Confidence Bounds & Intervals for Parameters Relating to the Binomial, Negative Binomial, Poisson and Hypergeometric Distributions

With Applications to Rare Events

Fritz Scholz[1]

February 28, 2008

1 Introduction and Overview

We present here by direct argument the classical Clopper-Pearson (1934) "exact" confidence bounds and corresponding intervals for the parameter p of the binomial distribution. The same arguments can be applied to derive confidence bounds and intervals for the negative binomial parameter p, for the Poisson parameter λ, for the ratio of two Poisson parameters, $\rho = \lambda_1/\lambda_2$, and for the parameter D of the hypergeometric distribution.

The 1-sided bounds presented here are exact in the sense that their **minimum** probability of covering the respective unknown parameter is equal to the specified target confidence level $\gamma = 1-\alpha$, $0 < \gamma < 1$ or $0 < \alpha < 1$. If $\hat\theta_L$ and $\hat\theta_U$ denote the respective lower and upper confidence bounds for the parameter θ of interest, this amounts to the following coverage properties

$$\inf_\theta P_\theta(\hat\theta_L \le \theta) = \gamma \qquad \text{and} \qquad \inf_\theta P_\theta(\hat\theta_U \ge \theta) = \gamma\ .$$

Such infima of coverage probabilities are also referred to as **confidence coefficients** of the respective bounds. For some parameters the coverage probability of these bounds is equal to or arbitrarily close to the desired confidence level while for the remaining parameters the coverage probability is greater than γ. In that sense these one-sided bounds are conservative in their coverage.

By combining such one-sided bounds for θ, each with confidence coefficient $1-\alpha/2$ or with maximum miss probability $\alpha/2$, we can use $[\hat\theta_L, \hat\theta_U]$ as confidence interval for θ with confidence coefficient $\ge 1-\alpha = \gamma$.

This derives from the fact that typically we have $P(\hat\theta_L \le \hat\theta_U) = 1$ when the confidence coefficients of the individual bounds are $1-\alpha/2 \ge 1/2$, namely

[1]Please report any errors or typos to me at `fscholz@u.washington.edu`

$$
\begin{aligned}
P_\theta(\hat{\theta}_L \le \theta \le \hat{\theta}_U) &= 1 - \left[P_\theta(\hat{\theta}_L > \theta \ \cup \ \theta > \hat{\theta}_U)\right] \\
&= 1 - \left[P_\theta(\hat{\theta}_L > \theta) + P_\theta(\hat{\theta}_U < \theta)\right] \ge 1 - [\alpha/2 + \alpha/2] = 1 - \alpha \ .
\end{aligned}
$$

Taking the infimum over all θ on both sides adds a further level of conservatism by yielding confidence intervals with minimum coverage probability or confidence coefficient $\ge \gamma = 1 - \alpha$. This minimum probability of interval coverage is typically $> \gamma$ since the parameters where the respective one-sided bounds achieve their maximum miss probability of $\alpha/2$ are usually not the same. This is illustrated in the context of the binomial distribution in Figures 4 and 5.

For confidence bounds in the hypergeometric situation the minimum achievable confidence level for one-sided bounds may be higher than the desired specified confidence level. However there are some confidence levels where the minimum achievable confidence level or confidence coefficient coincides with the specified confidence level. This issue is due to the discrete nature of the parameter D. Corresponding illustrations are given in Figures 22 and 23.

For the binomial situation Agresti and Coull (1998) have recently discussed advantages of alternate methods, where the actual coverage oscillates more or less around the target value γ, and not above it. The advantage of such intervals is that they are somewhat shorter than the Clopper-Pearson intervals. A recent similar discussion for the Poisson parameter can be found in Barker (2002). Given that we often deal with confidence bounds concerning rare events (accidents or undesirable defects) we prefer the conservative approach of Clopper and Pearson. Any conservative properties should be viewed as an additional bonus or precaution.

It is shown how such “exact” bounds for binomial, negative binomial, and Poisson parameters can be computed quite easily using either the functions `BETAINV` and `GAMMAINV` in the Excel spread sheet or using the functions `qbeta` and `qgamma` in the statistical packages R (R is available as Free Software under the terms of the Free Software Foundation’s GNU General Public License for various operating systems (Unix, Linux, Windows, MacOS X) at `http://cran.r-project.org/`) or S-Plus (commercial `http://www.insightful.com/`). However, the `GAMMAINV` function in the Excel spread sheet is not always stable in older versions of Excel. Thus care needs to be exercised.

As far as confidence bounds for the parameter D of the hypergeometric distribution are concerned Excel does not offer a convenient tool. However, it does have hypergeometric probability function `HYPGEOMDIST` but no cumulative counterpart. Some sufficiently capable person can presumably come up with a macro for Excel that produces such confidence bounds for D based on the recipe given here.

However, all confidence bounds developed here can be computed in R using straightforward commands or functions that are supplied as part of an R work space on the web at
`http://www.stat.washington.edu/fritz/Stat498B.html`.

We first give the argument for confidence bounds for the binomial parameter p. The development of lower and upper bounds are completely parallel and it suffices to get a complete grasp of only one such derivation. Even though it becomes repetitive we give the complete arguments again for all other distributions.

This is followed by the corresponding argument for the negative binomial parameter p and the Poisson parameter λ. For very small p it is pointed out how to use the very effective Poisson approximation to get bounds on p. Finally we give the classical method for constructing confidence bounds on the ratio $\rho = \lambda_1/\lambda_2$ based on two independent Poisson counts X and Y from Poisson distributions with parameters λ_1 and λ_2, respectively. This latter method nicely ties in with our earlier binomial confidence bounds and is quite useful in assessing relative accident or incident rates.

The last section deals with the topic of inverse probability solving for the binomial and Poisson distributions and shows how to accomplish this in Excel.

We point out that the confidence bounds for a binomial success probability involve two probabilities within the same confidence statement. The one probability concerns the target of interest, the success probability p, while the other is invoked as the confidence level γ, which gives us some probability assurance that the computed confidence bounds or intervals correctly cover the unknown p. In using such intervals we always allow for the "rare" possibility that the target may be missed and we will not know whether that is the case or not. Of course, one can always increase the confidence level γ.

A conceptual difficulty arises when trying to control risks in the case of very small values of p. By stating an upper confidence bound on p we also take the additional risk $(1-\gamma)$ of having missed p. How should one choose γ to balance out the two risks? The one risk (p) concerns future operations under the same set of conditions, and the other risk concerns the uncertainty in the collected data that were used in computing the upper bound.

There is not much we can do about p itself. It is a given, although unknown. Concerning confidence bounds we can raise γ to make the risk of missing the target, namely $1-\gamma$, as small as we like. If the upper bound becomes too conservative we can always suggest to take larger sample sizes n. In either case there are costs involved and ultimately it becomes an issue of balancing cost and one of practicality. As far as we know, this interplay of risks and their reconciliation has not been addressed in the research literature.

For a unique reference on confidence intervals we refer to the text by Hahn and Meeker (1991). Their scope of applications is obviously much wider than what is attempted here. In particular, it covers parameters for continuous distributions as well.

2 Binomial Distribution: Upper and Lower Bounds for p

2.1 Preliminaries about the Gamma and Beta Distributions

The gamma function is defined for $a > 0$ as

$$\Gamma(a) = \int_0^\infty \exp(-x)x^{a-1}\,dx$$

and thus

$$f_a(x) = \frac{1}{\Gamma(a)}\exp(-x)x^{a-1} \quad \text{for } x > 0 \quad \text{and} \quad f_a(x) = 0 \quad \text{otherwise}$$

is a density on $(0, \infty)$. It is called the gamma density with shape parameter $a > 0$.
The beta function is defined for $a > 0$ and $b > 0$ as

$$B(a, b) = \int_0^1 x^{a-1}(1-x)^{b-1}\,dx$$

and thus

$$g_{a,b}(x) = \frac{1}{B(a, b)}\,x^{a-1}(1-x)^{b-1} \quad \text{for } 0 < x < 1 \quad \text{and} \quad g_{a,b}(x) = 0 \quad \text{otherwise}$$

is a density on $(0, 1)$. It is called the beta density with parameters $a > 0$ and $b > 0$.
We have the following identities

$$\Gamma(a+1) = a\Gamma(a) \quad \text{for any } x > 0 \qquad \text{and} \qquad \Gamma(n+1) = n! = 1 \cdot 2 \cdot \ldots \cdot n \quad \text{for integer } n$$

The first is shown by integration by parts, integrating $\exp(-x)$ and differentiating $x^{(a+1)-1}$ with respect to x. The second is shown by induction using the previous identity and showing $\Gamma(1) = 1$.
Further, there is the following expression of $B(a, b)$ in terms of the gamma function

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}\ .$$

This can be shown by considering the convolution density $h_{a,b}(w)$ of $W = X + Y$ with independent gamma random variables $X \sim f_a(x)$ and $Y \sim f_b(y)$, i.e., $h_{a,b}(w) = \int_{-\infty}^{\infty} f_b(w-x)f_a(x)\,dx = \int_0^w f_b(w-x)f_a(x)\,dx$, and showing that it reduces to

$$h_{a,b}(w) = \frac{\exp(-w)w^{a+b-1}}{\Gamma(a+b)}\int_0^1 x^{a-1}(1-x)^{b-1}\,dx\,\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} = f_{a+b}(w)\int_0^1 x^{a-1}(1-x)^{b-1}\,dx\,\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$$

from which the identity follows by integrating $h_{a,b}(w)$ over $(0, \infty)$. At the same time this yields the fact that $W = X + Y \sim f_{a+b}(w)$, i.e., is a gamma random variable with shape parameter $a + b$.

2.2 A Binomial Identity

Suppose X is a binomial random variable, i.e., X counts successes in n independent Bernoulli trials with success probability p $(0 \le p \le 1)$ in each trial. Then we have

$$P_p(X \le k) = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i}$$

and $P_p(X \ge k) = 1 - P_p(X \le k-1)$. It is intuitive that $P_p(X \le k)$ is strictly decreasing in p for $k = 0, 1, \ldots, n-1$ and by complement $P_p(X \ge k)$ is strictly increasing in p for $k = 1, 2, \ldots, n$. Using the identities $i\binom{n}{i} = n\binom{n-1}{i-1}$ and $(n-i)\binom{n}{i} = n\binom{n-1}{i}$ a formal proof can be obtained by taking the derivative of $P_p(X \ge k)$ with respect to p and canceling all but one term in the difference of sums resulting from the differentiation, i.e., one gets

$$\begin{aligned}\frac{\partial P_p(X \ge k)}{\partial p} &= \sum_{i=k}^{n} \binom{n}{i} i p^{i-1}(1-p)^{n-i} - \sum_{i=k}^{n-1} \binom{n}{i} (n-i) p^i (1-p)^{n-i-1} \\ &= n\left[\sum_{i=k}^{n} \binom{n-1}{i-1} p^{i-1}(1-p)^{n-i} - \sum_{i=k}^{n-1} \binom{n-1}{i} p^i (1-p)^{n-i-1}\right] = k\binom{n}{k} p^{k-1}(1-p)^{n-k} > 0 \,.\end{aligned}$$

For $k > 0$ this immediately results in the following relationship between the binomial right tail summation and the Beta distribution function (also called incomplete Beta function), namely

$$P_p(X \ge k) = \sum_{i=k}^{n} \binom{n}{i} p^i (1-p)^{n-i} = I_p(k, n-k+1) \,, \tag{1}$$

where

$$I_y(a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^y t^{a-1}(1-t)^{b-1}\, dt = P(Y \le y)$$

denotes the cdf of a beta random variable Y with parameters $a > 0$ and $b > 0$. Note that by the Fundamental Theorem of Calculus the derivative of

$$I_p(k, n-k+1) = \frac{\Gamma(n+1)}{\Gamma(k)\Gamma(n-k+1)} \int_0^p t^{k-1}(1-t)^{n-k}\, dt = k\binom{n}{k} \int_0^p t^{k-1}(1-t)^{n-k}\, dt$$

with respect to p is

$$k\binom{n}{k} p^{k-1}(1-p)^{n-k}$$

which agrees with the previous derivative of $P_p(X \ge k)$ with respect to p. This proves the above identity (1) since for both sides of that identity the starting value at $p = 0$ is 0.

By complement we get for $k < n$ the following dual identity for the left tail binomial summation:

$$P_p(X \le k) = 1 - P_p(X \ge k+1) = 1 - I_p(k+1, n-k) \,. \tag{2}$$

2.3 A General Monotonicity Property

Suppose the function $\psi(x)$ defined on the integers $x = 0, 1, \ldots, n$ is monotone increasing (decreasing) and not constant, then the expectation $E_p(\psi(X))$ is strictly monotone increasing (decreasing) in p.
Proof: Using the identities $x\binom{n}{x} = n\binom{n-1}{x-1}$ and $(n-x)\binom{n}{x} = n\binom{n-1}{x}$ and changing the summation index to $y = x-1$ in one of the sums, it follows by differentiation as above

$$
\begin{aligned}
\frac{\partial E_p(\psi(X))}{\partial p} &= \frac{\partial}{\partial p}\sum_{x=0}^{n}\binom{n}{x}p^x(1-p)^{n-x}\psi(x) \\
&= \sum_{x=1}^{n}\binom{n}{x}xp^{x-1}(1-p)^{n-x}\psi(x) - \sum_{x=0}^{n-1}\binom{n}{x}(n-x)p^x(1-p)^{n-x-1}\psi(x) \\
&= n\sum_{x=1}^{n}\binom{n-1}{x-1}p^{x-1}(1-p)^{n-x}\psi(x) - n\sum_{x=0}^{n-1}\binom{n-1}{x}p^x(1-p)^{n-x-1}\psi(x) \\
&= n\sum_{y=0}^{n-1}\binom{n-1}{y}p^y(1-p)^{n-1-y}\psi(y+1) - n\sum_{x=0}^{n-1}\binom{n-1}{x}p^x(1-p)^{n-x-1}\psi(x) \\
&= n\sum_{y=0}^{n-1}\binom{n-1}{y}p^y(1-p)^{n-1-y}[\psi(y+1)-\psi(y)] > 0 \qquad \text{q.e.d.}
\end{aligned}
$$

Of course, the previous monotonicity property of $P_p(X \le k)$ ($P_p(X \ge k)$) follows from this result by taking $\psi(x) = 1$ for $x \le k$ and $\psi(x) = 0$ for $x > k$ ($\psi(x) = 0$ for $x < k$ and $\psi(x) = 1$ for $x \ge k$), however by going through the direct argument we also obtained the useful identities (1) and (2).

2.4 Upper Bounds for p

Consider testing the hypothesis $H(p_0) : p = p_0$ against the alternative $A(p_0) : p < p_0$. Small values of X can be viewed as evidence against the hypothesis $H(p_0)$. Thus we reject $H(p_0)$ at target significance level α when $X \le k(p_0, \alpha)$, where $k(p_0, \alpha)$ is chosen as the largest integer k for which $P_{p_0}(X \le k) \le \alpha$. Thus we have $P_{p_0}(X \le k(p_0, \alpha)) \le \alpha$ and $P_{p_0}(X \le k(p_0, \alpha) + 1) > \alpha$. When $X > k(p_0, \alpha)$ we accept $H(p_0)$.
When testing hypotheses it is often more informative to report the p-value corresponding to an observed value x of X. This p-value simply is the probability of seeing a result as extreme or more extreme pointing away from the hypothesis in the direction of the entertained alternative, when in fact the hypothesis is true. In our testing situation this p-value corresponding to x is $p(x, p_0) = P_{p_0}(X \le x)$. It is evident that the decision of rejecting or accepting $H(p_0)$ can be based simply on this p-value, namely reject $H(p_0)$ whenever $p(x, p_0) \le \alpha$ and accept $H(p_0)$ whenever $p(x, p_0) > \alpha$, see the illustration in Figure 1.

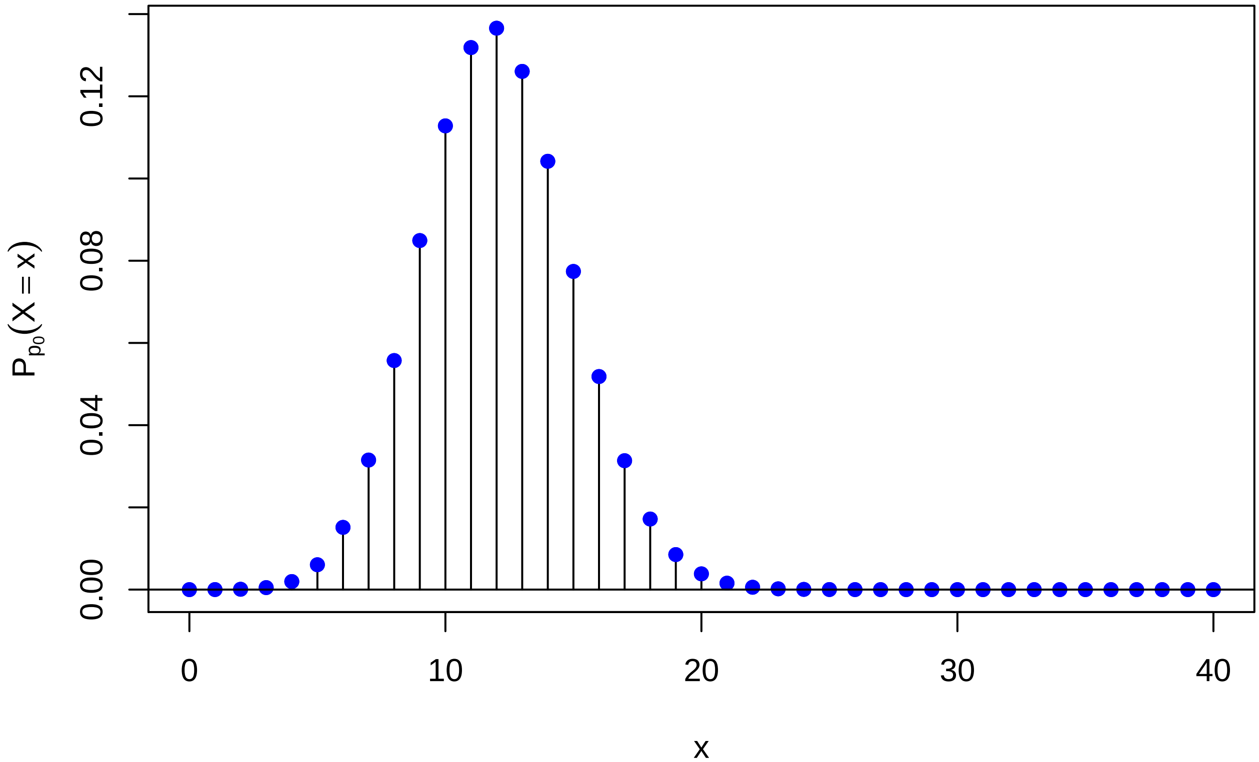
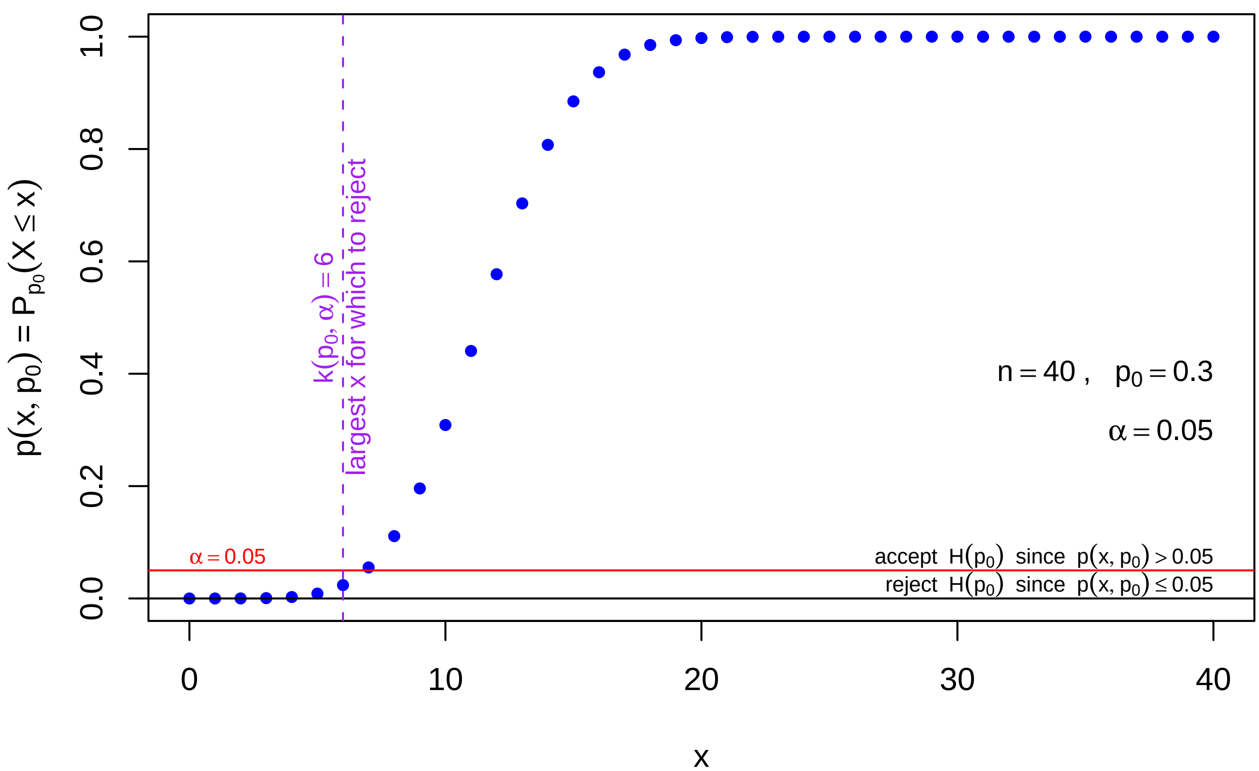


Figure 1: Binomial Distribution.



It is possible to establish a basic duality between testing hypotheses and confidence sets by defining such confidence sets as all values p_0 for which the observed value x of X results in an acceptance of the hypothesis H_{p_0}. Denote this confidence set as $C(x)$. We can also view the collection of such confidence sets (defined for each $x = 0, 1, \ldots, n$) as a random set $C(X)$ before having realized any observed value x for X. The following coverage probability property then holds for $C(X)$:

$$P_{p_0}\left(p_0 \in C(X)\right) = 1 - P_{p_0}\left(p_0 \notin C(X)\right) = 1 - P_{p_0}\left(X \leq k(p_0, \alpha)\right) \geq 1 - \alpha \quad \text{for all } p_0 \,, \tag{3}$$

i.e., this random confidence set has coverage probability $\geq \gamma = 1 - \alpha$ no matter what the value p_0 is. Thus we do not need to know p_0 and we can consider $C(X)$ as a $100\gamma\%$ confidence set for the unknown parameter p_0.

We point out that the inequality $\geq 1 - \alpha$ in (3) becomes an equality for some values of p_0, because there are values p_0 for which we have $P_{p_0}(X \leq k(p_0, \alpha)) = \alpha$. This follows since for any integer $k < n$ the probability $P_{p_0}(X \leq k)$ decreases continuously from 1 to 0 as p_0 increases from 0 to 1. Thus the confidence coefficient $\bar{\gamma}$ (the minimum coverage probability) of the confidence set $C(X)$ is indeed $1 - \alpha = \gamma$.

It remains to calculate the confidence set $C(x)$ explicitly for all possible values x that could be realized, i.e., for $x = 0, 1, \ldots, n$. According to the above definition of $C(x)$ it can be expressed as follows

$$C(x) = \{p_0 : p(x, p_0) = P_{p_0}(X \leq x) > \alpha\} \ .$$

Since $P_{p_0}(X \leq x)$ is strictly decreasing in p_0 for $x = 0, 1, \ldots, n-1$ we see that the confidence set $C(x)$ for such x values consists of the interval $[0, \hat{p}_U(\gamma, x, n))$, where $\hat{p}_U(\gamma, x, n) = \hat{p}_U(1 - \alpha, x, n)$ is the value p that solves

$$P_p(X \leq x) = \sum_{i=0}^{x} \binom{n}{i} p^i (1-p)^{n-i} = \alpha = 1 - \gamma \ . \tag{4}$$

Since $P_p(X \leq k)$ is continuous and strictly decreasing from 1 to 0 as p increases from 0 to 1, there is a unique value p satisfying equation (4). For $x = n$ equation (4) has no solution since $P_p(X \leq n) = 1$ for all p. However, according to the above definition of $C(x)$ we get $C(n) = [0, 1]$. We define $\hat{p}_U(\gamma, n, n) = 1$ in this special case.

Thus we can treat $\hat{p}_U(\gamma, X, n)$ as a $100\gamma\%$ upper confidence bound for p. Rather than finding $\hat{p}_U(\gamma, x, n)$ for $x < n$ by solving (4) for p we use the identity (2) involving the Beta distribution

$$P_p(X \leq x) = 1 - I_p(x+1, n-x) = 1 - \gamma \qquad \text{or} \qquad I_p(x+1, n-x) = \gamma \ .$$

We see that the solution p is the γ-quantile of the Beta distribution with parameters $x + 1$ and $n - x$. This value p can be obtained from Excel by invoking BETAINV($\gamma, x+1, n-x$) and in R or S-Plus by the command qbeta(γ,$x+1$,$n-x$).

As a check example use the case $k = 12$ and $n = 1600$ with $\gamma = .95$, then one gets $\hat{p}_U(.95, 12, 1600) =$ $\texttt{qbeta}(.95, 12 + 1, 1600 - 12) = .01212334$ as 95% upper bound for p.
Using (4) it is a simple exercise to show that the sequence of upper bounds is strictly increasing in x, i.e., $0 < \hat{p}_U(\gamma, 0, n) < \hat{p}_U(\gamma, 1, n) < \ldots < \hat{p}_U(\gamma, n-1, n) < \hat{p}_U(\gamma, n, n) = 1$. One immediate consequence of this is that $P_p(p < \hat{p}_U(\gamma, X, n)) \geq P_p(p < \hat{p}_U(\gamma, 0, n)) = 1$ for all $p < \hat{p}_U(\gamma, 0, n)$. Figure 2 shows the set of all $n + 1$ upper bounds for $\gamma = .95$ and $n = 100$ superimposed on the corresponding estimates $\hat{p}(x) = x/n$.

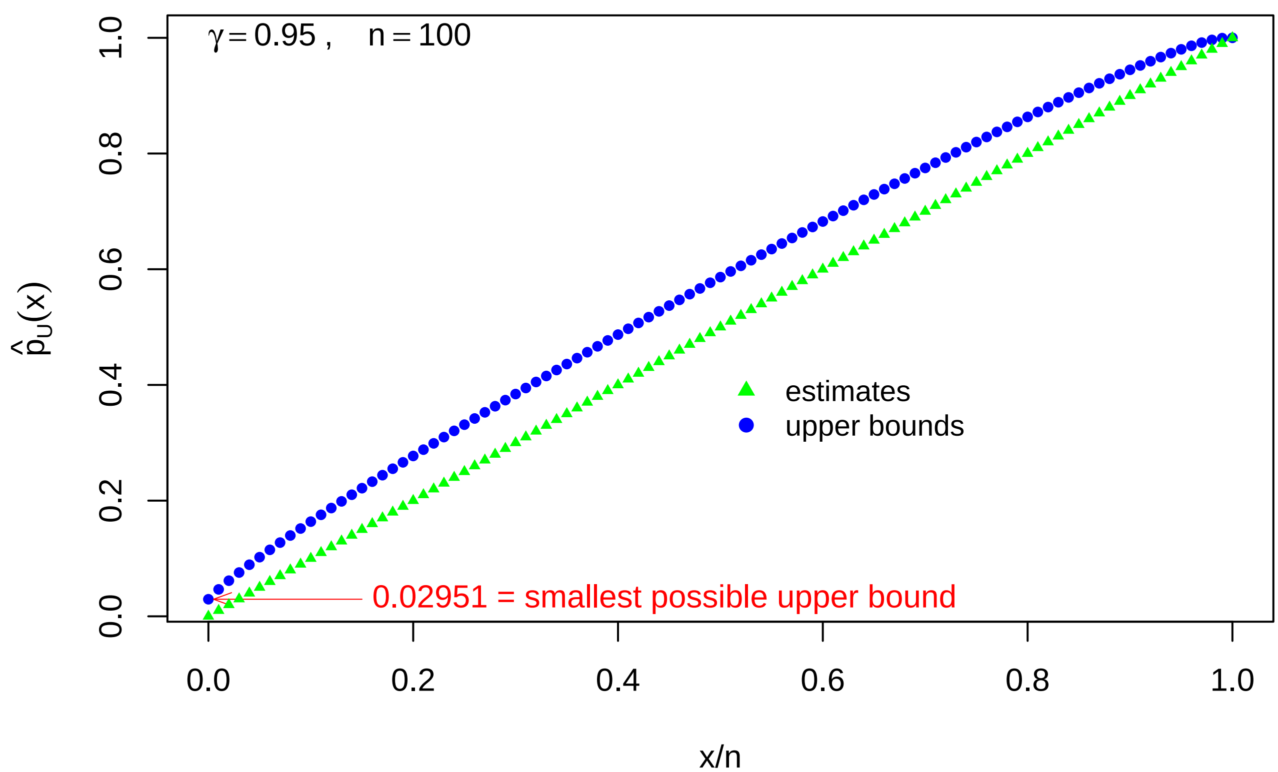


Figure 2: Binomial Upper Bounds in Relation to the Corresponding Estimates.

2.4.1 The Confidence Coefficient $\bar{\gamma}$

The confidence coefficient of the upper bound is defined as

$$\bar{\gamma} = \inf_p \{P_p(p \in C(X))\} = \inf_p \{P_p(p < \hat{p}_U(\gamma, X, n))\} \ .$$

The confidence coefficient is $\bar{\gamma} = \gamma = 1 - \alpha$ since $P_p(p \in C(X)) = \gamma$ for some p, or $P_p(p \notin C(X)) = P_p(X \leq k(p, \alpha)) = 1 - \gamma = \alpha$ for some p.
Proof: Recall that for $x = 0, 1, \ldots, n-1$ the value $p = \hat{p}_U(\gamma, x, n)$ solved

$$P_p(X \leq x) = \alpha$$

Thus for $p_i = \hat{p}_U(\gamma, i, n)$, $i = 0, 1, \ldots, n-1$, with $k(p_i, \alpha) = i$ we have

$$P_{p_i}(X \leq k(p_i, \alpha)) = P_{p_i}(X \leq i) = \alpha \ , \tag{5}$$

i.e., the infimum above is indeed attained at $p = p_0, p_1, \ldots, p_{n-1}$.

2.4.2 Special Cases

For $x = 0$ and $x = n-1$ the upper bounds defined by (4) can be expressed explicitly as

$$\hat{p}_U(\gamma, 0, n) = 1 - (1-\gamma)^{1/n} \qquad \text{and} \qquad \hat{p}_U(\gamma, n-1, n) = \gamma^{1/n} \ .$$

Obviously, the bound for $x = 0$ is of more practical interest than the bound in the case of $x = n-1$. For $\gamma = .95$ and approximating $\log(.05) = -2.995732$ by -3 the upper bound for $x = 0$ becomes

$$\hat{p}_U(.95, 0, n) = 1 - (.05)^{1/n} = 1 - \exp\left[\frac{\log(.05)}{n}\right] \approx 1 - \exp(-3/n) \approx \frac{3}{n} \ ,$$

which is sometimes referred to as the **Rule of Three**, because of its mnemonic simplicity, see van Belle (2002). Here the last approximation is valid only for large n, say $n \geq 100$.
One common application of this Rule of Three is to establish the appropriate sample size n when it is desired to establish with 95% confidence that p is bounded above by $3/n$. For this to work one has to be quite sure to see 0 "successes" (bad or undesirable outcomes) in n trials. For example, if one wants to establish $.001 = 3/n$ as 95% upper bound for p one should conduct $n = 3/.001 = 3000$ trials and hope for the best, namely 0 events.

