*Proc. of the European Safety and Reliability Conference (ESREL 2026)*
*Edited by* José C. Matos, Paulo B. Lourenço, Michael Beer, Daniel Oliveira,
Edoardo Patelli, Hélder S. Sousa, Rui A. Silva and Monica Santamaria

doi: 10.3850/ESREL2026061419_esrel26-p27819-cd

# Correcting Over-Conservatism in Exact Two-Sided Confidence Curves for Poisson Rates and Sequential Binary Tests

Alexander Wimbush

*Cancer Research Clinical Trials Unit, Birmingham, B15 2TT, United Kingdom.*
*E-mail: a.wimbush.1@bham.ac.uk*

Exact confidence intervals (such as Garwood or Clopper-Pearson) for discrete data problems are guaranteed to bound the true parameter at some minimum desired rate given the random distribution of the data. However, the discrete nature of the data necessarily means that there are only a finite number of confidence levels which meet the nominal coverage level exactly. An acceptable compromise for one-sided intervals perhaps, but an unfortunate source of over-conservatism when this gets compounded in the creation of two-sided intervals. Many solutions to this problem have been presented over the years, either by relaxing the coverage requirements, or producing disjoint or non-nested confidence intervals. For safety critical applications these compromises may be undesirable or impractical.

This manuscript describes a method to correct the over-conservatism in confidence intervals drawn for the Poisson distribution for estimating event rates, as well as a sequential binary test for efficiently accepting or rejecting a test of competing hypotheses about a probability. In both cases, the result is an exact, two-sided confidence curve that eliminates the over-conservatism present in conventional two-sided confidence intervals for these problems. These curves also protect against the issue of false confidence that stems from the use of additive belief structures such as precise probability distributions.

*Keywords*: Confidence, Inference, Exact, Poisson, Sequential

## 1. Conservatism in Exact Confidence Intervals

Exact Confidence Intervals (CIs) derived from the likelihoods of discrete distributions have a long history of use in statistics, such as those described by Clopper and Pearson (1934) for the binomial distribution, or by Garwood (1936) for the Poisson distribution. These intervals must be conservative to a certain extent due to the nature of the discrete distribution and the requirement of 'exact' coverage, where

$$\mathrm{P}(\mathrm{CI}^{-1}(I; x, \eta) \ni \theta_0; X \sim \mathrm{F}_{\theta_0}) \geq |I|, \quad \forall I \subseteq [0,1], \theta_0 \in \Theta, \quad (1)$$

with $I$ representing an $\alpha$-interval $I = [\underline{\alpha}, \overline{\alpha}] \subseteq [0,1]$, $\mathrm{CI}^{-1}$ representing a procedure for drawing a CI based on $I$, known parameters $\eta$, and the observed data $x$ which follows a known distribution $\mathrm{F}_{\theta_0}$ which takes $\theta_0$ as a parameter. This requirement ensures that the rate at which a CI bounds the true value given $\mathrm{F}_{\theta_0}$ can never fall below the specified confidence level. Naturally for discrete distributions where the values $X$ may take are countable, this requires some confidence levels to have coverage rates greater than specified.

Exact CIs are often criticized as a result of this conservatism. Since the performance of a CI is only realized upon repeated applications of the procedure, Agresti and Coull (1998) argue in favour common alternatives that relax the strict coverage requirement of Eq. (1) to instead offer coverage over distributions of $\theta_0$ rather than for precise values. This produces tighter CIs that facilitate efficient decision-making, with average coverage rates closer to nominal levels than Clopper-Pearson CIs. However, exact methods may still be preferable in cases where risks must be controlled for each individual inference. For example, if the risks and benefits stemming from each application of a procedure are borne by distinct groups, it may not be equitable to allow that one group should bear greater risks than the other on the basis that their aggregated risk is controlled at nominal levels.

For one-sided CIs conservatism is 'necessary', in that the coverage rate for any CI could not be reduced without compromising the exact nature of the CI. However, two-sided CIs are formed as intersections of one-sided CIs, and this necessary conservatism now compounds to produce two-sided CIs that are no longer merely necessarily conservative, they are 'over-conservative'. Wimbush et al. (2022) describe the Singh plot, a method for visualizing the coverage properties of a confidence procedure. Given that this case is for inference over a Poisson distribution, the Singh plot should ideally show steps of a height equal to the probability mass of each outcome, following the uniform at the base of each of these steps, as is the case for the one-sided CIs. The fact that this is not the case for two-sided Garwood CIs is a reflection of their over-conservatism.

