One All But One

1 One

The context of this discussion is where there is a test of hypothesis of the Null hypothesis that none of the K key word sets have ELSs that are in more compact arrangement than expected by chance against the Alternative Hypothesis that one of K key word sets have ELSs that are in a more compact arrangement than expected by chance.

The experiment has N trials, trial 1 being the trial involving the Torah text and trials 2 through N involving monkey texts. Each trial produces K compactness values, one for the ELSs of each of the K key word sets. These compactness observations can be arranged in a matrix of N rows by K columns. Each row corresponds to the K compactness values observed in a text. Each column corresponds to the compactness values of a key word set observed over the N trials. We denote the entry in row n column k by c_{nk}.

Let us take the convention that small compactness values means very good compactness. Now create a new N by K matrix, where the entry p_{nk} is computed as the normalized relative rank[1] of c_{nk} taken over all the N entries, $c_{1k}, \ldots, c_{Nk}$ of the k^{th} column. Assuming no ties, for each k, $p_{1k}, \ldots, p_{Nk}$, is distributed uniformly on the discrete set $\{1/2N, 3/2N, \ldots, (2N-1)/2N\}$.

Under the Null hypothesis, the K normalized relative ranks of the first text, the Torah text, $p_{11}, \ldots, p_{1K}$ are uniformly distributed on the discrete set $\{1/2N, 3/2N, \ldots, (2N-1)/2N\}$. Under the Alternative hypothesis, one of these normalized relative ranks must have a small relative rank and the remaining ones are uniformly distributed on the discrete set. To be uniformly distributed on the discrete set $\{1/2N, 3/2N, \ldots, (2N-1)/2N\}$ means that each possible value in the set has equal probability, that probability being $\Delta = 1/N$. One value having small relative rank means that it has higher probability for taking smaller values in the set $\{1/2N, 3/2N, \ldots, (2N-1)/2N\}$.

Let Δ be the constant probability that a relative rank comes from population 1, governed by the Null hypothesis. Let $P_2(v)$ be the probability

[1] The normalized relative rank is the number of values less than it plus one half the number of values equal to it divided by the total number of values.

that a relative rank comes from population 2, governed by the Alternative hypothesis.

Define

$$\Phi_1 = \{\phi \subset \{1, \ldots, K\} \mid \#\phi = 1\}$$

To motivate the definition for the test statistic we derive, we make an assumption in the derivation: that the key word sets are independent. We know this is assuredly not true. So the test statistic we derive must be just considered as a score whose distribution is not known. However, this score is computed for the text of each trial. Hence it is possible to compute the relative rank of the score to determine the p-value associated with the hypothesis test.

To simplify our notation in the derivation, let $x_1, \ldots, x_K$ be the K normalized relative ranks associated with the Torah text: $x_k = p_{1k}, k = 1, \ldots, K$

$$\begin{aligned}
P(x_1, \ldots, x_K \mid \Phi_1) &= \frac{P(x_1, \ldots, x_K, \Phi_1)}{P(\Phi_1)} \\
&= \sum_{\phi \in \Phi_1} \frac{P(x_1, \ldots, x_K, \phi)}{1} \\
&= \sum_{\phi \in \Phi_1} P(x_1, \ldots, x_K \mid \phi) P(\phi) \\
&= \sum_{\phi \in \Phi_1} P(x_1, \ldots, x_K \mid \phi) \frac{1}{K} \\
&= \sum_{\phi \in \Phi_1} \prod_{j \in \phi^c} \Delta \prod_{j \in \phi} P_2(x_j) \\
&= \frac{\Delta^{K-1}}{K} \sum_{k=1}^{K} P_2(x_k)
\end{aligned}$$

When smaller values are more probable, and when x is known to lie in a range that is greater than 0 and less than or equal to 1, then probability distributions that monotonically decrease with increasing value include the harmonic distribution $P_2(x) = \frac{\alpha}{x}$ and the geometric distribution $P_2(x) = -\beta log(x)$