2.4.3 Side Comment on Treatment of $X = 0$

When one observes $X = 0$ successes in n trials, especially when n is large, one is still not inclined to estimate the success probability p by $\widehat{p}(0) = 0/n = 0$, since that is a very strong statement. When $p = 0$ then we will never see a success in however many trials. Since $p = 0$ is such a strong statement, but one still thinks that p is likely to be very small if $X = 0$ in a large number n of trials, one common practice to get out of this dilemma is to "conservatively" pretend that the first success is just around the corner, i.e., happens on the next trial. With that one would estimate p by $\tilde{p} = 1/(n+1)$ which is small but not zero. There are other (and statistically better) rationales for justifying $\tilde{p} = 1/(n+1)$ as an estimate of p but we won't enter into that here, since they have no bearing on the issue of confidence that some construe out of the above "conservative" step.

As an estimate of the true value of p the use of $\tilde{p}$ is not entirely unreasonable but somewhat conservative. It is however quite different from our 95% upper confidence bound of $3/n$, namely by roughly a factor of 3. One could ask: what is the actual confidence associated with $\tilde{p}$?

We can assess this by solving $1 - (1-\gamma)^{1/n} = 1/(n+1)$ for γ, which leads to

$$\gamma = 1 - \left(\frac{n}{n+1}\right)^n = 1 - \left(1 - \frac{1}{n+1}\right)^n \approx 1 - \exp\left(-\frac{n}{n+1}\right) \approx 1 - \exp(-1) = .6321 \, .$$

This means that we can treat $\tilde{p} = 1/(n+1)$ only as a 63.21% upper confidence bound for p. This is substantially lower than the 95% which led to the factor 3 in $3/n$ as upper bound.

2.4.4 Closed Intervals or Open Intervals?

So far we have given the confidence sets in terms of the right open intervals $[0, \hat{p}_U(\gamma, x, n))$ with the property that

$$P_p(p < \hat{p}_U(\gamma, X, n)) \geq \gamma \qquad \text{and} \qquad \inf_p \{P_p(p < \hat{p}_U(\gamma, X, n))\} = \gamma \, ,$$

where the value γ is achieved at $p = \hat{p}_U(\gamma, x, n)$ for $x = 0, 1, \ldots, n-1$.

The question arises quite naturally: why not take instead the right closed interval $[0, \hat{p}_U(\gamma, x, n)]$, again with property

$$P_p(p \leq \hat{p}_U(\gamma, X, n)) \geq \gamma \qquad \text{and} \qquad \inf_p \{P_p(p \leq \hat{p}_U(\gamma, X, n))\} = \gamma \, .$$

The difference is that the infimum is not achieved at any p.

Whether we use the closed interval or not, the definition of $\hat{p}_U(\gamma, x, n)$ stays the same. Any value p_0 equal to it or greater would lead to rejection of $H(p_0)$ when tested against $A(p_0) : p < p_0$. By closing the interval we add a single p to it, namely $p = \hat{p}_U(\gamma, x, n)$, that is not acceptable.

2.4.5 Coverage Probability Continuity Properties

Let $p_i = \hat{p}(\gamma, i, n)$ for $i = 0, 1, \ldots, n$. Recall $0 < p_0 < p_1 < \ldots < p_n = 1$.
For $p \in [p_{i-1}, p_i)$ the set $A = \{j : p < \hat{p}(\gamma, j, n)\} = [i, n]$ does not change.

$$\Longrightarrow \quad P_p(X \in A) = P_p(X \geq i) \quad \text{increases continuously in } p \text{ over } [p_{i-1}, p_i)\ .$$

When $p = p_i$ the set A loses the value $j = i$ and $P_p(p < \hat{p}(\gamma, X, n))$ drops by $P_{p_i}(X = i)$ to $P_{p_i}(X \geq i + 1) = 1 - P_{p_i}(X \leq i) = 1 - \alpha = \gamma$ see (5), and so on. This behavior is illustrated in Figure 3.

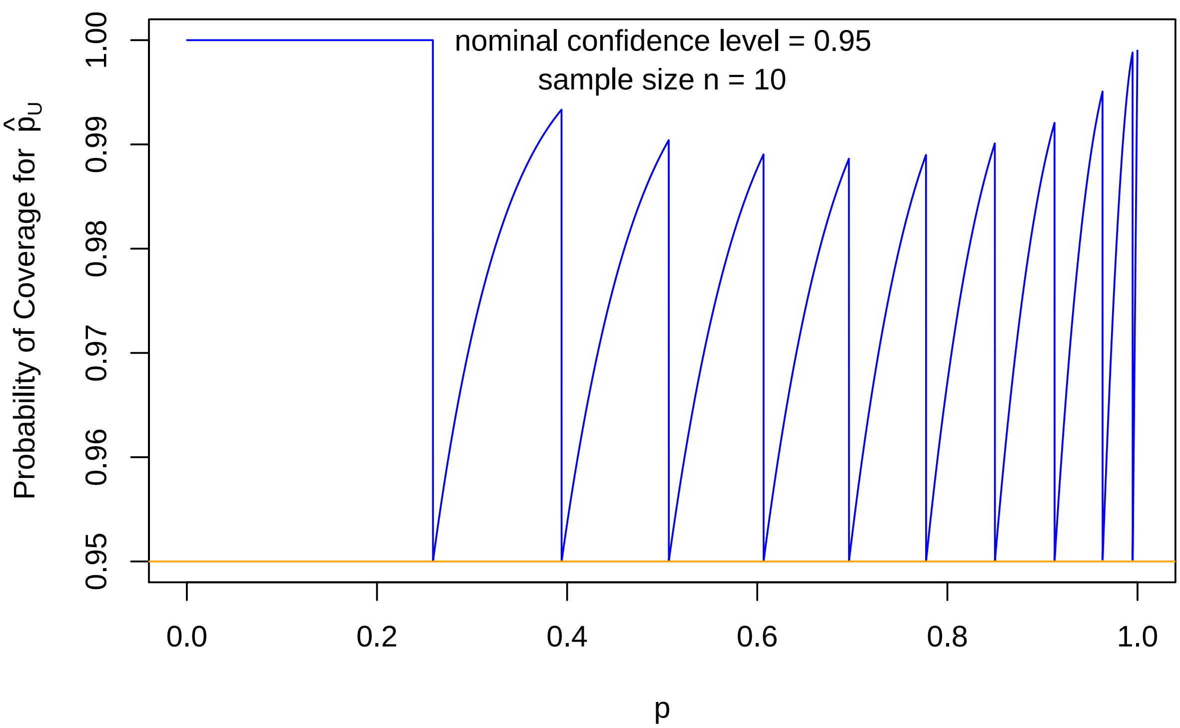


Figure 3: Coverage Probability Behavior of Upper Bound

For $p \in (p_{i-1}, p_i]$ the set $B = \{j : p \le \hat{p}(\gamma, j, n)\} = [i, n]$ does not change.

$$\Longrightarrow \quad P_p(X \in B) = P_p(X \ge i) \quad \text{increases continuously in } p \text{ over } (p_{i-1}, p_i] \ .$$

As $p \searrow p_{i-1}$ we have $P_p(X \ge i) \searrow P_{p_{i-1}}(X \ge i) = 1 - P_{p_{i-1}}(X \le i-1) = 1 - \alpha = \gamma$ see (5), but at $p = p_{i-1}$ we have a jump since

$$P_{p_{i-1}}(p_{i-1} \le \hat{p}(\gamma, X, n)) = P_{p_{i-1}}(X \ge i-1) > P_{p_{i-1}}(X \ge i) \ ,$$

i.e., γ is never attained.

2.5 Lower Bounds for p

Large values of X can be viewed as evidence against the hypothesis $H(p_0) : p = p_0$ when testing it against the alternative $A(p_0) : p > p_0$. Again one can carry out the test in terms of the observed p-value $p(x, p_0) = P_{p_0}(X \ge x)$ by rejecting $H(p_0)$ at level α whenever $p(x, p_0) \le \alpha$ and accepting it otherwise. Invoking again the duality between testing and confidence sets we get as confidence set

$$C(x) = \{p_0 : p(x, p_0) = P_{p_0}(X \ge x) > \alpha\} \ .$$

Since $P_{p_0}(X \ge x)$ is strictly increasing and continuous in p_0 we see that for $x = 1, 2, \ldots, n$ this confidence set $C(x)$ coincides with the interval $(\hat{p}_L(\gamma, x, n), 1]$, where $\hat{p}_L(\gamma, x, n)$ is that value of p that solves

$$P_p(X \ge x) = \sum_{i=x}^{n} \binom{n}{i} p^i (1-p)^{n-i} = \alpha = 1 - \gamma \ . \tag{6}$$

Invoking the identity (1) this value p can be obtained by solving

$$I_p(x, n - x + 1) = \alpha = 1 - \gamma$$

for p, i.e., it is the α-quantile of a Beta distribution with parameters x and $n - x + 1$. For $x = 0$ we have $P_{p_0}(X \ge 0) = 1 > \alpha$ and we cannot use (6), but the original definition of $C(x)$ leads to $C(0) = [0, 1]$ and we define $\hat{p}_L(\gamma, 0, n) = 0$ in that case.

In Excel we can obtain $\hat{p}_L(\gamma, x, n)$ for $x > 0$ by invoking `BETAINV`$(1-\gamma, x, n-x+1)$ and from R or S-Plus by the command `qbeta(`$1-\gamma$`,`x`,`$n-x+1$`)`.

As a check example take $x = 4$ and $n = 500$ with $\gamma = .95$, then one gets $\hat{p}_L(.95, 4, 500) = .002737$ as 95% lower bound for p.

Using (6) it is a simple exercise to show that the sequence of lower bounds is strictly increasing in x, i.e., $0 = \hat{p}_L(\gamma, 0, n) < \hat{p}_L(\gamma, 1, n) < \ldots < \hat{p}_L(\gamma, n-1, n) < \hat{p}_L(\gamma, n, n) < 1$. Again it follows that $P_p(\hat{p}_L(\gamma, X, n) < p) \ge P_p(\hat{p}_L(\gamma, n, n) < p) = 1$ for $p > \hat{p}_L(\gamma, n, n)$.

As in the case of the upper bound one may prefer to use the closed confidence set $[\hat{p}_L(\gamma, x, n), 1]$, with the corresponding commentary. In particular, the lower endpoint $p_0 = \hat{p}_L(\gamma, x, n)$ does not represent an acceptable hypothesis value, while all other interval points are acceptable values.

2.5.1 Special Cases

For $x = 1$ and $x = n$ the lower bounds defined by (6) can be expressed explicitly as

$$\hat{p}_L(\gamma, 1, n) = 1 - \gamma^{1/n} \qquad \text{and} \qquad \hat{p}_L(\gamma, n, n) = (1 - \gamma)^{1/n}$$

For obvious reasons the explicit lower bound in the case of $x = n$ is of more practical interest than the lower bound for $x = 1$.
For $\gamma = .95$ and $x = n$ the lower bound becomes

$$\hat{p}_L(.95, n, n) = (1 - .95)^{1/n} \approx \exp(-3/n) \approx 1 - \frac{3}{n} ,$$

a dual instance of the **Rule of Three**. Here the last approximation is only valid for large n. This duality should not surprise since switching the role of successes and failures with concomitant switch of p and $1 - p$ turns upper bounds for p into lower bounds for p and vice versa.

2.5.2 Side Comment on Treatment of $X = n$

When observing $X = n$ successes in n trials, especially when n is large, one is still not inclined to estimate p by $\hat{p} = n/n = 1$, because of the consequences of such a strong statement. Because of the just mentioned duality when switching successes with failures we will not repeat the parallel discussion of Section 2.4.3.

2.6 Confidence Intervals for p

As outlined in the Introduction, such lower and upper confidence bounds, each with respective confidence coefficient $1 - \alpha/2$, can be used simultaneously as a $100(1 - \alpha)\%$ confidence interval $(\hat{p}_L(1-\alpha/2, X, n), \hat{p}_U(1-\alpha/2, X, n))$, provided we show $P_p(\hat{p}_L(1-\alpha/2, X, n) < \hat{p}_U(1-\alpha/2, X, n)) = 1$ for any p.
We demonstrate this by assuming that we have $\hat{p}_U(1 - \alpha/2, x, n) \le \hat{p}_L(1 - \alpha/2, x, n)$ for some x and thus $\hat{p}_U(1 - \alpha/2, x, n) \le p_0 \le \hat{p}_L(1 - \alpha/2, x, n)$ for some p_0 and some x. This means that the p-values from the respective hypothesis tests linked to the one-sided bounds would cause us to reject the hypothesis $H(p_0)$ in each case, i.e.,

$$P_{p_0}(X \le x) \le \alpha/2 \qquad \text{and} \qquad P_{p_0}(X \ge x) \le \alpha/2$$

and by adding those two inequalities we get

$$1 + P_{p_0}(X = x) \le \alpha < 1 , \qquad \text{i.e., a contradiction.}$$

Using $\hat{p}_L(1-\alpha/2, x, n) < \hat{p}_U(1-\alpha/2, x, n)$ for any p and any $x = 0, 1, \ldots, n$ we have

$$\begin{aligned}
&P_p(\hat{p}_L(1-\alpha/2, X, n) < p < \hat{p}_U(1-\alpha/2, X, n)) \\
&= 1 - P_p(p \le \hat{p}_L(1-\alpha/2, X, n) \cup \hat{p}_U(1-\alpha/2, X, n) \le p) \\
&= 1 - [P_p(p \le \hat{p}_L(1-\alpha/2, X, n)) + P_p(\hat{p}_U(1-\alpha/2, X, n) \le p)] \\
&\ge 1 - [\alpha/2 + \alpha/2] = 1 - \alpha\ .
\end{aligned}$$

2.7 Coverage

We examine here briefly the issue of coverage, namely, what is the actual probability for the upper or lower bounds or intervals to be above or below p or to contain p, respectively, for the various values of p. Figures 4 and 5 show the results in the example case of $n = 100$ and nominal confidence level $\gamma = .95$ computed over a very fine grid of p values.

These plots were produced in R by calculating in the case of the lower bound coverage probability

$$P\left(\hat{p}_L(\gamma, X, n) \le p\right) = (1-p)^n + \sum_{x=1}^{n} I_{\{\texttt{qbeta}(1-\gamma, x, n-x+1) \le p\}} \binom{n}{x} p^x (1-p)^{n-x}\ ,$$

where $I_A = 1$ whenever A is true and $I_A = 0$ otherwise. For the upper bound one computes the coverage probabilities via

$$P\left(\hat{p}_U(\gamma, X, n) \ge p\right) = \sum_{x=0}^{n-1} I_{\{\texttt{qbeta}(\gamma, x+1, n-x) \ge p\}} \binom{n}{x} p^x (1-p)^{n-x} + p^n\ ,$$

while for the confidence interval one calculates the coverage probability as

$$\begin{aligned}
&P\left(\hat{p}_L((1+\gamma)/2, X, n) \le p\ \cap\ \hat{p}_U((1+\gamma)/2, X, n) \ge p\right) \\
&= \sum_{x=1}^{n-1} I_{\{\texttt{qbeta}((1-\gamma)/2, x, n-x+1) \le p\ \cap\ \texttt{qbeta}((1+\gamma)/2, x+1, n-x) \ge p\}} \binom{n}{x} p^x (1-p)^{n-x} \\
&\qquad + (1-p)^n I_{\{\texttt{qbeta}((1+\gamma)/2, 1, n) \ge p\}} + p^n I_{\{\texttt{qbeta}((1-\gamma)/2, n, 1) \le p\}}
\end{aligned}$$

Note that in the above calculations the closed form of the one-sided bounds or confidence intervals was used. Close examination of the plots shows that the minimum coverage probability seems to go as low as the target $\gamma = .95$ for the one-sided bounds. The value .95 appears to be (almost) achieved at every dip, except for p near 0 or 1. The dips correspond to all possible confidence bound values. As discussed previously, for the closed one-sided bounds the coverage probability should get arbitrarily close to .95 at all of these dips. The fact that .95 does not appear to be fully approximated for p near 0 or 1 is due to the grid over which these coverage probabilities were evaluated.

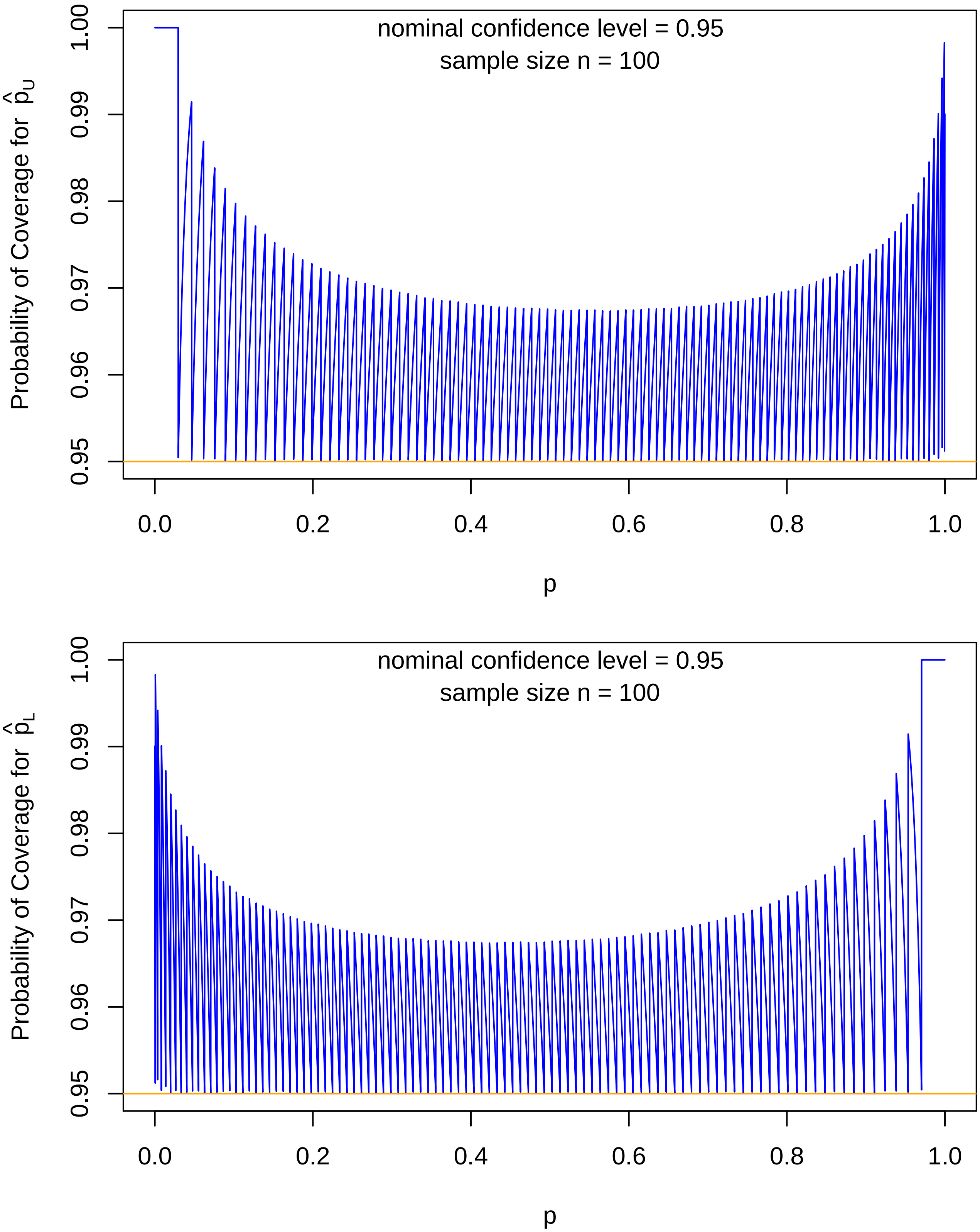


Figure 4: Binomial Coverage Probabilities for Upper and Lower Confidence Bounds

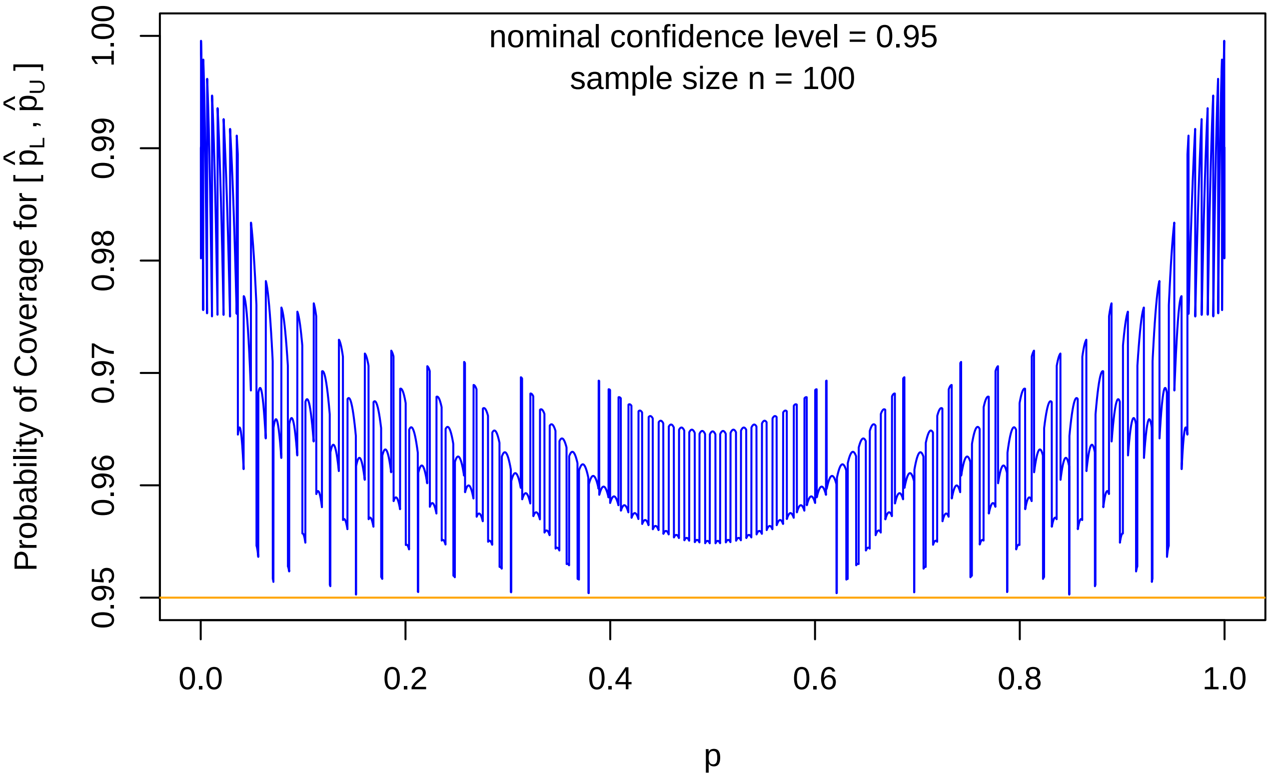
For the confidence interval the coverage probability appears to reach not quite as far down as .95, see Figure 5. To illustrate this more clearly we also show the corresponding plot for $n = 11$ in the bottom part of Figure 5. This conservative coverage is due to the fact that the supremum of a sum is typically smaller than the sum of the suprema over the respective summands, because the point where one summand comes close to its supremum does not necessarily coincide with the point where the other summand comes close to its supremum, i.e.,

$$
\begin{aligned}
&\inf_p P_p(\hat{p}_L(1-\alpha/2, X, n) \le p \le \hat{p}_U(1-\alpha/2, X, n)) \\
&\qquad = 1 - \sup_p P_p(p < \hat{p}_L(1-\alpha/2, X, n) \cup \hat{p}_U(1-\alpha/2, X, n) < p) \\
&\qquad = 1 - \sup_p \{P_p(p < \hat{p}_L(1-\alpha/2, X, n)) + P_p(\hat{p}_U(1-\alpha/2, X, n) < p)\} \\
&\qquad \ge 1 - \left\{\sup_p P_p(p < \hat{p}_L(1-\alpha/2, X, n)) + \sup_p P_p(\hat{p}_U(1-\alpha/2, X, n) < p)\right\} \\
&\qquad = 1 - \{\alpha/2 + \alpha/2\} = 1 - \alpha\ .
\end{aligned}
$$

Here the inequality $\ge$, for reasons explained above, typically takes the strict form $>$.
Furthermore, for p close to 0 or 1 the coverage probability rises effectively to $1 - \alpha/2 = .975$. This is a consequence of the previously noted coverage with probability 1, for upper bounds for $p < \hat{p}_U(1-\alpha/2, 0, n)$ and for lower bounds for $p > \hat{p}_L(1-\alpha/2, n, n)$, so that

$$
\begin{aligned}
&P_p(\hat{p}_L(1-\alpha/2, X, n) \le p \le \hat{p}_U(1-\alpha/2, X, n)) \\
&\qquad = 1 - P_p(\hat{p}_L(1-\alpha/2, X, n) > p) - P_p(\hat{p}_U(1-\alpha/2, X, n) < p) \\
&\qquad = 1 - P_p(\hat{p}_L(1-\alpha/2, X, n) > p) \ge 1 - \alpha/2 \qquad \text{for } p < \hat{p}_U(1-\alpha/2, 0, n) \\
&\qquad = 1 - P_p(\hat{p}_U(1-\alpha/2, X, n) < p) \ge 1 - \alpha/2 \qquad \text{for } p > \hat{p}_L(1-\alpha/2, n, n)
\end{aligned}
$$

It would appear that for small p there is little sense in working with confidence intervals, allocating half of the miss probability ($\alpha/2$) to the lower bound and raising the confidence level of the righthand interval endpoint to $1-\alpha/2$. This leads to a conservative assessment of the smallness of p. If small p are of main concern (typical for risk situations) one should focus on upper bounds for p, i.e., use $\hat{p}_U(1-\alpha, X, n)$ instead of the higher righthand endpoint $\hat{p}_U(1-\alpha/2, X, n)$ of the interval.



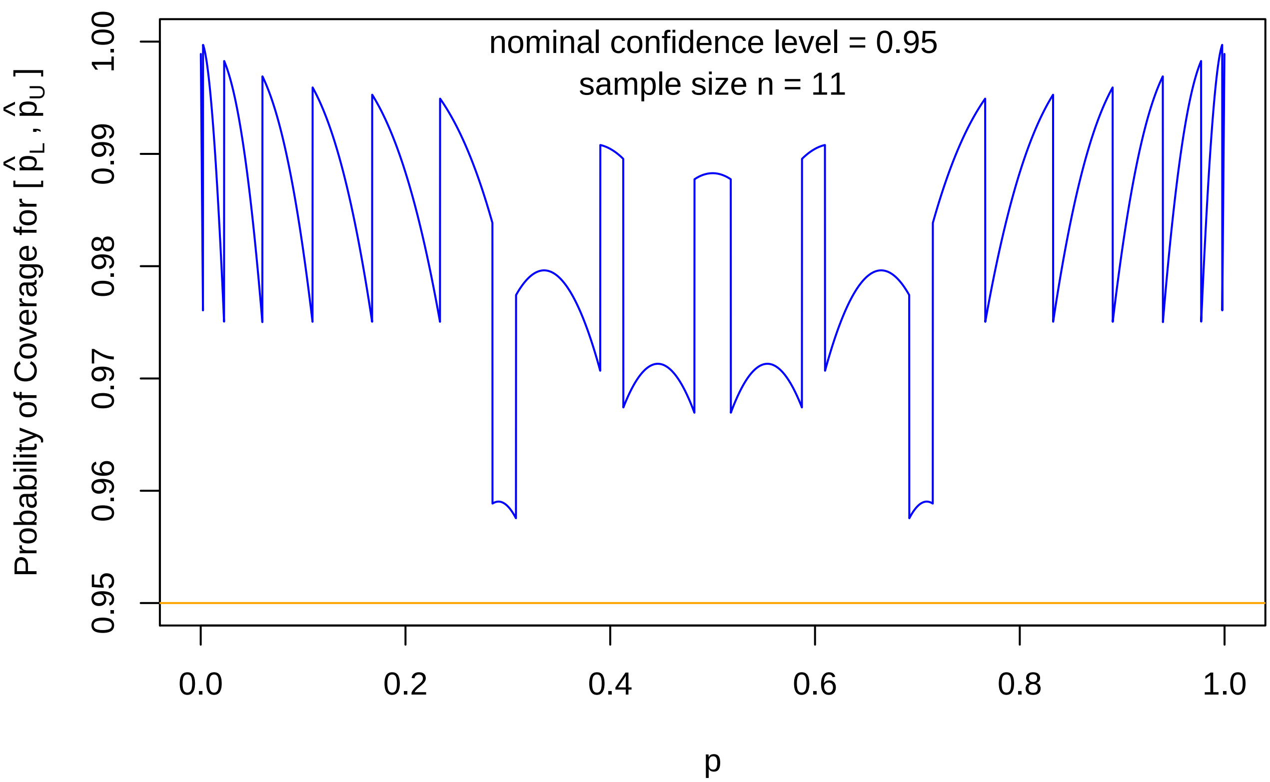


Figure 5: Binomial Confidence Interval Coverage Probabilities

To get some sense of what might be given away in using the latter rather than the former we can ask: How large a sample size m is needed to equate $\hat{p}_U(1-\alpha/2, 0, m) = \hat{p}_U(1-\alpha, 0, n)$ when in fact we see 0 events in either case. This leads to the following equation

$$1 - (\alpha/2)^{1/m} = 1 - \alpha^{1/n} \qquad \text{or} \qquad m = n\,\frac{\log(\alpha) - \log(2)}{\log(\alpha)} \ .$$

For $\alpha = .05$ this becomes $m = 1.2314\, n$, i.e., a 23% increase in sample size over n with the additional stipulation that we also see no event during the additional trials.

A corresponding argument can be made for using the lower bound in place of an interval when p near 1 is of primary concern. This arises typically in reliability contexts where the chance of successful operation is desired to be high.

2.8 Simulated Upper Bounds

To get a better understanding of the nature of confidence bounds we simulated 1000 instances of X from a binomial distribution $\mathsf{Bin}(n, p)$ for $n = 100$ and several different values of p. Figures 6-9 show the resulting upper confidence bounds in relation to the respective true values of p.

The $n+1$ possible upper bounds are solely determined by the value n (recall $\texttt{qbeta}(\gamma, \texttt{x}+\texttt{1}, \texttt{n}-\texttt{x})$ for $x = 0, 1, \ldots, n-1$ and $\hat{p}_U(\gamma, n, n) = 1$) and as the true p changes the distribution of X changes. As p increases we tend to see more high values of X. Note that Figure 6 shows no upper bounds below the target p, while Figure 7 shows 3.8% of the upper bounds barely below the target p. We would have expected 5% of the bounds below p with $\gamma = .95$, but due to the randomness of the particular set of 1000 generated X values this should not surprise us.

As we raise the value of p in Figures 8 and 9, but still staying below the next possible upper bound value, we see that the percentage of upper bounds below p decreases. That is precisely due to the fact that we see fewer cases with $X = 0$ as p increases.

