4 あなたを改善するための新しい方法
単一母集団の推定
Minitab の比率

基本統計学における一般的な問題は、母集団における特定の関心のある特性を持つ個人の割合の推定です。たとえば、品質エンジニアは、大量生産されたユニットの大規模なバッチの特定の日の欠陥の割合を推定できます。医学者は、特定の病原体に対してワクチンを接種したが、それにもかかわらず関連する病気を経験したコミュニティ内の個人の割合を調査したいと思うかもしれません。キャンペーンマネージャーは、候補者に投票する予定の登録有権者の割合に関心がある場合があります。

この問題の最もよく知られている区間推定法は、Wald 信頼区間(CI)および ClopperPearson 正確確率(1934)CI と呼ばれる教科書的な正規近似法です。一方で、Wald CI は、CI の実際の信頼水準(またはカバレッジ確率)が目標の名義水準をはるかに下回っているという点で、特に真の比率が 0 または 1 に近いという点で、非常にリベラルです(図 1 を参照)。一方、正確な Clopper-Pearson 信頼区間は、CI の実際の信頼水準(またはカバレッジ確率)が目標名義水準をはるかに上回っているという点で、過度に保守的です。これらの方法はどちらも、もはや実用的なアプリケーションには使用しないでください (Agresti-Coull, 1998;Brown et al., 2001)。

しかし、近年では、中間カバレッジの確率が向上する、より優れた CI メソッドの開発に大きな役割を果たしています。たとえば、Agresti-Coull 近似 CI は WaldCI の調整です。Blaker (2000, 2001) の正確信頼区間では、反復数値アルゴリズムの開始推定値として Clopper-Pearson 信頼限界を使用します。Minitab では、これらの新しく改善された方法を念頭に置いて、単一の母集団比率を推定するための統計ツールを更新し、調整された Blaker CI と検定法、Wilson/score CI と検定法(連続性補正ありとなし)、Agresti-Coull CI と検定法の 4 つの方法を追加しました。さらに、Minitab では、これらの各方法について、信頼区間と検定で一貫した結果が得られることが保証されます。

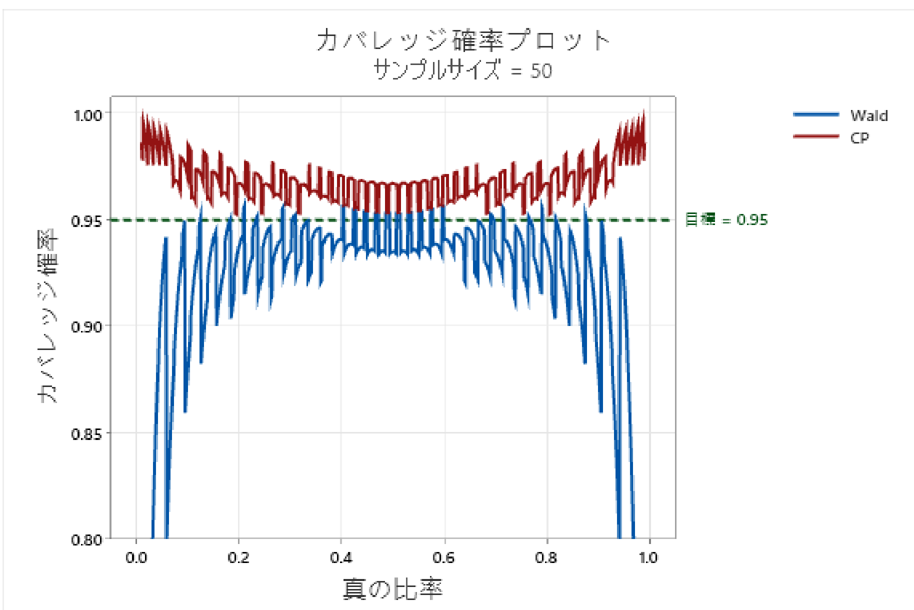

図 1: サンプルサイズが 50 の場合の真の比率の関数としての Wald 信頼区間と Clopper-Pearson(CP)信頼区間のカバレッジ確率の比較。このグラフは、Wald CI と Clopper Pearson CI がそれぞれ過度にリベラルと保守的であることを示しています。真の比率が区間(0, 1)に一様に分布していると仮定すると、サイズ 50 のサンプルに基づく平均カバレッジ確率は、Wald CI と Clopper-Pearson CI でそれぞれ 0.901 と 0.969 です。

