



Maximum Ignorance Probability, with application to surgery's error rates



Nassim Nicholas Taleb
September 7, 2021
Medicine, Probability

Introduction and Result

A maximum entropy alternative to Bayesian methods for the estimation of independent Bernoulli sums.

Let $X = \{x_1, x_2, \dots, x_n\}$, where $x_i \in \{0, 1\}$ be a vector representing an n sample of independent Bernoulli distributed random variables $\sim B(p)$. We are interested in the estimation of the probability p .

We propose that the probability that provides the best statistical overview, p_m (by reflecting the *maximum ignorance* point) is

$$p_m = 1 - I_{\frac{1}{2}}^{-1}(n - m, m + 1), (1)$$

where $m = \sum_{i=1}^n x_i$ and $I(\cdot, \cdot)$ is the beta regularized function.

Comparison to Alternative Methods

EMPIRICAL: The sample frequency corresponding to the “empirical” distribution $p_e = E(\frac{1}{n} \sum_{i=1}^n x_i)$, which clearly does not provide information for small samples.

BAYESIAN: The standard Bayesian approach is to start with, for prior, the parametrized Beta Distribution $p \sim \text{Beta}(\alpha, \beta)$, which is not trivial: one is constrained by the fact that matching the mean and variance of the Beta distribution constrains the shape of the prior. Then it becomes convenient that the Beta, being a conjugate prior, updates into the same distribution with new parameters. Allora, with n samples and m realizations:

$$p_b \sim \text{Beta}(\alpha + m, \beta + n - m) \quad (2)$$

with mean $p_b = \frac{\alpha+m}{\alpha+\beta+n}$. We will see below how a low variance beta has too much impact on the result.

Derivations

Let $F_p(x)$ be the CDF of the binomial $\text{Bin}(n, p)$. We are interested in $\{p : F_p(x) = q\}$ the maximum entropy probability. First let us figure out the target value q .

To get the maximum entropy probability, we need to maximize $H_q = -(q \log(q) + (1 - q) \log(1 - q))$. This is a very standard result: taking the first derivative w.r. to q , $\log(q) + \log(1 - q) = 0$, $0 \leq q \leq 1$ and since H_q is concave to q , we get $q = \frac{1}{2}$.

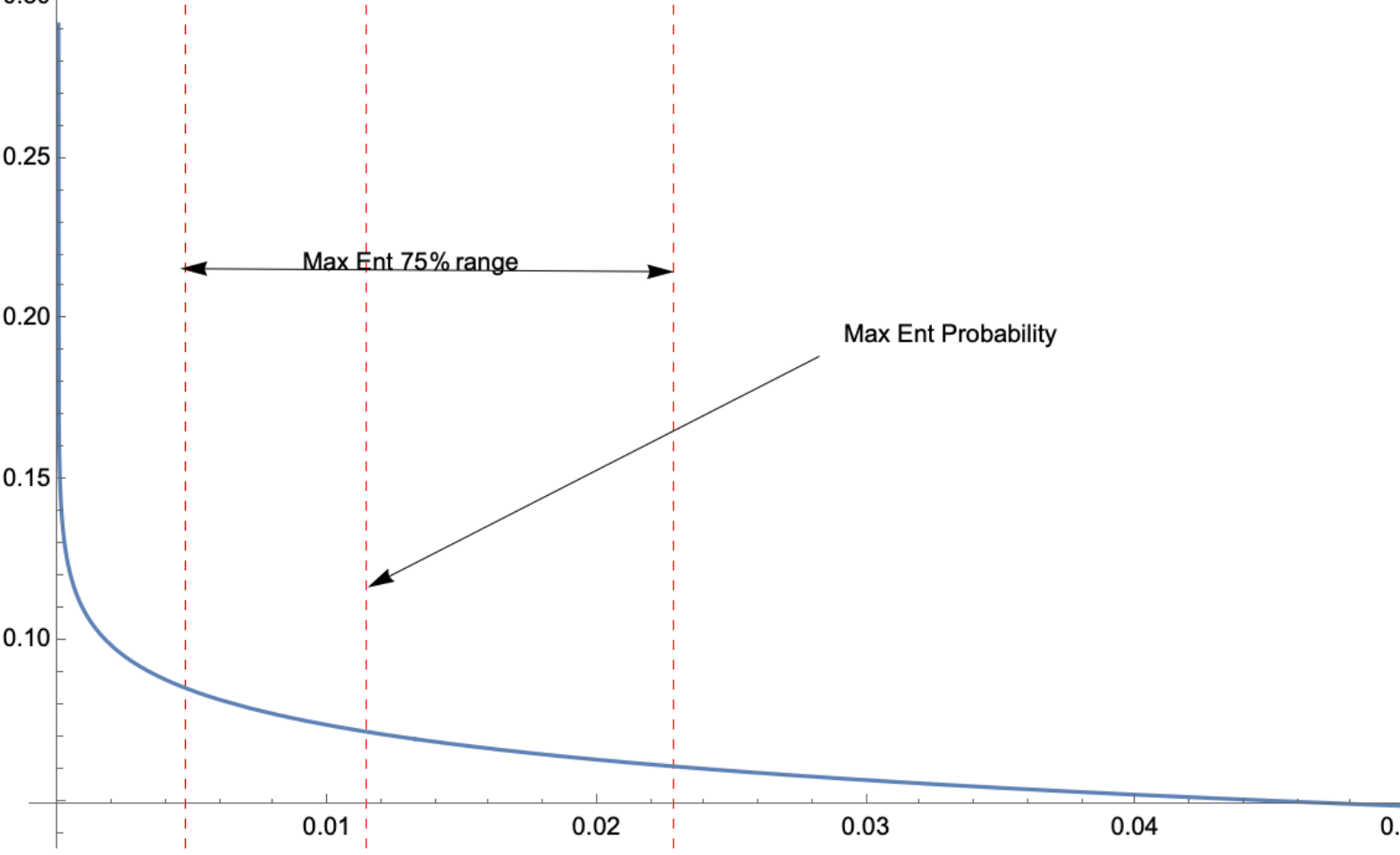
Now we must find p by inverting the CDF. Allora for the general case,

