

Supplementary File 2

Two ways of computing E-values in HMMER2: Extreme-value distribution and logistics function

In HMMER2, the routine “PValue” as part of the source file “Mathsupport.c” is responsible for calculating the significance estimate P for a hit with a given score s (see commented source code below). The significance is calculated in two ways, first via a logistics function (comment 1)

$$P_{\text{logistics}}(S \geq s) = \frac{1}{1 + e^{s \log 2}}$$

and then via an extreme value distribution function (comment 2)

$$P_{\text{EVD}}(S \geq s) = 1 - \exp(-e^{-\lambda(s-\mu)}).$$

The smaller of the two values (comment 3) is return to the calling program.

```
/* Function: PValue()
 * Date:      SRE, Mon Oct 27 12:21:02 1997 [Sanger Centre, UK]
 *
 * Purpose:   Convert an HMM score to a P-value.
 *            We know P(S>x) is bounded by 1 / (1 + exp_2^x) for a bit score of x.
 *            We can also use EVD parameters for a tighter bound if we have
 *            them available.
 *
 * Args:      hmm - model structure, contains EVD parameters
 *            sc  - score in bits
 *
 * Returns:   P value for score significance.
 */
double
PValue(struct plan7_s *hmm, float sc)
{
    double pval;
    double pval2;

    /* the bound from Bayes */
    if (sc >= sreLOG2(DBL_MAX)) pval = 0.0;
    else if (sc <= -1. * sreLOG2(DBL_MAX)) pval = 1.0;

    /* Comment 1 : The logistic function is first used to compute the initial p-value.
    */

    else
        pval = 1. / (1.+sreEXP2(sc));
        /* try for a better estimate from EVD fit */

    /* Comment 2 : After hmmlcalibrate, the 8th bit in hmm->flags is set. When hmm-
    >flags is bitwise AND-ed with PLAN7_STATS (with value binary 10000000), the final
    hmm->flags is non-zero. To be exact, the 8th bit in the variable and constant will
    evaluate to one at the 8th bit while the other bits evaluate to zero. Thus,
    following will return true. */

    if (hmm != NULL && (hmm->flags & PLAN7_STATS))
        {

    /* Comment 3 : The EVD (extreme value distribution) is used to compute the second
    p-value. */

        pval2 = ExtremeValueP(sc, hmm->mu, hmm->lambda);
```

```
/* Comment 4 : The smaller p-value is always chosen for the subsequent computation  
of the final E-value. */
```

```
    if (pval2 < pval) pval = pval2;  
    }  
    return pval;  
}
```