4 つの新しい推定方法の紹介

4 つの新しい方法は、調整ブレイカー法と呼ばれる 1 つの正確な CI および検定法と、ウィルソン/スコア(Wilson)法、イェーツの連続性補正(WilsonCC)法による Wilson/score、および Agresti-Coull(AC)法を含む 3 つの近似 CI および検定法で構成されています。このコンテキストでの厳密法とは、ある種の正規近似手順を使用して取得される近似法とは対照的に、法の導出に使用される近似がないことを意味します。

1.調整されたブレイカー法

Klaschka and Reiczigel (2021) による調整された Blaker 法は、Blaker (2000, 2001) の正確な CI および試験法を修正したものです。この変更により、元の Blaker アルゴリズムの計算集約型の性質と、その CI とテスト結果の間に時折発生する不整合に対処します。元の Blaker CI と同様に、結果として得られる調整済み CI は正確で、ネストされており、Clopper-Pearson CI に含まれています。その結果、調整済み Blaker CI は Clopper-Pearson CI よりも保守的ではありません。CI は、信頼度の高い CI に常に信頼度が低い CI が含まれるという意味でネストされます。たとえば、両側 95%(調整済み)の Blaker CI には、対応する両側 90%信頼区間が常に含まれています。入れ子性は、二項分布などの離散分布から導出された厳密な CI 法の魅力的な特性です。たとえば、Clopper-Pearson CI は入れ子になっています。ただし、必ずしもネストされているとは限らない正確な CI メソッドが使用可能です。例えば、いわゆる Blyth-Still-Casella CI(Blyth and Still, 1983;Casella, 1986)は、最短の正確な CI であることが保証されていますが、ネストされていません。Crow (1956) CI も入れ子になっていません。Blaker または調整済み Blaker 法に基づく CI の計算は、数値アルゴリズムを必要とするため、前述の古典的な CI 法よりも複雑です。しかし、現在のコンピューティング技術の革新により、より良い結果をもたらす複雑なアルゴリズムの実装を躊躇するべきではありません。図 2 は、Clopper-Pearson CI に対する調整済み Blaker CI の改善を示しています。

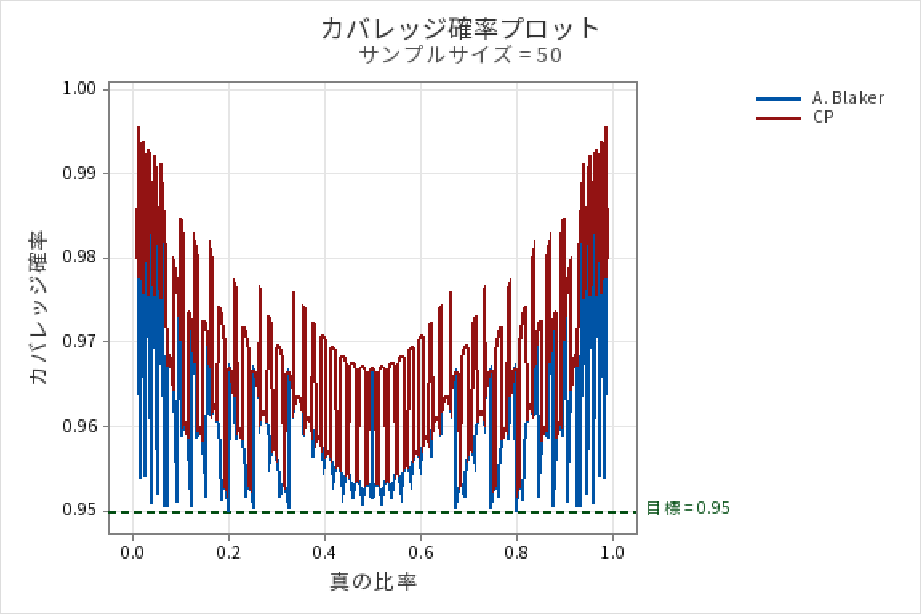

図 2: サンプルサイズが 50 の場合の真の比率の関数としての調整済み Blaker(A. Blaker)信頼区間と Clopper-Pearson(CP)信頼区間のカバレッジ確率の比較。このグラフは、Clopper-Pearson CI のカバレッジ確率が、少なくとも調整済み Blaker CI のカバレッジ確率であることを示しています。これは、調整後の Blaker CI が Clopper-Pearson CI に含まれることと一致しています。サイズ 50 の任意のサンプルについて、平均カバレッジ確率(真の比率が単位区間で一様に分布していると仮定)は、調整済み Blaker CI と Clopper-Pearson CI でそれぞれ 0.960 と 0.969 です。

