CA Risk Trending using a Bayesian Framework With a Zero-Inflated Beta Distribution

J.J. Vallejo (primary author)	Department of Statistics, Baylor University P.O. Box 97140 Waco, TX 76798 402-740-1019 jonathon_vallejo@baylor.edu
M.D. Hejduk	Astrorum Consulting LLC
J.D. Stamey	Department of Statistics, Baylor University

The probability of collision (Pc) value, which determines the likelihood of two satellites coming within a very small distance of each other by considering their propagated state estimates and uncertainties, is the single parameter used by most satellite conjunction risk assessment organizations to encapsulate conjunction risk. However, because both the primary and secondary objects are typically receiving sensor tracking, the state estimates are updated as the event unfolds, and the propagated covariances tend to shrink due to the fact that the propagation interval to the time of the two satellites' closest approach (TCA) becomes smaller; because of this evolution, the Pc value is not static but changes as the event develops. There is a certain "canonical" evolution of the Pc that is due merely to the expected contraction of the covariance as TCA is approached, and it is a function of the ratio of the size of the covariance to the miss distance. Typically the Pc begins as a small value when the covariance is very large compared to the miss distance, then as the covariance shrinks the Pc increases until it reaches a maximum when the ratio of the miss distance to the one-sigma covariance size is $\sqrt{2} / 1$. Then, as the covariance further shrinks, the Pc rapidly drops off to essentially zero. However, the miss distance between the satellites also changes, both in magnitude and geometry, with each update, and the satellites' covariances are also changing in unpredictable ways; so overlaid upon this canonical behavior is often quite a bit of "noise"—changes to the knowledge of the states (and their uncertainties) that are not predictable. Despite this somewhat polluted situation, what would be extremely helpful operationally would be the ability to separate out the canonical behavior from the noise as best one can in order to determine whether the event's maximum Pc is likely already to have occurred or can still be expected. The ability to inform an owner/operator that there is a high degree of likelihood that the peak Pc has already passed and therefore that risk assessment can be made reasonably against the most recently received Pc value can allow many events to be evaluated as extremely unlikely to require active remediation. What is most needed, therefore, is the ability to predict where the peak Pc is likely to occur in an event's time history. An added bonus would be a reliable prediction of the value of that peak.

A previous effort investigated the prediction of the peak Pc location within an event using a Bayesian framework with an inverted parabola as the characteristic function. Although initially attempted as a naïve model simply for orientation and investigative purposes, once properly tuned it yielded surprising predictive force, prompting a search for a more sophisticated characteristic function that might produce even better performance. As a distribution that is typically used to characterize distributions over the [0 1] bounded interval (which is the interval of possible Pc values), the beta distribution suggested itself as a possible candidate. Given that actual conjunction events produce a large number of zero or near-zero Pc values, the beta distribution used as the characteristic function requires alteration ("zero-inflation") to account for and predict the likelihood of zero and near-zero values. Additionally, as the beta distribution will not naturally inflect to manifest or predict peak values, additional stratagems for approximating peak location based on distribution functional behavior are necessary. The theory of the zero-inflated beta distribution will be presented, focusing specifically on the parameters appropriate to the present problem. The selection of parameter prior distributions and parameter estimation for these prior distributions will be discussed, and model performance will be evaluated against an expansive test-set of operational event Pc histories (over 1000 actual events). The “inverted parabola” characteristic function model will also be run against the same test-set of event histories, which will give an opportunity to examine comparative performance statistics for both models.