted so that an entropy estimating tool for claiming conformance can be developed in a traceable manner.

1.2 Corrections to 6.3.4 The Compression Estimate

1. Correction to step 4-b-ii

> If $dict[s_i]$ is zero, add that value to the dictionary, i.e., $dict[s_i'] = i$. Let $D_{i-d} = i$.

should be replaced by the following:

> If $dict[s_i{}']$ is zero, add that value to the dictionary, i.e., $dict[s_i'] = i$. Let $D_{i-d} = i$.

2. Correction to the expression of $G(z)$
The Eq.(1.1) should be replaced by Eq.1.2.

$$G(z) = \frac{1}{\nu} \sum_{t=d+1}^{L} \sum_{u=1}^{t} \log_2(u) F(z,t,u) \tag{1.1}$$

$$G(z) = \frac{1}{\nu} \sum_{t=d+1}^{\lfloor L/b \rfloor} \sum_{u=1}^{t} \log_2(u) F(z,t,u) \tag{1.2}$$

This correction makes sense if the summation over t starts from $d+1$ then it should end at $\lfloor L/b \rfloor$. Also the factor $\frac{1}{\nu}$ coincides with this argument as $\nu = \lfloor L/b \rfloor - d$.
The Eq.(1.2) can be further rewritten to as Eq.(1.3), as $\log_2(1) = 0$.

$$G(z) = \frac{1}{\nu} \sum_{t=d+1}^{\lfloor L/b \rfloor} \sum_{u=2}^{t} \log_2(u) F(z,t,u) \tag{1.3}$$

Note that the r.h.s. of Eq.(1.3) can be optimized further for algorithmic efficiency.

1.3 Corrections to 6.3.10 The LZ78Y Prediction Estimate

1. Missing step
   The variable C is used in step 4 without its definition. So the following new step should be introduced just before step 4, and steps 4 to 6 should be renumbered accordingly.

> Let C be the number of ones in *correct*.

2. Corrections to step 3-a-ii

> ii: If $(s_{i-j-1}, \ldots, s_{i-2})$ is in D,
> Let $D[s_{i-j-1}, \ldots, s_{i-2}][s_{i-1}] = D[s_{i-j-1}, \ldots, s_{i-2}][s_{i-1}] + 1$

should be replaced by the following:

> ii: **if** $(s_{i-j-1}, \ldots, s_{i-2})$ is in D, **then**
> iii: **if** $[(s_{i-j-1}, \ldots, s_{i-2}), s_{i-1}]$ is in D, **then**
> iv: Let $D[s_{i-j-1}, \ldots, s_{i-2}][s_{i-1}] = D[s_{i-j-1}, \ldots, s_{i-2}][s_{i-1}] + 1$
> v: **else**
> vi: **if** $dictionarySize < maxDictionarySize$ **then**
> vii: Let $D[s_{i-j-1}, \ldots, s_{i-2}][s_{i-1}] = 0$
> ▷ With this step, the issue can be resolved that the value of $D[s_{i-j-1}, \ldots, s_{i-2}][s_{i-1}]$ is undefined, when $(s_{i-j-1}, \ldots, s_{i-2})$ is in D but $[(s_{i-j-1}, \ldots, s_{i-2}), s_{i-1}]$ is not in D.
> viii: $dictionarySize = dictionarySize + 1$
> ▷ The value $dictionarySize$ is equal to the size of dictionary D, or the number of $(\boldsymbol{x}, y)$ pairs in $D[\boldsymbol{x}][y]$.
> ix: **end if**
> x: **end if**
> xi: **end if**

3. Column header in Example
   Also the column header

> **Max $\boldsymbol{D}[\boldsymbol{prev}]$ entry**

in the table for *Example* should be replaced by

> $\arg\max_y \boldsymbol{D}[\boldsymbol{prev}][y]$

.

Bibliography

[1] Meltem Sönmez Turan, Elaine Barker, John Kelsey, Kerry A. McKay, Mary L. Baish, Mike Boyle *Recommendation for the Entropy Sources Used for Random Bit Generation*, NIST Special Publication 800-90B, Jan. 2018

[2] Franck W. J. Oliver, Daniel W. Lozier, Ronald F. Boisvert, Charles W. Clark, *NIST Handbook of Mathematical Functions*, National Institute of Standards and Technology, 2010