2.ウィルソン法とウィルソン CC 法

Wilson (1927) CI 法は、スコア検定に対応する CI として導出され、検定統計量の分母に古典的な標準誤差 $\sqrt{\hat{p}(1-\hat{p})/n}$ とは対照的に、帰無標準誤差 $\sqrt{p_0(1-p_0)/n}$ を使用する検定です。このため、Wilson/score CI とも呼ばれます。実際のカバレッジ確率は、目標とする公称レベルよりも小さくなったり大きくなったりする可能性がありますが、真の比率が 0 または 1 に近い場合を

除き、公称レベルに近いままです (図 3 を参照)。 イェーツの連続性補正を使用して、Wilson/スコア信頼区間を保守的に調整できます。Minitab では、CI バージョン(イェーツの連続性補正あり/なし)と一致仮説検定の両方が提供されます。

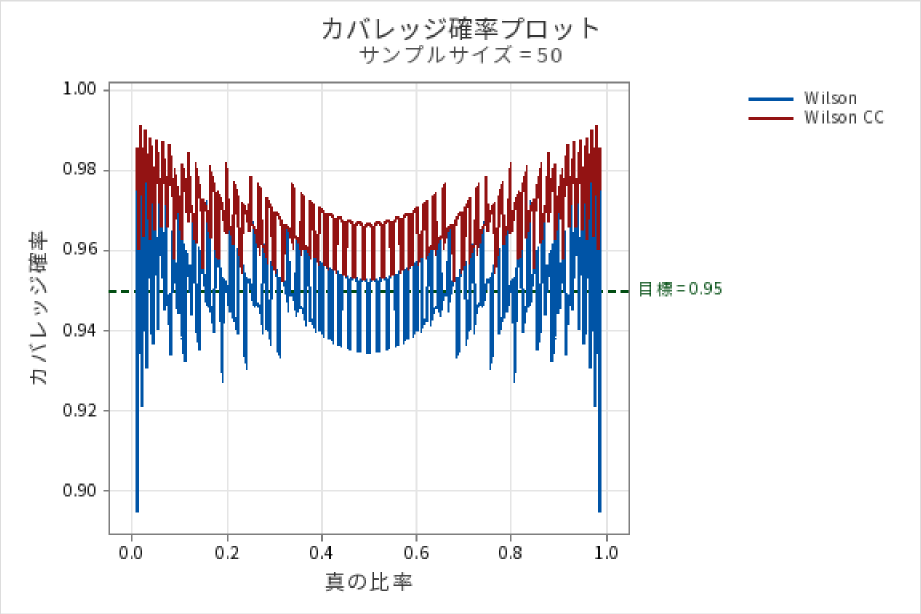

図 3: サンプルサイズが 50 の場合の真の比率の関数としての Wilson/スコア(Wilson)信頼区間と Wilson/スコアとイェーツの連続性補正(Wilson CC)のカバレッジ確率の比較。グラフは、ウィルソン CC の CI は常に保守的であるのに対し、ウィルソンの CI は真の比率の大きさに応じて保守的でリベラルであることを示しています。特に、ウィルソン CI は、真の比率が 0 または 1 に非常に近い場合、リベラルすぎる傾向があります。サイズ 50 の任意のサンプルについて、平均カバレッジ確率は Wilson CI と Wilson CC CI でそれぞれ 0.952 と 0.969 です。

3.Agresti-Coull 法

Agresti-Coull CI は、過度にリベラルな古典的 Wald CI の調整から得られます。結果として得られる CI のカバレッジ プロパティは Wilson CI と似ていますが、一般的には少し保守的です。さらに、2 つのタイプの CI は同じ中間点を持ちますが、Wilson CI は常に Agresti-Coull CI に含まれています。図 4 に示すように、真の比率が中程度の場合、カバレッジ確率は基本的に同じです。ただし、Agresti Coull CI は、真の比率が 0 または 1 に近い場合、一般的にリベラルではありません。図 4 に示すように、サイズ 50 のサンプルでは、真の比率が 0 または 1 に近づくにつれて、Agresti-Coull CI は保守的になります。 Agresti-Coull CI のもう一つの魅力は、Wald CI から受け継いだ実装のシンプルさです。さらに、特に信頼度が 95%の場合、教えやすく、覚えやすいです。この信頼水準では、Wald CI に対して行われた調整の覚書として、一般に「2 つの成功と 2 つの失敗を追加」CI 法と呼ばれます。

