MATH 341 - ARTICLE REVIEW - NOVEMBER 19

CHAD BROWN

A Derivation of the Won-Loss Formula in Baseball - Steven J. Miller

In the above paper, a theoretical justification is provided for Bill James' Pythagorean Win-Loss Formula, which has been shown to be a strong predictor of a Major League Baseball team's future performance. The formula is as follows:

$$Win\% = \frac{RS_{obs}^{\gamma}}{RS_{obs}^{\gamma} + RA_{obs}^{\gamma}} \tag{0.1}$$

where RA_{obs} is the number of runs allowed up to that point in the season and RS_{obs} is the number of runs scored thus far. There are two key assumptions made. Namely, that Runs Scored and Runs Allowed are independent and that both are drawn from a three parameter Weibull distribution, shown below.

$$f(x;\alpha,\beta,\gamma) = \begin{cases} \frac{\gamma}{\alpha}(x-\beta/\alpha)^{\gamma-1}e^{-(x-\beta/\alpha)^{\gamma}} & \text{if } x \geq \beta \\ 0 & \text{otherwise.} \end{cases} \tag{0.2}$$

We might worry that RS and RA are not independent, since there are no ties in baseball and thus the runs allowed and runs scored in any particular game are necessarily dependent. Statistical evidence in the form of Chi-Squared tests of independence are provided to ease our fears, however. Our second assumption is chosen because the Weibull is a flexible distribution and can be fitted to the data well. More importantly, though, the Weibull facilitates the theoretical justification due its computationally friendly mean and the ease with which its terms can be integrated.

Given these assumptions, X is defined as $Weibull(\alpha_{RS}, \beta, \gamma)$ and Y is defined by $Weibull(\alpha_{RS}, \beta, \gamma)$. The probability $P(X > Y)$ is then shown to be equivalent to the Pythagorean Win-Loss Formula above. After providing thorough theoretical justification, the paper also goes on to provide strong numerical support for the formula using real data from the American League 2004 MLB season. Finally, the paper concludes by providing estimates for γ using this same data. These estimates fall in line with the observed best value of gamma ($\gamma = 1.82$) found by fitting the Won-Loss formula to observed winning percentages. Though the paper focuses its work specifically on baseball, care is taken to provide a framework that can be easily extended to other sports, as well.