Balch (2020) introduced a novel confidence curve to adhere to the Martin-Liu validity criterion described in Martin (2019), protecting an analyst from the false confidence that Balch et al. (2019) describes, where false assertions can be frequently assigned a high degree of belief. This was achieved through a two-sided, consonant, unimodal confidence curve, following the terminology of Birnbaum (1961). The resulting curve also corrects the over-conservatism of the Clopper-Pearson CIs. A similar structure could be created from the 'direct' two-sided intervals used to generate the green line in Figure 1, but the over-

**Coverage of Exact CIs for Poisson Distribution with Rate $\theta_0 = 4$**

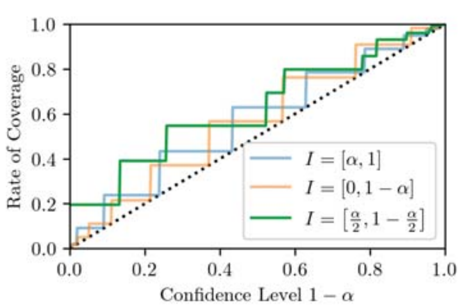


Fig. 1. Singh plot showing the coverage rates for Poisson CIs with $I$ defined for left-sided, right-sided, and two-sided $\alpha$-intervals. The dotted black line indicates the minimum coverage requirement for Eq. (1).

conservatism inherent in these intervals makes this an unappealing solution. This manuscript demonstrates the extension of this approach to the Poisson distribution in Section 2, and an example of corrected CIs for a binary sequential sampling distribution.

***Novel Contributions***

This manuscript defines an analytical solution for a corrected, two-sided confidence curve for the Poisson distribution. An approximate solution is also described for sequential binary tests that can be defined with common root-finding algorithms.

## 2. Over-Conservatism

Exact CIs for the Poisson distribution can be generated using a confidence box (c-box, Ferson et al. (2013)) calculated from the $\chi^2$ distribution,

$$\begin{aligned} \mathrm{C}_P(\lambda; x) &= \left[\underline{\mathrm{C}}_P(\lambda; x), \overline{\mathrm{C}}_P(\lambda; x)\right] \\ &= \left[\mathrm{F}_{\chi^2}(2\lambda; 2(x+1)), \mathrm{F}_{\chi^2}(2\lambda; 2x)\right], \end{aligned} \quad (2)$$

where $\mathrm{F}_{\chi^2}(\lambda; k)$ is the CDF of a $\chi^2$ distribution evaluated at $\lambda$ with $k$ degrees of freedom. This may be used to draw CIs as

$$\mathrm{CI}_P\left(\left[\underline{\alpha}, \overline{\alpha}\right]; x\right) = \left[\frac{\overline{\mathrm{C}}_P^{-1}(\underline{\alpha}; x)}{2}, \frac{\underline{\mathrm{C}}_P^{-1}(\overline{\alpha})}{2}\right], \quad (3)$$

where $\mathrm{CI}_P\left(\left[\underline{\alpha}, \overline{\alpha}\right]; x\right)$ is an $\overline{\alpha} - \underline{\alpha}$-level CI taken from a c-box given observed outcome $x$ known to follow a Poisson distribution. Note that the lower bound of the c-box defines the upper bound of the CI, and vice versa. This produces the Singh plots shown in Figure 1, where the one-sided CIs exhibit only necessary conservatism, while the two-sided CIs exhibit over-conservatism. In each of the three cases presented, there is a single $1-\alpha$ level required to bound the true value for each outcome. The one-sided cases are free from over-conservatism because this $1-\alpha$ level can be expressed as a cumulative sum of ordered likelihoods, or simply a CDF (or SF) in this case.