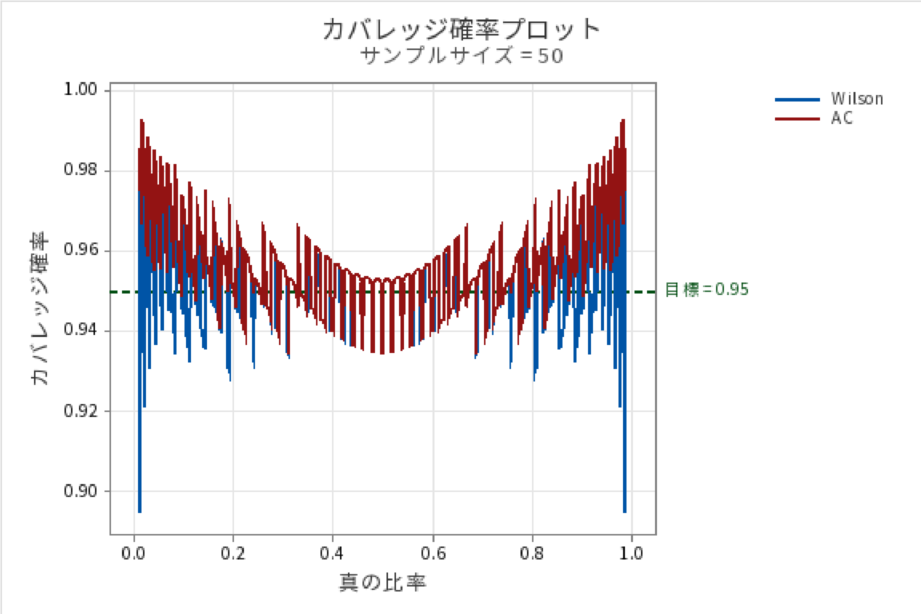

図 4: サンプルサイズが 50 の場合の真の比率の関数としての Wilson/score (Wilson)信頼区間と Agresti-Coull(AC)信頼区間のカバレッジ確率の比較。このグラフは、区間内の真の比率 (0.3, 0.75) の中程度の値に対して、2 つの手法でカバレッジ確率が基本的に同じであることを示しています。ただし、真の比率が 0 または 1 に近い値の場合、Wilson CI はリベラルで、Agresti-Coull 信頼区間は保守的です。これは、Wilson CI が Agresti-Coull CI に含まれるという事実と一致しています。サイズ 50 の任意のサンプルについて、Wilson CI と Agresti-Coull 信頼区間の平均カバレッジ確率は、それぞれ 0.952 と 0.958 です。

簡単な実例

大量生産工場の品質エンジニアが、ある日に大量生産された 1465 個の商品を無作為に抽出しました。1465 項目を独自に検査した結果、53 項目に欠陥があることが判明しました。技師は、特定の日に生産された不良品の比率が 2.5%と有意に異なるかどうかを知りたいと考えています。

この問題は統計的仮説検定の質問として投げかけられますが、応用統計学者は、検定の p 値とともに点推定値と信頼区間を分析結果に含めることがますます奨励されています。Minitab では、特に基本統計量モジュールにおいて、可能な限りこの規則に従います。たとえば、Minitab を使用すると、調整されたブレイカー法に基づく上記の質問の分析結果は次のようになります。

1 サンプルの比率の検定および信頼区間

方法
p 事象比率
方法:ブレイカーの正確な方法を調整しました

記述統計量

N 事象サンプル p p の 95%信頼区間
1465、530.036177(0.027353、0.046822)

検定
帰無仮説:p=0.025
対立仮説 H_1: p ≠ 0.025

p 値
0.009

4 つのメソッドのそれぞれについて、同様の出力を生成できます。各方法の特徴と利点は次の表にまとめられています。

方法	95%	一致検定の p 値
調整済みブレイカー	(2.74%,4.68%)	0.009
イェーツの訂正とウィルソン	(2.75%,4.74%)	0.008
ウイルソンかスコア	(2.78%,4.70%)	0.006
アグレスティ・コール	(2.77%,4.74%)	0.007

