COMPARANDO AS TAXAS DE COBERTURA PARA DIFERENTES INTERVALOS DE CONFIANÇA PARA UMA PROPORÇÃO

Palavras-chave: taxa de cobertura, intervalo de confiança, distribuição binomial

Carlos Henrique Trigo Nasser Felix - UNICAMP

Profº Dr. Rafael Pimentel Maia – UNICAMP

INTRODUÇÃO

O cálculo de intervalos de confiança para uma proporção é um problema clássico de estatística, presente em livros didáticos como Mood, Graybill, and Boes (1963). O problema se baseia na criação de um intervalo, obtido a partir do número total de sucessos observados em uma amostra aleatória simples de tamanho *n* de uma população, que contenha a verdadeira proporção de sucessos da população com uma probabilidade próxima de um nível de confiança estabelecido. Uma das primeiras propostas é registrada em Laplace (1820), e posteriormente outros métodos surgiram em Wilson (1927) e Clopper e Pearson (1934). Métodos para comparação entre os intervalos estão presentes em Vollset (1993), Agresti and Coull (1998) e Brown et al. (2002), mostrando que este tópico ainda está em desenvolvimento até os dias de hoje. Por meio dos métodos e comparações presentes nessa seleção da literatura mostrarei que a utilização de propriedades que não dependem da verdadeira proporção de sucessos da população geram resultados mais estáveis, e apresento outro método de se obter intervalos de confiança e sugestões de suas utilizações.

REVISÃO DA LITERATURA

Primeiramente vejo que intervalos de confiança para uma proporção tem como objetivo a estimação intervalar da proporção de sucessos de uma população por meio do total de sucessos de ensaios de Bernoulli independentes obtidos de uma amostra aleatória simples. Para diminuir a repetição desses termos adotei a seguinte representação: p é a verdadeira proporção de sucessos da população e pertence ao espaço paramétrico $[0;1]$, $q = 1 - p$, n é o tamanho da amostra, X é o total de sucessos da amostra, $\hat{p} = \frac{X}{n}$, α é o nível de significância que nesse caso pode ser interpretado como a probabilidade desejada do intervalo não conter p. Por meio dessas representações defino os intervalos presentes na literatura que são utilizados na comparação desse texto, divididos em dois grupos: os que utilizam a aproximação da distribuição binomial pela normal e os que utilizam diretamente a binomial.

Intervalos que utilizam a aproximação pela normal

Os seguintes intervalos tem como base a utilização do teorema central do limite aplicado sobre a distribuição de Bernoulli, ou binomial, como pode ser visto no exemplo 5 do capítulo 8 de Mood, Graybill, and Boes (1963), obtendo a seguinte inequação.

$$\hat{p} - Z_{\frac{\alpha}{2}}\sqrt{\frac{pq}{n}} \leq p \leq \hat{p} + Z_{\frac{\alpha}{2}}\sqrt{\frac{pq}{n}}$$

Intervalo de Wald

O intervalo de Wald aparece brevemente em Laplace (1820) e Wilson (1927). Uma demonstração de sua obtenção está presente no exemplo 5 do capítulo 8 de Mood, Graybill, and Boes (1963). Com um nível de confiança de (1 − α) × 100% o intervalo é dado por

$$\left[\hat{p} - Z_{\frac{\alpha}{2}}\sqrt{\frac{pq}{n}}; \hat{p} + Z_{\frac{\alpha}{2}}\sqrt{\frac{pq}{n}}\right]$$

substituindo *p* por $\hat{p}$ *e* q por $1-\hat{p}$; $\hat{p}$ é a estimativa obtida a partir de uma amostra de tamanho *n* e $Z_{\alpha/2}$ é o $1-\alpha/2$ percentil de uma variável Z com distribuição normal padrão. Uma outra versão pode ser obtida substituindo *p* e *q* por 0.5.

Intervalo de Wilson

O intervalo de Wilson é proposto em Wilson (1927), uma demonstração de sua obtenção está presente no exemplo 5 do capítulo 8 de Mood, Graybill, and Boes (1963). com um nível de confiança de (1 − α) × 100% o intervalo é dado por

$$\left[\frac{\hat{p}+\frac{k^2}{2}-k\sqrt{\hat{p}(1-\hat{p})+\frac{k^2}{4}}}{1+k^2}; \frac{\hat{p}+\frac{k^2}{2}+k\sqrt{\hat{p}(1-\hat{p})+\frac{k^2}{4}}}{1+k^2}\right]$$

em que $k = \frac{Z_{\alpha/2}}{\sqrt{n}}$, $\hat{p}$ é a estimativa obtida a partir de uma amostra de tamanho *n* e $Z_{\alpha/2}$ é o $1-\alpha/2$ percentil de uma variável Z com distribuição normal padrão.