This is not the case for $[\alpha/2, 1-\alpha/2]$ CIs. In this case, the $1-\alpha$ level required to bound the true value of $\lambda$ can be calculated from a possibility distribution over $\Lambda$ defined as

$$\pi_D(\lambda; x) = 1 - 2\min\left(\overline{\mathrm{C}}_P(\lambda; x), 1 - \underline{\mathrm{C}}_P(\lambda; x), 0.5\right), \quad (4)$$

where the required $1-\alpha$ level would be calculated with $\alpha = \pi_D(\lambda; x)$. This approach is here referred to as the 'direct' method for conversion into a two-sided confidence curve. Since this possibility distribution is no longer a CDF or an SF, as is the case for one-sided CIs, or a simple sum of likelihoods, it is prone to over-conservatism.

To eliminate the over-conservatism, an appropriate criterion that determines which likelihoods to sum together to calculate the possibility must be defined. A simple, and valid, solution would be to order outcomes based on their likelihoods, summing only dominated likelihoods to produce a distribution of $\alpha$-levels analogous to a possibility distribution,

$$\pi(\lambda; x) = \sum_{k \in X} \left\{\mathrm{f}_P(k; \lambda); \mathrm{f}_P(k; \lambda) \leq \mathrm{f}_P(x; \lambda)\right\}.$$

A two-sided $1-\alpha$-level CI given observed data $x$ could then be defined as the set of $\lambda$ values with a possibility exceeding $\alpha$,

$$\mathrm{CI}_{P,\pi}(1-\alpha; x) = \{\lambda \in \Lambda; \pi(\lambda; x) \geq \alpha\}. \quad (5)$$

This would eliminate the over-conservatism, but produces a possibility distribution that is not unimodal. Balch describes the limitations of such an approach, and developed a novel unimodal structure for the binomial distribution. The same approach can be applied to other discrete distributions to produce confidence curves suitable for robust inference and valid assignments of belief for a variety of applications.

### 2.1. *Summation criteria*

Consider the portion of the distribution to the left of the maximum possibility region, where the gradient must be positive. Any left bound of a CI cannot possibly be greater than that of the Garwood CI taken from $\overline{\mathrm{C}}_P$, as it is known to be exact and only to exhibit necessary conservatism. The CDF $\overline{\mathrm{C}}_P(\lambda; x)$ represents the sum of likelihoods for all outcomes following a given outcome $X \geq x$. This then serves as a base for the summation to construct the left side of the confidence curve. Similarly, the SF $1 - \underline{\mathrm{C}}_P(\lambda; x)$ represents the sum of likelihoods for outcomes preceding a given outcome $X \leq x$. So if the summation given an observed outcome $x_k$ begins from the base of $\overline{\mathrm{C}}_P(\lambda; x_k)$, likelihoods for outcomes preceding $x_k$ can be added in order as

$$\overline{\mathrm{C}}_P(\lambda; x_k) + 1 - \underline{\mathrm{C}}_P(\lambda; x_i).$$

The likelihoods of outcomes up to and including some $x_i < x_k$ can then be added to the sum while maintaining unimodality so long as

$$\frac{\partial}{\partial\lambda}\left(\overline{\mathrm{C}}_P(\lambda; x_k) + 1 - \underline{\mathrm{C}}_P(\lambda; x_i)\right) \geq 0.$$

Now, rather than proceeding with summation based on dominated likelihoods as proposed earlier, summation can proceed on the basis of dominated c-box bound density functions, where

$$\frac{\partial}{\partial\lambda}\Big(\overline{\mathrm{C}}_P(\lambda; x_k)\Big) \geq \frac{\partial}{\partial\lambda}\Big(\underline{\mathrm{C}}_P(\lambda; x_i)\Big)$$

is now the criteria for adding the SF of $x_i$ to the sum while maintaining unimodality.

A similar process may be applied for the right bound of the confidence curve, though now the SF of the lower bound of the Garwood c-box is used as the base, and likelihood of alternative outcomes are added on the basis of

$$\frac{\partial}{\partial\lambda}\Big(\underline{\mathrm{C}}_P(\lambda; x_k)\Big) \geq \frac{\partial}{\partial\lambda}\Big(\overline{\mathrm{C}}_P(\lambda; x_i)\Big).$$

### 2.2. *Imprecise likelihood*

However, a problem arises with this criterion. Outcomes directly before or after $x_k$ share the c-box boundaries that are used for comparison, for example

$$\overline{\mathrm{C}}_P(\lambda; x_k) = \underline{\mathrm{C}}_P(\lambda; x_{k-1}).$$

Clearly the gradient will be 0 wherever all likelihoods are summed together, so using this as the basis of summing likelihoods leads to a vacuous structure.

