March 7, 2016

1 conditioning

Consider the original collection D consisting of m ones and $N-m$ zeros. Denote probability of number of successes for that collection as $P(S|D)$. The probability ratio at s is given by:

$$R_s = \frac{P(s|D)}{P(s-1|D)}$$

Denote expectation of s as μ:

$$\mu = mp + (N-m)q$$

For simplicity, denote probabilities at s for D as:

$$P(s|D) = P_s$$

It's known that for all $s < \mu$, the ratio R_s is greater than 1 and increasing:

Property 1.

$$R_{s-1} = \frac{P_{s-1}}{P_{s-2}} > R_s = \frac{P_s}{P_{s-1}} \tag{1.1}$$

$$P_{s-1}^2 > P_s P_{s-2} \tag{1.2}$$

Create two collections by adding to D one 1 and one 0. Call them D_1 and D_0 respectively. The probability of observing s from D_1 the is given by:

$$P(s|D_1) = pP_{s-1} + qP_s$$

Similarly for the second collection (with extra 0):

$$P(s|D_0) = qP_{s-1} + pP_s$$

Now consider the probability ratio for the collections D_1 and D_0 collections at some s:

$$R_s(D_1) = \frac{pP_{s-1} + qP_s}{pP_{s-2} + qP_{s-1}} \tag{1.3}$$
$$R_s(D_0) = \frac{qP_{s-1} + pP_s}{qP_{s-2} + pP_{s-1}} \tag{1.4}$$

Lemma 1.

$$R_s(D_1) > R_s(D_0) \text{ for } s < \mu$$

proof:

$$\frac{pP_{s-1} + qP_s}{pP_{s-2} + qP_{s-1}} > \frac{qP_{s-1} + pP_s}{qP_{s-2} + pP_{s-1}} \tag{1.5}$$
$$(pP_{s-1} + qP_s)(qP_{s-2} + pP_{s-1}) > (qP_{s-1} + pP_s)(pP_{s-2} + qP_{s-1}) \tag{1.6}$$
$$(p^2 - q^2)(P_{s-1}^2 - P_sP_{s-2}) > 0 \tag{1.7}$$

The above holds because $p > q$ and $P_{s-1}^2 > P_sP_{s-2}$ of **Property 1**.

Lemma 2.

$$R_{s-1}(D_0) > R_s(D_1) \text{ for } s < \mu$$

This lemma essentially says that if we step one point to the left of s, the ratio for the distribution with extra 0 is always greater.

proof:

$$\frac{qP_{s-2} + pP_{s-1}}{qP_{s-3} + pP_{s-3}} > \frac{pP_{s-1} + qP_s}{pP_{s-2} + qP_{s-1}} \tag{1.8}$$
$$qp(P_{s-2}^2 - P_{s-3}P_{s-1}) + qp(P_{s-1}^2 - P_{s-2}P_s) + q^2(P_{s-2}P_{s-1} - P_{s-3}P_s) > 0 \tag{1.9}$$
$$\tag{1.10}$$

Each expression in the first two parenthesis is greater than 0 by Property 1. The last parenthesis is greater than zero, because:

$$\frac{P_{s-2}}{P_{s-3}} > \frac{P_{s-1}}{P_{s-2}} > \frac{P_s}{P_{s-1}} \tag{1.11}$$
$$P_{s-2}P_{s-1} > P_{s-3}P_s \tag{1.12}$$

1.1 Differences

Consider the difference between probabilities in s and $s-1$ for both collections:

$$P(s|D_1) - P(s-1|D_1) = pP_{s-1} + qP_s - pP_{s-2} - qP_{s-1} = p(P_{s-1} - P_{s-2}) + q(P_s - P_{s-1}) \quad (1.13)$$
$$P(s|D_0) - P(s-1|D_0) = qP_{s-1} + pP_s - qP_{s-2} - pP_{s-1} = q(P_{s-1} - P_{s-2}) + p(P_s - P_{s-1}) \quad (1.14)$$

Denote probabilistic difference is s is δ, we can write:

$$\delta(s|D_1) = P(s|D_1) - P(s-1|D_1) = p(P_{s-1} - P_{s-2}) + q(P_s - P_{s-1}) \quad (1.15)$$
$$\delta(s|D_0) = P(s|D_0) - P(s-1|D_0) = q(P_{s-1} - P_{s-2}) + p(P_s - P_{s-1}) \quad (1.16)$$

Note that $\delta(s|D_0) > \delta(s|D_1)$ for $s < \mu$, since

$$\delta(s|D_0) - \delta(s|D_1) = (p-q)[(P_s - P_{s-1}) - (P_{s-1} - P_{s-2})] \quad (1.17)$$

Assumption 1

I believe that:

$$P_s - P_{s-1} > P_{s-1} - P_{s-2} \text{ for } s < \mu$$

I think it can be proved by induction, since it's true for the collection D of all zeros (or ones), it should be true the mixture. I have verified that empirically this statement appears to hold.

1.2 Higher order differences

This may not be all that useful, but recording it for the facts collection.

Suppose that instead of adding a single bit the original collection D, we add k zeros and k ones to arrive to the collections D_0^k and D_1^k. The conditional probabilities for finding s successes will be:

$$P(s|D_1^k) = p^k P_{s-k} + \binom{k}{1} p^{k-1} q P_{s-k+1} + \cdots + \binom{k}{k-1} q^{k-1} p P_{s-1} + q^k P_s \quad (1.18)$$
$$P(s|D_0^k) = q^k P_{s-k} + \binom{k}{1} q^{k-1} p P_{s-k+1} + \cdots + \binom{k}{k-1} p^{k-1} q P_{s-1} + p^k P_s \quad (1.19)$$

Expressing the differences between probabilities in s and $s-1$, we have:

$$\delta(s|D_1^k) = \sum_{i=0}^{k} \binom{k}{i} p^{k-i} q^i \cdot (P_{s-k+i} - P_{s-k+i-1}) \quad (1.20)$$
$$\delta(s|D_0^k) = \sum_{i=0}^{k} \binom{k}{i} q^{k-i} p^i \cdot (P_{s-k+i} - P_{s-k+i-1}) \quad (1.21)$$

Taking the difference, we arrive to:

$$\delta(s|D_0^k) - \delta(s|D_1^k) = \sum_{i=0}^{k/2} \binom{k}{i} (p^{k-i}q^i - q^{k-i}p^i) \cdot [(P_{s-i} - P_{s-i-1}) - (P_{s-k+i} - P_{s-k+i-1})] \quad (1.22)$$

Given that the differences in square brackets are always greater than 0, the $\delta(s|D_0^k)$ is always greater than $\delta(s|D_1^k)$.

Observation

The statement 1.17 and (1.22) implies that probabilities fall the fastest when we add zero to D. Since probabilities at every μ_m are about the same, faster reduction should result in higher ratio for distribution of $m = 0$. Perhaps this may lead to a sufficient proof.

IGNORE

Consider different collections D_m where m represents number of ones. This collections correspond to different distributions of number of successes S with respective expectations μ_m. Consider values of s_m equidistant from each respective μ_m by same number of steps l.

$$s_m = \mu_m - l$$

Denote $R_{m,l}$ as probability ratio at $s_m = \mu_m - l$ for specific m.

We should be able to show that:

$$R_{0,l+1} > R_{m,l} \text{ for any } m \text{ and } l$$

Because $s_{0,l+1}$ is the smallest value of s for given m and l. But the discreetness is a problem here, because how do we round s_m? Need your advice.