Maxim Zhilyaev David Zeber

March 22, 2016

1 thoughts

$$R = \frac{P(S|D' + 0)}{P(S|D' + 1)} \tag{1.1}$$

$$R - 1 = \frac{P(S|D' + 0)}{P(S|D' + 1)} - 1 = \frac{P(S|D' + 0) - P(S|D' + 1)}{P(S|D' + 1)} \tag{1.2}$$

1.1 Univariate case

$$P(S|D' + 0) = qP(S - 1) + pP(S) \tag{1.3}$$

$$P(S|D' + 1) = pP(S - 1) + qP(S) \tag{1.4}$$

$$R = 1 + \frac{P(S|D' + 0) - P(S|D' + 1)}{P(S|D' + 1)} = 1 + \frac{P(S) - P(S - 1)}{pP(S - 1) + qP(S)} \cdot (p - q) \tag{1.5}$$

Assume

$$\delta(S) = P(S) - P(S - 1)$$

$$R = 1 + \frac{P(S) - P(S - 1)}{pP(S - 1) + qP(S)} \cdot (p - q) = 1 + \frac{\delta(S)}{pP(S - 1) + q(P(S - 1) + \delta(S))} \cdot (p - q) \tag{1.6}$$

$$R = 1 + \frac{\delta(S)(p - q)}{P(S - 1) + q\delta(S)} \tag{1.7}$$

As m changes from 0 to N, and S is chosen to be $S = \mu_m - 3 * \sigma$, it appears that $\delta(S)$ increases slightly, while $P(S - 1)$ increades significantly when m grows. Basically, the reason why R_0 is the largest at the left end of range is simply because probabilities at the end-point of the rage are smaller for $m = 0$ than for $m > 0$. Perhaps we can bound $\delta(S)$, the differences shouldn't be very large especially for large N, and perhaps we can achieve good bound using normal assumption, and then simply show that denominator is smallest when $m = 0$, which I think is the case.

1.2 a variation of same idea

Consider R_m at the left end of the range (denoted as S_m) when m is changing from 0 to N.

$$R_0 = \frac{P(S_0|D_0+0)}{P(S_0|D_0+1)} \tag{1.8}$$

$$R_1 = \frac{P(S_1|D_1+1)}{P(S_1|D_1+1)} \tag{1.9}$$

$$... \tag{1.10}$$

$$R_N = \frac{P(S_N|D_N+0)}{P(S_N|D_N+1)} \tag{1.11}$$

$$\tag{1.12}$$

Let's consider a situation when all numerators are equal:

$$P(S_0|D_0+0) = P(S_1|D_1+1) = ... = P(S_N|D_N+0)$$

Then the ratio R_m will only depend on the denominator. Suppose we choose a quantile probability of 0.005. Now, let's fin the corresponding PDF probability for that quintile for $m = N$ distribution. Call this probability y. The corresponding S_N for that y will be l units away from the mean μ_N.

I believe it can be shown for all distributions with $m < N$ that a) S_m corresponding to that y is at least l units (or further) away from μ_m, and b) S_m falls into smaller quintile, and c) $P(S_m|D_m+1)$ is smallest when $D' = 0$. Basically, we fix the numerator probability for all m, and then prove that denominator will be smallest at $m = 0$ and that point will be bellow the local-privacy range for the corresponding D'.

1.3 multivariate case

in the collection of N vectors of length L, we replace a unit vector 1 with a zero vector 0. Try argue same thing:

$$R = 1 + \frac{P(S|D'+0) - P(S|D'+1)}{P(S|D'+1)} \tag{1.13}$$

Assuming that for large N the numerator is reasonably bounded (that is doesn't change with D'), then R is dominated by the value of denominator.

$$P(S|D'+1) = P(s_1-1, s_2, \dots, s_{2^L}|D')q^L + P(s_1, s_2-1, \dots, s_{2^L}|D')q^{L-1}p + \dots + P(s_1, s_2, \dots, s_{2^L}-1|D')p^L \tag{1.14}$$

It may be so, that if $D' = 0$, then $P(S|D')$ is simply smaller compared to all other D', hence the linear combination of slight variations of $P(S)$ also comes out smallest.