University of Washington

STATISTICS

STAT 498 B

Industrial Statistics

Confidence Bounds for Binomial, Poisson and Hypergeometric Distributions

Fritz Scholz

Spring Quarter 2008

Confidence Bounds and Intervals

We will develop exact confidence bounds for parameters θ associated with various discrete distributions, such as

- Binomial distribution, $\theta =$ success probability

- Poisson distribution, $\theta =$ mean

- Hypergeometric distribution, $\theta = $ # of defective items in sampled population

- Comparing two Poisson means λ and μ, $\theta = \lambda/\mu$.

Equivalence of Confidence Sets and Hypothesis Tests

Let $\mathcal{X}$ denote the set of all possible data outcomes X (counts or pairs of counts).

Let $A(\theta_0) \subset \mathcal{X}$ denote the acceptance region of a level α test for testing a hypothesis $H(\theta_0)$ indexed by θ_0. Thus $\mathcal{X} - A(\theta_0) = R(\theta_0)$ is the rejection region.

Possible hypotheses: $H(\theta_0) : \theta = \theta_0$ or $H(\theta_0) : \theta \leq \theta_0$ or $H(\theta_0) : \theta \geq \theta_0$

Thus $P_\theta(X \in R(\theta_0)) \leq \alpha$ or $P_\theta(X \in A(\theta_0)) \geq 1 - \alpha$ for $\theta \in H_{\theta_0}$, e.g., for $\theta = \theta_0$.

Define $C(X) = \{\theta_0 : X \in A(\theta_0)\} =$ set of all θ_0 for which H_{θ_0} is acceptable via X.

$$\implies \{\theta_0 \in C(X) \iff X \in A(\theta_0)\} \implies P_{\theta_0}(\theta_0 \in C(X)) = P_{\theta_0}(X \in A(\theta_0)) \geq 1 - \alpha$$

This is true for any θ_0 $\implies$ $P_\theta(\theta \in C(X)) \geq 1 - \alpha$ $\forall \theta$ (for all θ).

Note the randomness of $C(X)$ via X and that θ is treated as fixed and unknown.

We will revisit this abstract scenario over and over in concrete settings.

How to Construct Reasonable Hypothesis Tests?

There was disagreement over the issue of hypothesis testing between the founding fathers of statistics, Fisher on one side and Neyman and Pearson on the other.

See Lehmann, E.L. (2008) *Reminiscences of a Statistician, The Company I Kept* and Salsburg, D. (2001) *The Lady Tasting Tea, How Statistics Revolutionized Science in the Twentieth Century*.

Fisher viewed hypothesis testing as examining whether the data provided sufficient evidence (small enough P-value) for rejecting the hypothesis, based on some "obvious" or "intuitive" test criterion computed from the data.

Neyman and Pearson took alternatives to the hypothesis into account when constructing appropriate tests. They introduced the concept of type I and type II error probabilities and developed an optimality theory within this framework, often resulting in "obvious" or "intuitive" test criteria.

Confidence Sets

Neyman subsequently developed the theory of confidence sets with its close connection to hypothesis tests (slide # 2).

Confidence sets acknowledge that estimates of parameters are hardly ever correct. At best they are approximately correct. How to gauge "approximately"?

We do this by giving confidence intervals of various widths, each with its assurance or confidence of covering (capturing) the unknown parameter.

In nested intervals the wider interval has higher confidence.

Sometimes only one end point of such an interval is important, e.g., we only care about a risk assessment (probability of a serious event happening) in terms of the estimated risk and how much higher it might be with a specified confidence.

What We Will Do

We will start with intuitively appealing test criteria, without justifying them as optimal in any way. As it turns out, these are often optimal tests.

We then take these tests and convert them into confidence bounds and intervals according to the previously indicated equivalence relationship.

We will examine the operational properties of these confidence sets and compare them with other ones as they are frequently taught in elementary statistics courses.

We then illustrate these confidence sets with several examples that I encountered in practice. Some of these require the building of appropriate bridges to make the link to these confidence sets.

Binomial Distribution

A binomial random variable X counts the number of successes in n independent Bernoulli trials with success probability $p, \quad 0 \le p \le 1$

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

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 $k = 1, 2, \ldots, n$.

$$
\begin{aligned}
\frac{\partial P_p(X \ge k)}{\partial p} &= \sum_{i=k}^{n} \binom{n}{i} i p^{i-1} (1-p)^{n-i} - \sum_{i=k}^{n-1} \binom{n}{i} (n-i) p^i (1-p)^{n-i-1} \\
&= n \sum_{i=k}^{n} \binom{n-1}{i-1} p^{i-1} (1-p)^{n-i} - n \sum_{i=k}^{n-1} \binom{n-1}{i} p^i (1-p)^{n-i-1} \\
\stackrel{j=i-1}{\Longrightarrow} \quad &= n \sum_{j=k-1}^{n-1} \binom{n-1}{j} p^j (1-p)^{n-1-j} - n \sum_{i=k}^{n-1} \binom{n-1}{i} p^i (1-p)^{n-i-1} \\
&= n \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} = k \binom{n}{k} p^{k-1} (1-p)^{n-k} > 0.
\end{aligned}
$$

Gamma and Beta Densities

$$\Gamma(a) = \int_0^\infty \exp(-x) x^{a-1} \, dx \qquad \text{is called the gamma function for } a > 0$$

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 $(0, \infty)$, called the gamma density with shape parameter $a > 0$.

$$B(a,b) = \int_0^1 x^{a-1} (1-x)^{b-1} \, dx \qquad \text{is called the beta function for } a > 0 \text{ and } b > 0$$

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 $(0, 1)$, called the beta density with parameters $a > 0$ and $b > 0$.

$$\Gamma(n+1) = n! = 1 \cdot 2 \cdot \ldots \cdot n \quad \text{for integer } n \quad \text{and} \quad B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

Useful Identities

For $k > 0 \implies$ binomial tail summations $\sim$ beta cdf or incomplete beta function,

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

$$P_p(X \leq k) = 1 - P_p(X \geq k+1) = 1 - I_p(k+1, n-k). \qquad (2)$$

where $\quad I_y(a,b) = \dfrac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \displaystyle\int_0^y t^{a-1}(1-t)^{b-1} \, dt = P(Y \leq y)$

denotes the cdf of a beta random variable Y with parameters $a > 0$ and $b > 0$.

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

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

(since $\Gamma(n+1) = n!$ for integer n) same as $\partial P_p(X \geq k)/\partial p$

$$I_p(k, n-k+1) - P_p(X \geq k) \to 0 - 0 = 0 \quad \text{as} \quad p \to 0 \quad \implies \quad (1)$$

8

General Binomial Monotonicity Property

If the function $\psi(x)$, defined for $x = 0, 1, \ldots, n$, is monotone increasing (decreasing) and not constant, then the expectation $E_p(\psi(X))$ is strictly monotone increasing (decreasing) in p. Intuitively appealing: $p \nearrow \Rightarrow X \nearrow \Rightarrow \psi(X) \nearrow \Rightarrow E_p(\psi(X)) \nearrow$

$$\frac{\partial E_p(\psi(X))}{\partial p} = \frac{\partial}{\partial p} \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} \psi(x)$$

$$= \sum_{x=1}^{n} \binom{n}{x} x p^{x-1} (1-p)^{n-x} \psi(x) - \sum_{x=0}^{n-1} \binom{n}{x} (n-x) p^x (1-p)^{n-x-1} \psi(x)$$

$$= n \sum_{x=1}^{n} \binom{n-1}{x-1} p^{x-1} (1-p)^{n-x} \psi(x) - n \sum_{x=0}^{n-1} \binom{n-1}{x} p^x (1-p)^{n-x-1} \psi(x)$$

$$= n \sum_{y=0}^{n-1} \binom{n-1}{y} p^y (1-p)^{n-1-y} \psi(y+1) - n \sum_{x=0}^{n-1} \binom{n-1}{x} p^x (1-p)^{n-x-1} \psi(x)$$

$$= n \sum_{y=0}^{n-1} \binom{n-1}{y} p^y (1-p)^{n-1-y} [\psi(y+1) - \psi(y)] > 0 \qquad \text{q.e.d.}$$

9

Testing Hypotheses Concerning p

Testing the hypothesis $H(p_0) : p = p_0$ against the alternative $A(p_0) : p < p_0$ we would view small values of X as evidence against the hypothesis $H(p_0)$.

