THE ESTIMATE OF $\chi^2$ DISTANCE BETWEEN BINOMIAL AND GENERALIZED BINOMIAL DISTRIBUTIONS

VYTAS ZACHAROVAS
Institute of Informatics, Vilnius University, Naugarduko 24, Vilnius, Lithuania
e-mail: vytas.zacharovas@mif.vu.lt

We investigate the distribution of a sum
$$S_n = I_1 + I_2 + \cdots + I_n$$

of $n$ independent indicators $I_j$ taking value 1 with probability $p_j = \mathbb{P}(I_j = 1)$ and 0 with probability $1 - p_j$. We will denote by $p$ the arithmetic average of the probabilities $p_j$, that is,
$$p = \frac{1}{n} \sum_{j=1}^{n} p_j.$$

In a special case when all indicators take values 1 and 0 with the same probabilities $p$ and $q = 1 - p$, respectively, the sum $S_n$ will assume value $j$ with probability $b(n, p; j) = \binom{n}{j} p^j q^{n-j}$ for all $0 \leq j \leq n$. Our goal is to estimate the $\chi^2$-distance between $\mathbb{P}(S_n = j)$ and $b(n, p; j)$ defined as
$$\chi^2(\mathbb{P}(S_n), b(n, p)) = \sum_{j=0}^{n} \left( \frac{\mathbb{P}(S_n = j)}{b(n, p; j)} - 1 \right)^2 b(n, p, j).$$

Our main result is formulated in the following theorem.

THEOREM 1. For all $n \geq 2$ and $\delta < 1$, the following inequalities hold:
$$\frac{n}{2(n-1)} \delta^2 \leq \chi^2(\mathbb{P}(S_n), b(n, p)) \leq \frac{n}{2(n-1)} \delta^2 \left( 1 + O(\frac{\delta^2}{1 - \delta^2} + \frac{1}{\sqrt{n} \delta^3}) \right)$$
and
$$\chi^2(\mathbb{P}(S_n), b(n, p)) \leq n \delta_2.$$

Here we have used the notation
$$\delta_m := \frac{1}{n(pq)^m} \sum_{j=1}^{n} |p_j - p|^m \quad \text{for all} \quad m \geq 2.$$

The constants in the symbol $O(\ldots)$ can be made explicit and are shown to be quite small.