While in Figures 6-9 we were operating with values of p that we controlled, in the real life practical situation we do not know p and it is not clear whether we were lucky enough ($\geq 95\%$ of the time) to have gotten an upper bound above the unknown p or whether we were unlucky ($\leq 5\%$ of the time). As Myles Hollander once put it: “Statistics means never having to say you’re certain.”

2.9 Elementary Classical Bounds Based on Normal Approximation

We point out here that the often and popularly given upper confidence bound for p, based on the estimate $\hat{p} = X/n$ and its estimated approximate normal distribution, is

$$\hat{p} + z_\gamma \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \ , \qquad \text{with } z_\gamma \text{ being the standard normal } \gamma\text{-quantile.}$$

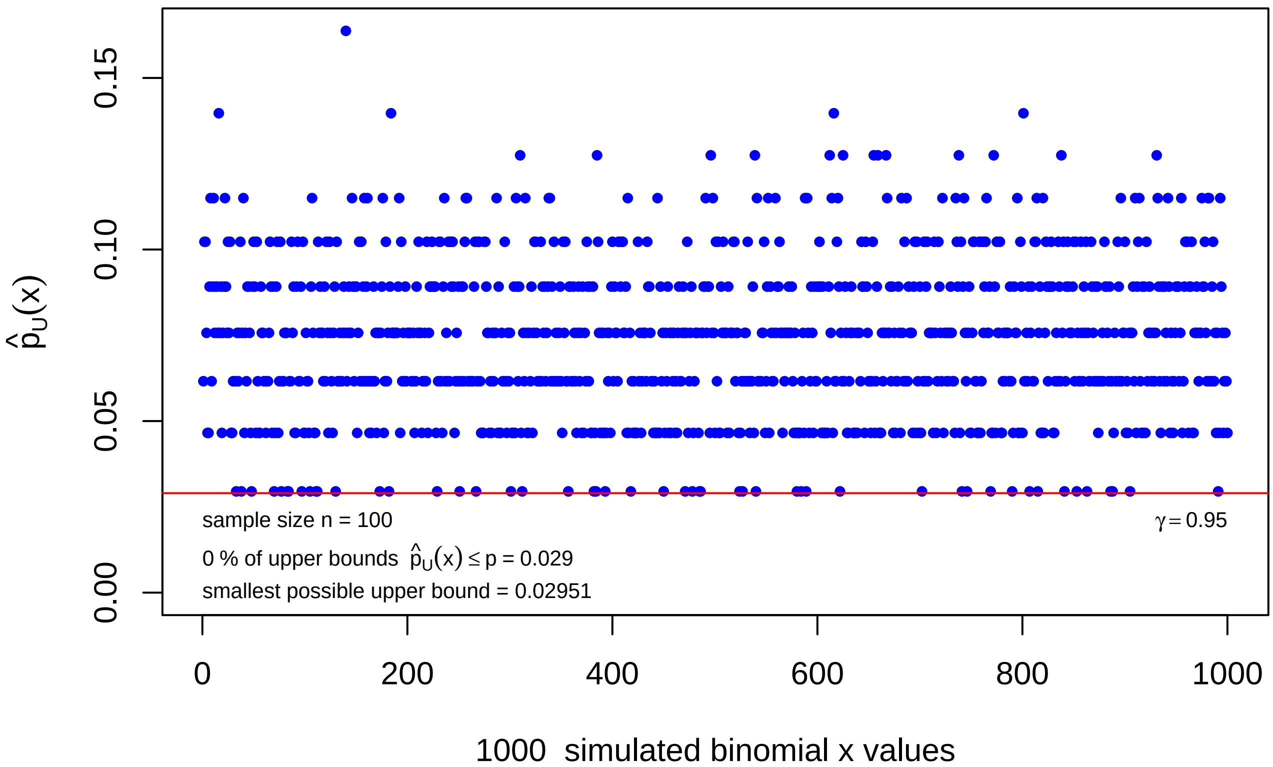


Figure 6: 1000 Simulated Binomial Upper Bounds for $p = .029$.

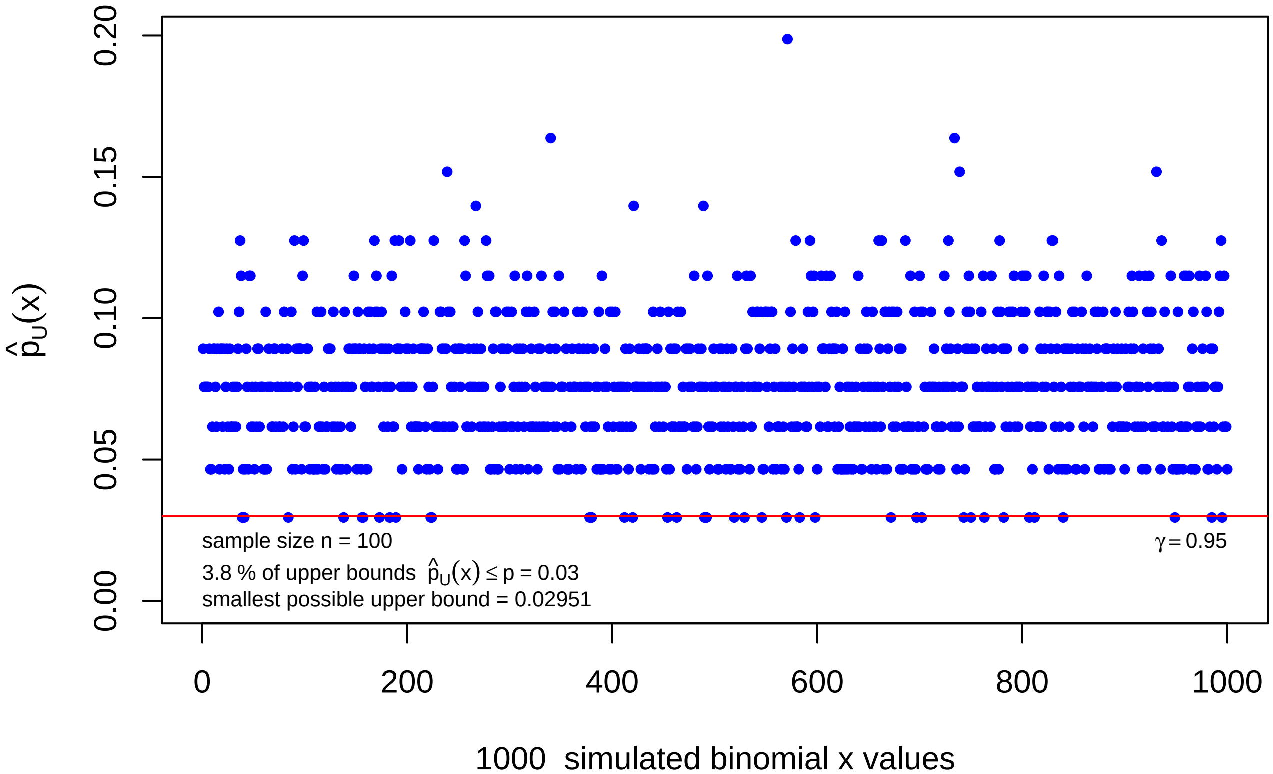


Figure 7: 1000 Simulated Binomial Upper Bounds for $p = .030$.

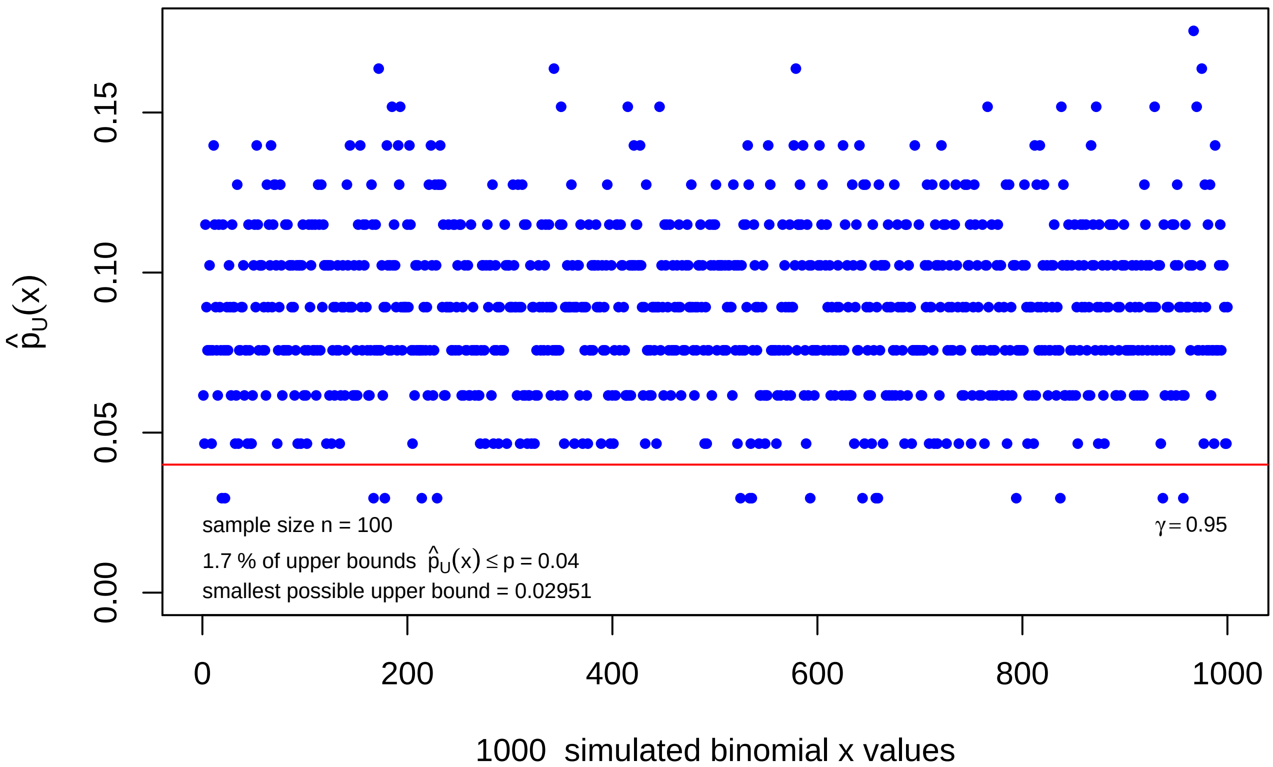


Figure 8: 1000 Simulated Binomial Upper Bounds for $p = .040$.

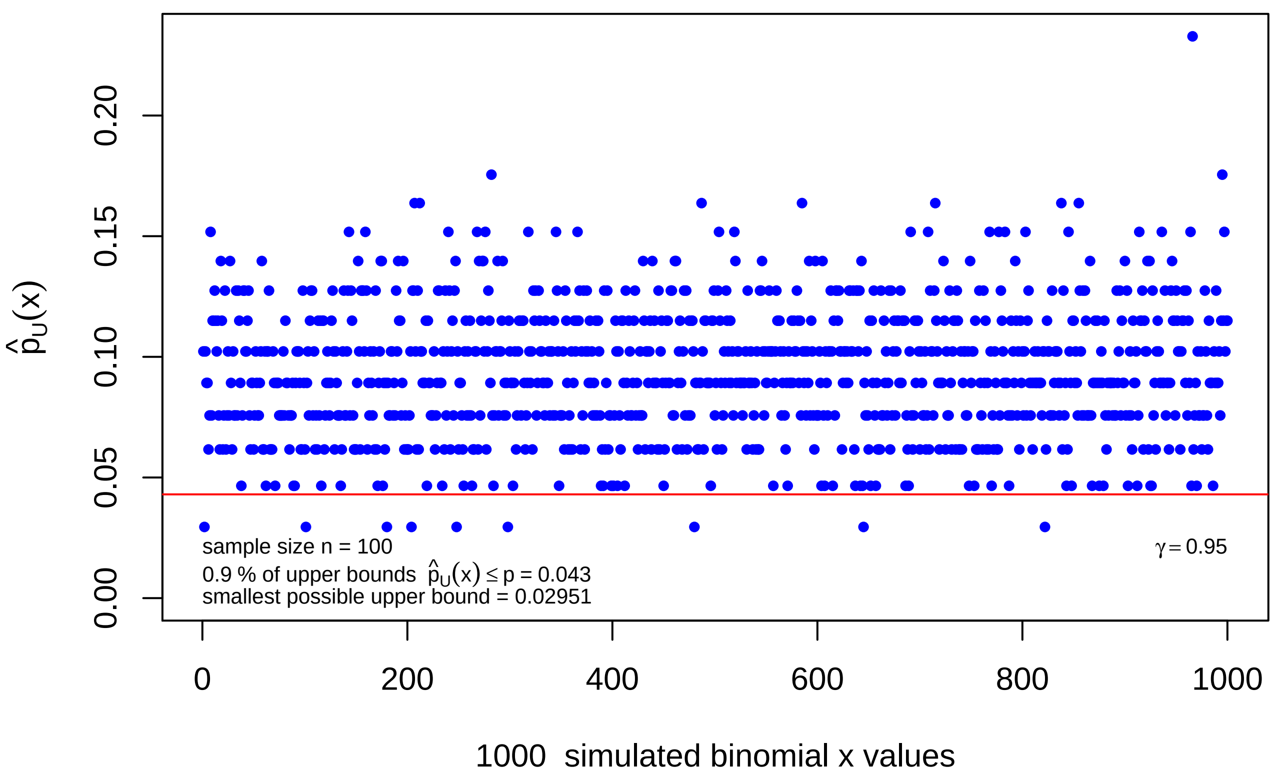


Figure 9: 1000 Simulated Binomial Upper Bounds for $p = .043$.

This upper bound has minimum coverage zero over all $p \in (0,1)$ and not the intended confidence level γ. This can be seen from the fact that for $p > 0$ very close to zero the chance that $\hat{p} = 0$ gets closer and closer to one. But $\hat{p} = 0$ implies that the above upper bound is zero as well and can thus no longer be $\geq p(> 0)$, with probability closer and closer to one. Hence the minimum coverage probability is zero.

The reason for this failure of such a popularly cited procedure is that it is based on the approximate normality of $(\hat{p} - p)/\sqrt{\hat{p}(1-\hat{p})/n}$. This approximation becomes problematic when $p \approx 0$ or $p \approx 1$, in which case a larger and larger sample size n is required to make this approximation reasonable. However, even such a large n will not cover all contingencies w.r.t. $p \in (0,1)$. As $p \searrow 0$ or $p \nearrow 1$ this would lead to the requirement that $n \to \infty$, which is hardly practical. A similar case can be made in the case of lower bounds and intervals.

Figure 10 shows the coverage probability behavior of these elementary upper bounds. It makes quite clear that these upper bounds are reasonable only in the vicinity of $p = .5$.

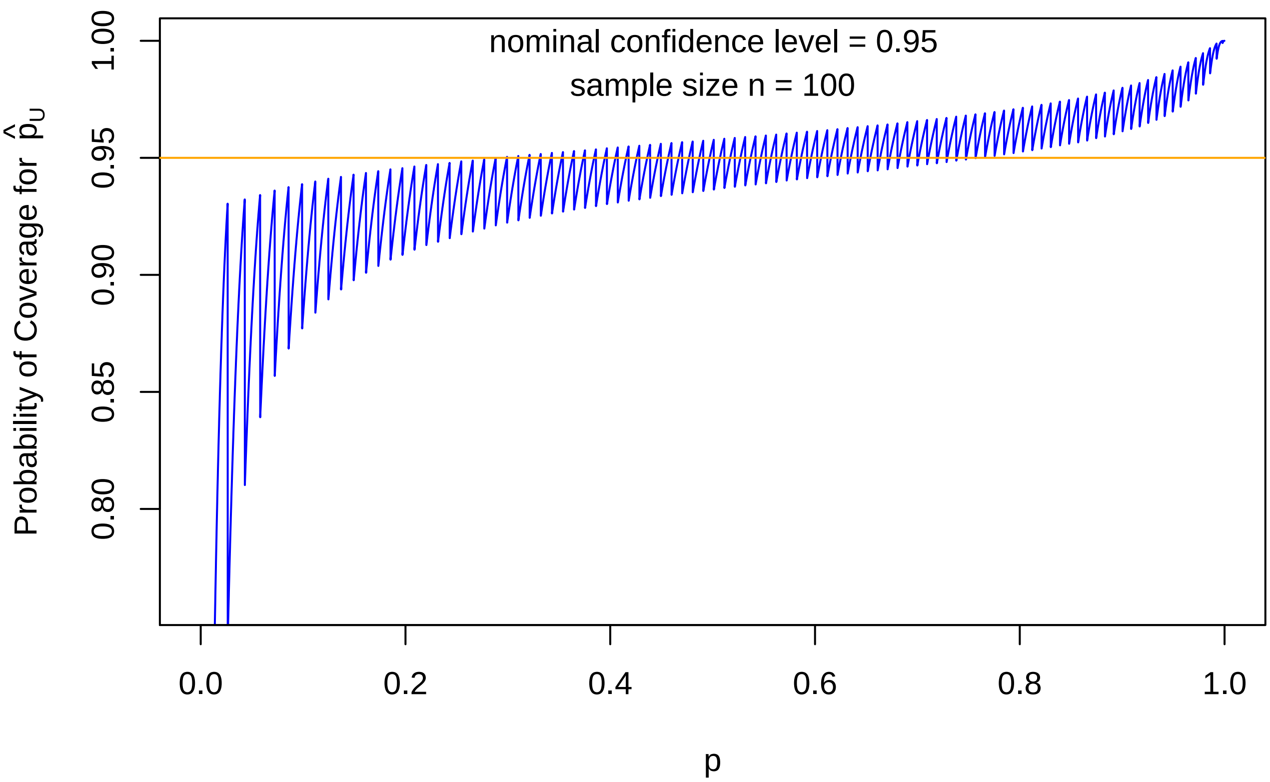


Figure 10: Coverage Probabilities for Binomial Elementary Upper Confidence Bounds

2.10 Score Confidence Bounds

Agresti and Coull (1998) make a case for using score confidence intervals based on inverting score tests. In the case of upper confidence bounds this simply amounts to solving equation (4) for p by approximating the left hand side of (4) using the normal distribution, i.e., solve

$$\alpha = P_p(X \leq x) \approx \Phi\left(\frac{x - np}{\sqrt{np(1-p)}}\right) ,$$

where Φ is the standard normal cdf. This approximation is done without using a continuity correction, which would amount to replacing x by $x + .5$. The result is a rather unwieldy expression that does not hold much mnemonic appeal, namely

$$\tilde{p}_U(\gamma, x, n) = \frac{z^2/(2n) + \hat{p}}{1 + z^2/n} + \frac{z}{1 + z^2/n}\sqrt{\frac{z^2}{4n^2} + \frac{\hat{p}(1-\hat{p})}{n}} ,$$

where z is the γ-quantile of the standard normal distribution and $\hat{p} = x/n$ is the classical estimate of p. The corresponding lower bounds use instead the $(1-\gamma)$-quantile or equivalently change the sign in front of the square root in the above upper bound expression. For intervals lower and upper bounds are combined with respective $(1+\gamma)/2$ confidence levels.

Figure 11 shows the coverage behavior of the score upper bounds. The case made by Agresti and Coull in favor of these score bounds is that their average coverage probability is closer to the target nominal. However, the minimum coverage probability can be quite a bit lower, although it occurs for p near 1, which is of not much concern in applications. When comparing the zigzag behavior in this plot with that in Figure 4 it is difficult to see much of a difference in the range of amplitudes. It would appear that lowering the confidence level of the Clopper-Pearson bounds appropriately would bring the average coverage to the same level. In my opinion the criterion of average coverage is not appealing when trying to bound risks. It has too much the flavor of a statistician with his head in the oven and his feet in an ice bucket, but feeling fine on average.

Out of curiosity we also computed the coverage probabilities for score upper bounds when using in their derivation a continuity correction in the normal approximation. Such upper bounds are basically the same as given previously, except that $\hat{p}$ is replaced by $\tilde{p} = (x + .5)/n$ throughout. Figure 12 shows the result. While these bounds appear more conservative than the bounds without such a correction, they still do not guarantee the desired confidence level.

In all fairness to Agresti and Coull we point out that they made their case in the context of intervals and not for one-sided bounds. Figure 13 shows the coverage probabilities for such intervals (without using a continuity correction). While the average coverage behavior is as advocated, the coverage probability deteriorates significantly near 0 or 1.

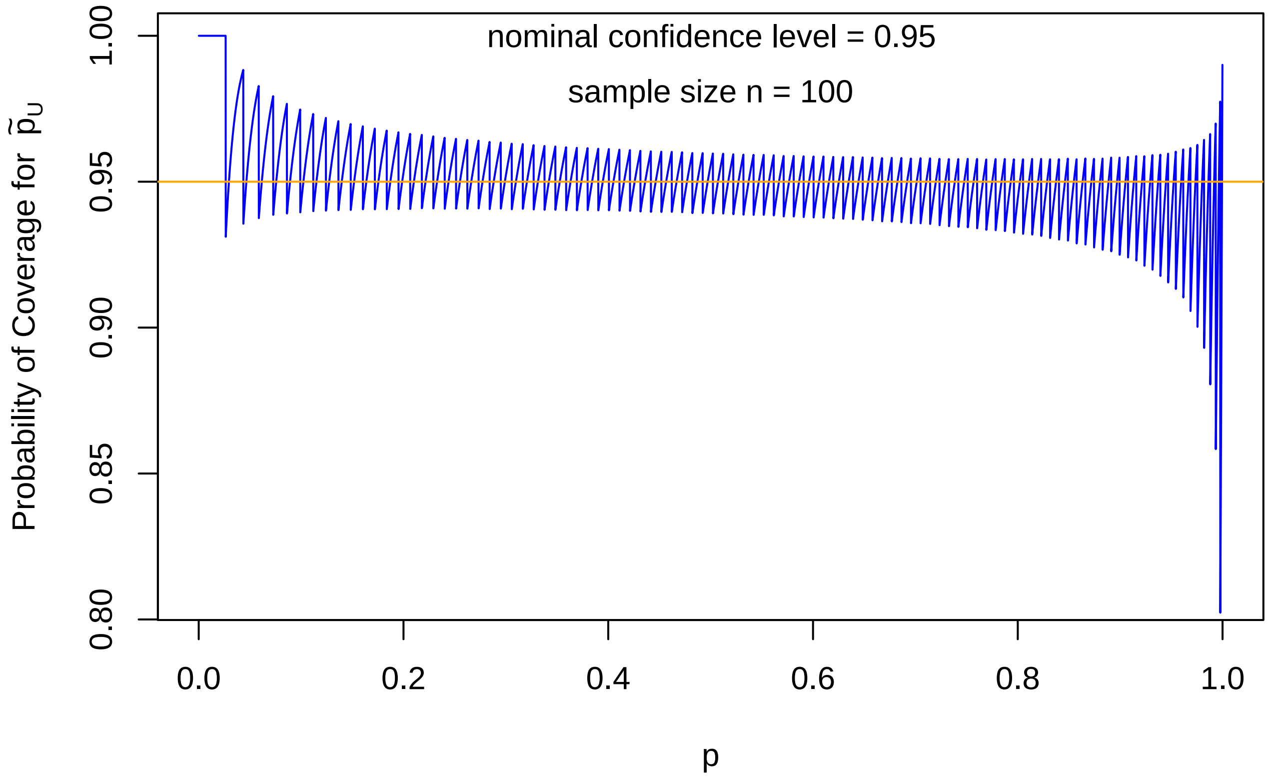


Figure 11: Coverage Probabilities for Score Upper Confidence Bounds

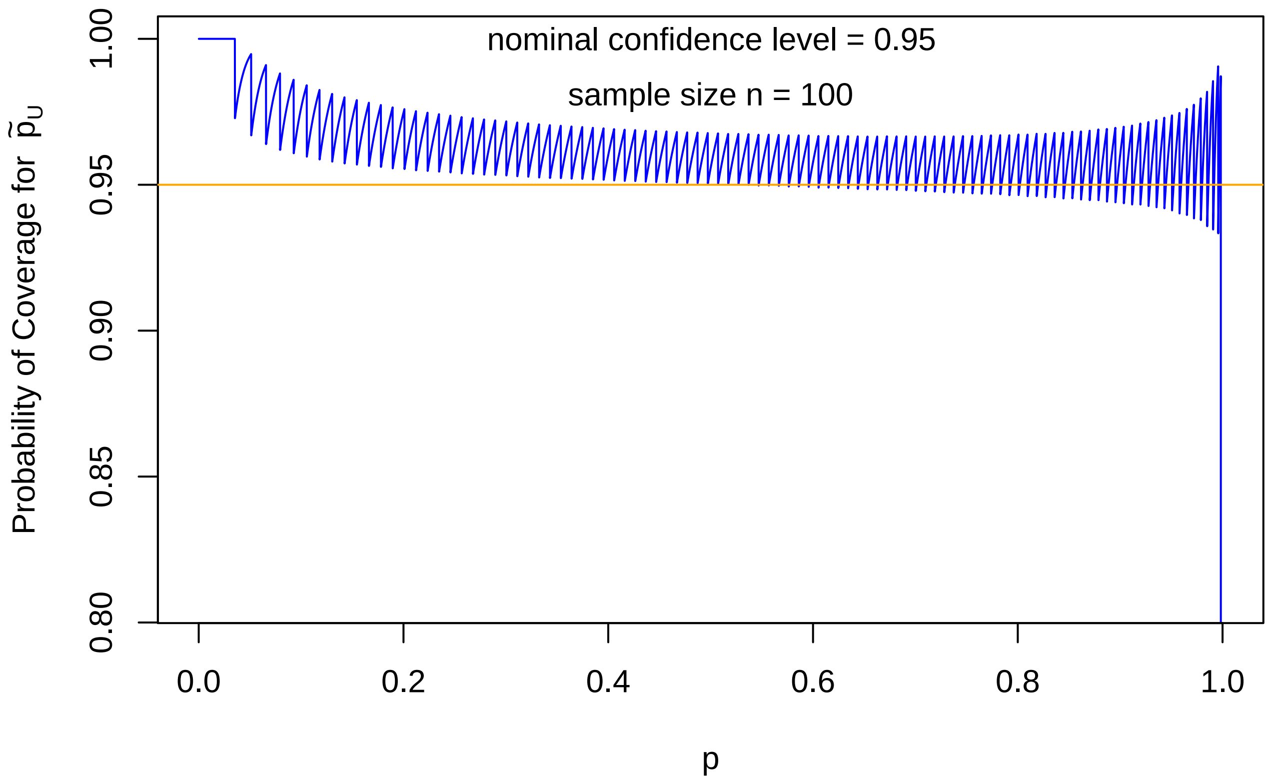


Figure 12: Coverage Probabilities for Score Upper Confidence Bounds using continuity correction

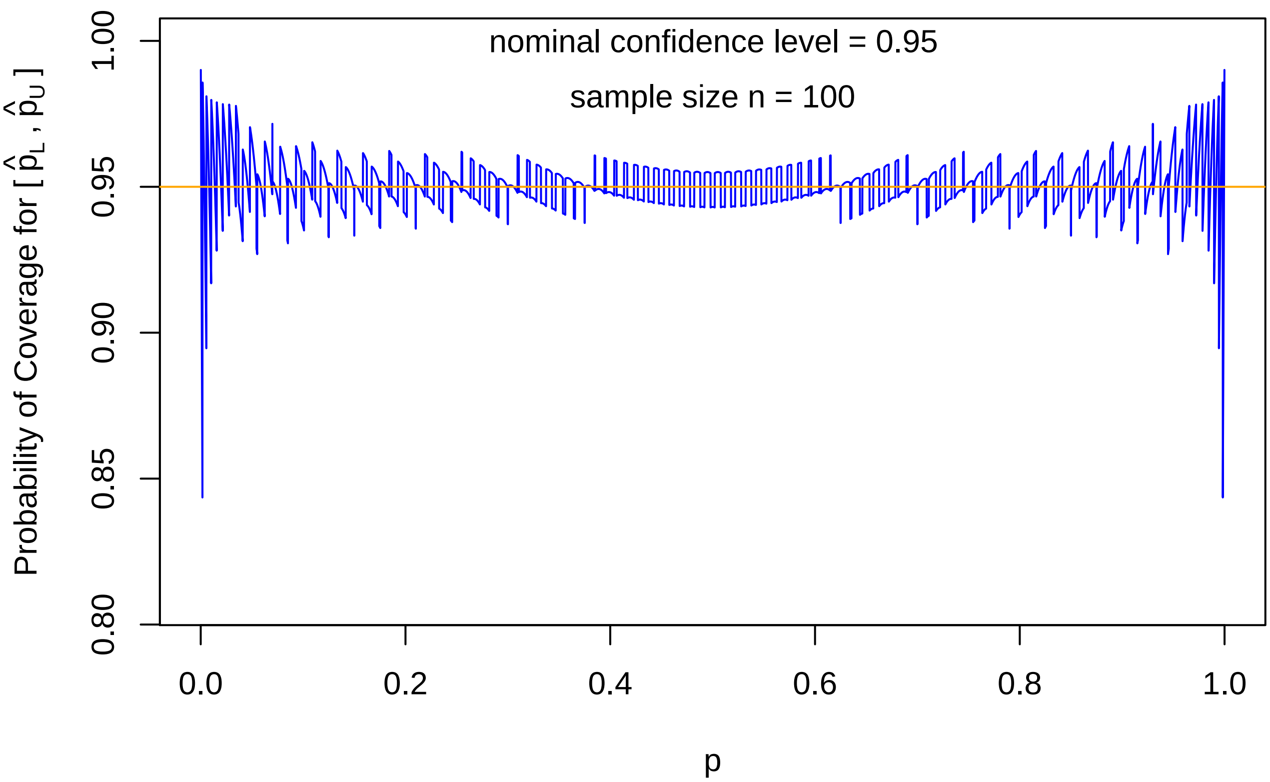


Figure 13: Coverage Probabilities for Score Confidence Intervals

2.11 Corrosion Inspection

Airplanes are inspected for corrosion at a 10 year inspection interval. Based on the outcome of n such inspections the customer wants 95% lower confidence bounds on the probability of an airplane passing its corrosion inspection without any findings. So far all n inspected aircraft made it through corrosion inspection without any findings. The customer did not tell me how many aircraft had been inspected, he only told me that 2.5% of the fleet had been inspected. How do we deal with this?

We can view each airplane's corrosion experience over a 10 year exposure window as a Bernoulli trial with probability p of surviving 10 years without corrosion. Thus the number X of aircraft without any corrosion found at their respective inspections is a binomial random variable with parameters n and success probability p. Based on $X = n$ the 95% lower confidence bound for p is $\hat{p}_L(.95, n, n) = (1 - .95)^{1/n}$. Without knowing n we can only plot $\hat{p}_L(.95, n, n)$ against n, as in Figure 14, and the customer can take it from there.

I suspect that this request arose because a particular airline, after finding no corrosion in n inspections, is wondering about necessity of the onerous task of stripping the interior of an airplane for the corrosion inspection. Maybe it is felt that this is a costly waste of resources or that the time to inspection should be increased so that cost is spread out over more service life. On the other hand, if corrosion is detected early it is a lot easier to fix than when it has progressed beyond extensive or impossible repair.

For the sake of argument, if this airline has 200 aircraft of which 2.5% (or $n = 5$) had their 10 year check, all corrosion free, then the 95% lower bound on p is about .55. This is not very reassuring. Jumping to any strong conclusions based on 5 successes in 5 trials has no basis. Most people don't have a good grounding in matters of probability and that is why such requests arise again and again.