Thus we reject $H(p_0)$ at target significance level α when $X \leq k(p_0, \alpha)$, where the critical value $k(p_0, \alpha)$ is chosen as the largest integer k for which
$$P_{p_0}(X \leq k) \leq \alpha \qquad \implies \qquad P_{p_0}(X \leq k(p_0, \alpha) + 1) > \alpha.$$
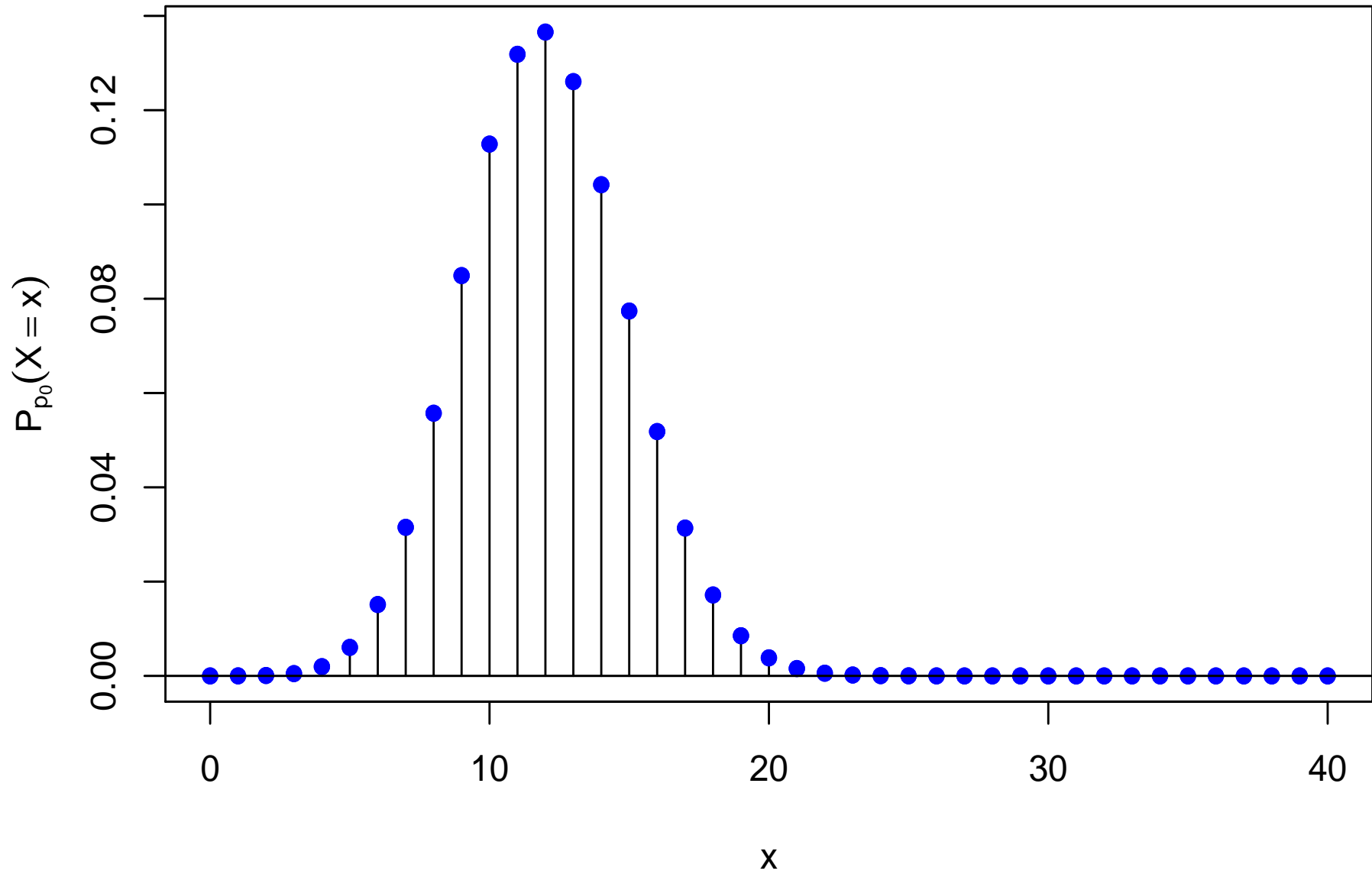
This controls the probability of type I error (rejecting $H(p_0)$ when true) at $\leq \alpha$.

For an observed value x of X we reject $H(p_0)$ at level α whenever the p-value $p(x, p_0) = P_{p_0}(X \leq x) \leq \alpha$. That happens exactly when $x \leq k(p_0, \alpha)$.
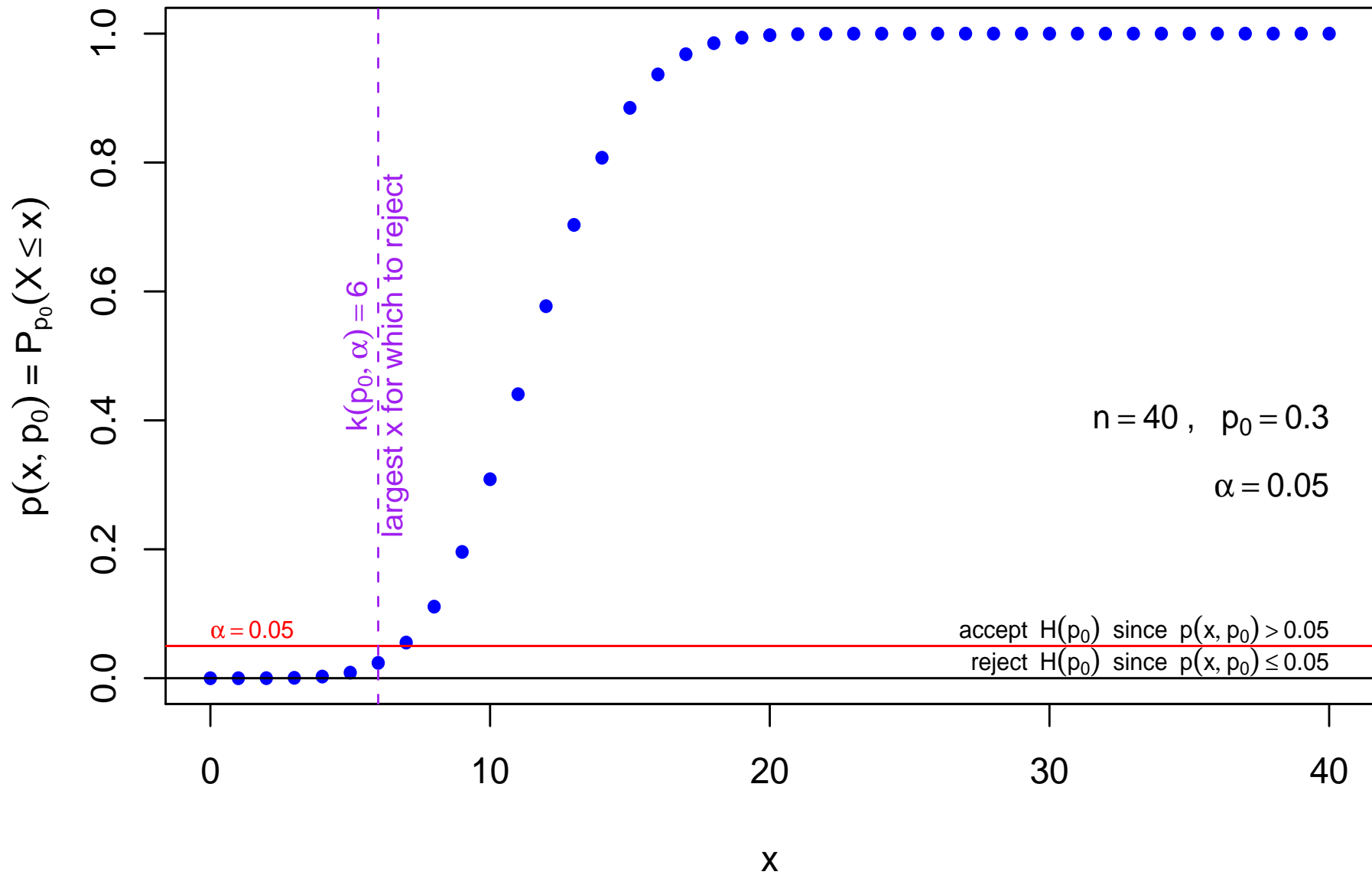
This view avoids the search for $k(p_0, \alpha)$ and tells us how strongly we reject $H(p_0)$, i.e., for how small an α we would have rejected $H(p_0)$ for that x.

$p(x, p_0)$ is also called the observed significance level.

Binomial Distribution: Density

Binomial Distribution: CDF

$p(x, p_0) = P_{p_0}(X \le x)$

$k(p_0, \alpha) = 6$
largest x for which to reject

$n = 40 , \quad p_0 = 0.3$

$\alpha = 0.05$

$\alpha = 0.05$

accept $H(p_0)$ since $p(x, p_0) > 0.05$
reject $H(p_0)$ since $p(x, p_0) \le 0.05$

x

12

Duality of Testing Hypotheses and Confidence Sets for p

We accept $H(p_0)$ whenever the p-value $p(x, p_0) = P_{p_0}(X \leq x) > \alpha$.