$$p = 1 - I_{\frac{1}{2}}^{-1}(n - x, x + 1).$$

And note that as in the graph below (thanks to comments below by überstatistician Andrew Gelman), we can have a “confidence band” (sort of) with

$$p_\alpha = 1 - I_{\alpha}^{-1}(n - x, x + 1);$$

in the graph below the band is for values of: $\alpha = \frac{1}{2}, \frac{3}{4}, \frac{1}{4}$.



Application: What can we say about a specific doctor or center's error rate based on n observations?

Case (Real World): A thoracic surgeon who does mostly cardiac and lung transplants (in addition to emergency bypass and aortic ruptures) operates in a business with around 5% perioperative mortality. So far in his new position in the U.S. he has done 60 surgeries with 0 mortality.

What can we reasonable say, statistically, about his error probability?

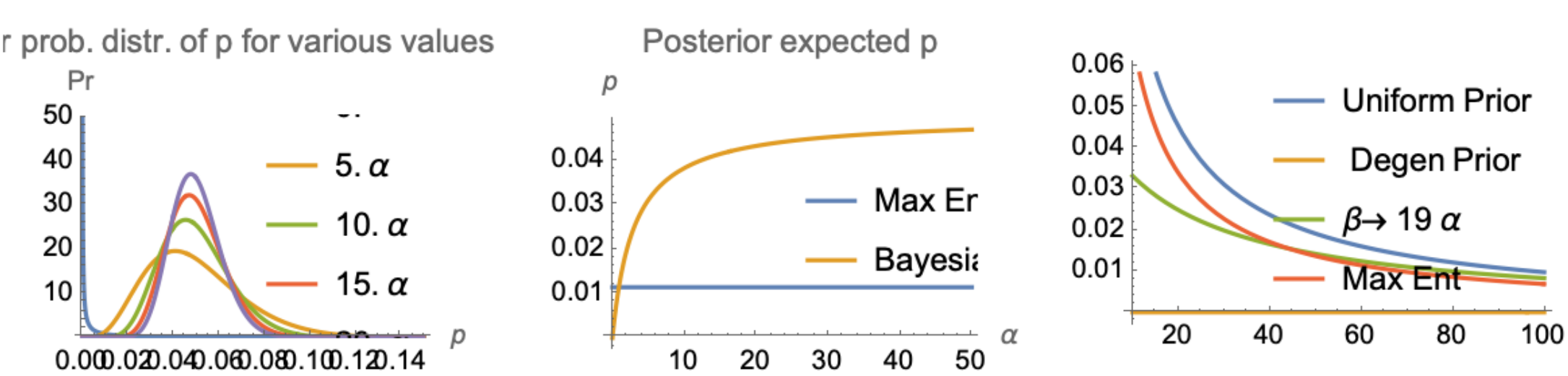
Note that there may be selection bias in his unit, which is no problem for our analysis: the probability we get is conditional on being selected to be operated on by that specific doctor in that specific unit.

Assuming independence, we are concerned with $Y = 0, 1, \dots, n$ a binomially distributed r.v. $\sim B(n, p)$ where n is the number of trials and p is the probability of failure per trial. Clearly, we have no idea what p and need to produce our best estimate conditional on, here, $y = 0$.

Here applying (1) with $m = 0$ and $n = 60$, we have $p = 0.01148$.

Why is this preferable to a Bayesian approach when, say, n is moderately large?

A Bayesian would start with a prior expectation of, say, .05, and update based on information. But it is highly arbitrary. Since the mean is $\frac{\alpha}{\alpha+\beta}$, we can eliminate one parameter. Let us say we start with $\text{Beta}(\alpha, 19\alpha)$ and have no idea of the variance. As we can see in the graph below there are a lot of shapes to the possible distribution: it becomes all in the parametrization.