Balch resolves this for the binomial distribution with an imprecise likelihood function, which also has the benefit of neatly bringing together the point of comparison for both sides of the confidence curve. Instead of comparing $\overline{\mathrm{C}}_P(\lambda; x_k)$

with $\underline{C}_P(\lambda; x_i)$ for the left side of the curve, and $\underline{C}_P(\lambda; x_k)$ with $\overline{C}_P(\lambda; x_i)$ for the right side, an imprecise likelihood function is defined that is capable of representing both bounding distributions, and the relevant parametric set of distributions contained within those bounds. This follows the reasoning of the imprecise beta model described by Walley et al. (1996), where the bounding distributions represent a set of (in this case) $\chi^2$ distributions corresponding to a range of priors defined to ensure good frequentist properties on the resulting set of posterior distributions.

In the case of the Poisson distribution, the imprecise likelihood function is

$$\mathcal{L}_P(\lambda; x) = \sup_{\varepsilon \in [0,1]} \mathrm{f}_{\chi^2}(\lambda; x + \varepsilon), \tag{6}$$

where $\mathrm{f}_{\chi^2}(\lambda; x)$ is the PDF of a $\chi^2$ distribution with $x$ degrees of freedom evaluated at $\lambda$. Using this as the point of comparison for summation as

$$\pi_C(\lambda; x) = \sum_{k \in X} \left\{ \mathrm{f}_P(\lambda; k); \mathcal{L}_P(\lambda; k) \leq \mathcal{L}_P(\lambda; x) \right\}, \tag{7}$$

produces a corrected, unimodal confidence curve, as shown in Figure 2. The corrected curve is not strictly tighter, in that $\pi_C(\lambda; x) \leq \pi_D(\lambda; x)$ not uniformly on $X$, though it would generally be preferable over the direct method. More importantly, it corrects the over-conservatism of the direct confidence curve, as shown in Figure 3. The corrected curve has lower minimum and average

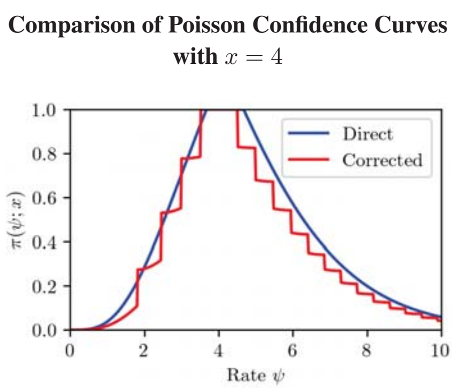


Fig. 2. Comparison of confidence curves for the Poisson distribution.

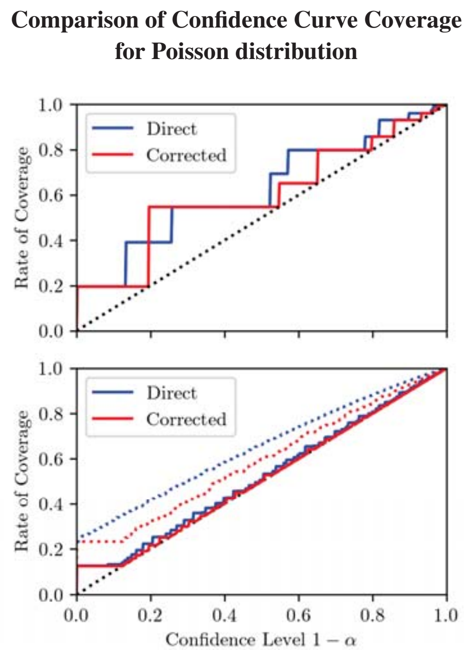


Fig. 3. (Top) Singh plots for $\lambda = 4$. (Bottom) Summaries of Singh plots for confidence curves given $\lambda_0 \in \{k/100; k \in 0, \ldots, 10\}$. Solid lines indicate minimum coverage, dotted lines indicate average coverage.

coverage rates while maintaining exact coverage. The lack of over-conservatism in the corrected CIs is apparent from the fact that bottom of each step follows the uniform, where they do not for the direct CIs.