For an observed value x of X define the confidence set $C(x)$ for p as all those values p_0 for which we would accept $H(p_0)$, i.e.,

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

$P_{p_0}(X \leq x)$ is continuous and strictly decreasing in p_0 for $x < n$. It follows that $C(x) = [0, \hat{p}_U(1 - \alpha, x, n))$, where $\hat{p}_U(1 - \alpha, x, n)$ is the unique value p that solves

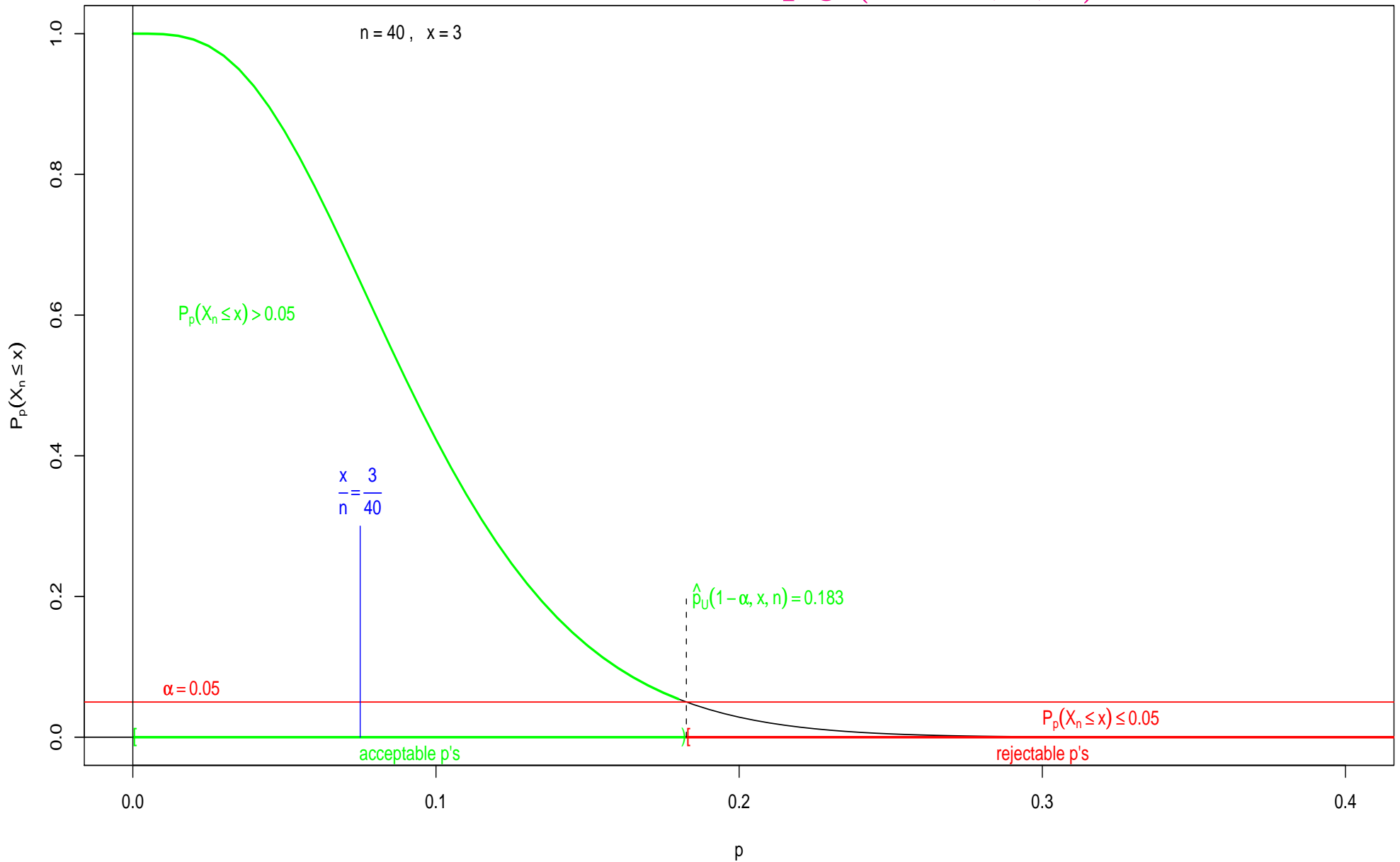$$P_p(X \leq x) = \sum_{i=0}^{x} \binom{n}{i} p^i (1-p)^{n-i} = 1 - I_p(x+1, n-x) = \alpha \qquad (3)$$

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

i.e., $\hat{p}_U(1 - \alpha, x, n) = \hat{p}_U(\gamma, x, n)$ is the γ-quantile of the beta distribution with parameters $x + 1$ and $n - x$ or using R/S-Plus or Excel, respectively:

$$\hat{p}_U(\gamma, x, n) = \texttt{qbeta}(\gamma, \texttt{x}+1, \texttt{n}-\texttt{x}) = \texttt{BETAINV}(\gamma, \texttt{x}+1, \texttt{n}-\texttt{x}) \quad \text{for } x < n$$

These bounds are also referred to as Clopper-Pearson bounds.

Illustration for $\hat{p}_U(1-\alpha, x, n)$

n = 40 , x = 3

$P_p(X_n \le x) > 0.05$

$\dfrac{x}{n} = \dfrac{3}{40}$

$\hat{p}_U(1-\alpha, x, n) = 0.183$

$\alpha = 0.05$

$P_p(X_n \le x) \le 0.05$

acceptable p's

rejectable p's

Special Case $x = n$

When $x = n$ we cannot use `qbeta` or `BETAINV`, since $n - x = 0$.

From the direct definition it follows that $C(n) = [0, 1]$, closed on the right.

since $P_{p_0}(X \leq n) = 1 > \alpha \quad \forall p_0$.

In concordance with that we define $\hat{p}_U(\gamma, n, n) = 1$ with the understanding that the confidence set should be closed on the right in that case.

Confidence Level of $C(X)$

We can view this collection of confidence sets $\{C(x) : x = 0, 1, \ldots, n\}$ also as a random set $C(X)$.

It has the following coverage probability property for any p_0

$$P_{p_0}(p_0 \in C(X)) = 1 - P_{p_0}(p_0 \notin C(X)) = 1 - P_{p_0}(\text{reject } H(p_0)) \geq 1 - \alpha = \gamma$$

Note that in this coverage probability it is the set $C(X)$ that is random, not p_0!

This holds for any p_0, whether we know it or not. We drop the subscript $_0$ and write

$$P_p(p \in C(X)) = P_p(p < \hat{p}_U(\gamma, X, n)) \geq \gamma \quad \forall p$$

$\hat{p}_U(\gamma, X, n)$ is called a $100\gamma\%$ upper confidence bound for the unknown p.

Here $p < \hat{p}_U(\gamma, X, n)$ should be read as $p \leq \hat{p}_U(\gamma, X, n)$ when $X = n$.

The Confidence Coefficient

The confidence coefficient of the upper bound is defined as

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

The confidence coefficient is $\bar{\gamma} = \underline{\gamma} = 1 - \alpha$ since $P_p \left(p \in C(X) \right) = \gamma$ for some p,

or $P_p \left(p \notin C(X) \right) = P_p(X \leq k(p, \alpha)) = 1 - \gamma = \alpha$ 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

$$P_{p_i}(X \leq k(p_i, \alpha)) = P_{p_i}(X \leq i) = \alpha , \tag{4}$$

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

An Example and Some Observations

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 confidence bound for p.

Using (3) it is a simple exercise to show that the sequence of upper bounds is strictly increasing in x, i.e.,

$$0 < \hat{p}_U(\gamma, 0, n) < \hat{p}_U(\gamma, 1, n) < \ldots < \hat{p}_U(\gamma, n-1, n) < \hat{p}_U(\gamma, n, n) = 1 .$$

$$\implies \quad P_p(p < \hat{p}_U(\gamma, X, n)) = 1 \quad \forall p < \hat{p}_U(\gamma, 0, n).$$
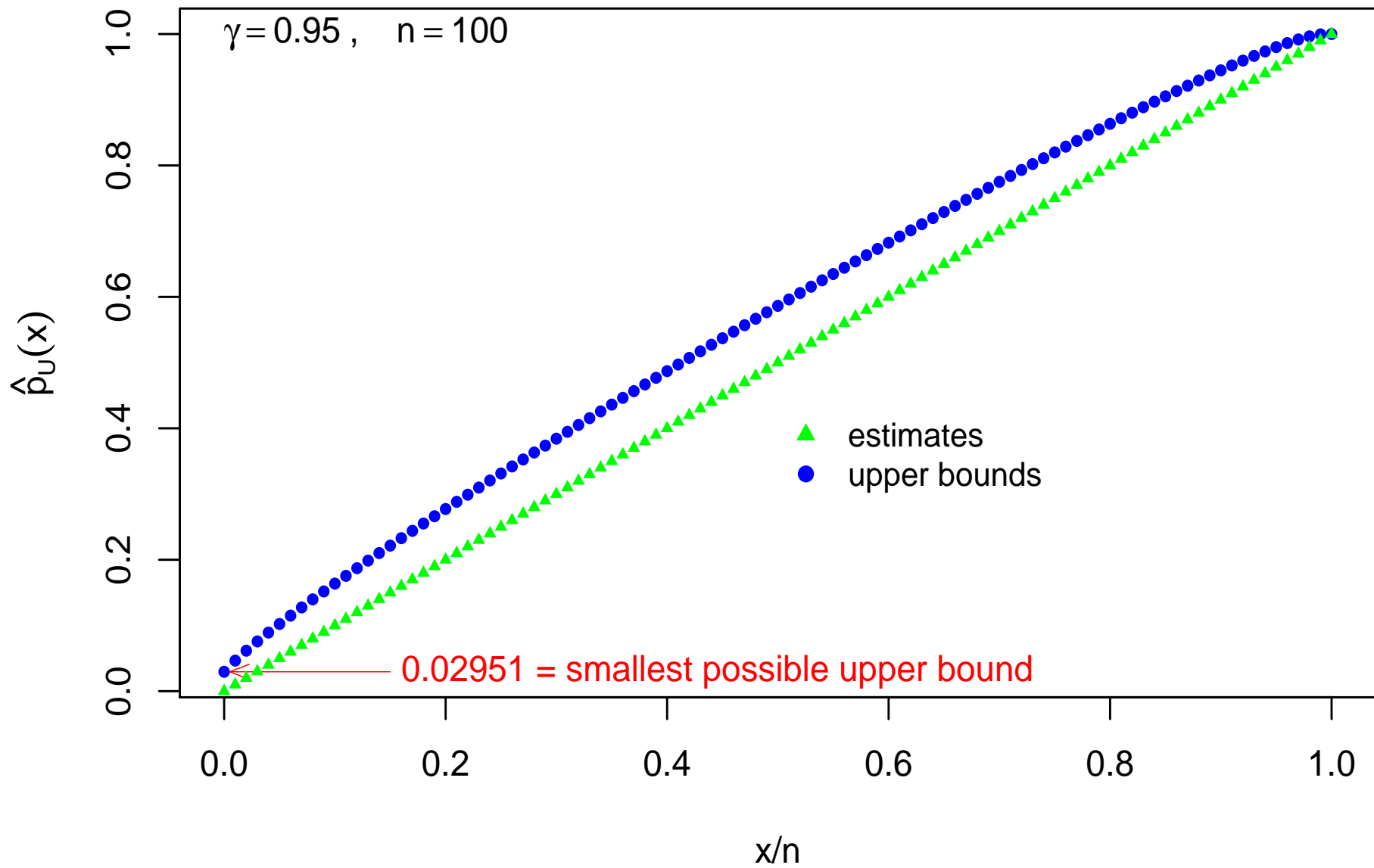
Explicit expression for $x = 0$:
Solving $P_p(X \leq 0) = (1-p)^n = 1 - \gamma$ for p we get $\hat{p}_U(\gamma, 0, n) = 1 - (1-\gamma)^{1/n}$