この例では、すべての p 値が 0.05 未満であるため、有意水準 0.05 で不良%が 2.5%と異なるという結論は、すべての方法で同じ結果が得られます。すべての方法の信頼区間と対応する p 値は、サンプルのサイズが非常に大きいことが一因で似ています。さらに、各方法の信頼区間は、一致する各仮説検定の p 値と一致する仮説比率値(2.5%)をカバーしていません。

上記の例では、品質エンジニアが 50 個の項目のみをテストし、2 個に欠陥があることがわかったとします。さらに、技師が不良品の比率が 1.0%と有意に異なるかどうかを知りたいとします。各方法の特徴と利点は次の表にまとめられています。

方法	95%	対応する検定の p 値
調整済みブレイカー	(0.72%,13.35%)	0.089
イェーツの訂正とウィルソン	(0.70%,14.86%)	0.155
ウイルソンかスコア	(1.10%,13.46%)	0.033
アグレスティ・コール	(0.34%,14.22%)	0.124

この場合、ウィルソン/スコア法のみが、有意水準 0.05 で不良率が 1.0%と異なるという点で有意な結論を導き出します。同じレベルの有意性では、他のすべての方法は、違いがあると判断するには証拠が不十分であるという反対の結論をもたらします。分析法間で結果にばらつきがあるのは、主にサンプルのサイズが中程度であることによるものです。平均して、これらの方法のカバレッジ確率は、サンプルサイズが大きくなるにつれて公称水準に近づきます(下の図 5 を参照)。ただし、小規模から中程度のサンプル計画では、各方法に関連する平均カバレッジ確率の差がより顕著になり、対応する CI の幅が著しく異なります。しかし、重要な問題は、どの結果を上司に報告するかということです。この質問に対する簡単な答えはありませんが、より良い答えは、真の比率の大きさやアプリケーションの領域に関する予備知識などの要因に依存する可能性があるためです。次のセクションでは、大まかな一般的なガイドラインを提供します。

結論

図 2、3、4 は、4 つの CI 手法、Adjusted Blaker、Wilson、Wilson CC、および Agresti-Coull のカバレッジ確率特性が異なることを示しています。ウィルソン CC が最も保守的であり、調整されたブレイカーがそれに続きます。アグレスティ・コール家とウィルソン家は、真の比率の大きさに応じて、リベラル派と保守派であることが多い。全体として、ウィルソン CI 法は 4 つの方法の中で最もリベラルです。さらに、図 5 は、平均して 4 つの方法すべてが保守的であり、ウィルソン CC が最も保守的であり、調整されたブレイカー法、アグレスティ・コール法、ウィルソン法がそれに続いていることを示しています。特定の問題に対して、選択する適切な方法は、特定のアプリケーション、サンプルサイズ、および真の比率の大きさに関する事前知識が利用可能かどうかによって異なります。

たとえば、規制当局は、消費者を保護するために保守的な方法に関心を持つことがよくあります。しかし、保守的すぎると規制が厳しくなり、リベラルすぎると規制が緩くなる可能性があります。一般に、中程度から大規模のサンプル設計では、調整済みブレイカー法またはウィルソン CC 法が適しています。小さなサンプル計画の場合、真の比率値に関する予備知識が利用できる場合は、Wilson または Agresti-Coull が適している可能性があります。このような事前知識は、多くの場合、以前の同様の経験または比率の大まかな推定値を得るために特別に設計されたパイロット小規模研究に基づいています。たとえば、欠陥の割合が通常 0 に近いアプリケーションの品質管理領域では、研究者が保守的な結果を望んでいるかどうかによって、選択する方法が異なります。保守的な結果は Agresti-Coull 法に基づいており、リベラルな結果は Wilson/score 法に基づいている可能性があります。最後に、より重要なことですが、タイプ II エラーから保護したり、CI の幅を制御したりするために、サンプルサイズを事前に計画することをお勧めします。Minitab には「そのためのアプリ」があります。Minitab には、仮説検定に関連する第 II 種過誤を制御したり、信頼区間の幅を制御したりするための適切なサンプルサイズを決定するツールが用意されています。

