

The Chernoff bound state that :

$$\mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n X_i \geq a\right) \leq e^{-n(\lambda a - f(\lambda))} \quad \text{with} \quad f(\lambda) = \log(\mathbb{E}[e^{\lambda x}]) \quad (MGF)$$

In the pooling case our X_i are i.i.d. Bernoulli Random Variables, equal 0 if the single person is in approval with the president, 0 otherwise. Indicating with " p " the fraction of people approving (that obviously will correspond with the probability of success) I define my estimator as the mean of these R.V. :

$$\hat{p} = \frac{1}{n} \sum_{i=1}^n X_i \quad ; \quad \mathbb{E}[X_i] = p \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i = p$$

In this case the Chernoff Bound become :

$$\mathbb{P}(\hat{p} \geq p + \epsilon) \leq \mathbb{E}[\lambda x]^n e^{-n\lambda(p+\epsilon)} = \frac{(pe^\lambda + (1-p))^n}{e^{n\lambda(p+\epsilon)}}$$

And minimizing this bound through :

$$\frac{\partial}{\partial \lambda} \frac{(pe^\lambda + (1-p))^n}{e^{n\lambda(p+\epsilon)}} = 0 \quad \Rightarrow \quad e^\lambda = \frac{(1-p)(p+\epsilon)}{p(1-p-\epsilon)}$$

Substituting it we find then :

$$\mathbb{P}[\hat{p} \geq p + \epsilon] \leq \left(\left(\frac{p}{p+\epsilon} \right)^{p+\epsilon} \left(\frac{1-p}{1-p-\epsilon} \right)^{1-p-\epsilon} \right)^n = e^{-nD_{KL}(p+\epsilon||p)}$$

With a similar calculation it can be verified also that :

$$\mathbb{P}[\hat{p} \leq p - \epsilon] \leq \left(\left(\frac{p}{p-\epsilon} \right)^{p-\epsilon} \left(\frac{1-p}{1-p+\epsilon} \right)^{1-p+\epsilon} \right)^n = e^{-nD_{KL}(p-\epsilon||p)}$$

From these two bounds we can find, at least in principle, the wanted N minimal in order to be sure to fulfill the asked condition. In fact :

$$\mathbb{P}[|\hat{p} - p| \leq \epsilon] = 1 - \mathbb{P}[\hat{p} \leq p - \epsilon] - \mathbb{P}[\hat{p} \geq p + \epsilon] \geq 1 - e^{-nD_{KL}(p-\epsilon||p)} - e^{-nD_{KL}(p+\epsilon||p)}$$

imposing now

$$\mathbb{P}[|\hat{p} - p| \leq \epsilon] = 0.95 \quad \text{and} \quad \epsilon = 0.01$$

We can determine the minimal number of people needed extracting n from the last inequality

$$-(0.95 - 1) = 0.05 \leq e^{-nD_{KL}(p+\epsilon||p)} \left(1 + e^{D_{KL}(p+\epsilon||p) - D_{KL}(p-\epsilon||p)} \right)$$

$$\Rightarrow \quad n_{min} = \frac{-\ln(0.05) + \ln \left(1 + e^{D_{KL}(p+\epsilon||p) - D_{KL}(p-\epsilon||p)} \right)}{D_{KL}(p+\epsilon||p)}$$

The problem is that we don't have p. However it can be solved defining n as a function of n and then take the maximal value in order to be sure, in any case, to make a correct pooling. Given that this function will not be of a simple form we can make a simple observation : the D_{KL} is a relative entropy between the estimator and the true distribution and what we want to find is the value that minimize it, then maximizing the n_{min} function. In the case of Bernoulli variables the variance and the entropy will be maximized when $p = 1 - p = 0.5$ and, given that the logarithm is monotonic, the D_{KL} will be minimized in that case. For p

$$= 0.5 \text{ we find } n_{min} = \frac{-\ln(0.05) + \ln \left(1 + e^{D_{KL}(0.5+0.01||0.5) - D_{KL}(0.5-0.01||0.5)} \right)}{D_{KL}(0.5+0.01||0.5)} \simeq 18443$$