Explicit expression for $x = n - 1$:
Solving $P_p(X \leq n-1) = 1 - p^n = 1 - \gamma$ for p we get $\hat{p}_U(\gamma, n-1, n) = \gamma^{1/n}$

The first case is more useful than the second.

Estimates in Relation to Upper Confidence Bounds

$\gamma = 0.95$, $\quad$ n = 100

estimates
upper bounds

0.02951 = smallest possible upper bound

x/n

$\hat{p}_U(x)$

The Rule of Three

For $\gamma = .95$ and approximating $\log(1 - \gamma) = \log(.05) = -2.995732$ by -3

the upper bound for $x = 0$ becomes

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

Here the last approximation is valid only for large n, say $n \geq 100$.

This is sometimes referred to as the **Rule of Three** (mnemonic simplicity!)

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

$\hat{p}_U(.95, 0, n) = 3/n$ is very useful for quick sample size planning,

based on the a priori conviction of seeing $x = 0$ successes in n trials.

Side Comment on Common Practice when $x = 0$

Observing $X = 0$ successes in n trials, even when n is large, we still are not inclined to estimate the success probability p by $\widehat{p}(0) = 0/n = 0$, since that is a very strong statement.

$p = 0 \implies$ we will never see a success in however many 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.
Thus estimate p by $\tilde{p} = 1/(n+1)$ which is small but not zero.

The use of $\tilde{p}$ as estimate of p is somewhat conservative. How conservative?

Find γ such that $\hat{p}_U(\gamma, 0, n) = 1 - (1-\gamma)^{1/n} = 1/(n+1)$

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

Can view $\tilde{p} = 1/(n+1)$ as a 63.21% upper confidence bound for p.

Closed 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

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

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

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

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 p is not an acceptable hypothesis $H(p)$ for this x.

Coverage Probability of Upper Confidence Bounds

nominal confidence level = 0.95
sample size n = 10

23
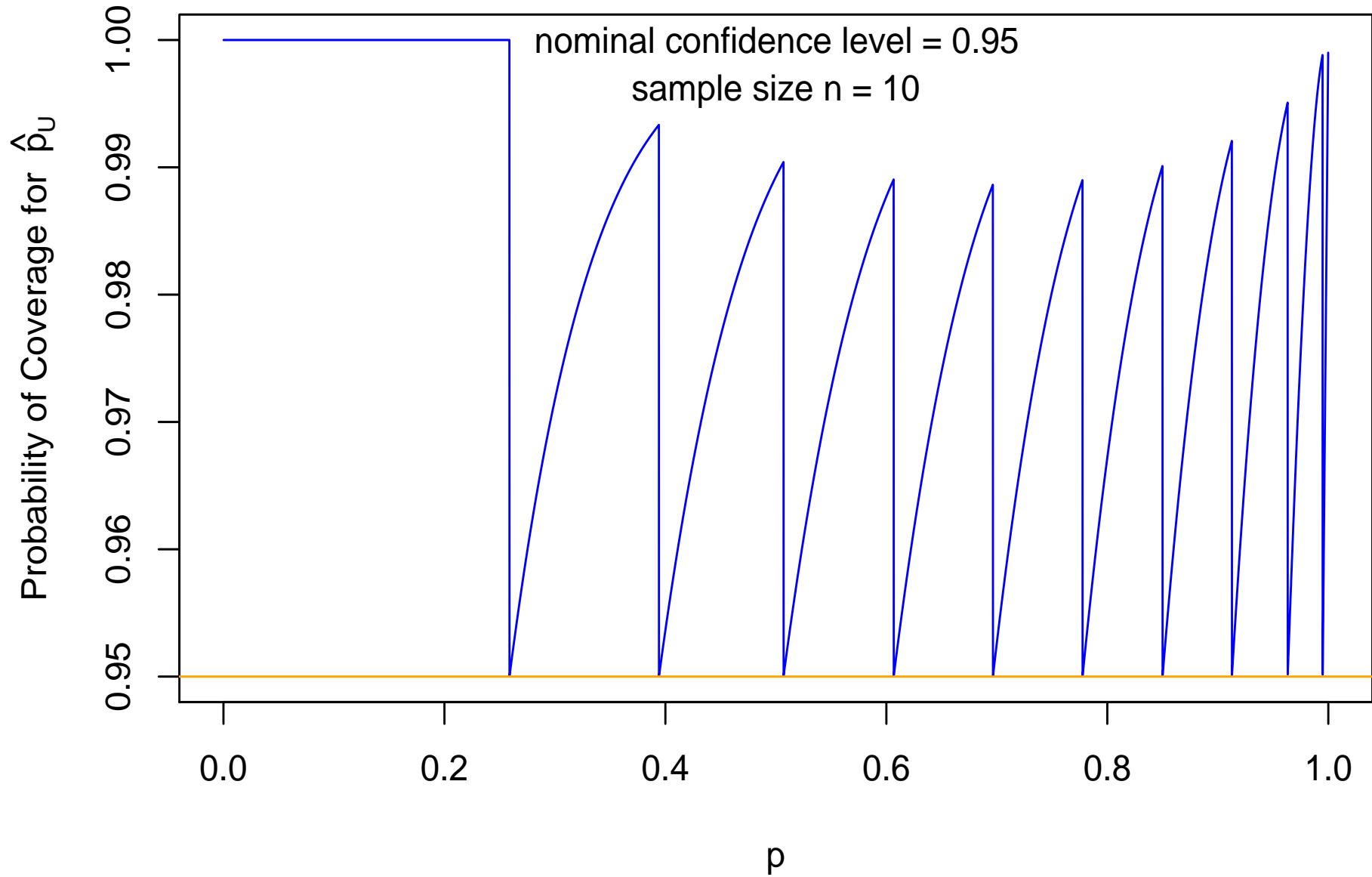
Lower Confidence Bounds on p

When testing $H(p_0) : p = p_0$ against the alternative $A(p_0) : p > p_0$
large values of X can be viewed as evidence against $H(p_0)$.