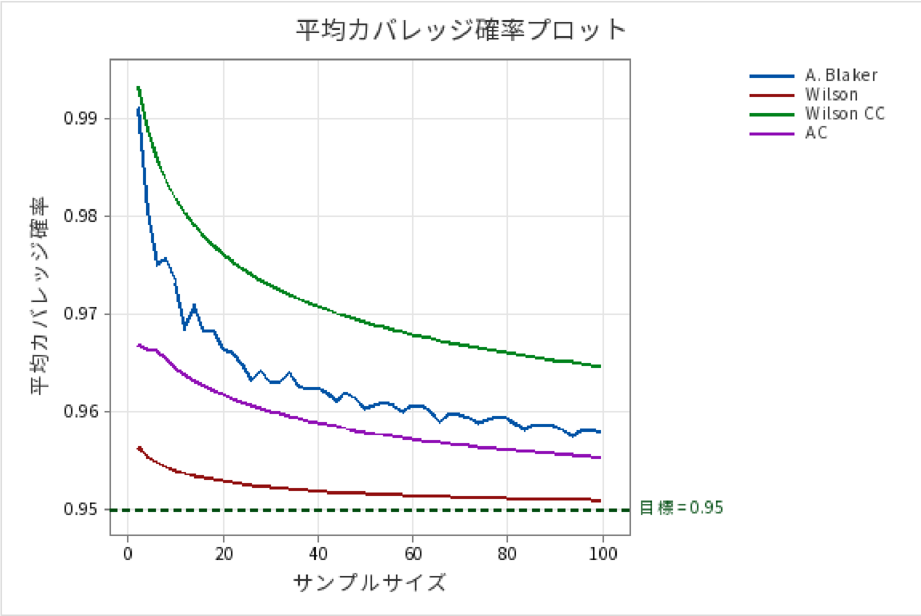

図 5: サンプルサイズの関数としての 4 つの CI 方法すべての平均カバレッジ確率。平均カバレッジ確率は、真の比率が単位区間で一様に分布していると仮定して計算されます。平均カバレッジ曲線は、平均して、イェーツの連続性補正(Wilson CC)CI 法による Wilson/スコアが最も保守的であり、次いで調整済みブレイカー法(A. Blaker)、Agresti-Coull(AC)、Wilson/スコア(Wilson)CI 法が続いていることを示しています。平均カバレッジ確率曲線は、サンプルサイズが大きくなるにつれて、目標の公称カバレッジレベルに近づきます。さらに、近似法(Wilson CC、Wilson、Agresti-Coull)の平均カバレッジ曲線は滑らかですが、厳密に調整された Blaker のカバレッジ曲線は、公称カバレッジに近づくにつれて多少の変動があります。これはおそらく、調整された Blaker CI メソッドをさらに改善できることを示しています。

Minitab

参考文献

Agresti, A. and Coull, B. A. (1998). Approximate is better than “Exact” for interval Estimation of Binomial Proportion. The American Statistician 52, 119–125

Blaker, H. (2000). Confidence Curves and Improved Exact Confidence Intervals for Discrete Distributions. The Canadian Journal of Statistics, 28, 783–798

Blaker, H. (2001). Corrigenda: Confidence curves and improves exact confidence intervals for discrete distributions. The Canadian Journal of Statistics, 29, 681.

Blyth, C. R. and Still, H. A. (1983). Binomial Confidence Intervals. Journal of the American Statistical Association 78, 108–116.

Brown, L. D., Cai, T. and Das Gupta, A. (2001). Interval Estimation for a Binomial Proportion. Statistical Science 16, 101–133.

Casella, G., 1986. Refining binomial confidence intervals. Canad. J. Statist. 14, 113-129.

Clopper, C. J. and Pearson, E. S. (1934). The Use of Confidence or Fiducial Limits Illustrated in the Case of Binomial. Biometrika 26, 404–413

Crow, E.L., 1956. Confidence intervals for a proportion. Biometrika 43, 423-435.

Klaschka, J. and Reiczigel, J. (2021). On matching confidence intervals and tests for some discrete distributions: methodological and computational aspects, Computational Statistics, Springer, vol. 36(3), 1775-1790.

Wilson E. B. (1927) Probable Inference, the Law of Successions and Statistical Inference. J. Amer. Statist. Assoc. 22, 209–21