## 2.3. *Analytical solution*

The possibility curve defined by Eq. (7) produces the desired result, but can be computationally expensive for CIs using Eq. (5). For the binomial distribution, Balch remedies this with the introduction of a piecewise formulation which reduces computational cost by summing not based on dominated imprecise likelihoods, but by splitting $\Lambda$ into regions where the appropriate likelihoods to sum are known. The curve can then be calculated by identifying which region contains the $\lambda$ to be evaluated, and then summing corresponding CDFs and SFs. This removes the need for repeated comparisons of imprecise likelihoods that can be costly to evaluate. The solution presented here fol-

$$\pi_C(\lambda; x) = \begin{cases} \mathrm{F}_{\chi^2}(\lambda; 2x) & \text{if} \quad \rho_P(\lambda, x) = 0 \\ \mathrm{F}_{\chi^2}(\lambda; 2x) + 1 - \mathrm{F}_{\chi^2}(\lambda; 2\rho_P(\lambda, x)) & \text{if} \quad 0 < \rho_P(\lambda, x) < x \\ 1 & \text{if} \quad \rho_P(\lambda, x) = x \\ 1 - \mathrm{F}_{\chi^2}(\lambda; 2(x+1)) + \mathrm{F}_{\chi^2}(\lambda; 2(\rho_P(\lambda, x) + 1)) & \text{if} \quad x < \rho_P(\lambda, x). \end{cases} \tag{8}$$

lows the same reasoning applied by Balch (2020), but for the Poisson distribution.

An intersection point between two imprecise likelihoods can be defined as

$$I_P(x, i) = \{\lambda \in \Lambda; \mathcal{L}_P(\lambda; x) = \mathcal{L}(\lambda; i)\}. \tag{9}$$

These can be calculated as

$$I_P(x, i) = \begin{cases} {\frac{i!}{(x-1)!}}^{(i-x+1)^{-1}} & \text{if} \quad i < x - 1, \\ e^{\digamma(x)} & \text{if} \quad i = x - 1, \\ e^{\digamma(x+1)} & \text{if} \quad i = x + 1, \\ {\frac{i!}{x!}}^{(i-x)^{-1}} & \text{if} \quad i > x + 1, \end{cases} \tag{10}$$

where $\digamma(x)$ is the digamma function evaluated at $x$. The position of $\lambda$ can then be assigned an integer as

$$\rho_P(\lambda, x, X) = \min\{i \in X \backslash x; I_P(x, i) > \lambda\}. \tag{11}$$

The summation can then take the piecewise formulation shown in Eq. (8) derived following the same procedure Balch used for the binomial case. Since the structure is unimodal, CIs can then be drawn and represented as

$$\mathrm{CI}_C(1-\alpha; x) = \Big[\inf\{\lambda \in \Lambda; \pi_P(\lambda; x) \geq \alpha\} \\ \sup\{\lambda \in \Lambda; \pi_P(\lambda; x) \geq \alpha\}\Big].$$

The maximum possibility value within each region between regions defined by Eq. (9) can be used to simplify the process of identifying these values.

## 3. Sequential Sampling C-Boxes

The procedure described by Balch is applied to reach a similarly refined confidence curve for the Poisson distribution above. This same approach can be extended to many other discrete distributions, though not all accept the analytical solution to reduce computational cost. But the base procedure of points of intersection between c-box bounds can be readily applied. As an extreme example, a corrected confidence curve can be defined for the Sequential Probability Ratio Test described by Wald (1992).

### 3.1. *Identifying c-boxes*

The appropriate set of c-boxes for the SPRT can be calculated by considering that the required 'height' of the c-box (i.e. $\overline{\mathrm{C}}_P(p_0; x, X) - \underline{\mathrm{C}}_P(p_0; x, X)$, with $p_0$ being the true probability of success) must be equal to the likelihood of $x$ given $X$ under $p_0$. The likelihoods resulting from the sequential test can be calculated following the direct method of Aroian (1968). These likelihoods can then be progressively summed to produce the required bounds for each c-box.