Again carry out the test in terms of the observed p-value $p(x, p_0) = P_{p_0}(X \geq x)$ by
rejecting $H(p_0)$ at level α whenever $p(x, p_0) \leq \alpha$ and accepting $H(p_0)$ otherwise.
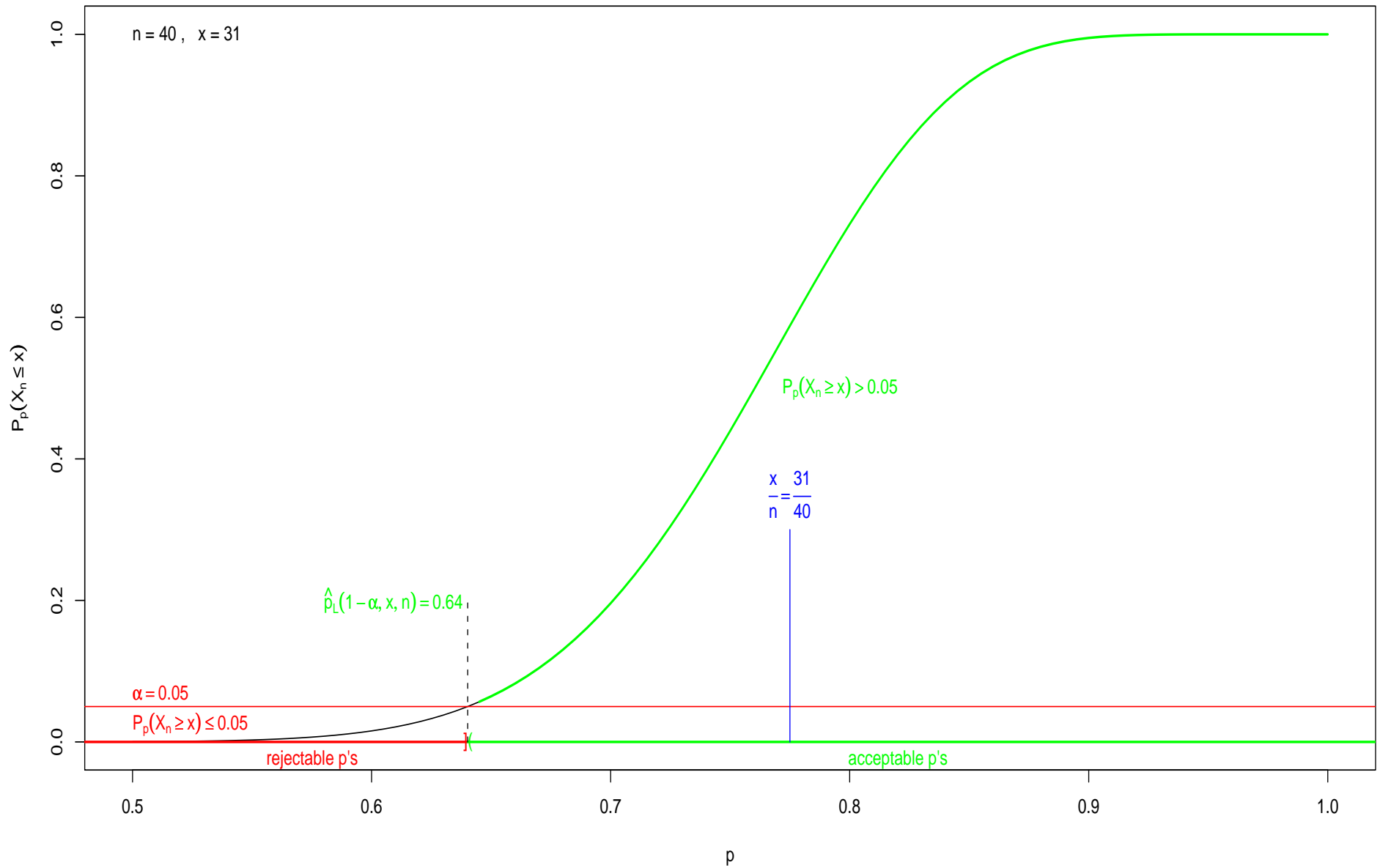
From the duality between testing and confidence sets we get as confidence set

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

Since $P_{p_0}(X \geq x)$ 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

$$P_p(X \geq x) = \sum_{i=x}^{n} \binom{n}{i} p^i (1-p)^{n-i} = \alpha = 1 - \gamma . \tag{5}$$

Illustration for $\hat{p}_L(1 - \alpha, x, n)$

n = 40 , x = 31

$P_p(X_n \geq x) > 0.05$

$\dfrac{x}{n} = \dfrac{31}{40}$

$\hat{p}_L(1 - \alpha, x, n) = 0.64$

$\alpha = 0.05$

$P_p(X_n \geq x) \leq 0.05$

rejectable p's

acceptable p's

$P_p(X_n \leq x)$

p

Lower Confidence Bounds on p (cont.)

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 $n-x+1$.

Using R/S-Plus or Excel we get it for $x > 0$ as

$$\hat{p}_L(\gamma, x, n) = \texttt{qbeta}(1-\gamma, \texttt{x}, \texttt{n}-\texttt{x}+1) = \texttt{BETAINV}(1-\gamma, \texttt{x}, \texttt{n}-\texttt{x}+1)$$

For $x = 0$ we have $P_{p_0}(X \geq 0) = 1 > \alpha$ and we cannot use (5), but the original definition of $C(x)$ leads to $C(0) = [0, 1]$.

We define $\hat{p}_L(\gamma, 0, n) = 0$ in that case, with the understanding that the left-open interval $(\hat{p}_L(\gamma, x, n), 1]$ becomes closed when $x = 0$.

An Example and Some Observations

As a check example take $x = 451$ and $n = 500$ with $\gamma = .95$, then one gets $\hat{p}_L(.95, 451, 500) = \texttt{qbeta}(.05, 451, 500 - 451 + 1) = 0.8773207$ as 95% lower bound for p.

Using (5) it is a simple exercise to show that the sequence of lower bounds is strictly increasing in x, i.e.,

$$0 = \hat{p}_L(\gamma, 0, n) < \hat{p}_L(\gamma, 1, n) < \ldots < \hat{p}_L(\gamma, n-1, n) < \hat{p}_L(\gamma, n, n) < 1 .$$

$$\implies \quad P_p(\hat{p}_L(\gamma, X, n) < p) = 1 \quad \text{for} \quad p > \hat{p}_L(\gamma, n, n).$$

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.

27

Special Cases

For $x = 1$ and $x = n$ the lower bounds defined by (5) 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**. The last approximation is 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.

When observing $X = n$ successes in n trials, especially when n is large, we still are not inclined to estimate p by $\hat{p} = n/n = 1$, because of the consequences of such a strong statement.

$$\hat{p}_L(1-\alpha/2,X,n) < \hat{p}_U(1-\alpha/2,X,n)$$

for $0 < \alpha < 1$ $\qquad P_p(\hat{p}_L(1-\alpha/2,X,n) < \hat{p}_U(1-\alpha/2,X,n)) = 1$ $\qquad$ for any p .

Suppose $\hat{p}_U(1-\alpha/2,x,n) \leq \hat{p}_L(1-\alpha/2,x,n)$ for some x

$\implies \hat{p}_U(1-\alpha/2,x,n) \leq p_0 \leq \hat{p}_L(1-\alpha/2,x,n)$ for some p_0 and some x.

With that x the hypothesis $H(p_0) : p = p_0$ would be rejected when testing against $A(p_0) : p < p_0$ or $\tilde{A}(p_0) : p > p_0$, i.e.,

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

and by adding those two inequalities we get

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

Confidence Intervals for p

Lower and upper confidence bounds, each with respective confidence coefficient $1 - \alpha/2$, can be used simultaneously as a $100(1 - \alpha)\%$ confidence interval

$$(\hat{p}_L(1 - \alpha/2, X, n), \ \hat{p}_U(1 - \alpha/2, X, n)) \,.$$

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

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

Confidence Coefficient of Confidence Intervals for p

$$\bar{\gamma} = \inf_p P_p(\hat{p}_L(1-\alpha/2,X,n) \leq p \leq \hat{p}_U(1-\alpha/2,X,n))$$

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

The inequality $\geq$ typically takes the strict form $>$

since the suprema are usually not attained or approached at the same p.

Coverage Probabilities

The coverage plots shown here were produced in R. In the case of the lower bound coverage probability calculate for fine grid of p values

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

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

$$P(\hat{p}_U(\gamma, X, n) \ge p) = \sum_{x=0}^{n-1} I_{\{\texttt{qbeta}(\gamma, x+1, n-x) \ge p\}} \binom{n}{x} p^x (1-p)^{n-x} + p^n,$$
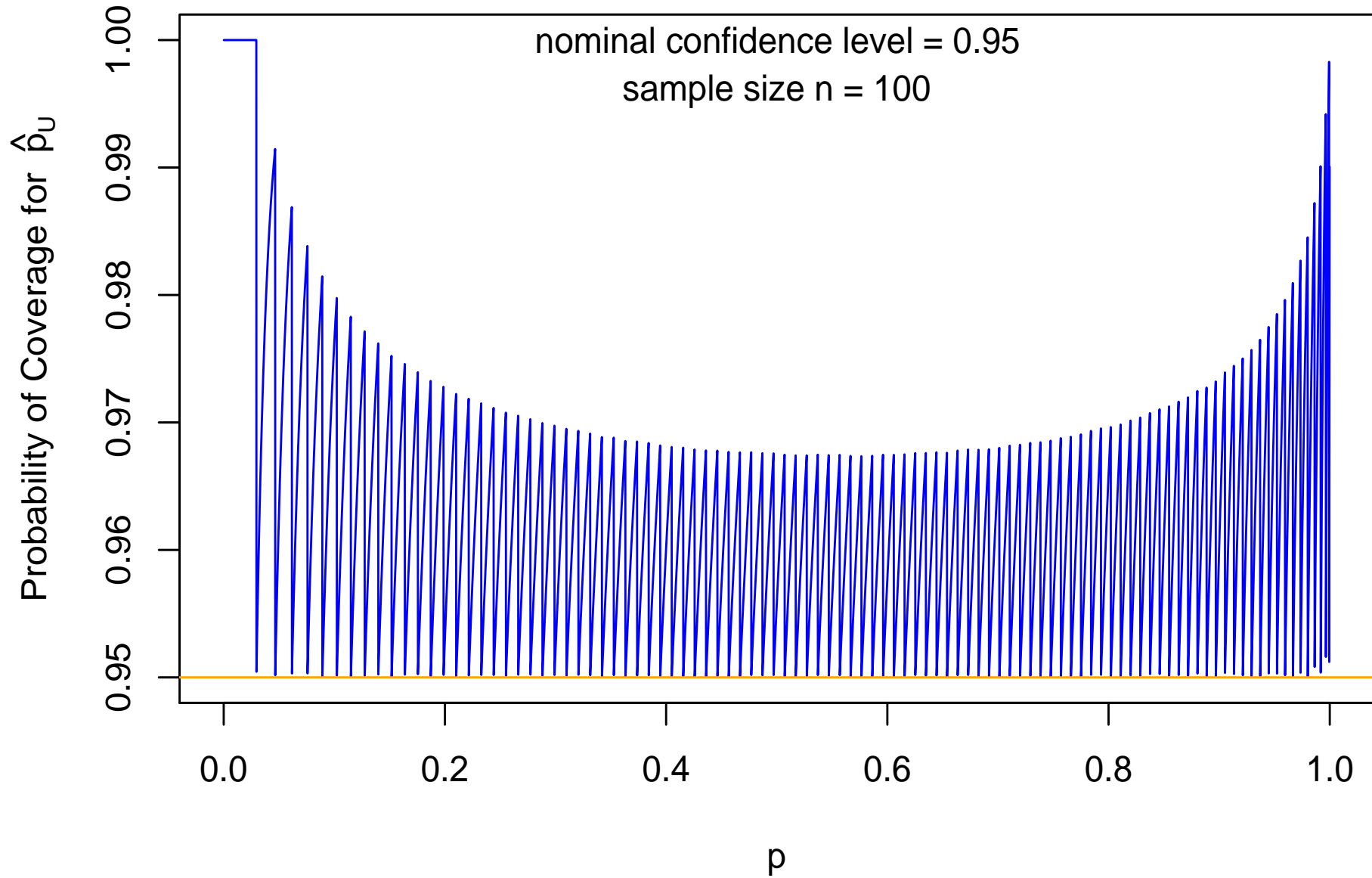
while for the confidence interval one calculates the coverage probability as