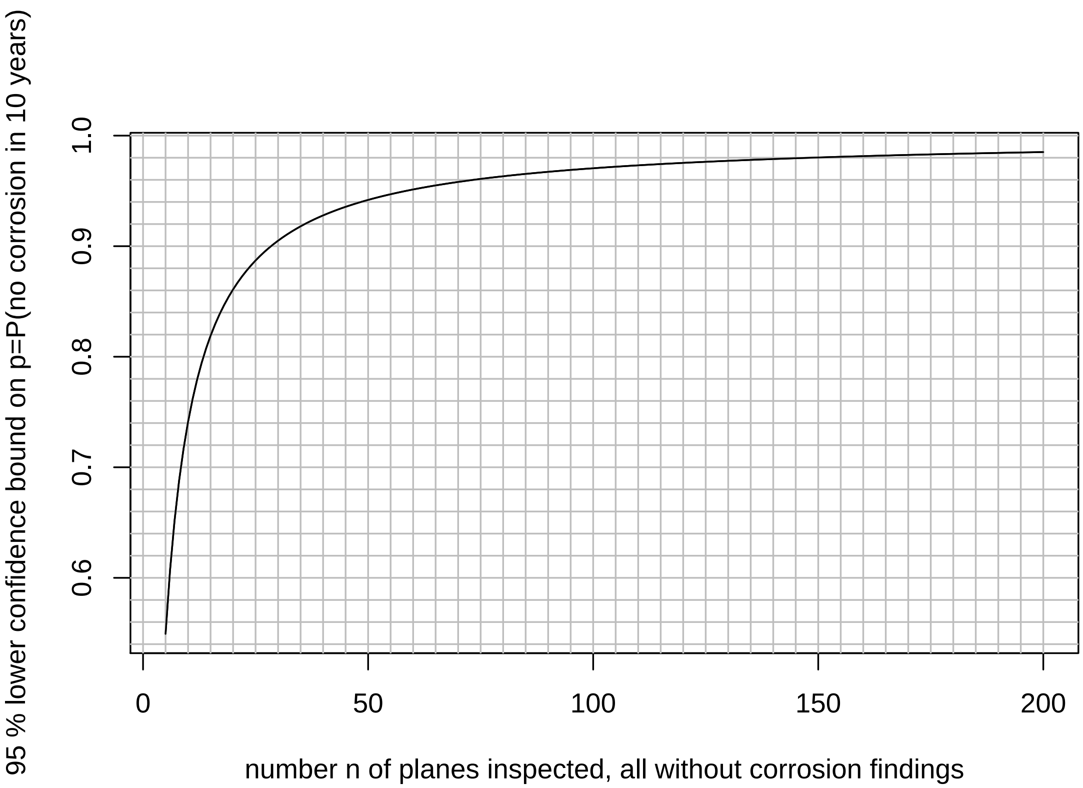


Figure 14: Plot of $\hat{p}_L(.95, n, n)$ against n.

2.12 Space Shuttle Application: Stud Hang-Up Issue

The two Solid Rocket Boosters (SRB) of the Space Shuttle are held down on the launch platform by 4 hold-down studs each, see Figure 15[2]. These studs are about 3.5″ in diameter and each is held in place by a frangible nut, that explodes at takeoff to let the stud fall through cleanly, thus releasing the Shuttle for takeoff. Due to various issues (explosion timing, etc) it sometimes happens that these studs do not fall away cleanly but hang up instead. See

`http://www.eng.uab.edu/ME/ETLab/HSC04/abstracts/HSC147.pdf`
and `http://www.nasa.gov/offices/nesc/home/Feature_1.html`

for some of the issues and consequence involved. It is possible that such a 3.5″ stud just gets snapped by the tremendous takeoff force.

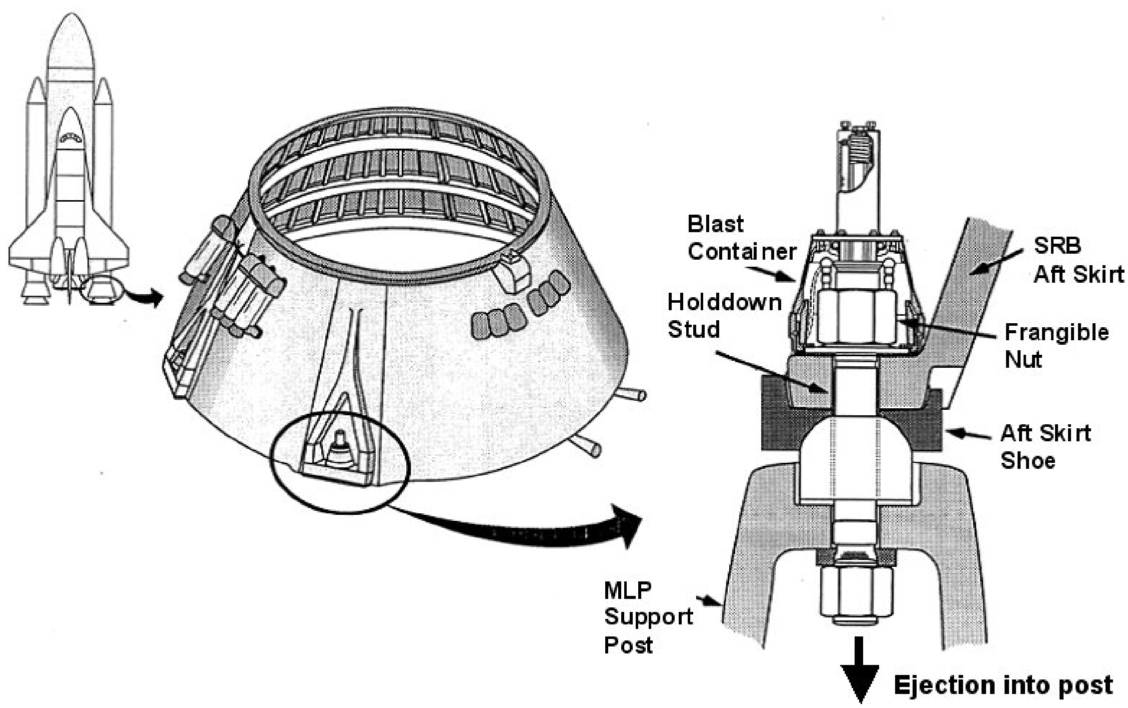


Figure 15: Solid Rocket Booster Hold-Down Stud

[2]taken from a NASA Presentation titled SRB HOLDDOWN POST STUD HANG-UP LOADS, Feb 6, 2004.

Out of the 114 lift-offs there were 21 with a one hang-up out of the eight studs and there were two lift-offs with two stud hang-ups. No lift-off experienced more than two stud hang-ups.

One can estimate the probability p of a stud hang-up by $\hat{p} = (21 \times 1 + 2 \times 2)/(114 \times 8) = 25/912 = 0.02741$. Based on this estimate and assuming that the events of hang-ups from stud to stud occur independently and with constant probability p one can estimate the probability p_i of i hang-ups among the 8 studs by the binomial probability formula

$$\hat{p}_i = \binom{8}{i} \hat{p}^i (1 - \hat{p})^{8-i} \ .$$

From this one gets $\hat{p}_0 = 0.8006$, $\hat{p}_1 = 0.1805$, $\hat{p}_2 = 0.0178$, and $\hat{p}_{3+} = 0.0010$, where $\hat{p}_{3+}$ is the estimate for 3 or more stud hang-ups. Based on this one would have expected $E_i = \hat{p}_i \times 114$ lift-offs with i stud hang-ups. These expected numbers are $E_0 = 91.27$, $E_1 = 20.58$, $E_2 = 2.03$ and $E_{3+} = 0.12$ which compare reasonably well with the observed counts of $O_0 = 91$, $O_1 = 21$, $O_2 = 2$, and $O_{3+} = 0$. Thus we do not have any indications that contradict the above assumption of independence and constant probability p of a stud hang-up.

The Clopper-Pearson 95% upper confidence bound on p is

$$\hat{p}_u(.95, 25, 912) = \texttt{qbeta}(.95, 25 + 1, 912 - 25) = 0.03807645 \ .$$

If X denotes the number of stud hang-ups during a single liftoff, we may be interested in an upper bound on the probability of seeing more than 3 hang-ups. Since $P_p(X \geq 4) = f_4(p)$ is monotone increasing in p we can use

$$f_4(\hat{p}_u(.95, 25, 912)) = 1 - \texttt{pbinom}(3, 8, 0.03807645) = 0.00013$$

as 95% upper bound for this risk $f_4(p)$.

Because of that small risk it was decided to run simulation models of such liftoffs: 800 simulations involving no hang-up, 800 simulations involving one hang-up, 800 simulations involving two hang-ups and 800 simulations involving three hang-ups. In those simulations many launch variables (stresses, deviations in angles, etc.) are monitored. If such a launch variable Y has a critical value y_0 it is desirable to get an upper confidence bound for $\bar{F}_p(y_0) = P_p(Y > y_0)$. By conditioning on $X = k$ we have the following expression for $\bar{F}_p(y_0)$

$$\begin{aligned}
\bar{F}_p(y_0) &= P(Y > y_0 | X = 0) P_p(X = 0) + P(Y > y_0 | X = 1) P_p(X = 1) \\
&\quad + P(Y > y_0 | X = 2) P_p(X = 2) + P(Y > y_0 | X = 3) P_p(X = 3) \\
&\quad + P(Y > y_0 | X \geq 4) P_p(X \geq 4)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{x=0}^{3} P(Y > y_0 | X = x) P_p(X = x) + P(Y > y_0 | X \geq 4) P_p(X \geq 4) \\
&\leq \sum_{x=0}^{3} P(Y > y_0 | X = x) P_p(X = x) + P_p(X \geq 4) = \bar{G}_p(y_0) \; .
\end{aligned}
$$

For many (but not for all) of the response variables one can plausibly assume that $P(Y > y_0|X = x)$ increases with x. Based on the monotonicity result in Section 2.3 we can get an upper confidence bound for $\bar{G}_p(y_0)$ (and thus conservatively also for $\bar{F}_p(y_0)$) by replacing p by the appropriate upper bound $\hat{p}_U(.95, 25, 912)$ in the expression for $\bar{G}_p(y_0)$. To do so would require knowledge of the exceedance probabilities $P(Y > y_0|X = x)$ for $x = 0, 1, 2, 3$. These would come from the four sets of simulations that were run. Of course such simulations can only estimate $P(Y > y_0|X = x)$ with their accompanying uncertainties, but that is an issue we won't address here.

In the Space Shuttle program it seems that, at least in certain areas of the program, the risk is managed to a 3σ limit. Since 3σ has different risk connotations for different distributions, such a 3σ risk specification usually means that the normal distribution is meant, since it is the most prevalent distribution in engineering applications. For a standard normal random variable Z we have $P(|Z| > 3) = .0027$. When faced with other than normal variability or uncertainty phenomena one should target this risk of .0023 (or .00135 when one-sided) for consistency. Thus we would like to show that $P(Y > y_0)$ is less than .00135 with 95% confidence.

As an illustration we show in Figure 16 for a particular indicator variable the empirical cdfs of the 800 simulations, as they were obtained under $0, 1, 2, 3$ hang-ups. These empirical cumulative distribution functions are defined as $\hat{F}_{800}(x) =$ proportion of simulated indicator values $\leq x$. These functions jump in increments of $1/800 = .00125$. For a threshold $y_0 = 55$ we show in the bottom plot (magnified version of the right tail of the top plot) of Figure 16 the fractions of values exceeding y_0. These fractions, i.e., $0, .00875, .03375$, and $.08375$, and the upper bound $\hat{p}_U(.95, 25, 912) = .03807645 = \hat{p}_U$ for p are then used to calculate the following 95% upper confidence bound for $\bar{G}_p(55) \geq P(Y > 55)$ as outlined previously

$$
0 \times \binom{8}{0} \hat{p}_U^0 (1 - \hat{p}_U)^8 + .00875 \times \binom{8}{1} \hat{p}_U (1 - \hat{p}_U)^7 + .03375 \times \binom{8}{2} \hat{p}_U^2 (1 - \hat{p}_U)^6
$$

$$
+ .08375 \times \binom{8}{j} \hat{p}_U^3 (1 - \hat{p}_U)^5 + \sum_{j=4}^{8} \binom{8}{j} \hat{p}_U^j (1 - \hat{p}_U)^{8-j} = 0.003459802
$$

This is larger than the one-sided .00135 risk associated with the 3σ requirement. We note that the last summation term in the expression above is $f_4(\hat{p}_u(.95, 25, 912)) = .00013$. It contributes less than 4% to the final result. Thus the upper bound use of $\bar{G}_p(55)$ for $P(Y > 55)$ does not represent a significant sacrifice.

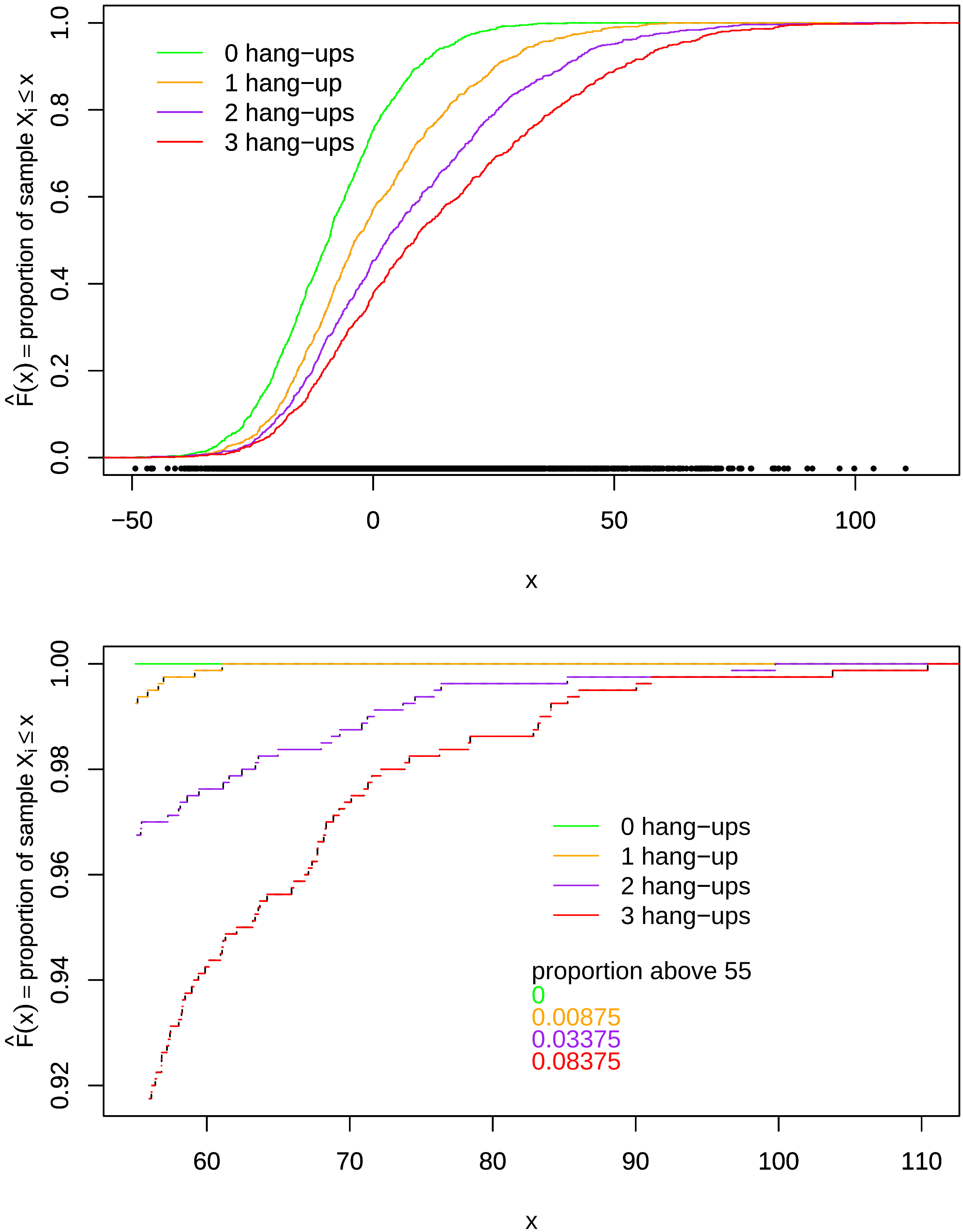


Figure 16: Simulation Results for a Specific Indicator Variable
Bottom Plot Shows Magnified Right Tail of Top Plot

We point out that the estimate of $\bar{G}_p(55)$ (using $\hat{p} = 25/912$ instead of $\hat{p}_U$) only yields

$$0 \times \binom{8}{0}\hat{p}^0(1-\hat{p})^8 + .00875 \times \binom{8}{1}\hat{p}(1-\hat{p})^7 + .03375 \times \binom{8}{2}\hat{p}^2(1-\hat{p})^6$$

$$+.08375 \times \binom{8}{j}\hat{p}^3(1-\hat{p})^5 + \sum_{j=4}^{8}\binom{8}{j}\hat{p}^j(1-\hat{p})^{8-j} = 0.002300867 \,.$$

Thus it is not the 95% confidence bound that is at issue here. However, while the data are somewhat real (transformed from real data) the threshold is entirely fictitious. It was chosen to point out what might happen in such (rare) situations. When the risk upper bound does not fall below the allowed 3σ value of .00135 there usually is an action item to revisit the requirements for the chosen threshold. Often these are chosen conservatively and an engineering review of the involved stresses or angular deviations may justify a relaxation of these thresholds.

We point out that the risks in the Space Shuttle program are of a different order of magnitude than those tolerated in commercial aviation. In the latter arena one often is confronted by the requirement of demonstrating a risk of 10^{-9} or less, i.e., expect at most one bad event in 10^9 departures. This appears to be based on the assumption that the number of flight critical systems is about 100 or less. If each such system is designed to the 10^{-9} requirement then the chance of at least one of these systems failing is at most $100 \times 10^{-9} = 10^{-7}$, which is still well below the current rate of hull losses and most of these are not due to any aircraft design deficiencies. See page 17 of

`http://www.boeing.com/news/techissues/pdf/statsum.pdf`

which attributes 17% of hull losses during the period 1996-2005 to the Airplane as the primary cause.

Establishing a 10^{-9} risk directly from system exposure data is difficult if not impossible, and certainly impracticable during the design phase. One usually builds up the exponent of such risks by multiplying verifiable risks of higher order and using redundancy in the design, like having at least two engines on a plane.

This difference in risk levels in the Space Program as compared to commercial aviation is based on many factors. Major drivers among these are the frequency with which the risk is taken and the consequences (loss of lives and financial losses) of a catastrophic event.

3 Negative Binomial Distribution: Upper and Lower Bounds for p

A negative binomial random variable N counts the number of required independent trials, with success probability p ($0 \le p \le 1$) in each trial, in order to obtain a predetermined number k of successes. Then we have

$$P_p(N = n) = \binom{n-1}{k-1} p^k (1-p)^{n-k}, \quad n = k, k+1, \ldots$$

based on the consideration that to realize $N = n$ we must have $k-1$ successes in the first $n-1$ trials and a success in the n^{th} trial. From a different angle we have

$$P_p(N \ge n) = \sum_{i=0}^{k-1} \binom{n-1}{i} p^i (1-p)^{n-1-i} ,$$

which is the probability of at most $k-1$ successes in the first $n-1$ trials.

On intuitive grounds it should get easier to satisfy the quota of k successes earlier as p increases. Thus we expect $P_p(N \ge n)$ to decrease as p increases. As in the binomial situation it is easy to show that

$$\frac{\partial P_p(N \ge n)}{\partial p} = -(n-1)\binom{n-2}{k-1} p^{k-1}(1-p)^{n-k-1} = \frac{\partial[1 - I_p(k, n-k)]}{\partial p} < 0 , \tag{7}$$

i.e., for $n > k$ the probability $P_p(N \ge n)$ decreases strictly from 1 to 0 as p increases from 0 to 1, while for $n = k$ we have $P_p(N \ge k) = 1$ for all p. For $n > k$ the identity (7) yields

$$P_p(N \ge n) = 1 - I_p(k, n-k) . \tag{8}$$

The negative binomial distribution is a useful vehicle in assessing success probabilities in Bernoulli trials when only a limited number of successes can be tolerated. For example, it could be that a "success" consists of a costly event, say destruction of the experimental unit, and only so many such units are available for running the trials. Shapiro and Gross (1981) cite the drilling for oil as an example, where success consists of finding a productive well and it is desired to drill as little as possible in a given area.

Another application is in random testing of software when working against a deadline. Such random testing can either involve random selection of inputs according to some usage profile from a well defined input space or it could consist of a session with a random person using the program (a spreadsheet or word processing program) in a given task. Each time a software run or session results in faulty behavior it requires finding and fixing the responsible bug. Such fixes take time and only so many (k) can be tolerated within a given time window.

3.1 Upper Bounds for p

Consider testing $H(p_0) : p = p_0$ against the alternative $A(p_0) : p < p_0$. Since it will take longer to fill the quota of k successes when p is small, large observed values $N = n$ can be viewed as evidence against the hypothesis. Thus we reject $H(p_0)$ at target significance level α when $N \geq n(p_0, \alpha)$, where $n(p_0, \alpha)$ is chosen as the smallest integer n for which $P_{p_0}(N \geq n) \leq \alpha$. Thus we have $P_{p_0}(N \geq n(p_0, \alpha)) \leq \alpha$ and $P_{p_0}(N \geq n(p_0, \alpha) - 1) > \alpha$.

As in the binomial situation we exploit the duality between confidence sets $C(n)$ and tests of hypotheses. Here n denotes the observed value of N. Again we express $C(n)$ in terms of p-values $p(n, p_0) = P_{p_0}(N \geq n)$, namely as the set of all acceptable hypotheses $H(p_0)$ or all p_0 for which the p-value $p(n, p_0) > \alpha$ leads to acceptance of $H(p_0)$:

$$C(n) = \{p_0 : p(n, p_0) = P_{p_0}(N \geq n) > \alpha\} \ .$$

When viewing the collection of confidence sets $C(n)$, $n = k, k+1, \ldots$ as a random set $C(N)$, we have again the following coverage probability property:

$$P_{p_0}(p_0 \in C(N)) = 1 - P_{p_0}(p_0 \notin C(N)) = 1 - P_{p_0}(N \geq n(p_0, \alpha)) \geq 1 - \alpha \qquad \text{for all } p_0 \ , \tag{9}$$

i.e., this random confidence set has coverage probability $\geq \gamma = 1 - \alpha$ no matter what the value p_0 is. Thus we may consider $C(N)$ as a $100\gamma\%$ confidence set for the unknown parameter p_0.

We point out that the inequality (9) becomes an equality for some values of p_0, because there are values p_0 for which $P_{p_0}(N \geq n(p_0, \alpha)) = \alpha$. This follows since for any integer $n > k$ the probability $P_{p_0}(N \geq n)$ decreases continuously from 1 to 0 as p_0 increases from 0 to 1, i.e., takes on the value α for some p_0. This can be accomplished for any value $n > k$. Thus the confidence coefficient (minimum coverage probability) of the confidence set $C(N)$ is indeed $\bar{\gamma} = 1 - \alpha = \gamma$.

It remains to calculate the confidence set $C(n)$ explicitly for all possible values $n = k, k+1, \ldots$. Since for $n > k$ the p-value $P_{p_0}(N \geq n)$ is continuous and strictly decreasing in p_0, it follows that the confidence set $C(n)$ is of the form $[0, \tilde{p}_U(\gamma, n, k))$, where $\tilde{p}_U(\gamma, n, k) = \tilde{p}_U(1 - \alpha, n, k)$ is the unique value p that solves

$$P_p(N \geq n) = \sum_{i=0}^{k-1} \binom{n-1}{i} p^i (1-p)^{n-1-i} = \alpha = 1 - \gamma \ . \tag{10}$$

For $n = k$ equation (10) has no solution since $P_p(N \geq k) = 1$ for all p. According to the definition of $C(n)$ it follows directly that $C(k) = [0, 1]$. In that special case define $\tilde{p}_U(\gamma, k, k) = 1$.

Thus we can treat $\tilde{p}_U(\gamma, N, k)$ as $100\gamma\%$ upper confidence bound for p. In the case $N = n > k$, it is convenient to use the Beta distribution identity (8) in place of the binomial summation when solving (10) for p, i.e., p is the γ-quantile of the Beta distribution with parameters k and $n - k$.

Thus $\tilde{p}_U(\gamma, n, k)$ can be obtained from Excel by invoking $\texttt{BETAINV}(\gamma, k, n-k)$ and in R or S-Plus by the command $\texttt{qbeta(}\gamma\texttt{,}k\texttt{,}n-k\texttt{)}$.
As a check example take $k = 25$ and $n = 1200$ with $\gamma = .95$, then $\tilde{p}_U(.95, 1200, 25) = .028036$ is the 95% upper confidence bound for p.
Using (10) it is a simple exercise to show that

$$1 = \tilde{p}_U(\gamma, k, k) > \tilde{p}_U(\gamma, k+1, k) > \tilde{p}_U(\gamma, k+2, k) > \ldots > \tilde{p}_U(\gamma, n, k) > \ldots \searrow 0 \quad \text{as } n \to \infty.$$

As argued previously, it is precisely at the points $p = \tilde{p}_U(\gamma, n, k)$, $n > k$, that the coverage probability is exactly equal to γ.
For the case $k = 1$ the upper bounds defined by (10) can be expressed explicitly as

$$\tilde{p}_U(\gamma, n, 1) = 1 - (1-\gamma)^{1/(n-1)} .$$

The case $k = 1$ becomes useful if n is large, i.e., it takes a long time to get the first success. For $\gamma = .95$ the upper confidence bound in this latter case then becomes

$$\tilde{p}_U(.95, n, 1) = 1 - (.05)^{1/(n-1)} = 1 - \exp\left[\frac{\log(.05)}{n-1}\right] \approx 1 - \exp\left[-\frac{3}{n-1}\right] \approx \frac{3}{n-1} ,$$

which can be viewed as another instance of the Rule of Three. The last invoked approximation is only valid for large n. The $n-1$ in the denominator can be viewed as the number of failures (the only random aspect here) that must precede the one success, that is assumed to happen a priori.
The case $n = k+1$ also entails an explicit form for the upper bound, namely $\tilde{p}_U(\gamma, k+1, k) = \gamma^{1/k}$. However, this case is of less practical interest.

3.2 Lower Bounds for p

When testing the hypothesis $H(p_0) : p = p_0$ against the alternative $A(p_0) : p > p_0$, small values of N will serve as evidence against the hypothesis, since under $A(p_0)$ the quota of k successes is going to be filled quicker than under $H(p_0)$. After observing $N = n$ one can carry out the test at nominal level α by rejecting $H(p_0)$ whenever the p-value $p(n, p_0) = P_{p_0}(N \le n) \le \alpha$. Invoking again the duality between tests and confidence sets we get as $\gamma = 1 - \alpha$ level confidence set

$$C(n) = \{p_0 : P_{p_0}(N \le n) > \alpha\} \quad \text{with} \quad P_{p_0}(p_0 \in C(N)) \ge 1 - \alpha \ \ \forall p_0 ,$$

where equality is achieved for some values of p_0. Since $P_{p_0}(N \le n)$ is continuous and strictly increasing in p_0 we see that $C(n)$ coincides with the interval $(\tilde{p}_L(\gamma, n, k), 1]$ where $\tilde{p}_L(\gamma, n, k)$ is the unique value of p that solves

$$P_p(N \le n) = 1 - P_p(N \ge n+1) = 1 - \sum_{i=0}^{k-1} \binom{n}{i} p^i (1-p)^{n-i} = \alpha = 1 - \gamma . \tag{11}$$

Thus we can treat $\tilde{p}_L(\gamma, n, k)$ as a $100\gamma\%$ lower confidence bound for p. Using again the identity (8) it can be obtained in Excel by invoking `BETAINV(`$1-\gamma, k, n-k+1$`)` and from R or S-Plus by the command `qbeta(`$1-\gamma$`,`k`,`$n-k+1$`)`. Note that this lower bound matches the binomial one for for fixed n, while this is not true for the upper bounds.

As a check example take $k = 5$ and $n = 30$ with $\gamma = .95$, then $\tilde{p}_L(.95, 5, 30) = .068056$ is the 95% lower confidence bound for p.

For $k = 1$ these lower bounds take the explicit form $\tilde{p}_L(\gamma, 1, n) = 1 - \gamma^{1/n}$ for $n = 1, 2, \ldots$, but they are not of much practical use, since they result in values close to 0. For $k = n$ the lower bound also has an explicit expression, i.e., $\tilde{p}_L(\gamma, n, n) = (1-\gamma)^{1/n}$. For $\gamma = .95$ this becomes

$$\tilde{p}_L(.95, n, n) = (1 - .95)^{1/n} \approx \exp(-3/n) \approx 1 - \frac{3}{n} ,$$

another instance of the Rule of Three. Here the last approximation is only valid for large $n = k$. Of course, using a large k only makes sense when $p \approx 1$, i.e., successes are accumulated quickly so that in the situation discussed we wind up with k successes in the first $n = k$ trials. This is akin to seeing a lot of failures before seeing the first success, when reversing the role of successes and failures, thus reducing this to the situation $k = 1$ discussed near the end of Section 3.1.

3.3 Confidence Intervals

The lower and upper bounds for p, i.e., $\tilde{p}_L(1-\alpha/2, n, k)$ and $\tilde{p}_U(1-\alpha/2, n, k)$, can be combined into a confidence interval $(\tilde{p}_L(1-\alpha/2, n, k), \tilde{p}_U(1-\alpha/2, n, k))$ at nominal level $\gamma = 1-\alpha$. The details are too close to those of Section 2.6 and are not repeated here.

3.4 Bounding N

The negative binomial random variable N is unbounded. This could easily lead to a very large N when p is very small. When the interest is focussed on bounding p from above one may want to put an integer limit $n^\star$ on N so that the possibility of a very small p does not degenerate into squandering valuable resources, namely time and experimental units involved in the Bernoulli trials. Having many trials without achieving the full quota should still lead to very effective upper bounds for p. Not much will change when deriving confidence bounds for p based on $\tilde{N} = \min(N, n^\star)$, where the choice of $n^\star$ should obviously satisfy $n^\star > k$ to be of any interest.

3.4.1 Upper Bounds for p

Clearly $\tilde{N}$ is limited to the integers from k to $n^\star$. We also note that for $n = k, \ldots, n^\star$

$$P_p(\tilde{N} \geq n) = P_p(\min(N, n^\star) \geq n) = P_p(\{N \geq n\} \cap \{n^\star \geq n\}) = P_p(N \geq n) , \tag{12}$$

which for $n > k$ is again a continuous and strictly decreasing function in p, and decreases from 1 to 0 as p increases from 0 to 1, while for $n = k$ we have $P_p(\tilde{N} \geq k) = 1$ for all p. As before, large values of $\tilde{N}$ (although bounded by $n^\star$) should cause us to reject $H(p_0) : p = p_0$ in favor of the alternative $A(p_0) : p < p_0$. In view of the equation (12) it should not surprise to find for p (as in the unbounded case) the same upper confidence bounds, namely for the degenerate case $n = k$ we get $\tilde{p}_U(\gamma, k, k) = 1$ and for $n = k+1, \ldots, n^\star$ we have

$$\tilde{p}_U(\gamma, n, k) = \mathtt{qbeta}(\gamma, \mathtt{k}, \mathtt{n} - \mathtt{k}) = \mathtt{BETAINV}(\gamma, \mathtt{k}, \mathtt{n} - \mathtt{k}) \; .$$

There are only a finite number of such bounds and they satisfy

$$1 = \tilde{p}_U(\gamma, k, k) > \tilde{p}_U(\gamma, k+1, k) > \tilde{p}_U(\gamma, k+2, k) > \ldots > \tilde{p}_U(\gamma, n^\star, k) > 0 \; .$$

Thus $\tilde{p}_U(\gamma, n^\star, k)$ is the smallest possible upper bound for p when using the bounded $\tilde{N}$. That is the price one has to pay for bounding N and it makes intuitive sense. Without bounding N the bounds can get arbitrarily small.