APPENDIX: JAYNES' BRANDEIS PROBLEM

When I worked on this problem, and posted the initial derivations, I wasn't aware of Jaynes' “Brandeis Problem”. It is not the same as mine as it ignores n and it led to a dead-end because the multinomial is unwieldy. But his approach would have easily let to more work if we had computational abilities then (maximization, as one can see below, can be seamless).

The Brandeis Problem (Jaynes)

While the method of *max ignorance probability* takes into account n , the number of trials, the problem by Jaynes know as “The Brandeis Problem”, applied to the Multinomial distr. does not take them into account.

Jaynes, E. (1979). *Where do we stand on maximum entropy? the maximum entropy formalism*. eds Levine, RE & Tribus, M., MIT Press, Cambridge, pages 15–118.

A die yields an average of 4.5 (instead of $\frac{7}{2}$ for flat probabilities).

What is the maximum entropy distribution?

Maximize $-\sum_{i=1}^6 p_i \log(p_i)$ under constraints $\sum_{i=1}^6 p_i = 1$ and $\sum_{i=1}^6 i p_i = 4.5$

$H = \text{Maximize} \left[\left(-\sum_{i=1}^6 p_i \log(p_i) \right), \left(\sum_{i=1}^6 p_i = 1, \sum_{i=1}^6 i p_i = 4.5 \right), \{p_1, p_2, p_3, p_4, p_5, p_6\} \right]$

$Pe = \text{Table}[\{p_i\}, \{1, 1, 6\}] /. N[2]$

$\text{BarChart}[Pe, \text{PlotRange} \rightarrow \{0, 1\}]$

$\{1.61359, p_1 \rightarrow 0.854354, p_2 \rightarrow 0.6787725, p_3 \rightarrow 0.114163, p_4 \rightarrow 0.165447, p_5 \rightarrow 0.239774, p_6 \rightarrow 0.347492\}$

$Hf = -\sum_{i=1}^6 p_i \log(p_i) /. Pe \rightarrow \frac{1}{6} /. N$

1.79176

What is the loss in entropy from the information?

$Hf - H[1]$

0.178174

Acknowledgments

Thanks to Saar Wilf for useful discussions.

5 thoughts on “Maximum Ignorance Probability, with application to surgery's error rates”



Andrew Gelman
September 9, 2021 at 1:16 am

Nassim:
It's not clear you can come up with much of a “best estimate” from data with y=0. Maybe all that can be given in general (without using some prior information on p) are reasonable bounds. I like the Agresti–Coull procedure which gives a reasonable 95% interval. I've used this in a consulting project (in that case, the data were y=75 out of n=75) and put it in Regression and Other Stories.

Reply



Nassim Nicholas Taleb
September 9, 2021 at 9:38 pm

Thanks, will update with a comment on bounds.

Reply



Nassim Nicholas Taleb
September 19, 2021 at 10:55 pm

The problem of bounds requires a knowledge of $\text{p}\$$. But there is such a thing as a maximum ignorance range, say between .25–.75. Replace $\frac{1}{2}$ by $\frac{1}{4}$ and $\frac{3}{4}$ and there is a band.

Reply



pechtygam
September 22, 2021 at 2:50 pm

Excuse me, Maestro, but don't you mean “where $\$m \leq \sum x_i \$$ ”, i.e. at most $\$m \$$ dead patients?

Reply

Carlos J. Gil Bellosta
January 10, 2022 at 3:00 am

Wouldn't make more sense to use $Sp_m = I^{-1}(\frac{1}{2})(m+1, n-m)$? Same results, more coherent with the fact that you are inverting the CDF.

Reply

Leave a Reply

Your email address will not be published. Required fields are marked *

COMMENT *

NAME *

EMAIL *

WEBSITE

POST COMMENT

PREVIOUS

The Lognormal at high variance: a probability paradox

NEXT

A nice Limit

Search ?

RECENT POSTS

- Detecting IS in Correlation Windows
- On Heart Rate Variability and Its Mismeasurement
- On weighted sums of Uniformly distributed random variables
- A nice Limit
- Maximum Ignorance Probability, with application to surgery's error rates

RECENT COMMENTS

- Carlos J. Gil Bellosta on Maximum Ignorance Probability, with application to surgery's error rates
- Black Paper de Nassim Taleb demonstra com argumentos e matemática que o BITCOIN vale 0. ~ Investimento para Iniciante on The “Bitcoin, Currencies and Fragility” Paper
- Doig Huffman on On Heart Rate Variability and Its Mismeasurement
- Waleed Addas on On Heart Rate Variability and Its Mismeasurement
- Interested party on On weighted sums of Uniformly distributed random variables

ARCHIVES

- November 2021
- October 2021
- September 2021
- June 2021
- April 2021
- January 2021
- December 2020
- November 2020

CATEGORIES

- Lognormal
- Medicine
- Probability
- Quant Finance
- Uncategorized

META

- Log in
- Entries feed
- Comments feed
- WordPress.org
- YAHOO! YAHOO! YAHOO!