$$P(\hat{p}_L((1+\gamma)/2, X, n) \le p \cap \hat{p}_U((1+\gamma)/2, X, n) \ge p)$$

$$= \sum_{x=1}^{n-1} I_{\{\texttt{qbeta}((1-\gamma)/2, x, n-x+1) \le p \cap \texttt{qbeta}((1+\gamma)/2, x+1, n-x) \ge p\}} \binom{n}{x} p^x (1-p)^{n-x}$$

$$+ (1-p)^n I_{\{\texttt{qbeta}((1+\gamma)/2, 1, n) \ge p\}} + p^n I_{\{\texttt{qbeta}((1-\gamma)/2, n, 1) \le p\}}$$
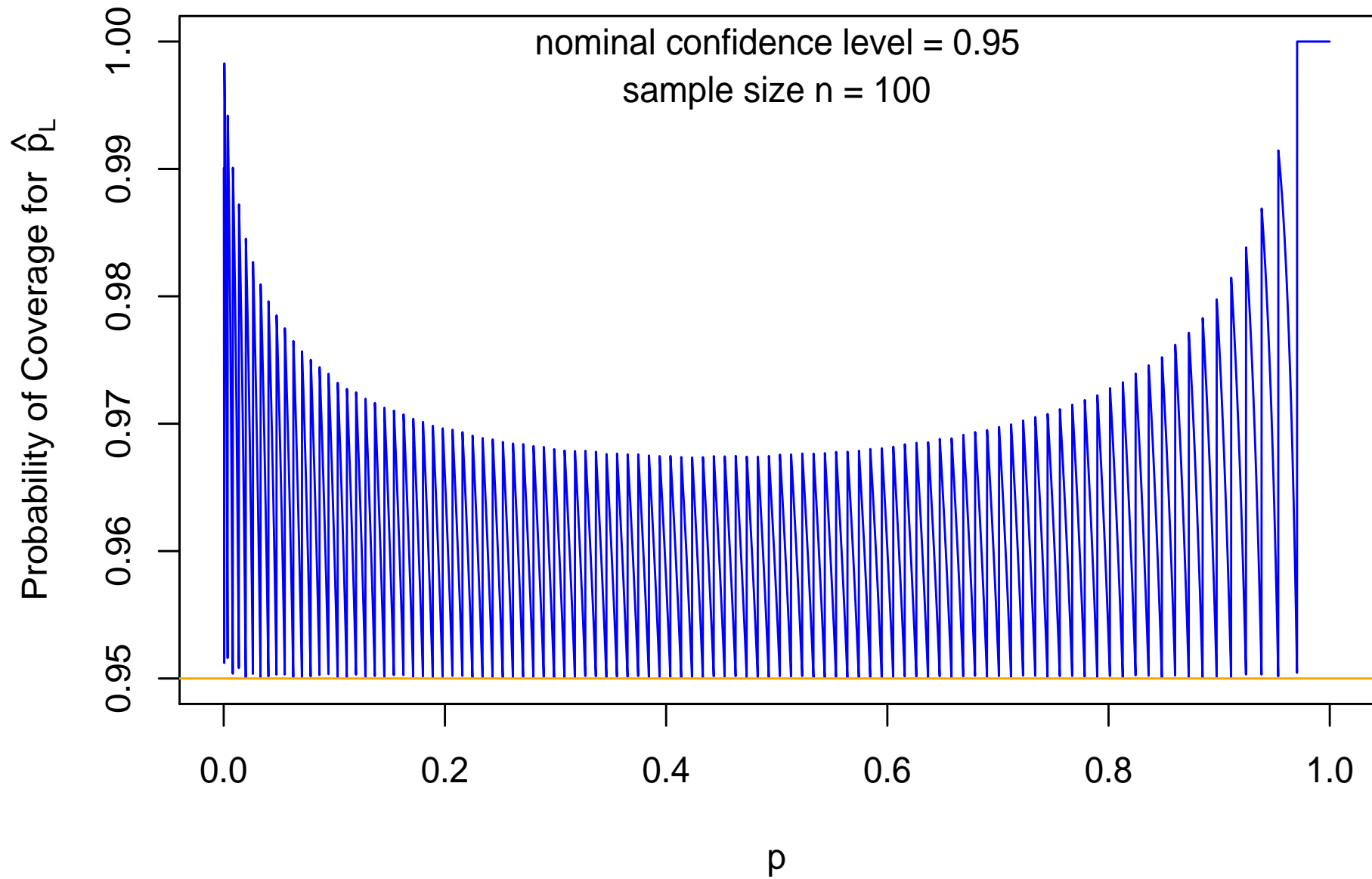
R Code for Upper Bound Coverage Probability

```
> bin.min.coverage.upper
function (n=100,gam=.95,pdf=F)
{# taking advantage of vectorized calculations over p
if(pdf==T) pdf(file="mincoverageupper.pdf",width=7,height=5)
p.coverage=0
p=seq(0.0001,.9999,.0001)
for(x in 0:(n-1)){
   p.coverage=p.coverage+(qbeta(gam,x+1,n-x)>=p)*dbinom(x,n,p)
}
p.coverage=p.coverage+dbinom(n,n,p)
plot(p,p.coverage,type="l",ylab=
  expression("Probability of Coverage for "~hat(p)[U]),
  col="blue",ylim=c(gam,1))
abline(h=gam,col="orange")
text(.5,1,paste("nominal confidence level =",gam),adj=.5)
text(.5,.996,paste("sample size n =",n),adj=.5)
if(pdf==T) dev.off()
}
```
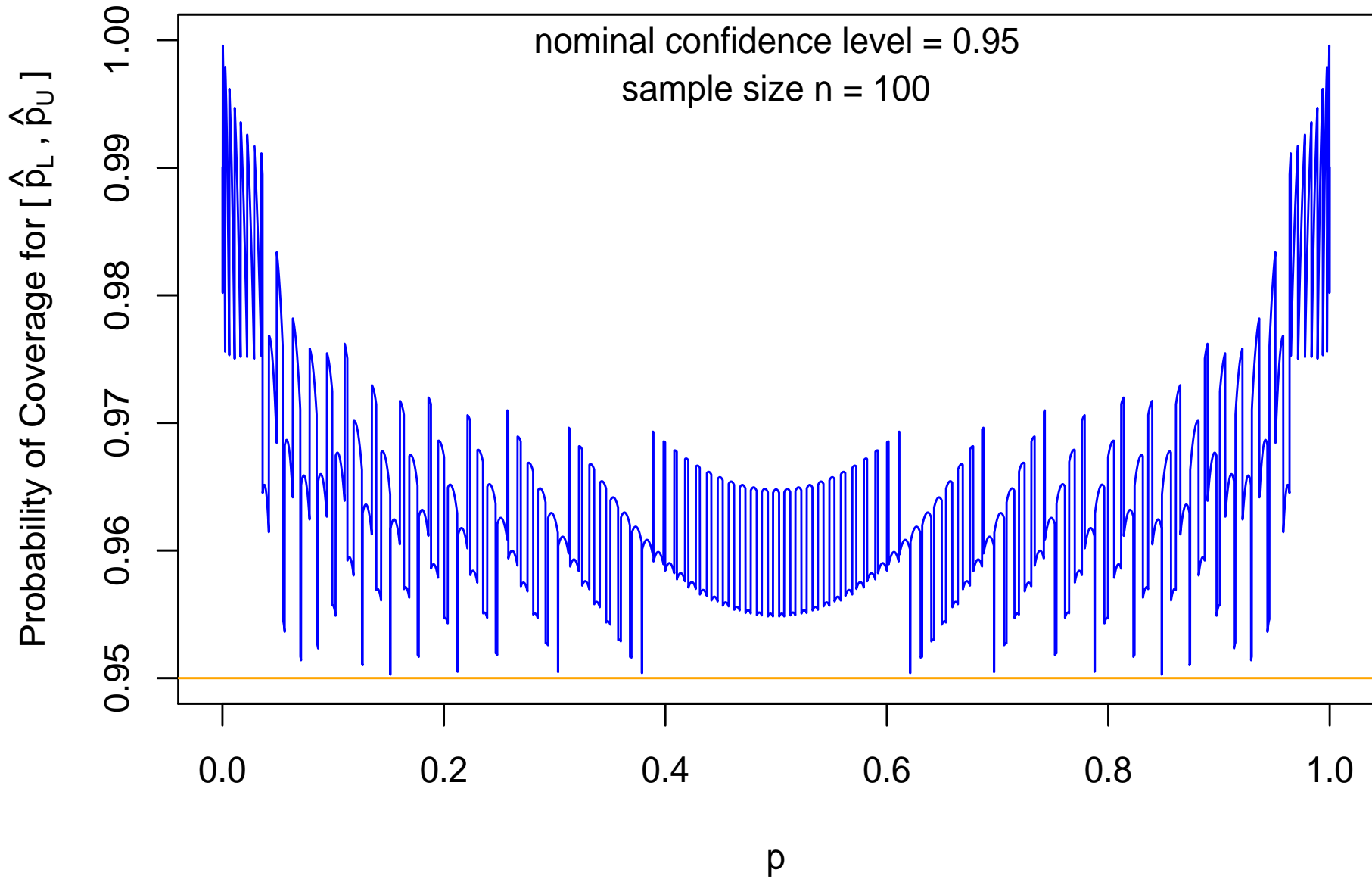
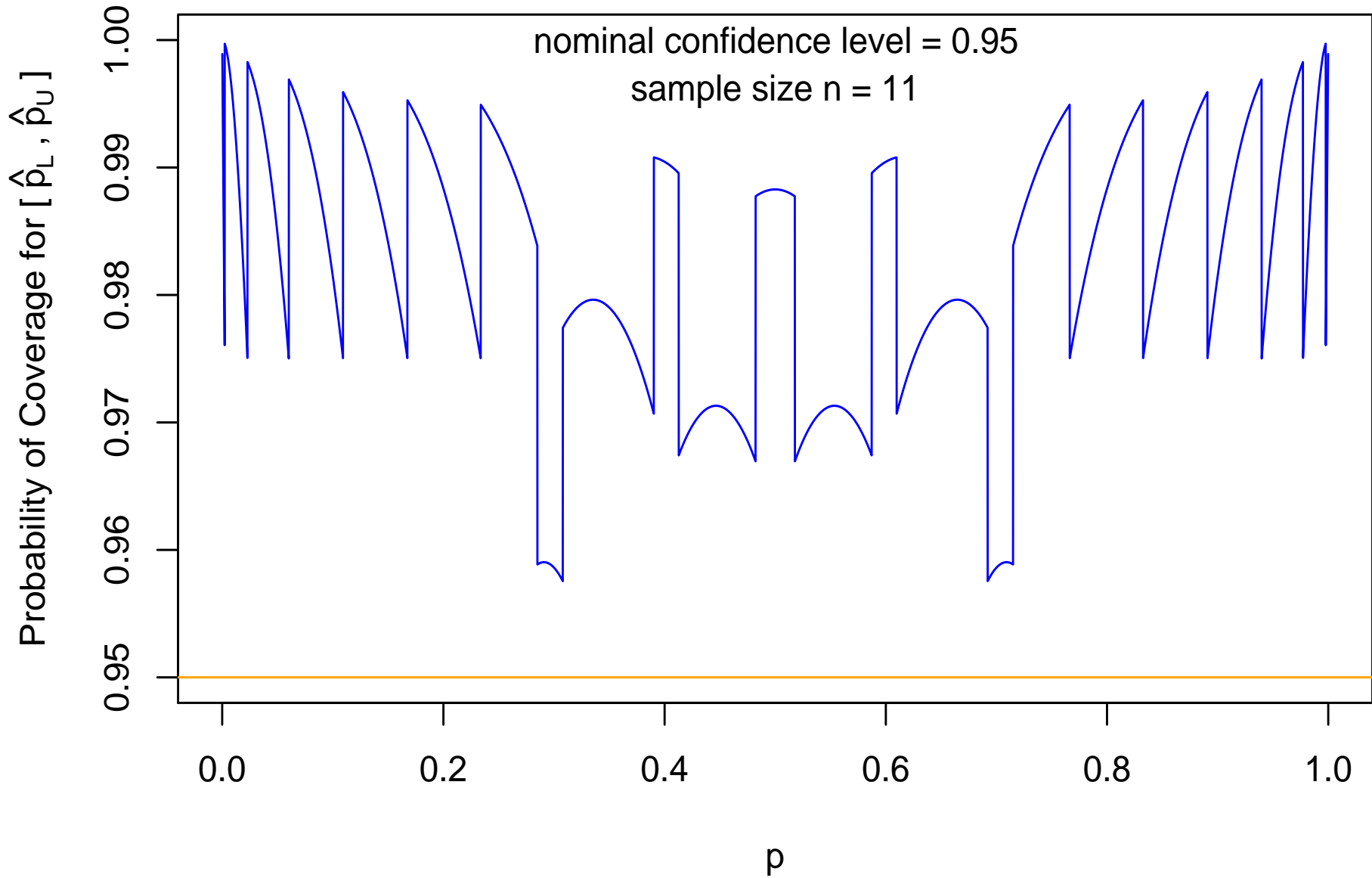Coverage Probability of Lower Confidence Bounds