Transformação do arco seno

A utilização da transformação do arco seno é comparada com outros métodos em Vollset (1993). com um nível de confiança de (1 − α) × 100% o intervalo é dado por

$$\left[sin^2\left(arcsin\left(\sqrt{\hat{p}}\right) - \frac{Z_{\frac{\alpha}{2}}}{2\sqrt{n}}\right); sin^2\left(arcsin\left(\sqrt{\hat{p}}\right) + \frac{Z_{\frac{\alpha}{2}}}{2\sqrt{n}}\right)\right]$$

em que $\hat{p}$ é a estimativa obtida a partir de uma amostra de tamanho *n* e $Z_{\alpha/2}$ é o $1-\alpha/2$ percentil de uma variável Z com distribuição normal padrão, além disso uso uma generalização da modificação presente em Vollset (1993), se $Z_{\alpha/2} \geq \pi\sqrt{n}$ o intervalo é substituído por $[0;1]$, se $\hat{p} \leq sin^2\left(\frac{Z_{\alpha/2}}{2\sqrt{n}}\right)$ o extremo inferior do intervalo é substituído por 0 e se $\hat{p} \geq 1 - sin^2\left(\frac{Z_{\alpha/2}}{2\sqrt{n}}\right)$ o extremo superior do intervalo é substituído por 1.

Intervalo de Agresti-Coull

O intervalo de Agresti-Coull é proposto em Agresti and Coull (1998), e é comparado com outros métodos em Brown et al. (2002). com um nível de confiança de (1 − α) × 100% o intervalo é dado por

$$\left[\bar{p} - Z_{\frac{\alpha}{2}}\sqrt{\left(\frac{\bar{p}(1-\bar{p})}{\bar{n}}\right)}; \bar{p} + Z_{\frac{\alpha}{2}}\sqrt{\left(\frac{\bar{p}(1-\bar{p})}{\bar{n}}\right)}\right]$$

em que $\bar{p} = \frac{2n\hat{p}+Z^2_{\alpha/2}}{2\bar{n}}$, $\bar{n} = n + Z^2_{\alpha/2}$, $\hat{p}$ é a estimativa obtida de uma amostra de tamanho n e $Z_{\alpha/2}$ é o $1-\alpha/2$ percentil de uma variável Z com distribuição normal padrão.

Intervalos que utilizam diretamente a distribuição binomial

Intervalo de Clopper-Pearson

O intervalo de Clopper-Pearson, ou Exato, é proposto em Clopper and Pearson (1934, com um nível de confiança de (1 − α) × 100% o intervalo é dado por [A; B] definido da seguinte forma.

$$\begin{cases} \frac{\beta_A(x,n-x+1)}{\beta_1(x,n-x+1)} = \frac{\alpha}{2} \\ \frac{\beta_B(x+1,n-x)}{\beta_1(x+1,n-x)} = 1-\frac{\alpha}{2} \end{cases}$$

em que x é o total de sucessos de uma amostra de tamanho n , $\beta_a(b,c) = \int_0^a t^{b-1}(1-t)^{c-1}dt$.

Intervalo de Jeffrey

O intervalo de Jeffreys é comparado com outros métodos em Brown et al. (2002). com um nível de confiança de (1 − α) × 100% o intervalo é dado por [A; B] definido da seguinte forma

$$\begin{cases} \frac{\beta_A\left(x+\frac{1}{2},n-x+\frac{1}{2}\right)}{\beta_1\left(x+\frac{1}{2},n-x+\frac{1}{2}\right)} = \frac{\alpha}{2} \\ \frac{\beta_B\left(x+\frac{1}{2},n-x+\frac{1}{2}\right)}{\beta_1\left(x+\frac{1}{2},n-x+\frac{1}{2}\right)} = 1-\frac{\alpha}{2} \end{cases}$$

em que x é o total de sucessos de uma amostra de tamanho n , $\beta_a(b,c) = \int_0^a t^{b-1}(1-t)^{c-1}dt$.