For example, consider an SPRT with $H_0 : p = 0.2$ and $H_1 : p = 0.5$, with error rates of $\alpha = 0.2$, $\beta = 0.2$. This can be achieved with a sample size of 10, so there are 11 c-boxes that must be calculated. The first opportunity to stop is after 2 samples, where observing 2 successes could lead to early termination. This produces the standard $k = 2, n = 2$ Clopper-Pearson c-box, and also provides the lower bound of the c-box for the next opportunity for early stopping due to efficacy, at $k = 3, n = 4$. The upper bound of this c-box can be calculated by summing the likelihoods of $k = 2, n = 2$ and $k = 3, n = 4$. This procedure is repeated until each c-box is defined. This produces the set of c-boxes shown in Figure 4. At the maximum sample size of $n = 10$, the test rejects $H_0$ if there are 4 successes or more, and accepts it otherwise, to maintain error rate control.

These c-boxes can be converted directly into confidence curves, but knowledge of the full set of possible c-boxes allows the corrected curve to be defined. In this case, the complexity of the c-box bounds makes an analytical solution imprac-

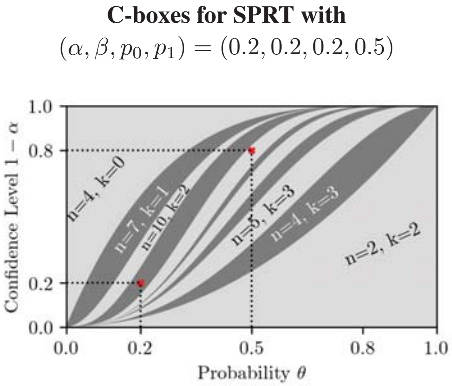


Fig. 4. C-boxes for possible outcomes of this SPRT. Hypotheses and corresponding error rates are marked with dotted lines and a red cross.

tical or impossible. Instead, points of intersection can be identified through root finding algorithms. The lower c-box bound for an observed outcome $x_i = (k_i, n_i)$ within the sequence of outcomes $X = (x_0, \ldots, x_n)$ is defined as

$$\underline{\mathrm{C}}_S(p; x_i, X) = \sum_{j \in \{i+1, \ldots, n\}} \mathrm{f}_S(x_j; p, X), \quad (12)$$

where $\mathrm{f}_S(x_j; p, X)$ is the probability mass assigned to outcome $x_j$ for the defined sequential test. The gradient of this CDF can then be defined as

$$\frac{\partial}{\partial p} \underline{\mathrm{C}}_S(p; x_i, X) = \sum_{j \in \{i+1, \ldots, n\}} \frac{\partial}{\partial p} \mathrm{f}_S(x_j; p, X). \quad (13)$$

For example, the intersection between the gradient of the lower bound of the c-box for outcome $x_i$ and that of the upper bound of $x_j$, where $x_j \succ x_{i+1}$ in $X$, can be identified as

$$I(i, j, X) = \{p \in (0, 1); g(p, i, j, X) = 0\},$$

where

$$g(p, i, j, X) = \sum_{k \in \{i+1, \ldots, j\}} \frac{\partial}{\partial p} \mathrm{f}_S(x_k; p, X). \quad (14)$$

Points of intersection with the upper bound and outcomes $x_j \prec x_{i-1}$ can be similarly identified by adjusting the set of outcomes over which to sum gradients accordingly.

Maximum possibility regions cannot be identified analytically, as was the case for the Poisson distribution, due to the nature of the function defining the gradients of the c-box bounds. Instead, the maximum possibility region is defined for an outcome $x_i$ within $X$ as

$$\overline{\pi}_S(i, X) = \left[ \frac{k_{i-1} n_i + k_i n_{i-1}}{2 n_i n_{i-1}}, \frac{k_i n_{i+1} + k_{i+1} n_i}{2 n_i n_{i+1}} \right]. \quad (15)$$

This is simply the average of the roots of the gradient functions for the observed outcome and preceding/subsequent outcomes. This is enabled by the ordering of the c-boxes imposed by the SPRT. The position of a given $p$ in relation to the points of intersection can then be defined as

$$\rho_S(p; x_i, X) = |\{x_j \in X \backslash x_i; I(i, j, X) < p\}|, \quad (16)$$

which will be contracted to $\rho_S^p$ (with the observed outcome and sequence implied) for notational simplicity. The piecewise definition of the corrected confidence curve is then defined as in Eq. (17). This produces the corrected confidence curve shown in Figure 5. The sampling method used in Figure 5 has a significant impact on the possibility assigned to each $p$ value. It is not necessarily preferable to the confidence curve that results from binomial sampling, but the SPRT is defined to minimize the expected sample size required to accept or reject each hypothesis, not to minimize CI width. Also, since the structure represents two-sided CIs, the possibility assigned to the hypotheses used to define the sequential test $p_0, p_1$ are not necessarily smaller than the required error rates for the appropriate accept/reject outcomes.