nominal confidence level = 0.95

sample size n = 100

Coverage Probability of Confidence Intervals

nominal confidence level = 0.95
sample size n = 100

Probability of Coverage for $[\hat{p}_L, \hat{p}_U]$

p

Coverage Probability of Confidence Intervals ($n = 11$)

nominal confidence level = 0.95

sample size n = 11

Confidence Intervals for $p \approx 0$ or $p \approx 1$

As the previous plots showed, for intervals the coverage probabilities stay above $1 - \alpha/2 = 1 - (1 - \gamma)/2 = (1 + \gamma)/2$ for p near 0 or 1.

This is a consequence of coverage probability 1 for upper bounds (lower bounds) when p is near 0 (near 1):

$$P_p(\hat{p}_L(1 - \alpha/2, X, n) \leq p \leq \hat{p}_U(1 - \alpha/2, X, n))$$

$$= 1 - P_p(\hat{p}_L(1 - \alpha/2, X, n) > p) - P_p(\hat{p}_U(1 - \alpha/2, X, n) < p)$$

$$= 1 - P_p(\hat{p}_L(1 - \alpha/2, X, n) > p) \geq 1 - \alpha/2 \qquad \text{for } p < \hat{p}_U(1 - \alpha/2, 0, n)$$

$$= 1 - P_p(\hat{p}_U(1 - \alpha/2, X, n) < p) \geq 1 - \alpha/2 \qquad \text{for } p > \hat{p}_L(1 - \alpha/2, n, n)$$

Confidence Bounds or Intervals?

For $p \approx 0$ it makes little sense to use lower bounds or intervals, wasting half the miss probability where it does not matter.

Most often the lower bound would be 0 anyway, a useless piece of information.

One should then allocate all of the miss probability α to the upper bound, i.e., use

$$\hat{p}_U(1-\alpha, X, n) \qquad \text{rather than} \qquad \hat{p}_U(1-\alpha/2, X, n) \qquad \text{which is higher.}$$

If $X = 0$ we can ask how large m should be so that

$$\hat{p}_U(1-\alpha/2, 0, m) = 1 - (\alpha/2)^{1/m} = \hat{p}_U(1-\alpha, 0, n) = 1 - (\alpha)^{1/n} \ ?$$

$$m = n \times \frac{\log(\alpha) - \log(2)}{\log(\alpha)} \qquad \text{For } \alpha = .05 \quad \text{we need} \quad m = 1.2314 \times n.$$

Provided $X = 0$ for both n and m, we would need 23% higher sample size with the more wasteful procedure in order to get the same upper bound.

Corresponding comments apply for $p \approx 1$ and lower bounds.