Métodos Comparativos

O primeiro método para comparações dos diversos intervalos propostos é a *probabilidade de cobertura real* presente em Agresti and Coull (1998), podendo ser calculado pelas expressões presentes na equação (1).

$$P[I(X,p)|p] = \sum_{k=0}^{n} P[I(X,p)|X=k,p]P[X=k|p] = \sum_{k=0}^{n} I(k,p)P[X=k|p] \quad (1)$$

com $I(k,p) = 1$ se o intervalo contém p para $X = k$, e $I(k,p) = 0$ caso contrário.

Na secção 2.2 do capítulo 8 de Mood, Graybill, and Boes (1963) esta presente as seguintes propriedades: valor esperado do tamanho do intervalo, calculado por meio das expressões presentes em (2), e a probabilidade do intervalo conter uma proporção dado o valor verdadeiro da proporção estão presentes, calculada por meio das expressões presentes em (3).

$$\int_0^1 P[I(X,p^*)|p]dp^* = \sum_{k=0}^{n} \int_0^1 I(k,p^*)dp^* P[X=k|p] \quad (2)$$

$$P[I(X,p^*)|p] = \sum_{k=0}^{n} I(k,p^*)P[X=k|p] \quad (3)$$

com $I(k,p) = 1$ se o intervalo contém p para $X = k$, e $I(k,p) = 0$ caso contrário.

Note que as equações (1) e (2) podem ser obtidas por meio de (3), utilizando $p = p^*$e integrando de 0 a 1 em função de p^* respectivamente. Além disso existem as seguintes comparações e conclusões presentes na literatura selecionada.

Vollset (1993) compara os métodos Wald, Wilson, arco seno e Clopper-Pearson com outros, e obtém que os métodos Clopper-Pearson e Wilson podem ser usados em qualquer circunstância, com o primeiro sendo mais conservador.

Agresti and Coull (1998) compara os métodos Wald, Wilson, Clopper-Pearson e Agresti-Coull, e obtém que os métodos Agresti-Coull e Wilson fornecem intervalos mais curtos, com probabilidade de cobertura real geralmente mais próxima da nominal e uma melhor desempenho operacional em termos da interpretação desses se comparados com Clopper-Pearson.

Brown et al. (2002) compara os métodos Wald, Wilson, Agresti-Coull e ,o intervalo Agresti-Coull, e obtém que embora possua intervalos maiores em média, pode ser recomendado para uso neste problema e se a simplicidade não é um questão primordial, o intervalo de Wilson ou de Jeffreys podem ser usado.

Intervalo Proposto

Por meio da implementação dos métodos citados, desenvolvo os seguintes intervalos, como a quantidade de intervalos que contém o valor *p* está relacionado com o tamanho do intervalo, é possível reduzir o segundo reduzindo o primeiro, mas é preciso levar em consideração o valor da taxa de cobertura, que deve ser superior ou próxima ao valor desejado. A Figura 1 ilustra uma forma de se realizar esse processo, adicionando o intervalo com maior probabilidade, até o critério de parada dado pelo método. Esse novo intervalo foi denominado de *minimalC*.

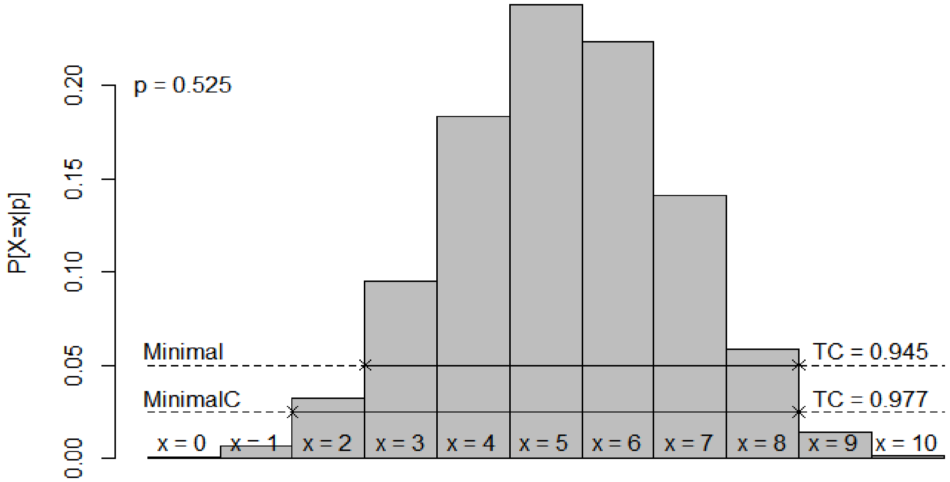

Figura 1: Probabilidade de se obter x sucessos pelo valor de x, com a probabilidade de sucesso igual a 0.525 e tamanho de amostra 10, as retas horizontais indicam os valores de x cujo intervalo contem a proporção p, para o nível de significância $\alpha = 0.05$ e com os respectivos critérios de parada.