Therefore, our score function S is given by

$$S = \frac{\Delta^{K-1}}{K} \sum_{k=1}^{K} P_2(x_k)$$

1.0.1 Testing the Null Hypothesis

Suppose that we are testing the Null Hypothesis that each $x_1 \ldots, x_k$ comes from the discrete uniform distribution against the Alternative Hypothesis that one of the observed values come from population 2 which is distributed by the harmonic distribution. Then the likelihood ratio of the Null hypothesis to the Alternative Hypothesis is given by

$$\begin{aligned} LR &= \frac{\Delta^K}{\frac{\Delta^{K-1}}{K}\sum_{k=1}^{K}\alpha/x_k} \\ &= \frac{\Delta}{\alpha}\frac{1}{\frac{1}{K}\sum_{k=1}^{K}\frac{1}{x_k}} \\ &= \frac{\Delta}{\alpha}\text{harmonic mean}(x_1, \ldots, x_K) \end{aligned}$$

This motivates the use of the harmonic mean of $x_1, \ldots, x_K$ as the test statistic. When the harmonic mean is sufficiently small, the Null hypothesis would be rejected.

Suppose that we are testing the Null Hypothesis that each $x_1 \ldots, x_k$ comes from the discrete uniform distribution against the Alternative Hypothesis that one of the observed values come from population 2 which is distributed by the geometric distribution. Then the likelihood ratio of the Null hypothesis to the Alternative Hypothesis is given by

$$\begin{aligned} LR &= \frac{\Delta^K}{\frac{\Delta^{K-1}}{K}\sum_{k=1}^{K}-\beta\log(x_k)} \\ &= -\frac{\Delta}{\beta}\frac{1}{\frac{1}{K}\sum_{k=1}^{K}\log(x_k)} \\ &= -\frac{\Delta}{\beta}\frac{1}{\frac{1}{K}\log(\prod_{k=1}^{K}x_k)} \\ &= -\frac{\Delta}{\beta}\frac{1}{\log\left((\prod_{k=1}^{K}x_k)^{\frac{1}{K}}\right)} \\ &= -\frac{\Delta}{\beta}\frac{1}{\log\left(\text{geometric mean}(x_1, \ldots, x_K)\right)} \end{aligned}$$

This motivates the use of the geometric mean as the test statistic. When the geometric mean of $x_1, \ldots, x_K$ is sufficiently small, the Null hypothesis would be rejected.

What this derivation implies is that we can take either the harmonic mean or geometric mean, corresponding to the assumed distributional form associated with the Alternative hypothesis, of the K scores for a text and we use the normalized relative rank of it as the p-value associated with the text. This procedure naturally normalizes out the unknown value for either the unknown α or β.

2 All But One

The context of this discussion is where there is a test of hypothesis of the Null hypothesis that none of the events has a key word set with its ELSs in more compact arrangement than expected by chance against the Alternative Hypothesis that all but one of J the events has a key word set whose ELSs are in a more compact arrangement than expected by chance.

The experiment has N trials, trial 1 being the trial involving the Torah text and trials 2 through N involving monkey texts. Each trial produces J scores, one for each of the J events. These compactness observations can be arranged in a matrix of N rows by J columns. Each row corresponds to the J event scores observed in a text. Each column corresponds to the scores of an event observed over the N trials. We denote the entry in row n column k by s_{nk}.

Each observed value either is sampled from population 1 or population 2. The probability of any observed value is the same constant in population 1; i.e. population 1 values are distributed as a discrete uniform. The probability of any observed value v from population 2 is given by probability function $P_2(v)$.

It is known a priori that all but one, of the sampled values come from population 2. As before, to simplify the notation, we let the event scores for some text be $x_1, \ldots, x_J$.

Assuming conditional independence, we can write,

$$
\begin{aligned}
Q &= \frac{1}{N}\frac{1}{J}\sum_{j=1}^{J}\prod_{\substack{i=1 \\ i\neq j}}^{K} P_2(x_i) \\
&= \frac{1}{N}\frac{1}{J}\prod_{j=1}^{J} P_2(x_j)\sum_{i=1}^{J}\frac{1}{P_2(x_i)}
\end{aligned}
$$

The p-value assocaited with the Torah text is then the normalized relative rank of is Q score.