This corrected confidence curve exhibits only necessary conservatism, as shown in Figure 6. Average coverage here is relatively high due to the large c-boxes for early stopping conditions. As before, the corrected curves have lower minimum and average coverage rates while maintaining exact coverage.

## 4. Discussion

These corrected confidence curves generally provide a more efficient means of drawing two-sided

$$\pi_S(p;x_i,X) = \begin{cases} \overline{\mathrm{C}}_S(p;x_i,X) & \text{if} \quad \rho_S^p = 0, \\ \overline{\mathrm{C}}_S(p;x_i,X) + 1 - \underline{\mathrm{C}}_S\left(p;x_{\rho_S^p-1},X\right) & \text{if} \quad 0 < \rho_S(p,x,X) < i, \\ 1 & \text{if} \quad \rho_S(p,x,X) = i, \\ 1 - \underline{\mathrm{C}}_S(p;x_i,X) + \overline{\mathrm{C}}_S\left(p;x_{\rho_S^p+1},X\right) & \text{if} \quad i < \rho_S(p,x,X) < |X|-1, \\ 1 - \underline{\mathrm{C}}_S(p;x_i,X) & \text{if} \quad \rho_S^p = |X|-1, \end{cases} \tag{17}$$

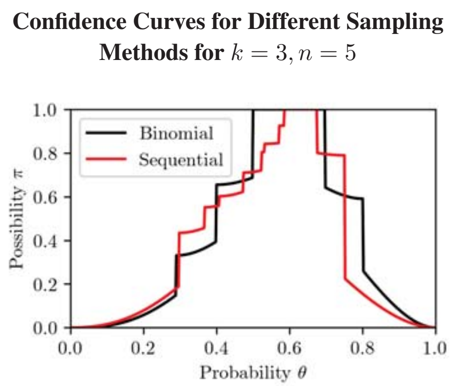


Fig. 5. Corrected confidence curves for $k = 3, n = 5$ following binomial sampling, and sequential sampling following the test defined in Section 3.1.

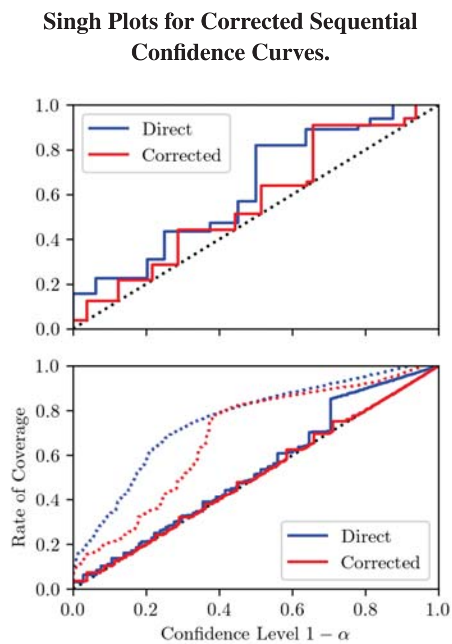


Fig. 6. (Top) Singh plots for $p = 0.5$ for direct and corrected confidence curves for the SPRT defined in Section 3.1. (Bottom) Summaries of Singh plots for confidence curves given $p \in \{k/100; k \in \{0,\ldots,100\}\}$. Solid lines indicate minimum coverage, dotted lines indicate average coverage.

CIs for inference with discrete data. The approach can also be extended to predictive cases while maintaining strict error rate control. These confidence curves will be of greatest value where inference must be performed with very small sample sizes, limiting the applicability of common alternatives. The exact coverage requirement also makes them particularly suitable for cases where risk must be controlled on a per-application basis rather than relying on average performance over repeated applications. The SPRT confidence curve allows efficient exact two-sided CIs to be drawn following a sequential test, further increasing the efficiency of inference.