Todos os métodos para calcular intervalos de confiança e compará-los foram implementados em linguagem R (por R Core Team, 2019). em um pacote denominado CIBP que está em fase final de implementação.

A Figura 2 apresenta os gráfico da probabilidade de cobertura real pela proporção *p* desses para cada um dos métodos apresentados.

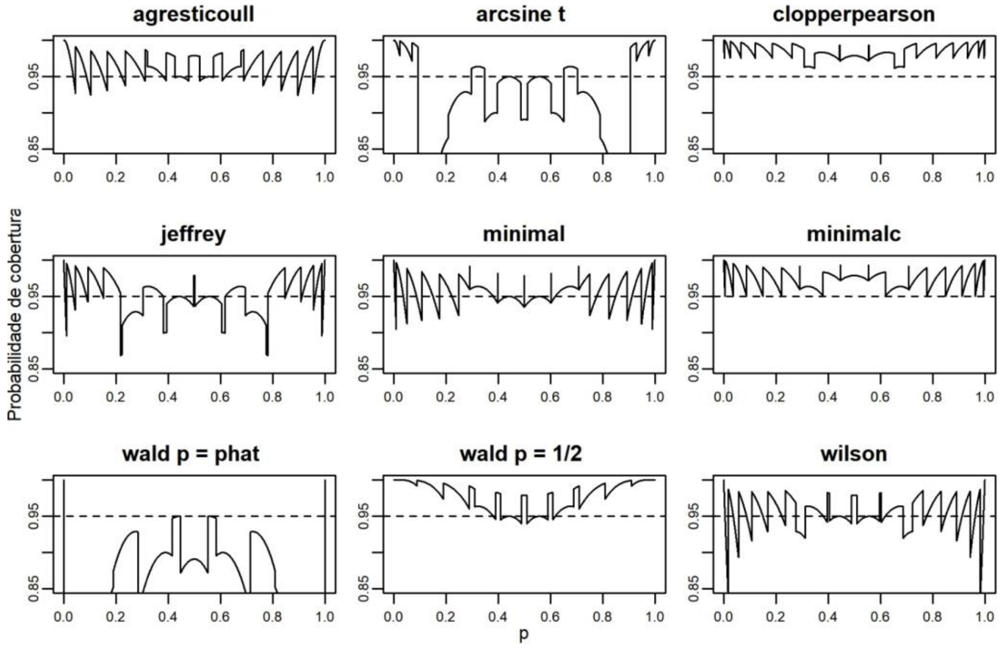

Figura 2: Probabilidade de cobertura real pela verdadeira proporção p variando entre 0 e 1, para os respectivos métodos com o tamanho de amostra n = 10 e nível de confiança α = 0.05, a linha tracejada horizontal é o valor da taxa de cobertura esperada 1 − α = 0.95.

Visto que a finalidade de um intervalo de confiança é obter uma estimativa da proporção desconhecida de sucessos, a comparação de propriedades que dependem dessa se tornam instáveis, então utilizo as seguintes modificações das propriedades: no lugar de utilizar o tamanho esperado do intervalo utilizo o tamanho máximo do intervalo $\max T(k)$, com k = 0, ..., n e T(k) =“Tamanho do intervalo quando X = k”; e ao invés de utilizar a taxa de cobertura utilizo a taxa de cobertura mínima dada por $\min P[I(X,p)|p]$, com p ∈ [0, 1].

RESULTADOS

Comparar diretamente a probabilidade de cobertura real dos métodos é instável, com isso as propriedades que utilizo para comparar as taxas de cobertura real dos métodos, para comparação do tamanho máximo utilizo o Teorema central do limite para obter que $\frac{Z_{\alpha/2}}{\sqrt{n}}$ é uma aproximação desse a medida que $n \to \infty$, então opto por utilizar $[max_k T(k)]\sqrt{n}/Z_{\alpha/2}$ ao invés de $max_k T(k)$, além disso removo os valores referentes a $n = 1$, pois o método “wald p = phat” retorna intervalos de tamanho 0, com essas ideias gero a Figura 3.