How would one use this in practice? Chose k, e.g. $k = 1$, and then $n^\star$ such that $\tilde{p}_U(\gamma, n^\star, k)$ is somewhat smaller than what is judged sufficiently small for p, i.e., $\tilde{p}_U(\gamma, n^\star, k)$ is sufficiently small for the application at hand. Obtaining smaller upper bounds, by raising the upper limit $n^\star$, would be overkill.

3.4.2 Lower Bounds for p

Note that for $k \leq n < n^\star$

$$P_p(\tilde{N} \leq n) = P_p(\min(N, n^\star) \leq n) = P_p(\{N \leq n\} \cup \{n^\star \leq n\}) = P_p(N \leq n) \qquad (13)$$

is again continuous and strictly increasing in p and increases from 0 to 1 as p increases from 0 to 1. For $n = n^\star$ we have $P_p(\tilde{N} \leq n^\star) = 1$ for all p. Small values of $\tilde{N}$ should induce us to reject the hypothesis $H(p_0) : p = p_0$ when testing it against the alternative $A(p_0) : p > p_0$. As in the unbounded case we obtain as lower bound

$$\tilde{p}_L(\gamma, n, k) = \mathtt{qbeta}(1 - \gamma, \mathtt{k}, \mathtt{n} - \mathtt{k} + 1) = \mathtt{BETAINV}(1 - \gamma, \mathtt{k}, \mathtt{n} - \mathtt{k} + 1)$$

for $k \leq n < n^\star$, while for $n = n^\star$ we get $\tilde{p}_L(\gamma, n^\star, k) = 0$ with the understanding that it results in the closed lower bound interval $[0, 1]$ since $P_p(\tilde{N} \leq n^\star) = 1$ for all p and thus the duality based confidence set in that case is

$$C(n^\star) = \left\{ p : P_p(\tilde{N} \leq n^\star) > \alpha \right\} = [0, 1] \; .$$

3.4.3 Confidence Intervals for p

As in the unbounded case these one-sided bounds can be combined into confidence intervals, i.e., $(\tilde{p}_L(1-\alpha/2, n, k),\ \tilde{p}_U(1-\alpha/2, n, k))$ can be considered as a $100\gamma\%$ $(\gamma = 1-\alpha)$ confidence interval for p. Further details are again omitted.

3.5 The Probability of a Typo or Other Error

Recently I read the book *The Calculus Wars, Newton and Leibniz, and the Greatest Mathematical Clash of All Time* by J.S. Bardi (2006). Although the book was definitely worth reading, I started to get annoyed by typos and other mistakes so that I decided to mark them, starting with the use of "German" in place of "Germany" on page 59. I recall other such errors on prior pages but I was not going to reread those pages to mark them as well. Since page 59 was my decision point to start marking these errors I will not count that error on page 59. I give here the pages on which I marked the errors that I marked after that (I may well have overlooked some): 69, 81, 82, 92, 104, 107, 113, 114, 121 (121), 123 (123), 130 (130), 135, 139, 144, 145, 147, 154, 182 (182) (182), 198, 204, 205, 207, 213, 222, 229, 250, 251, 253, 253, 255 (255), 256. Pages with several errors are recorded in parentheses for each additional error. The last page was 257.

If we treat each page as a Bernoulli trial with "success" probability p of containing an error (ignoring all the parenthetical additional counts) we wish to give a 95% confidence interval for p. We can do this based on the first 20 error pages found, starting with the first error on page 69 and the 20$^{\text{th}}$ on page 204. Thus we observed $N = 204 - 59 = 145$ and we get as .95% confidence interval

$$
\begin{aligned}
&(\tilde{p}_L(.975, 145, 20),\ \tilde{p}_U(.975, 145, 20)) \\
&\quad = \ (\texttt{qbeta}(.025, 20, 145-20+1),\ \texttt{qbeta}(.975, 20, 145-20)) = (0.0863,\ 0.1984)\ .
\end{aligned}
$$

If we had limited ourselves to the next $n^\star = 100$ pages or 20 errors, whichever comes first, we would have $\tilde{N} = 100 = n^\star$ and our 95% confidence interval becomes

$$
\begin{aligned}
&(\tilde{p}_L(.975, 100, 20),\ \tilde{p}_U(.975, 100, 20)) \\
&\quad = \ (0,\ \texttt{qbeta}(.975, 20, 100-20)) = (0,\ 0.2834)\ .
\end{aligned}
$$

Had we counted the number of errors in the first 100 pages, pp. 60-159 we could treat this count $(X = 17)$ as a binomial random variable with $n = 100$ and "success" probability p. The 95% confidence interval in that case is

$$
\begin{aligned}
&(\hat{p}_L(.975, 17, 100),\ \hat{p}_U(.975, 17, 100)) \\
&\quad = \ (\texttt{qbeta}(.025, 17, 100-17+1),\ \texttt{qbeta}(.975, 17+1, 100-17)) = (0.1023,\ 0.2582)\ .
\end{aligned}
$$

4 Some Afterthoughts

The above analysis resulting in the interval $(0,\ 0.2834)$ when reaching the upper limit of 100 pages leaves an uneasy feeling with regard to the lower bound of 0. Somehow we did not take advantage of the $X = 17$ count reached when we stopped at $n^\star = 100$, i.e., a whole lot of information was discarded by focussing on just $\tilde{N}$ when $\tilde{N} = n^\star$ occurs. This could have happened with $X = 0, 1, 2, \ldots, 20$. One should also use the number of successes accumulated when $\tilde{N} = n^\star$ is reached. On intuitive grounds a smaller count of successes at the stopping state $\tilde{N} = n^\star$ should speak more strongly for a small p. The same concern about ignored information should be raised with respect to upper bounds.

In performing such a truncated negative binomial experiment we should track two data quantities, namely $\tilde{N} = \min(N, n^\star)$ and $X_{\tilde{N}}$, where the latter counts the number of successes obtained by the time we stop observing. For $\tilde{N} = n = k, \ldots, n^\star - 1$ we have $X_{\tilde{N}} = k$. For $\tilde{N} = n^\star$ we could have have stopped with $X_{\tilde{N}} = k, k-1, \ldots, 0$ successes. In terms of evidence for a small p these situations can intuitively be ordered via

$$Y = \tilde{N} + k - X_{\tilde{N}}$$

with small Y indicating a high value of p and a large Y indicating a small value of p. When $\tilde{N} = k, k+1, \ldots, n^\star - 1$ then Y takes on the values $y = k, k+1, \ldots, n^\star - 1$ (since then $X_{\tilde{N}} = k$) and Y takes on the values $y = n^\star, n^\star + 1, \ldots, n^\star + k$ as $\tilde{N} = n^\star$ and $X_{\tilde{N}} = X_{n^\star} = k, k-1, \ldots, 1, 0$. When testing the hypothesis $H(p_0) : p = p_0$ against $A(p_0) : p > p_0$ we should reject when Y is too small. By the duality principle we get the following $\gamma = (1-\alpha)$-level confidence set

$$C(y) = \{p_0 : P_{p_0}(Y \le y) > \alpha\} \ .$$

Note that

$$P_p(Y \le y) = \begin{cases} P_p(N \le y) & \text{for } y = k, \ldots, n^\star \\ P_p(X_{n^\star} \ge n^\star + k - y) & \text{for } y = n^\star, n^\star + 1, \ldots, n^\star + k \end{cases}$$

is strictly increasing in p for $y < n^\star + k$ while $P_p(Y \le n^\star + k) = 1$ for all p. Note that the two forms for $P_p(Y \le n^\star)$ agree, the first being the probability of reaching the quota of k at or before $n^\star$ and the second being the probability of seeing at least k successes in the first $n^\star$ trials.

For $y < n^\star + k$ the γ-level confidence set $C(y)$ takes the form $(\tilde{p}_L(\gamma, y, k), 1]$ where $p = \tilde{p}_L(\gamma, y, k)$ solves $P_p(Y \le y) = \alpha$. Thus

$$\tilde{p}_L(\gamma, y, k) = \begin{cases} \texttt{qbeta}(1-\gamma, \texttt{k}, \texttt{y} - \texttt{k} + 1) & \text{for } y = k, \ldots, n^\star \\ \texttt{qbeta}(1-\gamma, \texttt{n}^\star + \texttt{k} - \texttt{y}, \texttt{y} - \texttt{k} + 1) & \text{for } y = n^\star, n^\star + 1, \ldots, n^\star + k \end{cases}$$

and as in previous situations take $\tilde{p}_L(\gamma, n^\star + k, k) = 0$ since $C(n^\star + k) = [0, 1]$.

To get the corresponding upper confidence bound consider testing the hypothesis $H(p_0) : p = p_0$ against $A(p_0) : p < p_0$. We should reject $H(p_0)$ in favor of $A(p_0)$ when Y is too large. By the duality principle we then get the following $\gamma = (1 - \alpha)$-level confidence set

$$C(y) = \{p_0 : P_{p_0}(Y \geq y) > \alpha\} \ .$$

By complement we have that $P_p(Y \geq y) = 1 - P_p(Y \leq y - 1)$ is strictly decreasing in p for $y > k$ and in that case we have $C(y) = [0, \tilde{p}_U(\gamma, y, k))$, where $p = \tilde{p}_U(\gamma, y, k)$ solves $P_p(Y \geq y) = \alpha$. Thus

$$\tilde{p}_U(\gamma, y, k) = \begin{cases} \texttt{qbeta}(\gamma, \texttt{k}, \texttt{y} - \texttt{k}) & \text{for } y = k+1, \ldots, n^\star \\ \texttt{qbeta}(\gamma, \texttt{n}^\star + \texttt{k} - \texttt{y} + 1, \texttt{y} - \texttt{k}) & \text{for } y = n^\star, n^\star + 1, \ldots, n^\star + k \end{cases}$$

For $y = k$ we take $\tilde{p}_U(\gamma, k, k) = 1$ since in that case we have $C(k) = [0, 1]$.
As before, such one-sided bounds $\tilde{p}_L(1 - \alpha/2, y, k)$ and $\tilde{p}_U(1 - \alpha/2, y, k)$ can be combined to form a $(1 - \alpha)$-level confidence interval for p, with confidence coefficient $\bar{\gamma}$ typically greater than the nominal level $\gamma = 1 - \alpha$.
Revisiting the previous example, where we limited ourselves to $n^\star = 100$ inspected pages with just $X = 17$ pages found with errors, we have as observed value $Y = y = 100 + 20 - 17 = 103$ and we now get as 95% confidence interval for p

$$\begin{aligned} & (\tilde{p}_L(.975, 103, 20), \tilde{p}_U(.975, 103, 20) \\ &= (\texttt{qbeta}(.025, 100 + 20 - 103, 103 - 20 + 1), \texttt{qbeta}(.975, 100 + 20 - 103 + 1, 103 - 20)) \\ &= (0.1022649, 0.2581754) \end{aligned}$$

This is certainly much improved over the interval $(0, 0.2834)$ based on $\tilde{N}$ and ignoring the information provided by $X_{\tilde{N}} = 17$.
We note that the above confidence interval based on Y is exactly as it was in the last treatment of the example with a fixed number on $n^\star = 100$ trials and observing 17 successes. However, here we had the possibility that the experiment could have stopped sooner if we had reached our quota of $k = 20$ prior to the set limit $n^\star = 100$ inspected pages.

5 Poisson Distribution: Upper and Lower Bounds for λ

Suppose X is a Poisson random variable with mean λ ($\lambda > 0$), i.e.,

$$P_\lambda(X \le k) = \sum_{i=0}^{k} \frac{\exp(-\lambda)\lambda^i}{i!} \,.$$

We also write $X \sim \mathsf{Poisson}(\lambda)$ to indicate that X has this distribution. On intuitive grounds a larger mean λ should entail a decrease in $P_\lambda(X \le k)$. Formally this is confirmed by taking the derivative of this probability with respect to λ and canceling all terms but one:

$$\begin{aligned} \frac{\partial P_\lambda(X \le k)}{\partial \lambda} &= -\sum_{i=0}^{k} \frac{\exp(-\lambda)\lambda^i}{i!} + \sum_{i=0}^{k} \frac{\exp(-\lambda)i\lambda^{i-1}}{i!} \\ &= -\sum_{i=0}^{k} \frac{\exp(-\lambda)\lambda^i}{i!} + \sum_{i=0}^{k-1} \frac{\exp(-\lambda)\lambda^i}{i!} = -\frac{\exp(-\lambda)\lambda^k}{k!} < 0 \end{aligned}$$

By the Fundamental Theorem of Calculus we have

$$\frac{\partial P_\lambda(X \le k)}{\partial \lambda} = -\frac{\exp(-\lambda)\lambda^k}{k!} = \frac{\partial}{\partial \lambda} \int_\lambda^\infty \frac{\exp(-x)x^{k+1-1}}{\Gamma(k+1)}\, dx$$

and thus obtain the following useful identity

$$P_\lambda(X \le k) = \sum_{i=0}^{k} \frac{\exp(-\lambda)\lambda^i}{i!} = \int_\lambda^\infty \frac{\exp(-x)x^{k+1-1}}{\Gamma(k+1)}\, dx \,, \tag{14}$$

since both the left and right side converge to 1 as $\lambda \to 0$. Here the righthand side of (14) is the right tail probability $P(V \ge \lambda)$ of a Gamma random variable V with scale 1 and shape parameter $k+1$.

The Poisson distribution is useful for random variables X that count incidents (accidents) over intervals of time $[0, T]$, or occurrences of defects or damage over surface areas A or volumes V (inclusions). Given the number x of such a count, the locations in time or surface or space of the x incidents can be viewed as having been distributed according to a uniform distribution.

In such a context (when counting incidents over a time interval $[0, T]$, or occurrences over an area A or volume V) one usually expresses the mean λ also as $\lambda = T \times \lambda_1$, or $\lambda = A \times \lambda_1$, or $\lambda = V \times \lambda_1$, where λ_1 would represent the mean of a count over a time interval of unit length, or over an area or volume of measure 1. This is useful when trying to project counts to time intervals of length different from T, or when comparing such counts coming from different time intervals, with corresponding adjustments when dealing with areas or volumes.

The Poisson distribution can also be viewed as a good approximation to a binomial distribution with small p. The approximating Poisson distribution would have parameter $\lambda = np$. In fact, assuming $np_n = \lambda$ the limiting behavior of the binomial probabilities

$$\binom{n}{x} p_n^x (1-p_n)^{n-x} = \frac{(np_n)^x (1-p_n)^n}{x!} \times \frac{(1-p_n)^{-x} n(n-1)\cdots(n-x+1)}{n^x} \longrightarrow \frac{\exp(-\lambda)\lambda^x}{x!}$$

as $n \longrightarrow \infty$, is very instrumental in motivating the Poisson distribution.

However, even though the above argument involves a limiting operation ($n \rightarrow \infty$) the actual approximation quality is governed by how small p is and not by the magnitude of n. In fact, there are much more general approximation results available.

Let $I_1, \ldots, I_n$ be independent Bernoulli random variables with $P(I_i = 1) = 1 - P(I_i = 0) = p_i$, $i = 1, \ldots, n$, i.e., they are not necessarily identically distributed. The distribution of $W = I_1 + \ldots + I_n$ is called the Poisson-Binomial distribution and its computation is quite challenging when the p_i are not the same.

For small p_i the distribution of W is very well approximated by the Poisson distribution with mean $\lambda = \sum_{i=1}^n p_i = n\bar{p}$. Thus, if $X \sim \mathsf{Poisson}(\lambda)$, we can approximate $P(W \in A)$ by $P(X \in A)$ for any set A. The total variation norm

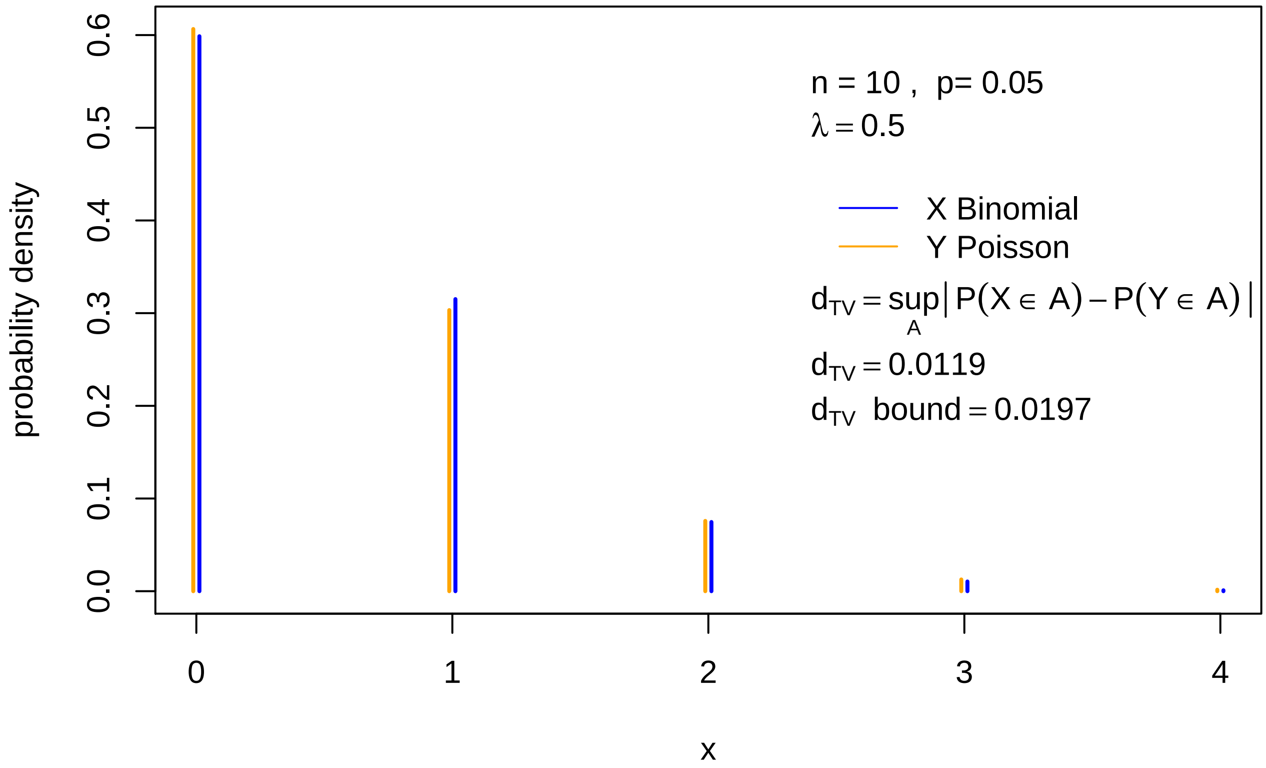
$$d_{\mathrm{TV}}(W, X) = \sup_A |P(W \in A) - P(X \in A)| = \frac{1}{2} \sum_{j=0}^{\infty} |P(W = j) - P(X = j)|$$

is used to measure the quality of this approximation. Barbour, Holst, and Janson (1992) give the following bound on this total variation norm

$$d_{\mathrm{TV}}(W, X) \le \frac{1 - \exp(-\lambda)}{\lambda} \sum_{i=1}^n p_i^2 \le \min(1, \lambda^{-1}) \sum_{i=1}^n p_i^2 \le \max(p_1, \ldots, p_n) \tag{15}$$

which through the last inequality links up with our previous statement concerning the approximation quality in the binomial case.

Figures 17 and 18 illustrate the approximation quality in the case of the binomial distribution ($p_1 = \ldots = p_n = p$). Note that n does not play much of a role, while the size of p does. Also shown are the actual total variation norms in each case and the first upper bound given in (15). Note that the upper bounds for d_{TV} are very close to p when $\lambda > 1$ and quite conservative (inflated by roughly a factor of 4) compared to the actually achieved d_{TV}. For the first plot in Figure 17 we have $\lambda = .5$ and the bound on d_{TV} is different from p and is not quite as conservative compared to the actually achieved value of d_{TV}.



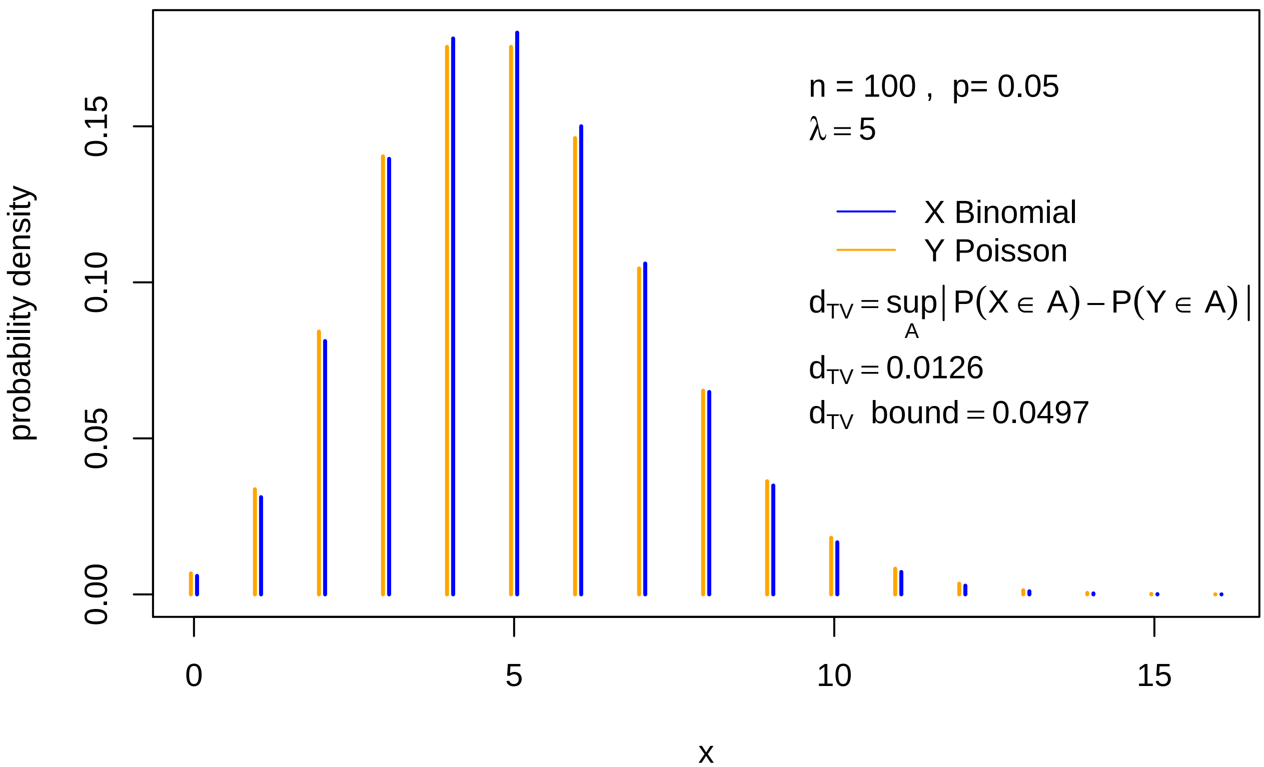
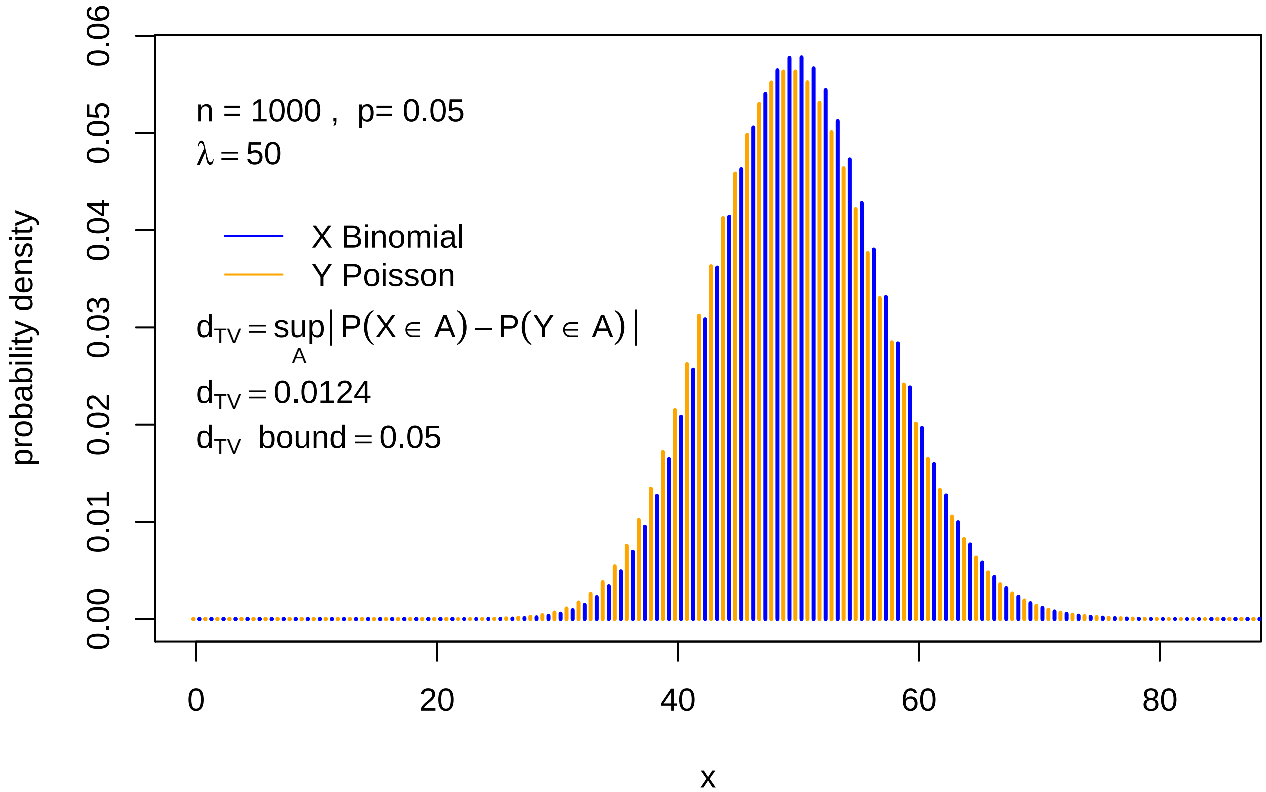


Figure 17: Poisson Approximation to Binomial(n, p): $p = .05, n = 10, 100$



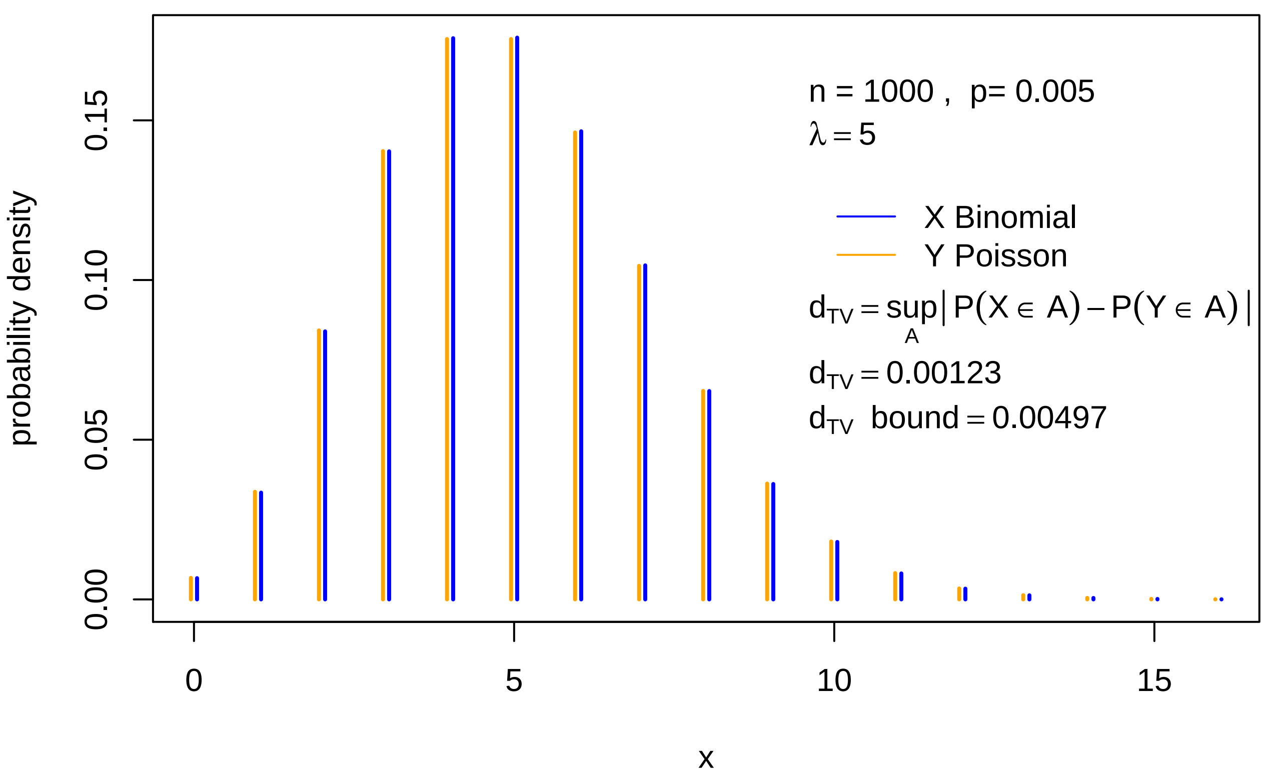


Figure 18: Poisson Approximation to Binomial(n, p): $p = .05, .005, n = 1000$

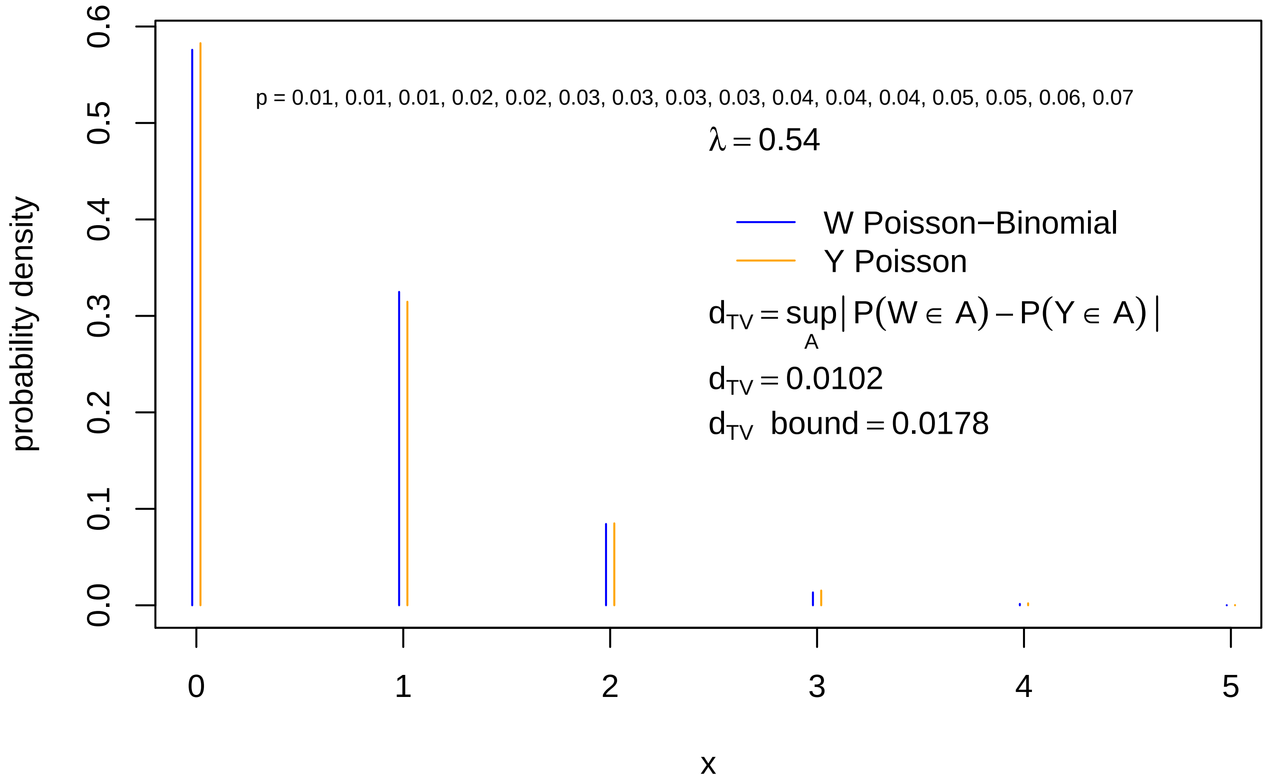


Figure 19: Poisson Approximation to Poisson-Binomial

In R a program to recursively compute the Poisson-Binomial probabilities $P(W = k)$ is quite short:

```
poisbin = function (k,p)
{
#================================================================
# this function computes the probability of k success in n trials
# with success probabilities given in the n-vector p.
#================================================================
  if(length(p)==0 | length(k)==0) return("invalid length of k or p")
  if(min(p*(1-p))<0) return("invalid p")
  if(k-round(k,0)>0) return(0)
  if(length(p)==1){
        if(k==0|k==1) {p^k*(1-p)^(1-k)}else{0}
     }else{
        p[1]*poisbin(k-1,p[-1])+(1-p[1])*poisbin(k,p[-1])
  }
}
```

However, note that there are about 2^n recursions, where n denotes the length of the vector p. Thus the time and memory requirements grow beyond practical limits very fast. Figure 19 shows the results for $n = 16$ with the following probabilities: .01, .01, .01, .02, .02, .03, .03, .03, .03, .04, .04, .04, .05, .05, .06, .07.