The benefits of these confidence curves comes at the cost of a significant increase in the computational cost of producing CIs for inference, though not to the point of being impractical.

While some of the confidence curves are relatively well specified, like the Poisson distribution example described above, others such as the SPRT curves are less concrete. For example, the ordering of outcomes has a significant impact on the c-boxes and confidence curves produced from such a test. The solution presented here orders outcomes by their corresponding point estimates. This ordering is imposed due to necessity, but valid c-boxes can be obtained with many alternative orderings, though not necessarily valid confidence curves. This makes the identification of the most efficient ordering difficult. For the SPRT in particular, the order of the c-boxes affects the appropriate point at which to stop sampling to maintain error rate control. However, the nature of the SPRT maintains an ordering consistent with that imposed by the point estimates used to construct the confidence curve in this manuscript.

## 5. Further Work

The curves identified so far are all for univariate problems, and confidence curves for continuous distributions can be developed for a range of multivariate cases following the methods described by Martin (2025). It remains to be seen whether it is practical to develop confidence curves for discrete multivariate distributions. The fundamental principle should remain consistent, that of summing over dominated imprecise likelihood functions. But this can be costly to evaluate for univariate cases, never mind multivariate problems.

That the ordering of the c-boxes affects the confidence curve raises some problems. The SPRT ensures that ordering by appearance and point estimates are consistent, but it is not difficult to produce alternative sampling methods where this ordering is not maintained. While these may not be realistic in practice, it suggests that there is considerable scope to affect the characteristics of the test, while maintaining coverage, through modifying the sampling methods. Similarly, the credal set used to define the imprecise likelihood function in the Binomial case can be altered, while maintaining unimodality, beyond simply considering $\varepsilon \in [0, 1]$. This does not appear to be the case for the Poisson distribution, though the constraints on this credal set for each case need to be considered in more detail.

**Acknowledgement**

This work was undertaken at the Cancer Research UK Clinical Trials Unit at University of Birmingham supported by a core funding grant from Cancer Research UK (CTUQQR-Dec22/100006), with further funding from the National Institute for Health and Care Research & the Medical Research Council (MR/Y008391/1).

## References

Agresti, A. and B. A. Coull (1998). Approximate is better than “exact” for interval estimation of binomial proportions. *The American Statistician 52*(2), 119–126.

Aroian, L. A. (1968). Sequential analysis, direct method. *Technometrics 10*(1), 125–132.

Balch, M. S. (2020). New two-sided confidence intervals for binomial inference derived using walley’s imprecise posterior likelihood as a test statistic. *International Journal of Approximate Reasoning 123*, 77–98.

Balch, M. S., R. Martin, and S. Ferson (2019). Satellite conjunction analysis and the false confidence theorem. *Proceedings of the Royal Society A 475*(2227), 20180565.

Birnbaum, A. (1961). Confidence curves: An omnibus technique for estimation and testing statistical hypotheses. *Journal of the American Statistical Association 56*(294), 246–249.

Clopper, C. J. and E. S. Pearson (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika 26*(4), 404–413.

Ferson, S., M. Balch, K. Sentz, and J. Siegrist (2013). Computing with confidence. In *Proceedings of the Eighth International Symposium on Imprecise Probability: Theory and Applications*.

Garwood, F. (1936). Fiducial limits for the poisson distribution. *Biometrika 28*(3/4), 437–442.

Martin, R. (2019). False confidence, non-additive beliefs, and valid statistical inference. *International Journal of Approximate Reasoning 113*, 39–73.

Martin, R. (2025). Possibilistic inferential models: a review. *Journal of the American Statistical Association* (just-accepted), 1–36.

Wald, A. (1992). Sequential tests of statistical hypotheses. In *Breakthroughs in statistics: Foundations and basic theory*, pp. 256–298. Springer.

Walley, P., L. Gurrin, and P. Burton (1996). Analysis of clinical data using imprecise prior probabilities. *Journal of the Royal Statistical Society Series D: The Statistician 45*(4), 457–485.

Wimbush, A., N. Gray, and S. Ferson (2022). Singhing with confidence: visualising the performance of confidence procedures. *Journal of Statistical Computation and Simulation 92*(13), 2686–2702.