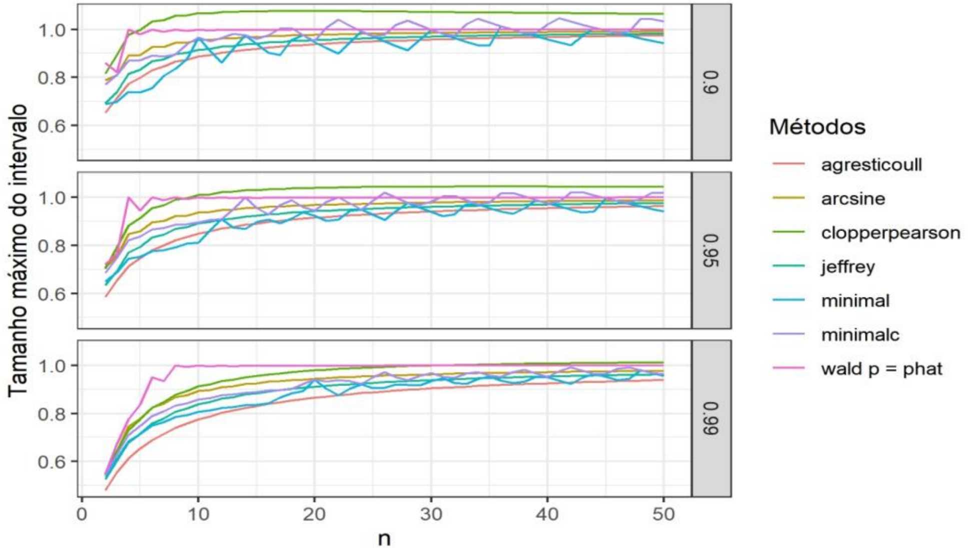

Figura 3: Tamanho máximo do intervalo multiplicado por $\sqrt{n}/Z_{\alpha/2}$ pelo tamanho da amostra n variando de 2 a 50 para taxa de cobertura nominal 1 − α = 0.9, 0.95, 0.99. Os métodos "wald p = 1/2" e "wilson" foram retirados por terem valores muito próximos dos métodos "wald p = phat" e "agrestcoull" respectivamente.

Portanto a partir de um certo valor de n a seguinte ordenação dos métodos do maior tamanho para o menor é obtida, Copper-Pearson, Wald p = 1/2, Wald p = phat, Arcoseno, Jeffrey, Agresti-Coull e Wilson, antes desse temos que os métodos Wald p = 1/2, Wald p = phat tendem a ser maior do que Copper-Pearson, além disso Minimal e Minimal C variam de posição ao variar o tamanho da amostra, com o primeiro variando entre os menores e sendo o menor para alguns valores, e o segundo entre os maiores e nunca é o maior.
De forma análoga, temos que a Probabilidade de cobertura tende a 1−α a medida que $n \to \infty$, com p ∈ (0, 1), portanto utilizo (taxa de cobertura mínima −1)/α ao invés da taxa de cobertura mínima, ilustrado pela Figura 4.

