Approximations to the Binomial distribution, its bounds, relative and absolute precision

Jorge Santos, Marília Pires, Russell Alpizar-Jara

Abstract

Approximations to the Binomial by a Gaussian distribution are not always consensus and can be tedious, especially when we are dealing with the probability of a range of values of extreme values of the random variate. In the last century, some research led to the use of cumulative normal distribution tables, as they are easily available. This approximation is best for large samples and evenly symmetric situations. We discuss the 3 usual criteria for the applicability of this method. We show that not only sample size and symmetry are important, but also that the error rates are crucial. Small values of the success probability p become unacceptable when we try to calculate tail probabilities that sometimes have the same order of magnitude as the error. Besides, we show that the most restrictive criterion demands more than 50 trials with a probability of success belonging to the range [0.1;0.9], the criteria based on a variance greater than 5 lead to a parabolic shape criterion and the most liberal ones lead to the intersection of two hyperbolic regions. With the increase of computer capabilities this is not a practical problem, but the main idea is helping to provide some guidelines to this subject that is pervasive for recommendations to usual introductory probability and statistics courses.

Keywords: binomial, continuity correction, normal approximation.

Acknowledgments

References

[1] Afonso, A., Nunes, C. *Estatística e Probabilidades: Aplicações e soluções em SPSS.*Escolar Editora, 2011.

[2] Agresti, A. *An Introduction to Categorical Data Analysis*, John Wiley and Sons, 2002.

[3] Rosner, B.A. *Fundamentals of Biostatistics*, Duxbury Press,1995.