Frequentist coverage study: intervals for a binomial parameter

Due for January 7, 2020

Homework counting for 20% of the final grade. You are welcome to help each other, especially on technicalities, but one document per person will need to be sent to `emilien.chapon@cern.ch`*. Special care should be given to providing short comments and interpretation of the results obtained. A computer format is preferred (Jupyter notebook, LaTeX or Word document, ...), with the computer part preferrably in C++ or python (using either numpy / scipy or ROOT). A scan of hand-written answers is also fine but less preferred.*

We are studying the case of a binomial law:

$$P(n) = B(n; N, p) = \binom{N}{n} p^n (1-p)^{N-n}$$

This corresponds to repeating an experiment N times, and counting n the number of times of "sucess", where the probability for this success is p. n is a random variable and follows this binomial law, of parameters N and p. One real-life example is the estimation of an efficiency: for instance, n could be the number of counts in a particle detector, or passing some selection, while N is the total number of particles or events considered.

Interval estimation of the binomial parameter p is a common problem, however nontrivial because the problem is discrete. There is abundant bibliography on the subject, and we will review some of the most popular methods available, as well as their frequentist coverage.

1 Prelude

1. Show that the mean is $E(n) = pN$ and the variance $V(n) = Np(1-p)$.
2. Assume that n counts are observed for a given N. What is the maximum likelihood estimate $\hat{p}$? What is its variance?
3. Propose a simple algorithm for generating n following a binomial law $B(n; N, p)$, assuming a uniform generator between 0 and 1.
4. Draw the binomial distribution $B(n; N, p)$ for $N = 10$, $p = 0.9$.

2 Interval estimation

For convenience, we will use the notation

$$Z_{\alpha/2} = \Phi^{-1}(1-\alpha/2) = -\Phi^{-1}(\alpha/2)$$

where

$$\Phi(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Z} \exp(-t^2/2)\mathrm{d}t = \frac{1+\mathrm{erf}(Z/\sqrt{2})}{2},$$

so that

$$Z = \sqrt{2}\,\mathrm{erf}^{-1}(1-\alpha).$$

Example: $Z_{\alpha/2} = 1$ for $1-\alpha = 0.6827$.

1. **Wald interval** This procedure (that we will see should be avoided) simply substitutes $\hat{p}$ for p in the variance of n, in a Gaussian approximation. Give explicitly the Wald interval for any C.L.

2. **Wilson score interval** The Wald intervals neglect the fact that $V(\hat{p})$ depends on p. E.B. Wilson proposed in 1927 to quote the largest interval $[p_1, p_2]$ such that $p_1 + Z_{\alpha/2}\sigma(p_1) \geq \hat{p}$ and $\hat{p} \geq p_2 - Z_{\alpha/2}\sigma(p_2)$. Derive the expression for such interval.

3. **Agresti and Coull** These authors proposed to use the midpoint $\tilde{p}$ of the Wilson interval in the formula for the Wald interval, instead of the MLE $\hat{p}$. Derive the corresponding interval.

4. **Clopper-Pearson** These intervals use the Neyman construction and ensure strict frequentist coverage. The ordering rule is as follows: the lower and upper limits are constructed separately, with C.L. $(1-\alpha)/2$. Write a function that builds the confidence belt in the (n, p) plane. You may first build the lower limits, then the upper limits, and combine the two together. Draw it for $N = 10$ and with a step size in p of 0.01.

2.1 Comparison of the intervals in particular cases

Compare the intervals on p from the above methods, in the following cases:

- $N = 10$, $n = 10$
- $N = 10$, $n = 9$
- $N = 10$, $n = 5$
- $N = 100$, $n = 90$

3 Study of frequentist coverage properties of various types of intervals

For each of the Wald, Wilson score, and Agresti and Coull intervals[1], study the frequentist properties (coverage) of the interval estimation for p. In practice this is done in the following way:

- Fix the true value of the parameters p and N;
- Generate n according to $B(n; N, p)$;
- Compute the interval knowing n and N;

[1] The study of Clopper-Pearson intervals, more difficult and much more computationnally intensive, is left as facultative exercise.

- Check if the true value of p is in this interval;
- Repat N_{trials} times (e.g. 1000 times);
- Report the fraction of times that the true value of p was inside the interval, and compare to the target confidence level (C.L.).

We will study 68.27% C.L. intervals, in 3 ways:

1. set $N = 10$ and check coverage as a function of p for $0 \leq p \leq 1$
2. make a 2-dimensional plot of coverage as a function of p and N, for $0 \leq p \leq 1$ and $2 \leq N \leq 20$
3. average the previous plot over all values of p, to get the average coverage as a function of N.

References

[1] R. D. Cousins, K. E. Hymes and J. Tucker, Nucl. Instrum. Meth. A **612** (2010) 388 doi:10.1016/j.nima.2009.10.156 [arXiv:0905.3831 [physics.data-an]].

[2] CERN ROOT's `TEfficiency` class documentation, `https://root.cern.ch/doc/master/classTEfficiency.html`.