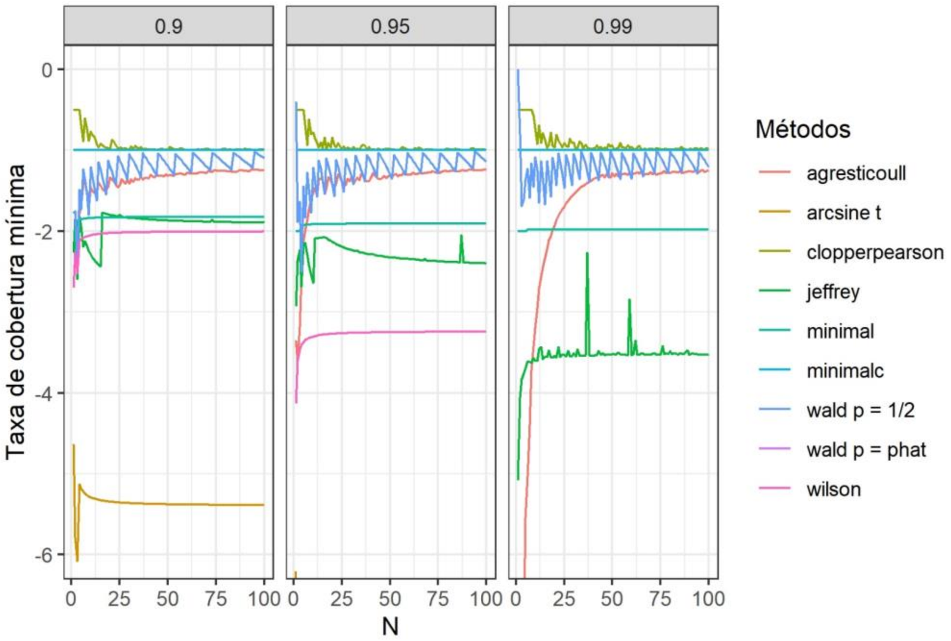

Figura 4: Taxa de cobertura mínima reescalonada pela função $f(x) = \frac{x-1}{\alpha}$ pelo tamanho de amostra n variando entre 1 e 100 para taxa de cobertura nominal 1 − α = 0.9, 0.95, 0.99.

De maneira análoga, a partir de um certo valor de n a seguinte ordenação dos métodos de acordo com sua convergência a 1 − α obtendo Minimal C, Copper-Pearson, Wald p = 1/2 / Agresti-Coull, Minimal, Jeffrey, Wilson, Arcoseno e Wald p = phat, observe que os métodos Minimal, Jeffrey, Wilson, Arcoseno e Wald p = phat não convergem para 1 − α, e são ordenados de acordo com a proximidade do respectivo valor que o método converge com 1 − α.

Juntando os resultados obtidos com os critérios discutidos anteriormente, se escolhido que a taxa de cobertura mínima deve ser maior que 1 − α, a melhor taxa de cobertura é a de MinimalC, seguida de Copper-Pearson e dos outros métodos ordenados do menor tamanho para o maior, se a taxa de cobertura mínima deve ser próxima de 1 − α, um método possui uma melhor taxa de cobertura que outro se tiver uma maior ou igual convergência da taxa de cobertura mínima e uma menor ou igual tamanho máximo do intervalo, caso um método não tiver uma taxa de cobertura melhor ou pior que outro, então é considerado igual.

A partir dos critérios desenvolvidos sugiro para obter uma probabilidade de cobertura maior que 1 − α a utilização do método Minimal C, caso opte por obter uma probabilidade de cobertura próxima de 1 − α se a taxa de cobertura mínima for superior ao igual ao do Minimal sugiro o método Agresti-Coull, caso contrário Minimal.

CONCLUSÕES

A utilização de propriedades que não dependam de *p*, como a taxa de cobertura mínima e o tamanho máximo do intervalo, na comparação de taxas de cobertura obter resultados mais estáveis, possibilitando uma seleção de métodos que geram intervalos de confiança para uma proporção binomial que independem de pequenas variações de *p, n e α*.

Comparando os resultados obtidos nesse texto com os presentes na literatura selecionada temos que em ambos a sugestão de intervalos independe do valor de p, os da literatura obtiveram essa utilizando valores médios das propriedades selecionadas, nesse texto foi utilizado valores máximos e mínimos dessas, que além de ser mais fácil de computar e interpretar é possível mensurar o controle do método sobre a propriedade o comparando com seu valor esperado. Outra diferença é o desenvolvimento de métodos que se baseiam no controle da taxa de cobertura real, Clopper and Pearson (1934) é o único texto presente na literatura selecionada que desenvolve um método baseado nesse conceito.

BIBLIOGRAFIA

Agresti, Alan, and Brent A Coull. 1998. “Approximate Is Better Than ‘Exact’ for Interval Estimation of Binomial Proportions.” The American Statistician 52 (2): 119–26.

Brown, Lawrence D, T Tony Cai, Anirban Dasgupta, and others. 2002. “Confidence Intervals for a Binomial Proportion and Asymptotic Expansions.” The Annals of Statistics 30 (1): 160–201.

Clopper, Charles J, and Egon S Pearson. 1934. “The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial.” Biometrika 26 (4): 404–13.

Laplace, Pierre Simon. 1820. Théorie Analytique Des Probabilités. Courcier.

Mood, AM, FA Graybill, and DC Boes. 1963. “Introduction to the Theory of Statistics. Mc-Graw Hill Book Company.” Inc., New York.

R Core Team. 2019. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Vollset, Stein Emil. 1993. “Confidence Intervals for a Binomial Proportion.” Statistics in Medicine 12 (9): 809–24.

Wilson, Edwin B. 1927. “Probable Inference, the Law of Succession, and Statistical Inference.” Journal of the American Statistical Association 22 (158): 209–12.