5.1 Upper Bounds for λ

Consider testing the hypothesis $H(\lambda_0) : \lambda = \lambda_0$ against the alternatives $A(\lambda_0) : \lambda < \lambda_0$.
Small values x of X can be viewed as evidence against the hypothesis $H(\lambda_0)$ and we would reject this hypothesis at significance level α when the p-value $p(x, \lambda_0) = P_{\lambda_0}(X \leq x) \leq \alpha$. Again we exploit the basic duality between testing hypotheses and confidence sets by defining the confidence set $C(x)$ as the set of all acceptable hypothesis values λ_0, i.e.,

$$C(x) = \{\lambda_0 : p(x, \lambda_0) = P_{\lambda_0}(X \leq x) > \alpha\} \ .$$

Treating this family of confidence sets $C(x), x = 0, 1, 2, \ldots$ equivalently as a random set $C(X)$ we have again the following coverage probability property

$$P_{\lambda_0}(\lambda_0 \in C(X)) = 1 - P_{\lambda_0}(\lambda_0 \notin C(X)) \geq 1 - \alpha = \gamma$$

for any $\lambda_0 > 0$. Equality is achieved for some values of λ_0. Since $P_\lambda(X \leq x)$ is continuous in λ and strictly decreasing from 1 to 0 it follows that $C(x)$ takes the following more explicit form $[0, \hat{\lambda}_U(\gamma, x))$ where $\hat{\lambda}_U(\gamma, x)$ is the unique value of λ that solves

$$P_\lambda(X \leq x) = \sum_{i=0}^{x} \frac{\exp(-\lambda)\lambda^i}{i!} = \alpha = 1 - \gamma \ . \tag{16}$$

Thus we can treat $\hat{\lambda}_U(\gamma, X)$ as a $100\gamma\%$ upper confidence bound for λ.
Rather than solving equation(16) for λ we exploit the identity (14) and get this upper bound as the $1 - \alpha = \gamma$ quantile of the Gamma distribution with scale 1 and shape parameter $x + 1$, also known as the incomplete Gamma function with parameter $x + 1$.
In Excel $\hat{\lambda}_U(\gamma, x)$ may be obtained by invoking `GAMMAINV`$(\gamma, x + 1, 1)$ and in R or S-Plus by the command `qgamma(`γ`,`$x + 1$`)`.
As a check example use the case $x = 2$ with $\gamma = .95$. Then $\hat{\lambda}_U(.95, 2) = 6.295794$ is obtained as 95% upper bound for λ.
For the special case $x = 0$ there is an explicit formula for the upper bound, namely $\hat{\lambda}_U(\gamma, 0) = -\log(1 - \gamma)$. For $\gamma = .95$ this amounts to $\hat{\lambda}_U(.95, 0) = 2.995732 \approx 3$, another instance of the Rule of Three.

5.2 Lower Bounds for λ

Here we consider testing $H(\lambda_0) : \lambda = \lambda_0$ against the alternatives $A(\lambda_0) : \lambda > \lambda_0$. Large values of X would serve as evidence against the hypothesis $H(\lambda_0)$. Upon observing $X = x$ we would reject $H(\lambda_0)$ whenever the p-value $p(x, \lambda_0) = P_{\lambda_0}(X \geq x) \leq \alpha$. For any observable x we define the confidence set $C(x)$ as consisting of all λ_0 corresponding to acceptable hypotheses $H(\lambda_0)$, i.e.,

$$C(x) = \{\lambda_0 : p(x, \lambda_0) = P_{\lambda_0}(X \geq x) > \alpha\} \ .$$

This collection of confidence sets $C(x), x = 0, 1, 2, \ldots$ when viewed equivalently as a random set $C(X)$ has the desired coverage probability

$$P_{\lambda_0}(\lambda_0 \in C(X)) = 1 - P_{\lambda_0}(\lambda_0 \notin C(X)) \geq 1 - \alpha = \gamma$$

for any $\lambda_0 > 0$. The inequality $\geq$ becomes an equality for some λ_0. Thus the confidence coefficient $\bar{\gamma}$ of $C(X)$ is γ.

Since $P_{\lambda_0}(X \geq 0) = 1$ for all λ_0 we see that $C(0) = (0, \infty)$. For $x > 0$ the probability $P_\lambda(X \geq x)$ is continuous in λ and strictly increasing from 0 to 1. It follows that $C(x)$ then takes the following form $(\hat{\lambda}_L(\gamma, x), \infty)$ where $\hat{\lambda}_L(\gamma, x)$ is the unique value of λ which solves

$$P_\lambda(X \geq x) = \sum_{i=x}^{\infty} \frac{\exp(-\lambda)\lambda^i}{i!} = 1 - \gamma \ . \tag{17}$$

For consistency we define $\hat{\lambda}_L(\gamma, 0) = 0$ to agree with the above form of $C(0)$.

Thus we can treat $\hat{\lambda}_L(\gamma, X)$ as a $100\gamma\%$ lower confidence bound for λ. In the case $x > 0$, rather than solving equation (17) for λ, we use again the identity (14) and find that it is the $(1-\gamma)$-quantile of the Gamma distribution with scale parameter 1 and shape parameter x.

In Excel it is obtained by invoking `GAMMAINV`$(1 - \gamma, x, 1)$ and in R or S-Plus by the command `qgamma(`$1 - \gamma$`,`x`)`. For $x = 1$ one can give the formula for the lower bound explicitly as $\hat{\lambda}_L(\gamma, 1) = -\log(\gamma)$, since

$$P_\lambda(X \geq 1) = 1 - P_\lambda(X = 0) = 1 - \exp(-\lambda) = 1 - \gamma \ \Rightarrow \ \exp(-\lambda) = \gamma \ \Rightarrow \ \lambda = -\log(\gamma) \ .$$

As a check example use the case $k = 30$ with $\gamma = .95$. Then one gets $\hat{\lambda}_L(.95, 30) = 21.59399$ as 95% lower bound for λ.

5.3 Confidence intervals

In order to combine lower and upper bounds for use as a $\gamma = 1 - \alpha$ level confidence interval $(\hat{\lambda}_L(1 - \alpha/2, x), \hat{\lambda}_U(1 - \alpha/2, x))$ for λ we need to show that $\hat{\lambda}_L(1 - \alpha/2, x) < \hat{\lambda}_U(1 - \alpha/2, x)$ for all $x = 0, 1, 2, \ldots$ for $0 < \alpha < 1$.

Suppose to the contrary that $\hat{\lambda}_L(1 - \alpha/2, x) \geq \hat{\lambda}_U(1 - \alpha/2, x)$ for some x. Then there exists a λ_0 with $\hat{\lambda}_L(1 - \alpha/2, x) \geq \lambda_0 \geq \hat{\lambda}_U(1 - \alpha/2, x)$. For that λ_0 this x would make us reject $H(\lambda_0)$ when testing against $A(\lambda_0) : \lambda > \lambda_0$ or when testing it against $\tilde{A}(\lambda_0) : \lambda < \lambda_0$. Thus we must have

$$P_{\lambda_0}(X \geq x) \leq \alpha/2 \quad \text{and} \quad P_{\lambda_0}(X \leq x) \leq \alpha/2$$

and adding these two inequalities we get

$$1 > \alpha/2 + \alpha/2 \geq P_{\lambda_0}(X \geq x) + P_{\lambda_0}(X \leq x) = 1 + P_{\lambda_0}(X = x) > 1$$

which leaves us with a contradiction. Hence our assumption can't be true.
Using $\hat{\lambda}_L(1-\alpha/2, x) < \hat{\lambda}_U(1-\alpha/2, x)$ for all $x = 0, 1, 2, \ldots$ for $0 < \alpha < 1$ we have

$$\begin{aligned}
&P_\lambda(\hat{\lambda}_L(1-\alpha/2, x) < \lambda < \hat{\lambda}_U(1-\alpha/2, x)) \\
&\qquad = 1 - [P_\lambda(\lambda \leq \hat{\lambda}_L(1-\alpha/2, x) \cup \hat{\lambda}_U(1-\alpha/2, x) \leq \lambda)] \\
&\qquad = 1 - [P_\lambda(\lambda \leq \hat{\lambda}_L(1-\alpha/2, x)) + P_\lambda(\hat{\lambda}_U(1-\alpha/2, x) \leq \lambda)] \\
&\qquad \geq 1 - [\alpha/2 + \alpha/2] = 1 - \alpha\ .
\end{aligned}$$

5.4 Poisson Approximation to Binomial

As shown previously, for very small p the binomial distribution of X can be well approximated by the Poisson distribution with mean $\lambda = np$. Thus confidence bounds for $p = \lambda/n$ can be based on those obtained via the Poisson distribution, namely by using $\hat{\lambda}_U(\gamma, k)/n$ and $\hat{\lambda}_L(\gamma, k)/n$.
A typical application would concern the number X of well defined, rare incidents (crashes or part failures) in n flight cycles in a fleet of airplanes. Here p would denote the probability of such an incident during a particular flight cycle. Typically p is very small and n, as accumulated over the whole fleet, is very large.

5.5 Aircraft Accidents

Accident data on commercial jet aircraft are updated yearly at

`http : //www.boeing.com/news/techissues/pdf/statsum.pdf`

and provide ample opportunities for applying the confidence bound methods developed so far. These reports do not show such bounds. Figure 20 shows a page on hull loss accidents from the 2001 report while Figure 21 shows a version with confidence bounds. Note how the confidence margins tighten as the number of accidents and the number of flights per aircraft model increase.

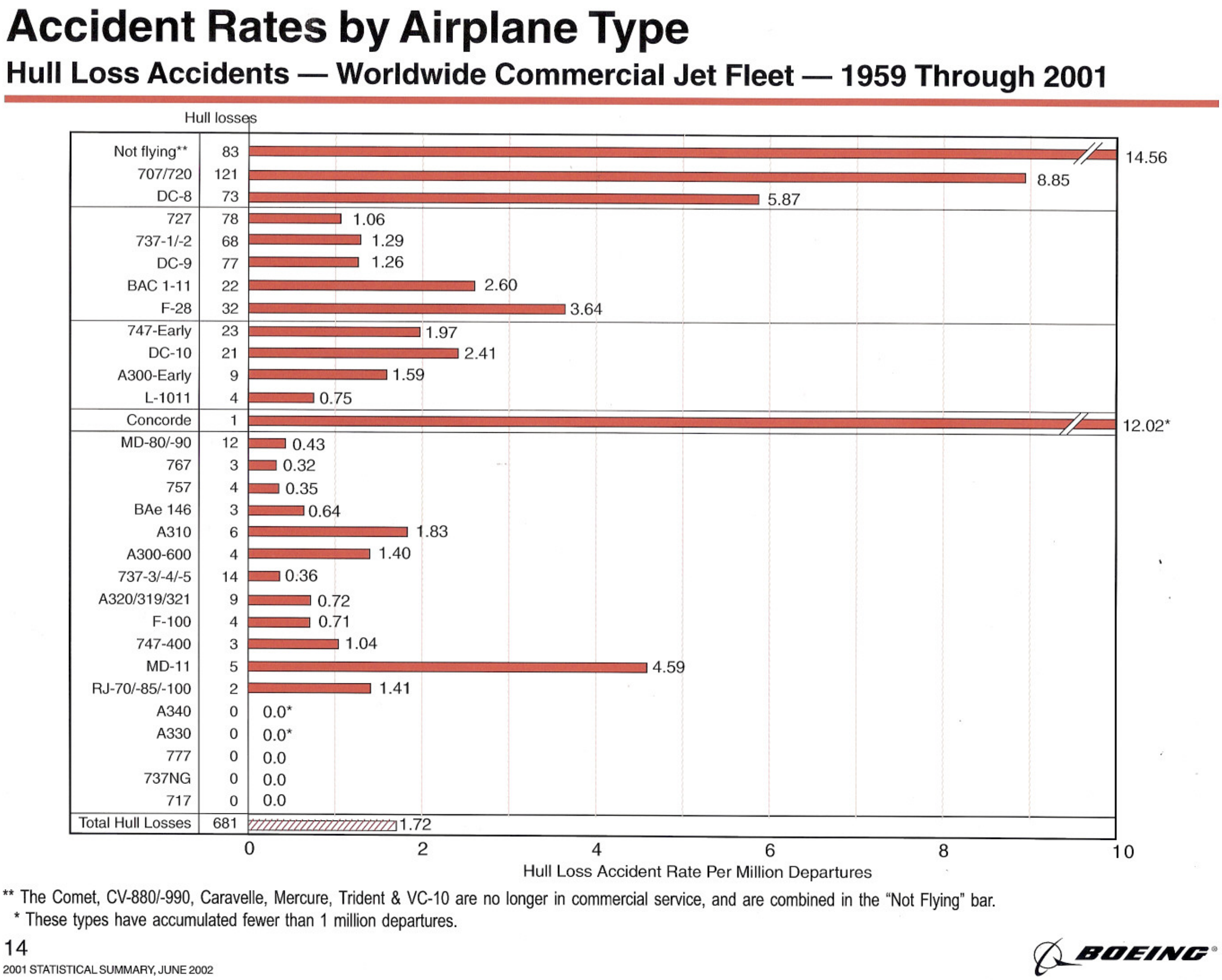


Figure 20: Hull Loss Accident Rates

from Statistical Summary of Commercial Jet Airplane Accidents
Worldwide Operations 1959-2001

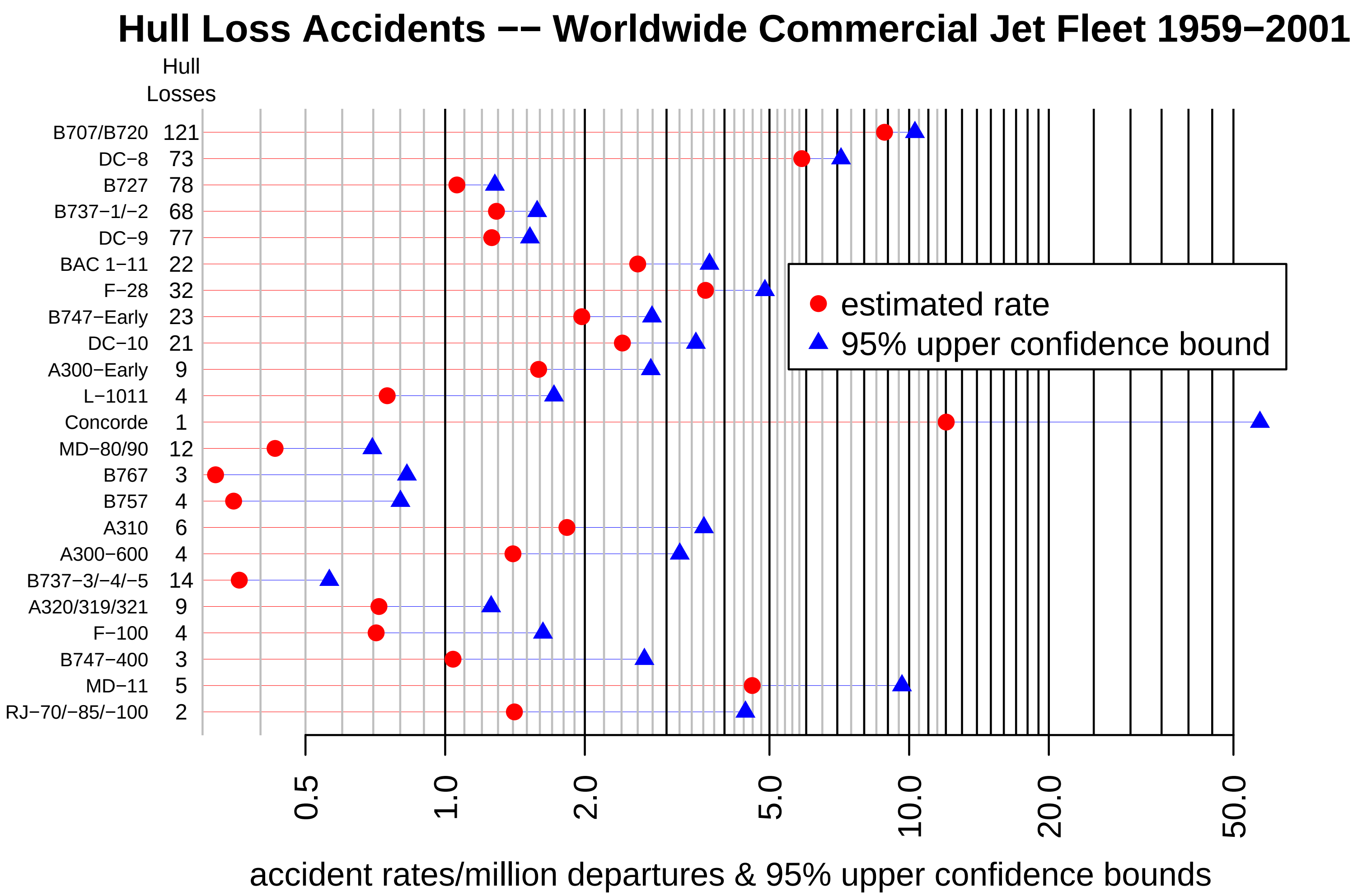


Figure 21: Hull Loss Accident Rates

With 95% Upper Confidence Bounds and on a $\log_{10}$-Scale.

In making such margin comparisons be mindful of the $\log_{10}$ scale. We calculated these 95% level confidence bounds based on the Poisson distribution and based on a binomial model as follows. From the given number x of hull losses and rate ρ (= number of losses per 1 million departures, as given in Figure 20) we backed out a reasonable number n of departures during which the losses were observed for each aircraft model, namely:

$$\frac{\rho}{1000000} = \frac{x}{n} \qquad \Longrightarrow \qquad n = \frac{x \times 1000000}{\rho} \qquad \text{with rounding to the nearest integer.}$$

For example, in the case of the Concorde we have $x = 1$, $\rho = 12.02$, hence $n = 1000000/12.02 = 83194.68 \approx 83195$. Denoting the probability of a hull loss per departure by p and using a Poisson approximation model with $\lambda = np$ we get as 95% upper bound for λ the value

$$\hat{\lambda}(.95, 1) = \texttt{qgamma}(.95, 1 + 1) = 4.743865$$

which converts to an upper bound

$$\hat{p}_U(.95, 1, n) = \frac{\hat{\lambda}(.95, 1)}{n} = \frac{\hat{\lambda}(.95, 1) \times \rho}{1000000} = \frac{57.02125}{1000000}$$

for p or to an upper bound for the rate per million departures of

$$\hat{\rho}_U(.95, 1, n) = 1000000 \times \hat{p}_U(.95, 1, n) = \hat{\lambda}(.95, 1) \times \rho = 57.02125\ .$$

If we use a binomial based approach directly we get

$$\hat{p}_U(.95, 1, n) = \texttt{qbeta}(.95, 1 + 1, \texttt{n} - 1) = 5.701975\texttt{e} - 05 = \frac{57.01975}{1000000}$$

or a rate of $\hat{\rho}_U(.95, 1, n) = 57.01975$ per million departures, which agrees with the previous value quite well. In fact, all such computations for binomial based 95% upper bounds and corresponding Poisson based upper bounds agree to two decimal places and thus it suffices to show just one upper bound column in Table 1.

Table 1: Hull Loss Rates per Million Departures
& 95% Upper Confidence Bounds

Aircraft Model	Hull Losses	ρ	$\hat{\rho}(.95)$
B707/B720	121	8.85	10.29
DC-8	73	5.87	7.13
B727	78	1.06	1.28
B737-1/-2	68	1.29	1.58
DC-9	77	1.26	1.52
BAC 1-11	22	2.60	3.71
F-28	32	3.64	4.89
B747-Early	23	1.97	2.79
DC-10	21	2.41	3.47
A300-Early	9	1.59	2.77
L-1011	4	0.75	1.72
Concorde	1	12.02	57.02
MD-80/90	12	0.43	0.70
B767	3	0.32	0.83
B757	4	0.35	0.80
A310	6	1.83	3.61
A300-600	4	1.40	3.20
B737-3/-4/-5	14	0.36	0.56
A320/319/321	9	0.72	1.26
F-100	4	0.71	1.62
B747-400	3	1.04	2.69
MD-11	5	4.59	9.65
RJ-70/-85/-100	2	1.41	4.44

6 Hypergeometric Distribution: Upper and Lower Bounds for D

When sampling n items without replacement from a finite population of N items containing D defective and $N-D$ non-defective items we observe the number X of defective items in the sample. Rather than focusing on defective items they may be special items of interest.

A simple counting argument gives the probability function $p(x)$ of the integer valued random variable X as

$$p(x) = P_D(X=x) = \frac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}} \qquad \text{for} \qquad \max(n-N+D,0) = D_m \le x \le D_M = \min(D,n)$$

with $p(x)=0$ otherwise. Its cumulative distribution function is

$$F_D(x) = P_D(X \le x) = \sum_{i=D_m}^{x} p(i) \qquad \text{for} \qquad D_m \le x \le D_M$$

while $F_D(x)=0$ for $x<D_m$ and $F_D(x)=1$ for $x \ge D_M$. On intuitive grounds it can be expected that $F_D(y)$ decreases or that $P_D(X \ge x)$ increases as D increases.

This is seen rigorously as follows. Think of an urn containing D red balls, 1 pink ball, and the remaining $N-D-1$ balls are white. Let X denote the number of red balls in the grab when the pink ball is viewed as white and let Y denote the number of red balls in the grab when the pink ball is viewed as red. There are $\binom{N}{n}$ possible grabs or samples of size n from this urn. We divide these different grabs into those (set A) that contain the pink ball and those (set B) that do not. We want to show $P_D(X \ge x) < P_D(Y \ge x) = P_{D+1}(X \ge x)$ provided $D_m + 1 \le x \le D_M$.

For all grabs $\omega \in B$ we have $X(\omega) = Y(\omega)$. For all grabs $\omega \in A$ we have $Y(\omega) = X(\omega)+1$.

$$\text{For} \qquad D_m + 1 \le x \le D_M \quad \Longrightarrow \quad P_D\left(\{X(\omega) = x-1\} \cap A\right) > 0$$

since for a grab $\omega \in \{X(\omega)=x\} \cap A$ we have not yet reached the minimum number of red balls, thus we can replace a red ball in ω with a white one from $\{\omega\}^c$, giving us the positive probability.

$$\Longrightarrow \quad P_D\left(\{X(\omega) \ge x\} \cap A\right) < P_D\left(\{X(\omega) \ge x-1\} \cap A\right)$$

$$\begin{aligned}
\Longrightarrow \quad P_D(X \ge x) &= P_D(\{\omega : X(\omega) \ge x\} \cap A) + P_D(\{\omega : X(\omega) \ge x\} \cap B) \\
&= P(\{\omega : X(\omega) \ge x\} \cap A) + P(\{\omega : Y(\omega) \ge x\} \cap B) \\
&< P_D(\{\omega : X(\omega) \ge x-1\} \cap A) + P_D(\{\omega : Y(\omega) \ge x\} \cap B) \\
&= P_D(\{\omega : Y(\omega) \ge x\} \cap A) + P_D(\{\omega : Y(\omega) \ge x\} \cap B) \\
&= P_D(Y \ge x) = P_{D+1}(X \ge x)
\end{aligned}$$

This argument aligns with the intuition and is more transparent than the usual abstract arguments that build on the monotone likelihood ratio (MLR) property underlying all of the strict monotonicity statements in this document. For this broader and more general treatment look at Lehmann and Romano (2005), Lemma 3.4.2 (p. 70). That Lemma only shows monotonicity but not strict monotonicity. Strict monotonicity is a byproduct of Theorem 3.4.1 (p. 65) in the same reference. However, that theorem is based on a fair amount of precursor machinery.

6.1 Upper Bounds for D

Following closely the derivation of the upper bound for p in the binomial case we view a small value of X as evidence against the hypothesis $H(D_0) : D = D_0$ when testing it against the alternative $A(D_0) : D < D_0$. Thus we reject $H(D_0)$ when the p-value $p(x, D_0) = P_{D_0}(X \le x) \le \alpha = 1-\gamma$. We define again the corresponding confidence set as all acceptable hypotheses $H(D_0)$ at level α, i.e.,

$$C(x) = \{D_0 : p(x, D_0) = P_{D_0}(X \le x) > \alpha\} \ .$$

Viewing the collection of all such confidence sets $C(x), x = 0, 1, \ldots, n$ as a random set $C(X)$ we have the following coverage probability

$$P_{D_0}(D_0 \in C(X)) = 1 - P_{D_0}(D_0 \notin C(X)) \ge 1 - \alpha = \gamma \ .$$

Because of the discrete nature of the parameter D the minimum coverage probability is usually not equal to the target value of γ for any D_0.

For $x < n$ this p-value $P_{D_0}(X \le x)$ is strictly decreasing in D_0, decreasing from 1 to 0. It follows that there is a unique largest D_0 for which $p(x, D_0) = P_{D_0}(X \le x) > \alpha$. We denote this value as $\hat{D}_U(\gamma, x)$. For $x = n$ we have $p(n, D_0) = P_{D_0}(X \le n) = 1 > \alpha$ for all D_0, thus $C(n) = [0, N]$ and we define $\hat{D}_U(\gamma, n) = N$ in that case. The confidence sets thus take the following form $C(x) = [0, \hat{D}_U(\gamma, x)]$. Note the closed form of the interval here.

Since we automatically have $x \le D$, i.e., the number of defective items in the population must be at least as large as what was seen in the sample, we modify the above upper bound confidence sets into intervals of the following type $[x, \hat{D}_U(\gamma, x)]$. This does not affect the confidence statement of the modified set, since the lower bound x on D is for sure. Such intervals can be computed by invoking the function `hypergeo.conf` provided at

`http://www.stat.washington.edu/fritz/Stat498B.html`

As a check example take $N = 2500$, $n = 50$, $x = 11$ and $\gamma = .95$ and obtain

```
> hypergeo.conf(11,50,2500,.95,type="upper",cc.flag=T)
$bounds
lower upper
   11   841

$confidence
  nominal   minimum
0.9500000 0.9500011
```

i.e., $\hat{D}_U = \hat{D}_U(\gamma, x) = 841$ as 95% upper confidence bound for D. However, we also get the automatic lower bound $x = 11$. Note that `hypergeo.conf` also computes the confidence coefficient (minimum coverage probability) when the logic flag `cc.flag=T` is set. Here it is 0.9500011, very close to the nominal level of .95.
This upper bound $\hat{D}_U$ yields in turn a 95% upper bound of $\hat{D}_U/N = 0.3364$ for the population proportion $p = D/N$ of defective items. As a contrast, the upper bound for p based on the binomial distribution (effectively assuming $N = \infty$ or sampling with replacement) computes to $\hat{p}_U = 0.3378$, which is not much different. This is understandable because $n = 50$ is relatively small compared to $N = 2500$, so that sampling without replacement is close to sampling with replacement.

6.2 Lower Bounds for D

The lower bound for D is obtained in dual fashion. When testing $H(D_0) : D = D_0$ against the alternative $A(D_0) : D > D_0$ high values of X can be viewed as evidence against the hypothesis $H(D_0)$. Upon observing $X = x$ we reject $H(D_0)$ at level α whenever the p-value $p(x, D_0) = P_{D_0}(X \geq x) \leq \alpha$. The corresponding confidence set is

$$C(x) = \{D_0 : p(x, D_0) = P_{D_0}(X \geq x) > \alpha\}$$

which again has coverage probability $\geq 1 - \alpha = \gamma$ for all D_0, possibly $> \gamma$ for all D_0, because of the discrete nature of the parameter D_0.
For $x > 0$ the p-value $P_{D_0}(X \geq x)$ is strictly increasing in D_0, from 0 to 1, as long as that probability is less than 1, it follows that there is a smallest value of D_0 for which $P_{D_0}(X \geq x) > \alpha$. Denote it by $\hat{D}_L(\gamma, x)$. For $x = 0$ we have $P_{D_0}(X \geq 0) = 1$ for all D_0, and in that case we define $\hat{D}_L(\gamma, 0) = 0$. Thus $C(x)$ takes the following form $[\hat{D}_L(\gamma, x), N]$, which again has a closed form.
Thus we can consider $\hat{D}_L = \hat{D}_L(\gamma, X)$ as a $100\gamma\%$ lower confidence bound for D. We can improve on the upper bound statement N by replacing it by $N - (n - x)$, since we saw $n - x$ non-defective

items in the sample. This limits the number D in the population by $N - (n - x)$ from above for sure, without impacting the confidence level of the lower bound.
As a check example take $N = 2500$, $n = 50$, $x = 11$, and $\gamma = .95$ and obtain

```
> hypergeo.conf(11,50,2500,.95,type="lower",cc.flag=T)
$bounds
lower upper
  324  2461

$confidence
  nominal   minimum
0.9500000 0.9500011
```

i.e., $\hat{D}_L = \hat{D}_L(\gamma, 11) = 324$ as 95% lower confidence bound for D. The automatic upper bound on D is $2500 - (50 - 11) = 2461$. Again the confidence coefficient $\bar{\gamma}$ was requested and is given as 0.9500011.
This lower bound converts to a 95% lower confidence bound of $324/2500 = .1296$ for $p = D/N$. The corresponding binomial based lower confidence bound (effectively assuming $N = \infty$ or sampling with replacement) computes to $\hat{p}_L = 0.1286$, which is not much different. This is again understandable because $n = 50$ is relatively small compared to $N = 2500$, so that sampling without replacement is close to sampling with replacement.

6.3 Confidence Intervals for D

If $\hat{D}_L(1-\alpha/2, x) \le \hat{D}_U(1-\alpha/2, x)$ for any x and $\alpha < 1$ we may use $[\hat{D}_L(1-\alpha/2, X), \hat{D}_U(1-\alpha/2, X)]$ as a confidence interval with coverage probability $\ge \gamma$, since

$$\begin{aligned}
&P_D(\hat{D}_L(1-\alpha/2, X) \le D \le \hat{D}_U(1-\alpha/2, X)) \\
&\quad = \; 1 - P_D(D < \hat{D}_L(1-\alpha/2, X) \cup \hat{D}_U(1-\alpha/2, X) < D) \\
&\quad = \; 1 - [P_D(D < \hat{D}_L(1-\alpha/2, X)) + P_D(\hat{D}_U(1-\alpha/2, X) < D)] \\
&\quad \ge \; 1 - [\alpha/2 + \alpha/2] = \gamma \, .
\end{aligned}$$

Thus it remains to show that the condition $\hat{D}_L(1-\alpha/2, x) \le \hat{D}_U(1-\alpha/2, x)$ holds quite generally. The argument presented below is different from the corresponding arguments given previously. The reason is that a negation of the above condition statement does not imply that there is a D such

$\hat{D}_U(1-\alpha/2, x) < D < \hat{D}_L(1-\alpha/2, x)$, which precludes us from using the same angle of attack here as was used previously.

Note that $D_0 = \hat{D}_U(1-\alpha/2, x)$ is by definition the largest D such that $P_D(X \leq x) > \alpha/2$ and $D_1 = \hat{D}_L(1-\alpha/2, x)$ is the smallest D such that $P_D(X \geq x) > \alpha/2$ or such that $P_D(X \leq x-1) < 1-\alpha/2$. Thus we have

$$\frac{\alpha}{2} \geq P_{D_0+1}(X \leq x) \geq P_{D_0}(X \leq x-1) < 1 - \frac{\alpha}{2} ,$$

where the last $<$ holds because we can't have $\alpha/2 \geq 1-\alpha/2$, the first $\geq$ follows from the definition of D_0, and the middle $\geq$ holds generally as will be argued below. The inequality $P_{D_0}(X \leq x-1) < 1-\alpha/2$ implies that $D_1 \leq D_0$, using the alternate definition of D_1.

We now argue that

$$P_{D-1}(X \leq x-1) \leq P_D(X \leq x)$$

holds quite generally for all x and D. To show this we use again the urn model with $D-1$ red balls, one pink ball, and $N-D$ white balls. The set A consists of all those grabs of n balls that contain the pink ball and B consists of all grabs of n balls that don't contain the pink ball. The random variable X counts the number of red balls in the grab, interpreting pink as white, while Y counts the number of red balls in the grab, interpreting pink as red. We have $X(\omega) = Y(\omega)$ for all grabs ω in B and $X(\omega) + 1 = Y(\omega)$ for all $\omega \in A$. Then

$$\begin{aligned}
P_{D-1}(X \leq x-1) &= P(X \leq x-1 \cap A) + P(X \leq x-1 \cap B) \\
&= P(Y \leq x \cap A) + P(Y \leq x-1 \cap B) \\
&\leq P(Y \leq x \cap A) + P(Y \leq x \cap B) \\
&= P(Y \leq x) = P_D(X \leq x) \qquad \text{q.e.d.}
\end{aligned}$$

One consequence of this inequality is that the upper bounds $\hat{D}_U(\gamma, x)$ are strictly increasing in x. To see this, recall that $\hat{D}_U(\gamma, x)$ is the largest D for which $P_D(X \leq x) > \alpha = 1-\gamma$. Using the above inequality we see that

$$P_{D+1}(X \leq x+1) \geq P_D(X \leq x) > \alpha \implies \hat{D}_U(\gamma, x+1) \geq \hat{D}_U(\gamma, x) + 1 > \hat{D}_U(\gamma, x) .$$

A corresponding property holds for $\hat{D}_L(\gamma, x)$, i.e., it is also strictly increasing with x.

6.4 Coverage Probabilities

As in the binomial case we computed coverage probabilities as a function of D for an example situation, namely for $n = 40$, $N = 400$, and $\gamma = .95$. The results are shown in Figures 22 and 23.

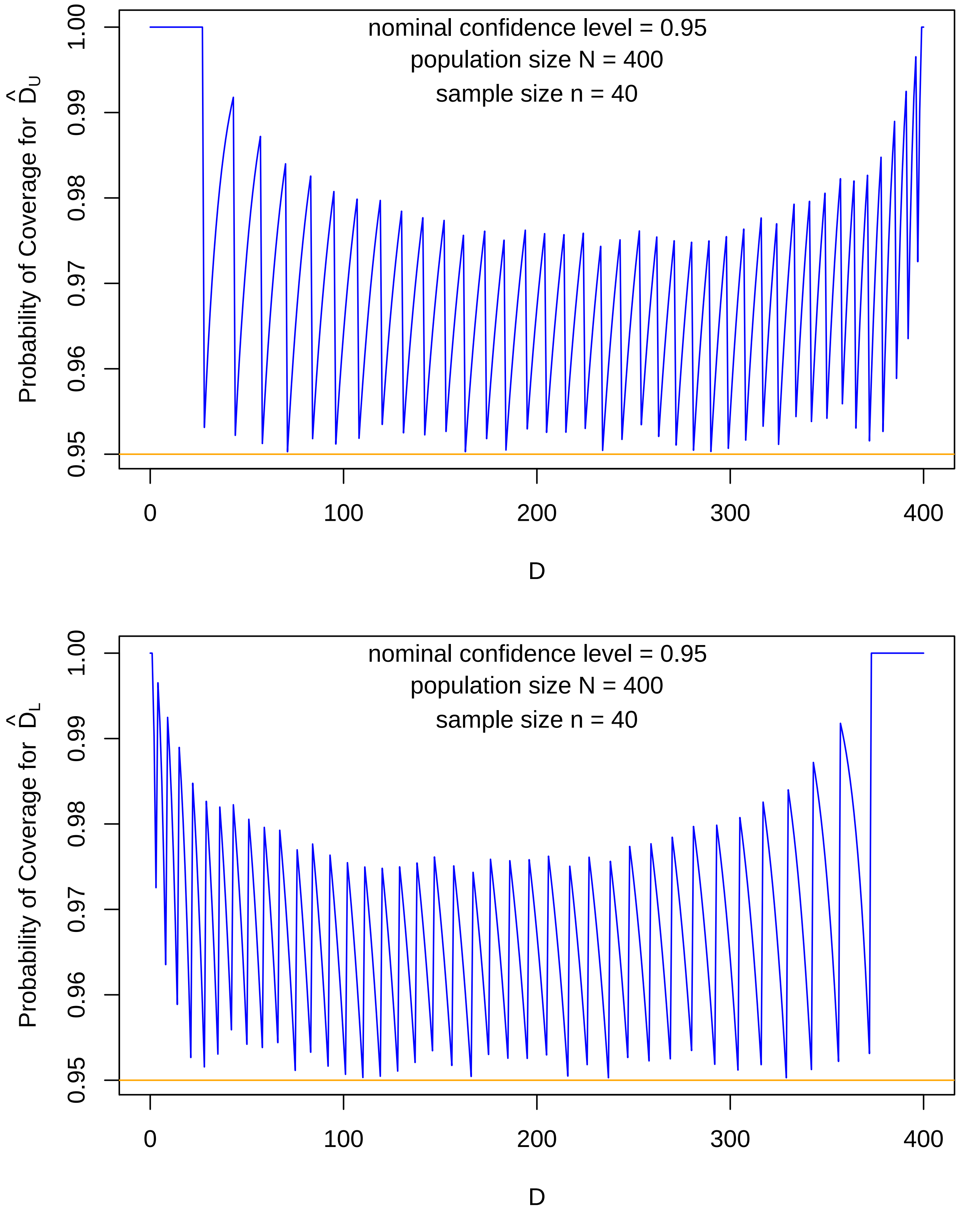


Figure 22: Hypergeometric Confidence Bounds Coverage Probabilities

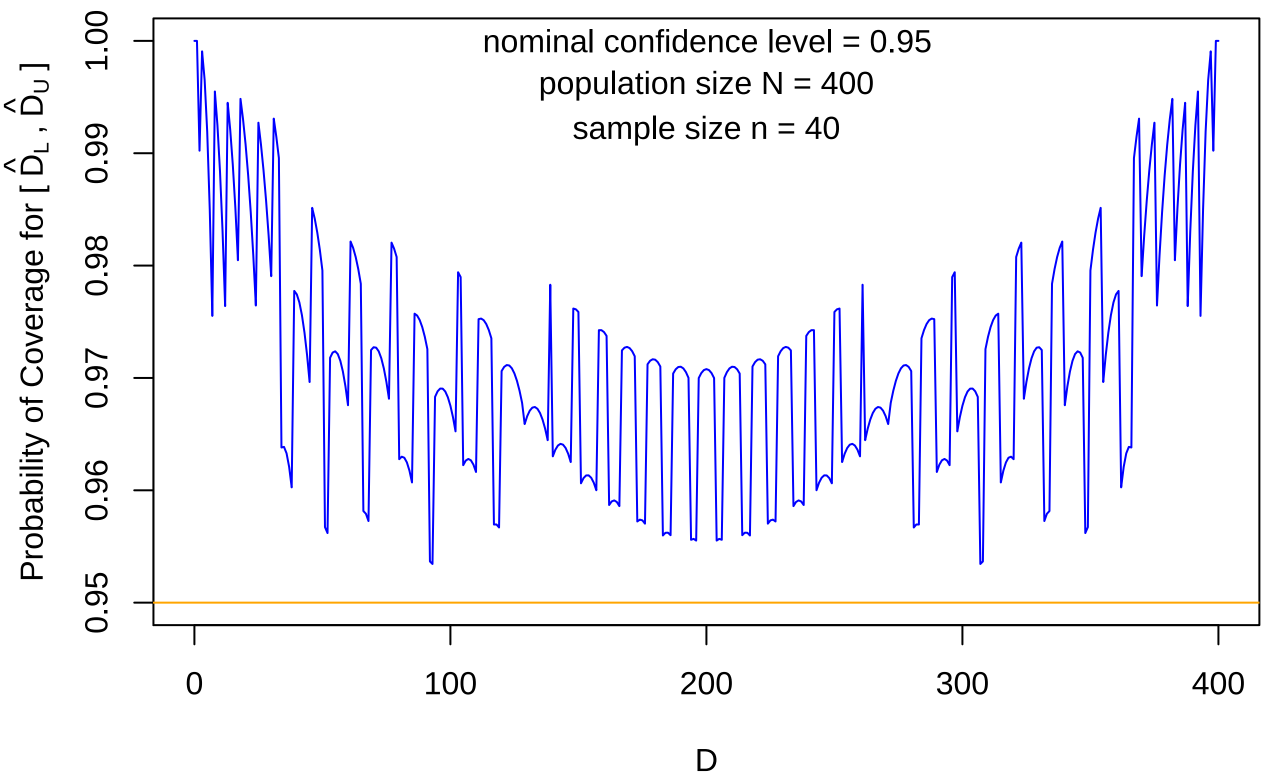


Figure 23: Hypergeometric Confidence Interval Coverage Probabilities

The functions used to produce these plots are

`D.minimum.coverage.upper`, `D.minimum.coverage.lower`, and `D.minimum.coverage.int` .

They can be found at the previously referenced R work space.

As pointed out before, there is no guarantee that the actual minimum coverage probability $\bar{\gamma}$ equals the nominal value of γ, although we know $\bar{\gamma} \geq \gamma$. The reason for this is that the defining inequalities for $\hat{D}_U$ and $\hat{D}_L$ may not be attained due to the discrete nature of the parameter D. Figure 22 illustrates this point, where $\bar{\gamma} = 0.9502894 > .95$, but only barely.

Figure 23 shows the corresponding plot for the confidence interval constructed from the appropriate lower and upper bounds for D. Here the minimum coverage probability is 0.9534429, which exceeds the target $\gamma = .95$ by a higher margin. The reason for this larger gap is that the minimum coverage probabilities for lower and upper bound may be attained at different values of D for each respective bound. In fact, the minimum coverage probability for an interval may be attained at a D that does not coincide with the minimum location for either one-sided bound.

In the case of one-sided bounds it can be shown that there is a distinct set of γ values for which the minimum coverage probability $\bar{\gamma}$ equals γ. We will show this here for upper bounds.

For each set of integers $D^\star$ and $x^\star$ with

$$0 < P_{D^\star+1}(X \leq x^\star) = \alpha(D^\star, x^\star) = \alpha^\star < 1$$

we define $\gamma(D^\star, x^\star) = \gamma^\star = 1 - \alpha^\star$. For such $\gamma^\star$ it follows that the corresponding upper bound at $x^\star$ equals $D^\star$, i.e.,

$$\hat{D}_U(\gamma^\star, x^\star) = D^\star \quad \text{since} \quad P_{D^\star+1}(X \leq x^\star) = \alpha^\star \quad \text{and} \quad P_{D^\star}(X \leq x^\star) > \alpha^\star \ ,$$

where the last $>$ comes from the strict monotonicity property of $P_D(X \leq x)$ with respect to D. Since $\hat{D}_U(\gamma^\star, X)$ is strictly increasing in X and since $\hat{D}_U(\gamma^\star, x^\star) = D^\star$ we have

$$P_{D^\star+1}\left(D^\star + 1 > \hat{D}_U(\gamma^\star, X)\right) = P_{D^\star+1}\left(D^\star \geq \hat{D}_U(\gamma^\star, X)\right) = P_{D^\star+1}\left(X \leq x^\star\right) = \alpha^\star$$

and hence

$$P_{D^\star+1}\left(D^\star + 1 \leq \hat{D}_U(\gamma^\star, X)\right) = \gamma(D^\star, x^\star) \implies \bar{\gamma}(D^\star, x^\star) = \gamma^\star \ .$$

Figure 24 shows $\bar{\gamma}$ in relation to γ for upper (and lower) bounds over a γ range of $[.97, .98]$ evaluated over 1000 equally spaced points using the function `hypergeo.conf.plot`. The discrepancy between $\bar{\gamma}$ and γ appears to be small. Thus the discrete nature of the parameter D does not appear to have a strong effect on the difference $\bar{\gamma} - \gamma$, at least not for the case $n = 40$, $N = 400$ examined here.

For intervals Figure 25, again produced by `hypergeo.conf.plot`, shows the corresponding relationship between $\bar{\gamma}$ and γ over the interval $[.94, .96]$. It is again a step function but with coarser steps. This may seem surprising at first but the explanation comes from the fact that the minimum coverage probabilities for the one-sided bounds do not necessarily occur at the same D. Thus the wider zigzag behavior in the coverage probabilities for the one-sided bounds comes into play.

This suggests the possibility of manipulating the nominal γ until $\bar{\gamma}$ takes on the desired value, e.g., choose γ somewhat less than .95 in order to attain $\bar{\gamma} = .95$ or at least to come closer to it. This may involve some trial and error. However, it may ultimately lead to making some form of compromise.

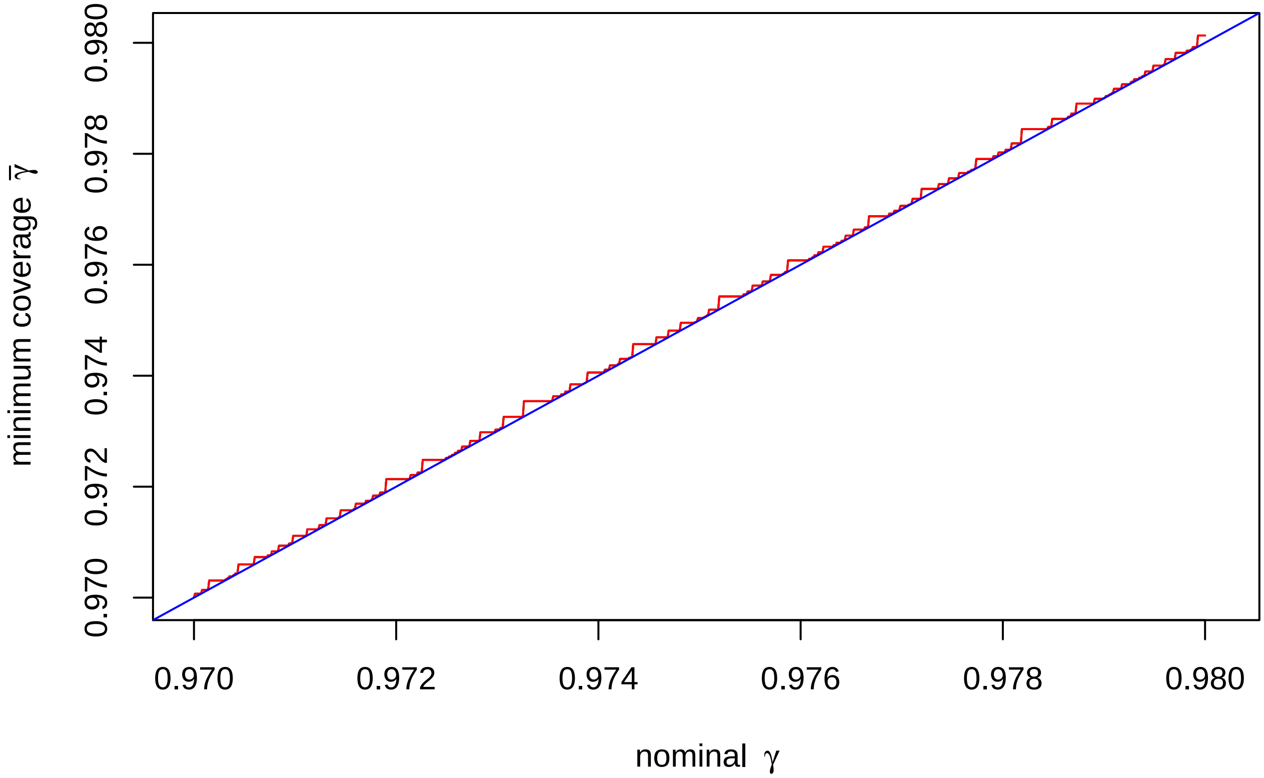


Figure 24: Confidence Coefficient $\bar{\gamma}$ as a Function of γ for Upper/Lower Bounds, $n = 40$, $N = 400$

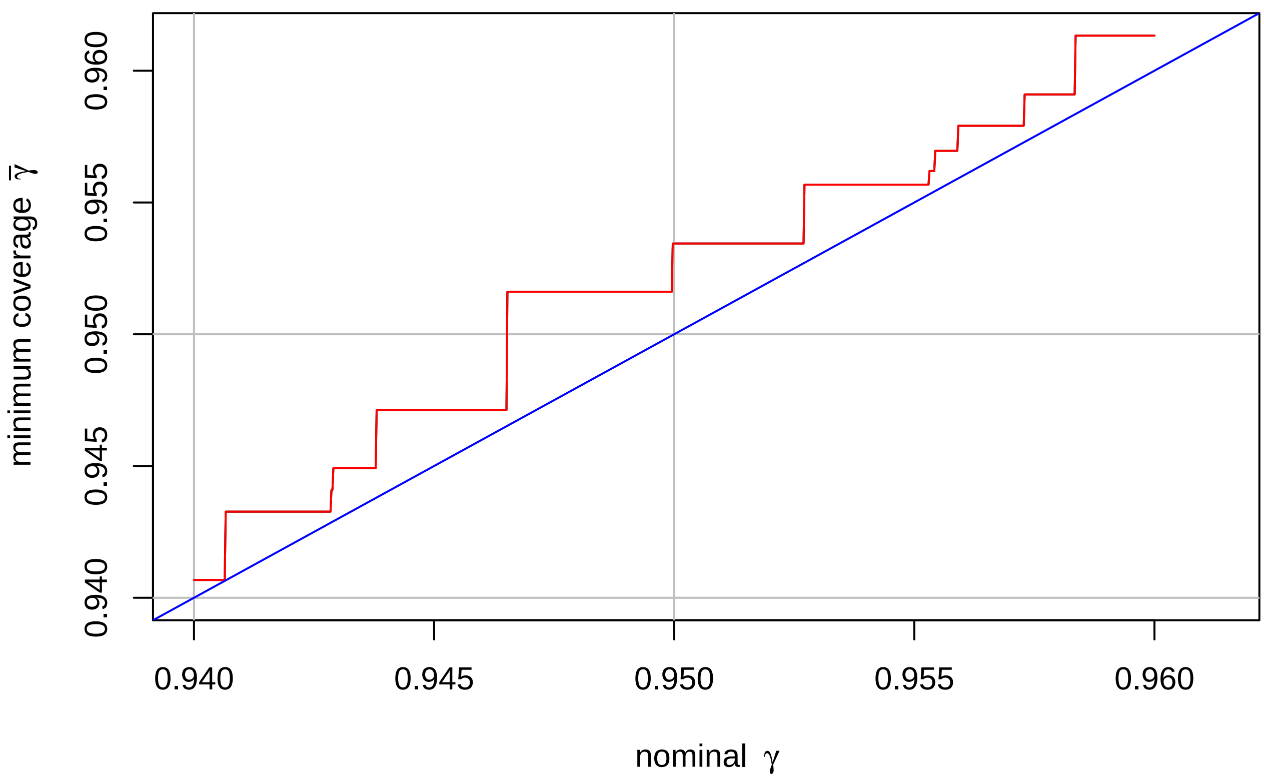


Figure 25: Confidence Coefficient $\bar{\gamma}$ as a Function of γ for Confidence Intervals, $n = 40$, $N = 400$

6.5 Hypergeometric Confidence Bounds with a Birthday Problem Twist

The classical birthday problem is often presented in elementary probability courses. The chance that n random persons have birthdays on different days in the year can be given as follows

$$P(D) = \frac{(365)_n}{365^n} = \frac{365(365-1)\cdots(365-n+1)}{365^n}$$

where the numerator counts the number of ways that n persons can choose n distinct birthdays and the denominator gives the total number of choices for all n persons without any restrictions. Thus the probability of at least one match is

$$P(M) = 1 - \frac{(365)_n}{365^n} \geq .5 \quad \text{for} \quad n \geq 23.$$

This is based on a non leap year and equal likelihoods of all 365 birthdays for all n persons.

Once one of my students asked the following question: What is the chance that in a group of n random persons we will have either at least one matching or one pair of adjacent birthdays?

We approach this by calculating the probability of the complementary event A: All birthdays in a group of n persons are at least one day apart. Here we would treat December 31 and January 1 as adjacent birthdays. Then

$$P(A) = \binom{365-n-1}{n-1} \frac{(n-1)!\; 365}{365^n} = \frac{(365-2n+1)(365-2n+2)\cdots(365-2n+n-1)}{365^{n-1}}$$

This is argued as follows: There are 365 ways to choose the birthday for person 1, hence the factor 365. There are $365-n$ non-birthdays since all n birthdays are distinct (the n birthdays have to be separated by at least one non-birthday). The position of the non-birthdays relative to the birthdays (without regard to which person gets each birthday, except for person 1 whose birthday is already chosen) is determined by choosing the remaining $n-1$ birthdays to each fill one of the remaining $365-n-1$ gaps between the non-birthdays, hence the factor $\binom{365-n-1}{n-1}$. The $(365-n)^{\text{th}}$ gap between non-birthdays is occupied by the birthday of person 1, who has to be bracketed by a non-birthday on either side. Finally, there are $(n-1)!$ ways to assigns these $n-1$ positioned remaining birthdays to the remaining $(n-1)$ persons, hence the factor $(n-1)!$. The denominator 365^n is as in the classical birthday problem, counting the number of ways that n birthdays can be chosen without any constraint.

Figure 26 illustrates this counting process by visualizing a year of $N = 20$ days with $n = 7$ people involved. Using $N = 365$ would make the plot too crowded and useless.

The smallest n for which we get $P(A^c) \geq .5$ is $n = 14$. In that case we get $P(A^c) = .5375$. Figure 27 gives a visual comparison of these probabilities for matching birthdays and for matching and adjacent birthdays.

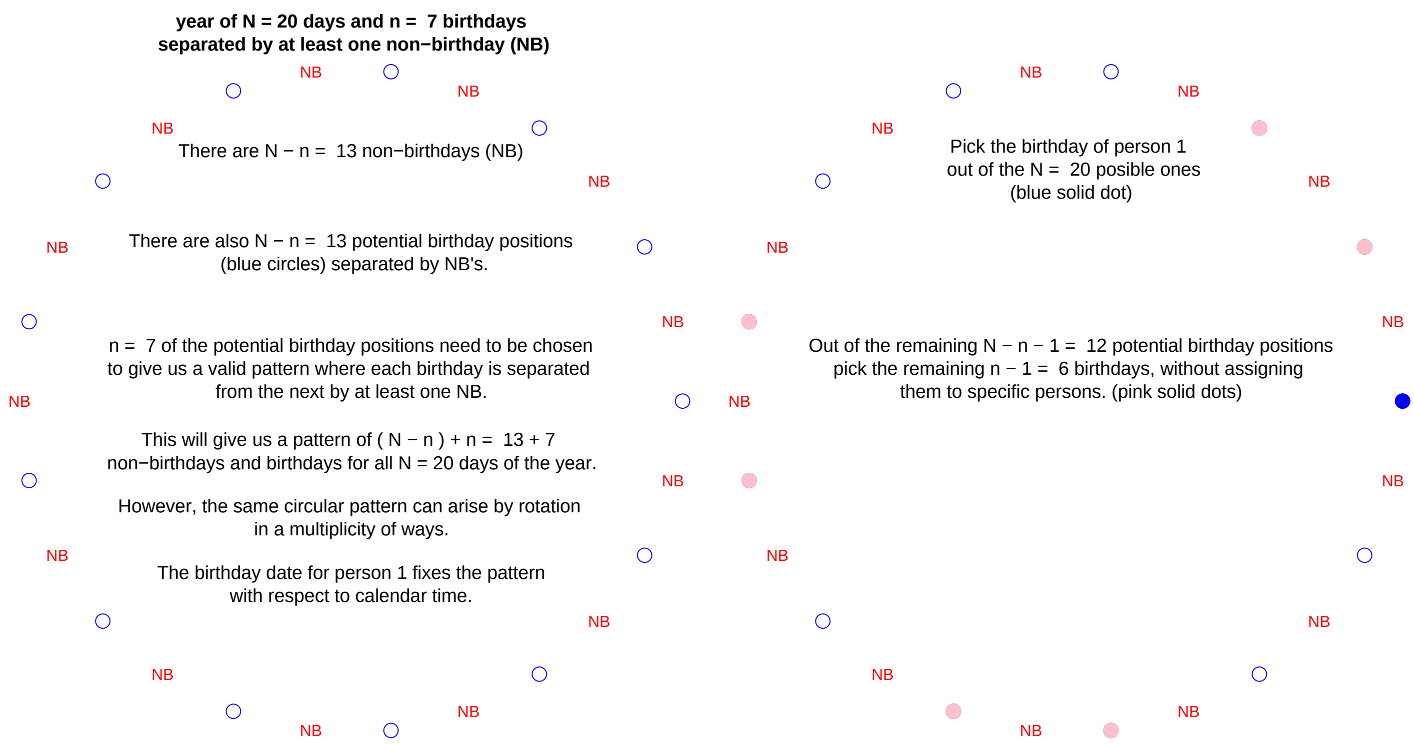


Figure 26: Visualization Diagram

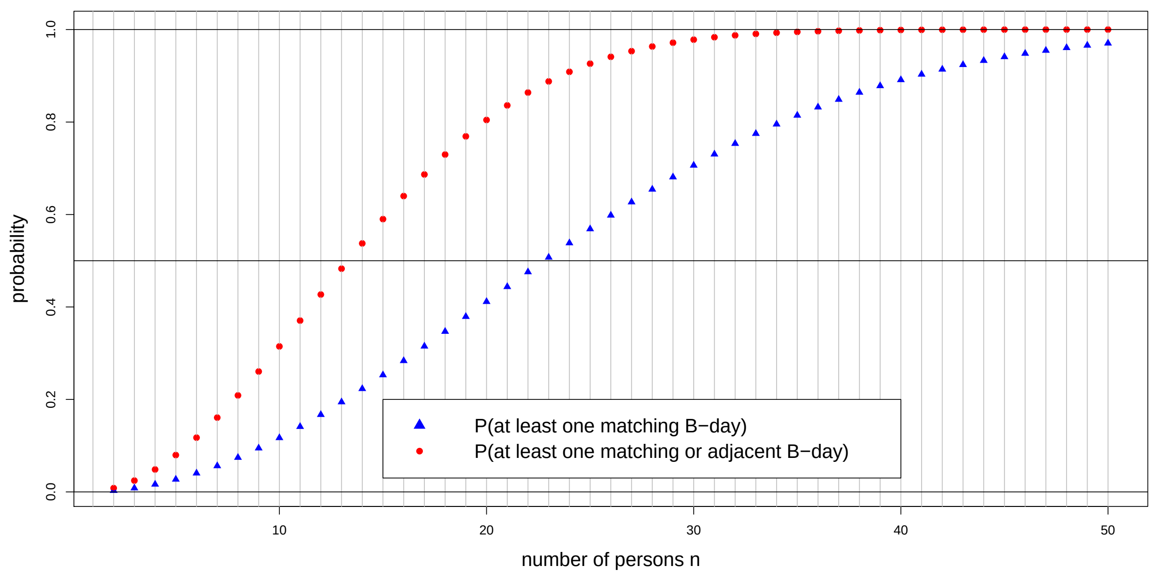


Figure 27: Birthday Problem Probabilities

6.5.1 Problem Description and Assumptions

At some point after dealing with the extension of the birthday problem the following problem arose in connection with the Inertial Upper Stage of the Military Space Program.

$$\texttt{http://www.fas.org/spp/military/program/launch/ius.htm}$$

We are dealing with a population of $N = 200$ parts (rocket motors) of which $n = 12$ are to be tested in some form. It is expected that all tested items will be non-defective. Based on this, what kind of upper confidence bounds $\hat{D}_U$ can we place on the number D of defective parts in the population. Furthermore, if $k = 16$ parts are selected at random from the population of N and are installed as a ship-set, what kind of lower bound $\hat{p}_1$ can be placed on the probability that this ship-set will contain at most one defective?

After addressing this problem it was pointed out that the 16 parts will be arranged in a circular pattern and all that matters is that no two defective parts be placed next to each other. What lower confidence bound $\hat{p}_2$ can be placed on the probability that no two defective parts be placed next to each other in any randomly chosen ship-set of 16?

This does not quite match the birthday problem since we cannot have two defective parts in the same position. However, it was sufficiently close to base the probability calculations on a similar counting trick.

6.5.2 Assumptions and Analysis

To make any statistical inference we have to make certain assumptions. We will assume that the $n = 12$ parts are selected randomly from the population. The same is assumed for the ship-set, where we assume that the $n = 12$ sampled ones were returned to the population before creating the ship-set by random selection. Furthermore, the $k = 16$ parts are arranged in random order in the circular pattern.

Let D denote the total number of defective parts in the population of N parts and X the number of defective items in the tested sample of $n = 12$. In our case we expect to see $X = 0$. Based on such an expected result what kind of upper confidence bounds $\hat{D}_U(\gamma, 0)$ can we place on D?

Recall that $\hat{D}_U(\gamma, 0) = D$ if and only if $P_D(X = 0) > \alpha = 1 - \gamma$ and $P_{D+1}(X = 0) \le \alpha$. By choosing the γ values judiciously as $\gamma_D = 1 - P_{D+1}(X = 0)$, i.e., as high as possible for each value of $D = \hat{D}(\gamma, 0)$, we can associate the highest possible confidence level with each such D. Note that any $\gamma > \gamma_D$ would result in $P_{D+1}(X \le 0) > 1 - \gamma$ and thus in $\hat{D}_U(\gamma, 0) \ge D + 1$. We also point out that this choice of γ_D ensures that $\bar{\gamma} = \gamma_D$, as was shown previously.

This was carried out and the results are displayed in Table 2. There the first column gives the upper confidence bounds on D and columns 2-7 give the respective confidence levels for sample sizes $n = 12, 20, 30, 40, 50, 60$. Column 8 gives the lower bound $\hat{p}_1$ on the probability that we see at most one defective in a shipset of size $k = 16$, with associated confidence levels as shown to the left for the respective strengths of evidence $n = 12, 20, \ldots, 60$ with zero defectives. The last column gives the lower bound $\hat{p}_2$ on the probability that we see no two defectives next to each other in the circular arrangement of the $k = 16$ parts, again with confidence levels as shown to the left for respective n with zero defectives.

The lower bounds $\hat{p}_1$ were arrived at in straightforward manner as follows. Let Y denote the random number of defectives in the shipset sample of $k = 16$. Then the probability of seeing at most 1 defective in that sample is

$$Q(D) = P_D(Y \le 1) = P_D(Y = 0) + P_D(Y = 1) = \frac{\binom{D}{0}\binom{N-D}{k-0}}{\binom{N}{k}} + \frac{\binom{D}{1}\binom{N-D}{k-1}}{\binom{N}{k}} .$$

Since this probability decreases with increasing D we can take $\hat{p}_1 = Q(\hat{D}_U(\gamma, 0))$ as lower bound on $Q(D)$ with same confidence level as was associated with $\hat{D}_U(\gamma, 0)$.

To motivate the lower bound $\hat{p}_2$ we need to derive a formula for the probability that no two defective parts will be placed next to each other in a circular arrangement of $k = 16$ parts. This probability can be expressed as follows

$$\begin{aligned} Q^\star(D) &= P_D(Y = 0) + P_D(Y = 1) + P_D(Y = 2)P(A_2) \\ &\quad + P_D(Y = 3)P(A_3) + \ldots + P_D(Y = 8)P(A_8) , \end{aligned} \tag{18}$$

where A_y is the event that no two defective parts will be placed next to each other when y defective parts and $k - y$ non-defective parts are arranged randomly in a circular pattern. We have

$$P_D(Y = y) = \frac{\binom{D}{y}\binom{N-D}{k-y}}{\binom{N}{k}} \qquad \text{and} \qquad P(A_y) = \frac{\binom{k-y}{y}}{\binom{k-1}{y}} .$$

The latter is arrived at as follows. Label the y defective parts $F_1, \ldots, F_y$ and the $m = k - y$ non-defective parts by $G_1, \ldots, G_m$. There are $k!$ ways of arranging these k parts in a circular pattern, all equally likely. Now we count how many of these patterns have no two F's next to each other and take the ratio of this over $k!$ as our probability $P(A_y)$. We start by realizing that G_1 can be in any of the k places. Next we can arrange the remaining $m - 1$ G's in $(m - 1)!$ ways to form some circular sequence order of G's with keeping only G_1 in its chosen place. The other G's are movable but stay in the circular order chosen.

Table 2: Confidence Bounds Based on Zero Defective Parts in a Sample of n

upper bound	confidence levels γ_D						lower bound	
$\hat{D}_U(\gamma_D, 0) = D$	$n = 12$	$n = 20$	$n = 30$	$n = 40$	$n = 50$	$n = 60$	$\hat{p}_1$	$\hat{p}_2$
1	0.117	0.190	0.278	0.361	0.438	0.511	1.000	1.000
2	0.170	0.272	0.388	0.490	0.580	0.659	0.994	0.999
3	0.221	0.346	0.481	0.594	0.687	0.763	0.983	0.998
4	0.268	0.413	0.560	0.676	0.767	0.836	0.967	0.995
5	0.313	0.473	0.628	0.743	0.827	0.886	0.948	0.992
6	0.356	0.527	0.685	0.796	0.871	0.921	0.925	0.988
7	0.396	0.576	0.734	0.838	0.905	0.946	0.900	0.984
8	0.434	0.620	0.776	0.872	0.929	0.963	0.873	0.978
9	0.469	0.660	0.811	0.899	0.948	0.974	0.845	0.972
10	0.503	0.696	0.841	0.920	0.962	0.983	0.815	0.966
11	0.534	0.728	0.866	0.937	0.972	0.988	0.784	0.958
12	0.564	0.757	0.888	0.950	0.979	0.992	0.753	0.950
13	0.592	0.783	0.906	0.961	0.985	0.994	0.721	0.942
14	0.618	0.806	0.921	0.969	0.989	0.996	0.690	0.932
15	0.643	0.827	0.934	0.976	0.992	0.997	0.658	0.923
16	0.666	0.846	0.944	0.981	0.994	0.998	0.627	0.913
17	0.688	0.863	0.954	0.985	0.996	0.999	0.596	0.902
18	0.709	0.878	0.961	0.989	0.997	0.999	0.566	0.890
19	0.728	0.891	0.968	0.991	0.998	0.999	0.536	0.879
20	0.746	0.904	0.973	0.993	0.998	1.000	0.508	0.867
21	0.763	0.914	0.978	0.995	0.999	1.000	0.480	0.854
22	0.779	0.924	0.981	0.996	0.999	1.000	0.453	0.841
23	0.794	0.933	0.985	0.997	0.999	1.000	0.427	0.828
24	0.808	0.940	0.987	0.998	1.000	1.000	0.401	0.814
25	0.821	0.947	0.989	0.998	1.000	1.000	0.377	0.800
26	0.834	0.953	0.991	0.999	1.000	1.000	0.354	0.786
27	0.845	0.959	0.993	0.999	1.000	1.000	0.332	0.772
28	0.856	0.963	0.994	0.999	1.000	1.000	0.311	0.757
29	0.866	0.968	0.995	0.999	1.000	1.000	0.291	0.742
30	0.876	0.971	0.996	0.999	1.000	1.000	0.272	0.727
31	0.884	0.975	0.997	1.000	1.000	1.000	0.254	0.712
32	0.893	0.978	0.997	1.000	1.000	1.000	0.237	0.696
33	0.900	0.980	0.998	1.000	1.000	1.000	0.220	0.681
34	0.908	0.983	0.998	1.000	1.000	1.000	0.205	0.665
35	0.914	0.985	0.998	1.000	1.000	1.000	0.190	0.649
36	0.921	0.987	0.999	1.000	1.000	1.000	0.177	0.634
37	0.926	0.988	0.999	1.000	1.000	1.000	0.164	0.618
38	0.932	0.990	0.999	1.000	1.000	1.000	0.152	0.602
39	0.937	0.991	0.999	1.000	1.000	1.000	0.141	0.587
40	0.942	0.992	0.999	1.000	1.000	1.000	0.130	0.571
41	0.946	0.993	1.000	1.000	1.000	1.000	0.120	0.555
42	0.950	0.994	1.000	1.000	1.000	1.000	0.111	0.540

Next we place the y F's between the G's so that only one F goes into any one gap between G's. There are m gaps and $\binom{m}{y}$ ways of designating the gaps to receive just one F each and there are $y!$ ways of arranging the order in which the y F's are placed in the designated gaps. All this amounts to the following total count of filling the k spots in the circular pattern so that no two F's are next to each other, namely

$$k(m-1)!\binom{m}{y}y! = k(k-y-1)!\binom{k-y}{y}y! \qquad \Longrightarrow \qquad P(A_y) = \frac{k(k-y-1)!\binom{k-y}{y}y!}{k!} = \frac{\binom{k-y}{y}}{\binom{k-1}{y}}\,.$$

Note that the formula gives the correct and appropriate value 1 in the two cases of $y = 0$ and $y = 1$ and $P(A_y) = 0$ for $y > k - y$. When $y > k - y$ there is no way of placing y defective items in such a way that only one defective item falls between any two non-defective items, since there are only $k - y$ gaps available.

Thus $Q^\star(D)$ in (18) can be written as the expected value $Q^\star(D) = E_D(\psi(Y))$ where $\psi(y) = P(A_y)$. Since $Q^\star(D)$ is again decreasing with increasing D (see the Appendix) we can use $\hat{p}_2 = Q^\star(\hat{D}_U(\gamma, 0))$ as lower bound for $Q^\star(D)$ with same confidence as was applied to $\hat{D}_U(\gamma, 0)$.

How do we interpret the results in Table 2? For the originally proposed sample size $n = 12$ for testing we can be 80.8% confident that $D \leq 24$ and that the probability of having less than two defective items in the ship-set of $k = 16$ is at least .401, not a very happy state of affairs. This is based on the assumption that we see indeed no defective items in the sample of $n = 12$. After qualifying that the more than one defective parts only matter when at least two are adjacent, we can be 80.8% confident that the probability of seeing no such adjacency is at least .814, which may still not be satisfactory.

That was the reason for showing the columns for higher sample sizes, still expecting no defective items among the n sampled items. Clearly this should provide us with a stronger feeling (confidence) and higher probability of avoiding the adjacency issue. For example, when choosing $n = 40$ we can be 95% confident that the probability of no adjacent defective parts is at least .95.

The function that computes $\hat{p}_1$ and $\hat{p}_2$ and the accompanying $\bar{\gamma}$ is `p1p2.hat` and is contained in the previously referenced `R` work space. Note that the tabulated confidence levels in Table 2 are only given to 3 decimals and thus should only serve as a rough guide. In fact, we get $\hat{p}_1 =$ `QD` and $\hat{p}_2 =$ `QD.star` by calling `p1p2.hat` with $\gamma = .95$

```
>  p1p2.hat(0,40,.950,200,16)
$D.U
upper
   12

$D.UX
upper
   12

$QD
[1] 0.7528636

$QD.star
[1] 0.9502322

$gam.bar
  minimum
0.9503716
```

When changing γ to a value that comes closer to $\bar{\gamma} = 0.9503716$ we get

```
>  p1p2.hat(0,40,.9503716,200,16)
```

... everything else staying the same

```
$gam.bar
  minimum
0.9503716
```

$\Longrightarrow$ the highest possible confidence level $\gamma = \bar{\gamma}$ for that value of $\hat{D}_U(0, \gamma) = \texttt{D.U} = \texttt{12}$.

Of course, if the observed result in the sample of size n is not zero but $x > 0$, one would substitute $\hat{D}_U(\gamma, x)$ for $\hat{D}_U(\gamma, 0)$ with $\hat{D}_U(\gamma, 0) < \hat{D}_U(\gamma, x)$ and the resulting lower bounds $\hat{p}_1$ and $\hat{p}_2$ would decrease while the confidence levels stay at the indicated γ.

In fact, if $x > 0$ we would not return the defective items into the population of size N. That way we would reduce N by x and D by x as well, i.e., $N \to N' = N - x$ and $D \to D' = D - x$. We then should use $\hat{D}'_U(X, \gamma) = \hat{D}_U(X, \gamma) - x = \texttt{D.UX}$ as the upper confidence bound for D' when calculating $\hat{p}_1$ and $\hat{p}_2$.

As an example, assume that we see $X = 1$ when testing $n = 40$ items. Table 2 suggests to take $\gamma = .95$ in order to get $\hat{p}_2 = .95$ when $X = 0$, but we previously saw that a better choice for γ is $\gamma = 0.9503716$, which gives us $\gamma = \bar{\gamma}$. Calling `p1p2.hat` with $x = 1$ and that value of γ yields

```
>  p1p2.hat(1,40,.9503716,200,16)
$D.U
upper
   20

$D.UX
upper
   19

$QD
[1] 0.5335137

$QD.star
[1] 0.8776381

$gam.bar
  minimum
0.9503716
```

While the confidence level remains unaffacted there is a decrease in the respective lower confidence bounds $\hat{p}_1 =$ `QD` and $\hat{p}_2 =$ `QD.star`. That is the penalty for not achieving $X = 0$ as predicted.

As a note of warning, one should not try to trade off some of the confidence level (lowering it) to increase the lower bounds $\hat{p}_1$ or $\hat{p}_2$. This would negate the theoretical underpinnings of confidence bounds. The confidence level should not become random (a function of X). This is not an empty warning, since I have experienced this kind of ingenuity from engineers[3].

7 Comparing Two Poisson Means

Suppose X and Y are independent Poisson random variables with respective means λ and μ. We are interested in confidence bounds for $\rho = \lambda/\mu$. If these Poisson distributions represent approximations of binomials for small "success" probabilities π_1 and π_2, i.e., $\lambda = n_1\pi_1$ and $\mu = n_2\pi_2$, then confidence bounds for the ratio $\rho = \lambda/\mu = (n_1\pi_1)/(n_2\pi_2)$ are equivalent to confidence bounds for π_1/π_2, since n_1 and n_2 are typically known.

The classical method for getting confidence bounds for ρ is to view Y in the context of what was

[3]According to Webster's, engineer and ingenuity have the same Latin origin ingenium.

observed for $X+Y$, i.e., consider the conditional distribution of Y given $T = X+Y = t$, which is

$$\begin{aligned} P(Y=k|X+Y=t) &= \frac{P(Y=k, X+Y=t)}{P(X+Y=t)} = \frac{P(Y=k, X=t-k)}{P(X+Y=t)} \\ &= \frac{P(Y=k)P(X=t-k)}{P(X+Y=t)} = \frac{[\exp(-\mu)\mu^k]/k! \times [\exp(-\lambda)\lambda^{t-k}]/(t-k)!}{\exp(-[\mu+\lambda])[\mu+\lambda]^t/t!} \\ &= \binom{t}{k} p^i (1-p)^{t-i} \ , \end{aligned}$$

where $p = \mu/(\lambda+\mu) = 1/(1+\rho)$. Here we used the fact that $X+Y$ has a Poisson distribution with mean $\mu+\lambda$. This is seen as follows

$$\begin{aligned} P(X+Y=t) &= \sum_{i=0}^{t} P(X=i) \times P(Y=t-i) = \sum_{i=0}^{t} \frac{\exp(-\lambda)\lambda^i}{i!} \times \frac{\exp(-\mu)\mu^{t-i}}{(t-i)!} \\ &= \frac{\exp(-[\lambda+\mu])(\lambda+\mu)^t}{t!} \sum_{i=0}^{t} \binom{t}{i} \left(\frac{\lambda}{\lambda+\mu}\right)^i \left(\frac{\mu}{\lambda+\mu}\right)^{t-i} = \frac{\exp(-[\lambda+\mu])(\lambda+\mu)^t}{t!} \end{aligned}$$

Hence the conditional distribution of Y given $X+Y=t$ is a binomial distribution with t trials and success probability $p = 1/(1+\rho)$, which is a monotone function of ρ.
Given the previous treatment in Section 2 we can get $100\gamma\%$ binomial upper confidence bounds $\hat{p}_U(\gamma, k, t)$ for p. By inverting the monotone relationship $p = 1/(1+\rho)$, i.e., $\rho = 1/p - 1$, we get from this a $100\gamma\%$ lower confidence bound for ρ in

$$\hat{\rho}_L(\gamma, k, t) = \frac{1}{\hat{p}_U(\gamma, k, t)} - 1 = \frac{1}{\mathtt{qbeta}(\gamma, \mathtt{k+1}, \mathtt{t-k})} - 1 \ .$$

Similarly one gets $100\gamma\%$ upper confidence bounds for ρ in

$$\hat{\rho}_U(\gamma, k, t) = \frac{1}{\hat{p}_L(\gamma, k, t)} - 1 = \frac{1}{\mathtt{qbeta}(1-\gamma, \mathtt{k}, \mathtt{t-k+1})} - 1$$

and $100\gamma\%$ confidence intervals for ρ in

$$(\hat{\rho}_L([1+\gamma]/2, k, t), \hat{\rho}_U([1+\gamma]/2, k, t)) = \left(\frac{1}{\hat{p}_U([1+\gamma]/2, k, t)} - 1, \frac{1}{\hat{p}_L([1+\gamma]/2, k, t)} - 1\right)$$

$$= \left(\frac{1}{\mathtt{qbeta}([1+\gamma]/2, \mathtt{k+1}, \mathtt{t-k})} - 1, \frac{1}{\mathtt{qbeta}([1-\gamma]/2, \mathtt{k}, \mathtt{t-k+1})} - 1\right)$$

In the context of the above Poisson approximations to binomial distributions one gets a $100\gamma\%$ lower confidence bound for $\kappa = \pi_1/\pi_2$ in

$$\begin{aligned}\hat{\kappa}_L(\gamma,k,t) &= \frac{n_2}{n_1}\times\hat{\rho}_L(\gamma,k,t) = \frac{n_2}{n_1}\times\left(\frac{1}{\hat{p}_U(\gamma,k,t)}-1\right)\\ &= \frac{n_2}{n_1}\times\left(\frac{1}{\texttt{qbeta}(\gamma,\texttt{k}+1,\texttt{t}-\texttt{k})}-1\right).\end{aligned}$$

Similarly one gets $100\gamma\%$ upper confidence bounds for $\kappa = \pi_1/\pi_2$ in

$$\begin{aligned}\hat{\kappa}_U(\gamma,k,t) &= \frac{n_2}{n_1}\times\hat{\rho}_U(\gamma,k,t) = \frac{n_2}{n_1}\times\left(\frac{1}{\hat{p}_L(\gamma,k,t)}-1\right)\\ &= \frac{n_2}{n_1}\times\left(\frac{1}{\texttt{qbeta}(1-\gamma,\texttt{k},\texttt{t}-\texttt{k}+1)}-1\right).\end{aligned}$$

and $100\gamma\%$ confidence intervals for $\kappa = \pi_1/\pi_2$ in

$$\begin{aligned}&(\hat{\kappa}_L([1+\gamma]/2,k,t),\hat{\kappa}_U([1+\gamma]/2,k,t))\\ &= \frac{n_2}{n_1}\times(\hat{\rho}_L([1+\gamma]/2,k,t),\hat{\rho}_U([1+\gamma]/2,k,t))\\ &= \frac{n_2}{n_1}\times\left(\frac{1}{\hat{p}_U([1+\gamma]/2,k,t)}-1,\frac{1}{\hat{p}_L([1+\gamma]/2,k,t)}-1\right)\\ &= \frac{n_2}{n_1}\times\left(\frac{1}{\texttt{qbeta}([1+\gamma]/2,\texttt{k}+1,\texttt{t}-\texttt{k})}-1,\frac{1}{\texttt{qbeta}([1-\gamma]/2,\texttt{k},\texttt{t}-\texttt{k}+1)}-1\right)\end{aligned}$$

As an example consider some (fictitious) accident data. Among Modern Wide Body Airplanes we had 0 accidents (substantial damage, hull loss, or hull loss with fatalities) during 11.128×10^6 flights. Among Modern Narrow Body Airplanes we had 5 such accidents during 55.6×10^6 flights. Thus we have $Y = k = 5$ and $T = X + Y = t = 0 + 5$ and $n_1 = 11.128\times10^6$ and $n_2 = 55.6\times10^6$. We find

$$\hat{p}_U(.95,5,5) = 1$$

and thus

$$\hat{\kappa}_L(.95,5,5) = \frac{n_2}{n_1}\times\left(\frac{1}{1}-1\right) = 0$$

and

$$\hat{p}_L(.95,5,5) = \texttt{qbeta}(.05,5,5-5+1) = (1-.95)^{1/5} = 0.5492803$$

resulting in

$$\hat{\kappa}_U(.95, 5, 5) = \frac{n_2}{n_1} \times \left(\frac{1}{0.5492803} - 1 \right) = 4.099871$$

Thus we can view 4.099871 as a 95% upper confidence bound for π_1/π_2, the ratio of rates for the two groups. Since this bound is above 1, one cannot rule out that the rates in the two groups may be the same. This is not surprising since the group with 5 accidents had about five times the exposure of the other group. Thus for a fifth of the exposures one might have expected to see one such accident in the second group. That is not all that different from zero in the realm of counting rare events.

7.1 Comparing Hull Loss Rates across Airplane Models

Based on the data in Figure 20 one may ask the question: How significant was the crash of the Concorde when comparing it to experiences of other airplanes? We made two comparisons, comparing it with the MD11 and the B767.

In the comparison with the MD11 we had $X = 5$ hull losses and a rate of 4.59 per million departures, while the Concorde had $Y = 1$ such loss and a rate of 12.02 per million departures. This translated into $n_1 = 1089325$ and $n_2 = 83195$ departures for the MD11 and Concorde, respectively. This gives us $(.04273, 18.06)$ as a 95% confidence interval for $\kappa = \pi_1/\pi_2$ which contains $\pi_1/\pi_2 = 1$ as a possible value. Thus one cannot conclude that the MD11 and the Concorde had sufficiently different experiences. However, the confidence intervals could be quite conservative in their coverage probabilities based on the fact that only $X + Y = 6$ accidents were involved in this comparison. Since we are mostly interested in the question $\pi_1/\pi_2 < 1$ we should focus more on the upper bound for π_1/π_2. The 95% upper confidence bound for π_1/π_2 is 8.896, which still exceeds 1 by a fair margin. However, here we know that the minimum coverage probability is .95, although its coverage behavior is coarsely fluctuating. Since the upper bound for π_1/π_2 is based on the lower bound $\hat{p}_L(\gamma, Y, 6)$ for $p = 1/(1+\rho) = 1/(1+n_1\pi_1/(n_2\pi_2))$ we show in Figure 28 the coverage behavior of this lower bound and indicate by vertical lines positions for various values of π_1/π_2. Depending on the unknown value of π_1/π_2 our coverage probability could be quite close to the desired level .95 or very much higher.

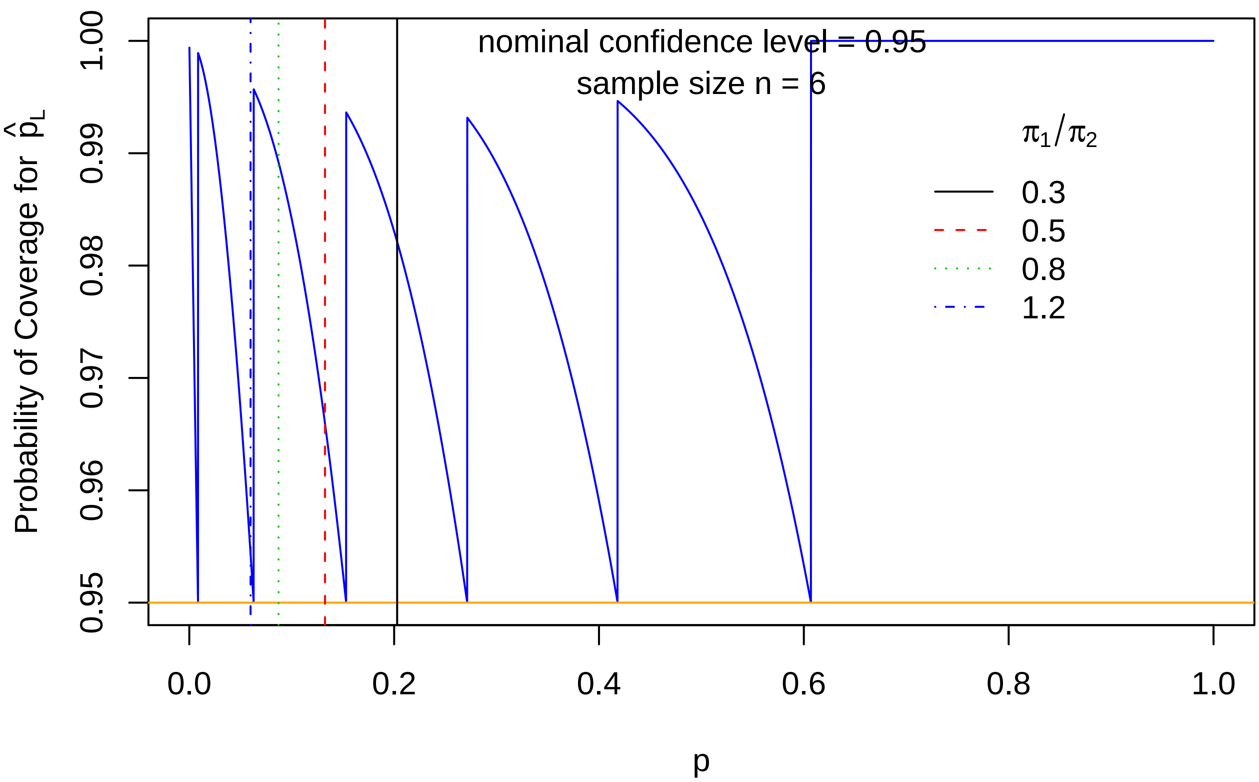


Figure 28: Coverage Behavior of Binomial Lower Bound for $n = 6$

In comparing the Concorde with the B767 we have $X = 3$ hull losses for the latter and a rate of .32 per million departures. This translates into $n_1 = 9375000$ departures for the B767. Here our 95% upper bound for π_1/π_2 is .6876, which is clearly below 1 and thus indicates a difference in experience between the B767 and the Concorde, in favor of the B767.

The Concorde ceased service in 2003. On November 5, 2003, the Concorde now on display at the Museum of Flight made its final landing at Boeing Field in Seattle. The photo below was taken while inspecting taxi ways at JFK.

Figure 29: Concorde Takeoff at JFK, June 18, 2003

8 Inverse Probability Solving

At times the question is asked: what is the smallest k such that $P(X \le k) \ge \gamma$, where X is a binomial random variable with parameters n and p and γ is a desired probability level. Clearly $P(X \le k)$ increases to one as k reaches n, so there is a smallest such number k. The question was asked whether one can solve this problem quickly in Excel without much iteration.

A similar problem arises when Y is a Poisson random variable with mean λ. What is the smallest k such that $P(Y \le k) \ge \gamma$.

The answer to the binomial problem is simple: yes, there is such a function in Excel. It is called `CRITBINOM` and

$$\texttt{CRITBINOM}(n, p, \gamma)$$

gives the smallest k such that $P(X \le k) \ge \gamma$. For example,

$$\texttt{CRITBINOM}(100, .1, .8) = 12$$

and one verifies that

$$\texttt{BINOMDIST}(12, 100, .1, TRUE) = .80182$$

while

$$\texttt{BINOMDIST}(11, 100, .1, TRUE) = .70303 \ .$$

The answer to the Poisson problem using Excel is not so direct but can be finessed by using `CRITBINOM`. This is possible since the Poisson distribution with mean λ is a very good approximation to the binomial distribution with parameters n and p with $\lambda = np$, provided p is very small. For fixed λ this means that we should choose n very large, say $n = 1000$ or $n = 10000$ and $p = \lambda/n$. Since this is an approximation it may not yield the exact solution k but one can then check the actual probability using `POISSON` in Excel.

As an example, suppose we have $\lambda = 4$ and $\gamma = .8$. Then

$$\texttt{CRITBINOM}(n, \lambda/n, \gamma) = \texttt{CRITBINOM}(1000, 4/1000, .8) = 6 \ .$$

We check the actual probability using the cumulative distribution function `POISSON` in Excel, namely:

$$\texttt{POISSON}(6, 4, TRUE) = .889326$$

while

$$\texttt{POISSON}(5, 4, TRUE) = .78513 \ .$$

Thus we were successful with the initial value given by `CRITBINOM`.

References

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

2. Barbour, A.D., Holst, L., and Janson, S. (1992), *Poisson Approximation*, Clarendon Press, Oxford.

3. Bardi, J.S. (2006), *The Calculus Wars, Newton and Leibniz, and the Greatest Mathematical Clash of All Time*, Thunder's Mouth Press, New York.

4. Barker, Lawrence (2002). "A comparison of nine confidence intervals for a Poisson parameter when the expected number of events is ≤ 5," *The American Statistician*, **56**, 2, 85-89.

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

6. Hahn, G.J. and Meeker, W.Q. (1991), *Statistical Intervals, A Guide for Practitioners*, John Wiley & Sons, New York.

7. Lehmann, E.L. and Romano, J.P. (2005), *Testing Statistical Hypotheses, Third Edition*, John Wiley & Sons, New York.

8. Shapiro, S.S. and Gross, A.J. (1981), *Statistical Modeling Techniques*, Marcel Dekker , Inc., New York.

9. van Belle, G. (2002) *Statistical Rules of Thumb*, John Wiley & Sons, New York.

Appendix: Monotonicity of $Q^\star(D)$

As a first step one shows that $\psi(y) = P(A_y)$ decreases as y increases from $y = 1$ to $y = k/2$ (assuming that k is even). This is seen by

$$\frac{P(A_{y+1})}{P(A_y)} = \frac{(k-2y)(k-2y-1)}{(k-y)(k-y-1)} < 1 \quad \Longleftrightarrow \qquad -2k + 3y + 1 < 0$$

which is true for $y = 1, \ldots, k/2$.
The second step consists in showing that $Q^\star(D+1) = E_{D+1}(\psi(Y)) \leq E_D(\psi(Y)) = Q^\star(D)$.
The proof is an adaptation of the argument given in Lehmann and Romano (2005), p.85.
First note that

$$p_{D+1}(y) = P_{D+1}(Y=y) = \frac{\binom{D+1}{y}\binom{N-D-1}{k-y}}{\binom{N}{k}} < \frac{\binom{D}{y}\binom{N-D}{k-y}}{\binom{N}{k}} = P_D(Y=y) = p_D(y) \;\Leftrightarrow\; y < \frac{k(D+1)}{N+1}\ .$$

Let $\qquad A = \{y : P_{D+1}(Y=y) < P_D(Y=y)\} \qquad$ and $\qquad B = \{y : P_{D+1}(Y=y) > P_D(Y=y)\}$

and note that the previous equivalence implies that A lies to the left of B with $k(D+1)/(N+1)$ in between. Because of the decreasing nature of $\psi(y)$ it follows that $a = \inf_A \psi(y) \geq \sup_B \psi(y) = b$. Hence

$$\begin{aligned} E_{D+1}[\psi(Y)] - E_D[\psi(Y)] &= \sum_{y\in A} \psi(y)\,(p_{D+1}(y) - p_D(y)) + \sum_{y\in B} \psi(y)\,(p_{D+1}(y) - p_D(y)) \\ &\leq a \sum_{y\in A} (p_{D+1}(y) - p_D(y)) + b \sum_{y\in B} (p_{D+1}(y) - p_D(y)) \\ &= (b-a) \sum_{y\in B} (p_{D+1}(y) - p_D(y)) \leq 0 \end{aligned}$$

since

$$0 = \sum_{y} (p_{D+1}(y) - p_D(y)) = \sum_{y\in A} (p_{D+1}(y) - p_D(y)) + \sum_{y\in B} (p_{D+1}(y) - p_D(y))$$

and thus

$$\sum_{y\in A} (p_{D+1}(y) - p_D(y)) = -\sum_{y\in B} (p_{D+1}(y) - p_D(y))\ .$$