Classical or Elementary Upper Bounds

In elementary statistics classes the following upper bound is usually presented

$$\hat{p} + z_\gamma \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \qquad \text{with} \qquad \hat{p} = \frac{x}{n} \quad \text{and} \quad z_\gamma = \gamma\text{-quantile of } Z \sim \mathcal{N}(0,1) \ .$$
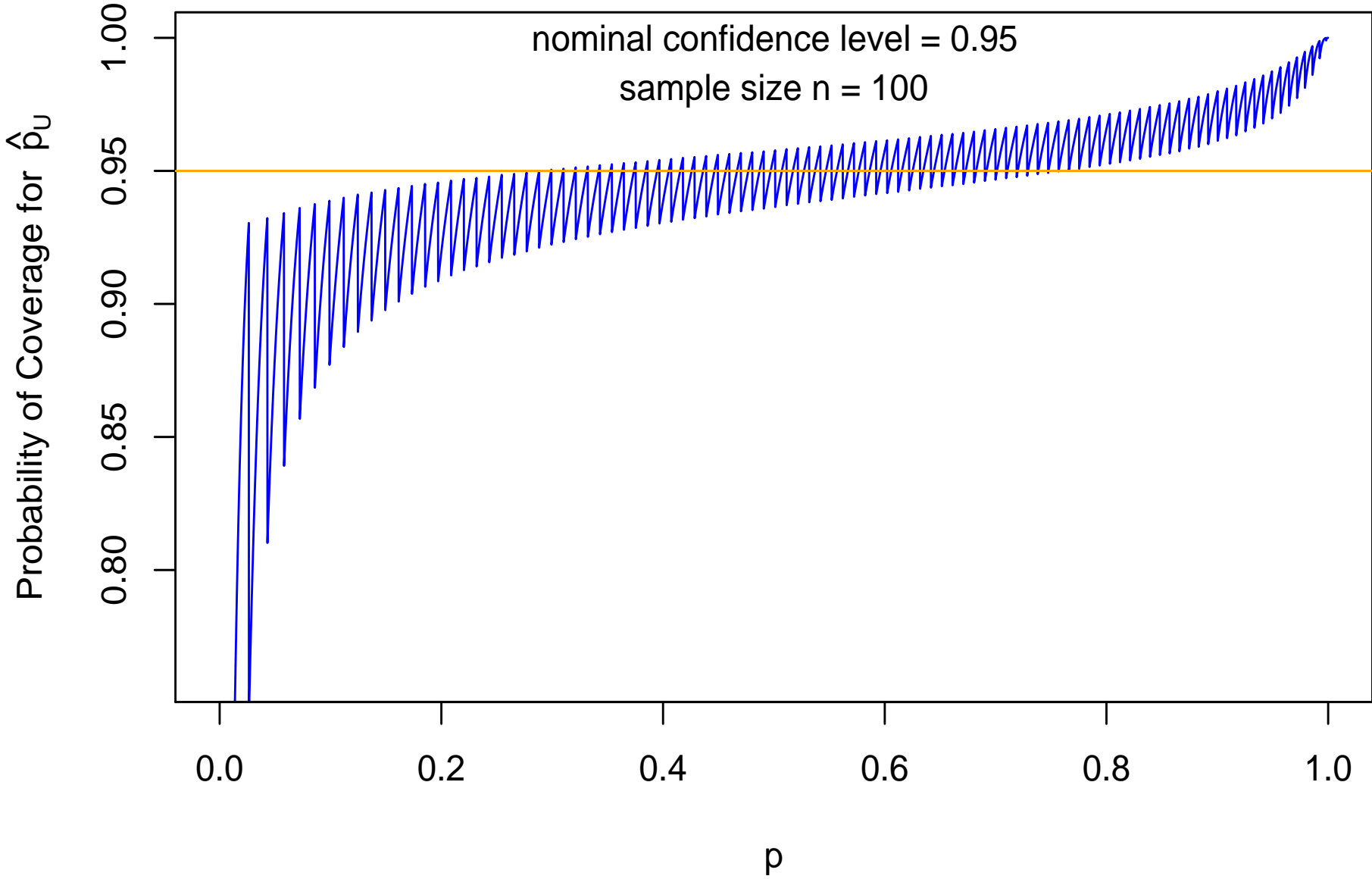
It is based on the approximation

$$\frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \ \sim \ \mathcal{N}(0,1) \quad \text{for large } n \qquad (\text{poor for } p \text{ near 0 or 1}) \ .$$

$$P\left(p \le \hat{p} + z_\gamma \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right) = P\left(-z_\gamma \le \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \right) \approx P\left(-z_\gamma \le Z \right) = P(Z \le z_\gamma) = \gamma$$

Confidence coefficient $\bar{\gamma} = 0$,

since for very small p the upper bound becomes 0 with probability approaching 1.

Coverage Probability of Elementary 95% Upper Bounds

nominal confidence level = 0.95
sample size n = 100

Score Test Upper Bounds

Uses the same testing/confidence bounds duality as did Clopper-Pearson.

Here use a normal approximation to calculate the p-value $p(x, p_0) = P_{p_0}(X \leq x)$

Use $X \sim \mathcal{N}(np_0, np_0(1-p_0))$ to find all acceptable p_0 with $P_{p_0}(X \leq x) > \alpha$, i.e., solve for p_0

$$\alpha = P_{p_0}(X \leq x) = P_{p_0}\left(\frac{X - np_0}{\sqrt{np_0(1-p_0)}} \leq \frac{x - np_0}{\sqrt{np_0(1-p_0)}} \right)$$

$$\approx P\left(Z \leq \frac{x - np_0}{\sqrt{np_0(1-p_0)}} \right) \implies \frac{x - np_0}{\sqrt{np_0(1-p_0)}} = z_\alpha = -z_{1-\alpha} = -z_\gamma$$

$$\implies p_0 = \tilde{p}_U(\gamma, x, n) = \frac{z^2/(2n) + \hat{p}}{1 + z^2/n} + \frac{z}{1 + z^2/n}\sqrt{\frac{z^2}{4n^2} + \frac{\hat{p}(1-\hat{p})}{n}}, \quad \text{with } z = z_\gamma$$

after some algebraic gyrations. Not very appealing or mnemonic.

Coverage Probability of Score Test 95% Upper Bounds

nominal confidence level = 0.95

sample size n = 100
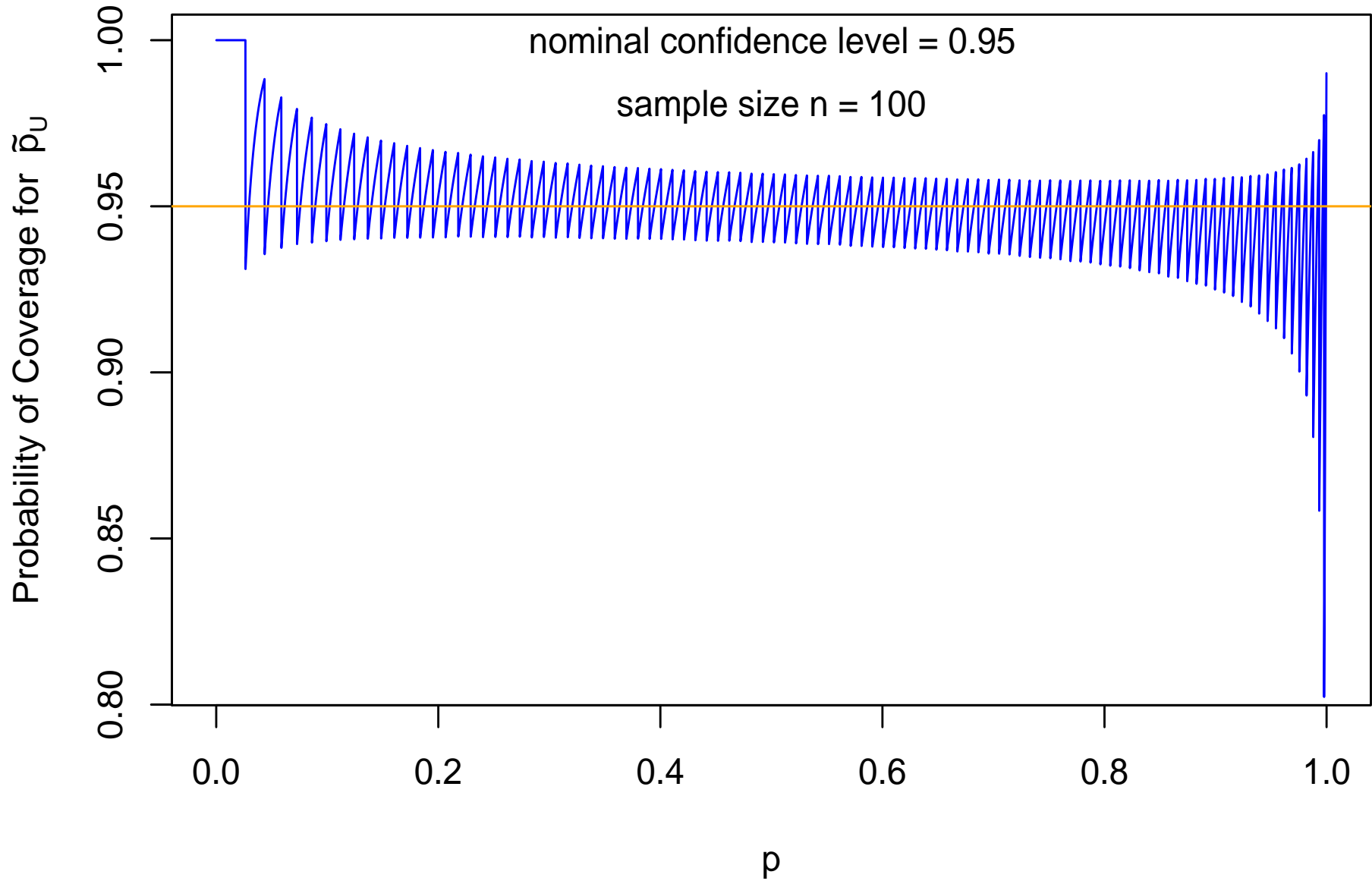
Commentary on Score Test Upper Bounds

Agresti and Coull (1998) claim that score test confidence intervals (combining lower and upper bounds) show better average coverage behavior.

No continuity correction, using $(x \pm .5)/n$ in the upper/lower bounds respectively.

We chose to examine this from the upper bound perspective. The claim seems OK, but for $p \approx 1$ the coverage deteriorates. That is of little interest.
However, it averages out against the coverage probability of 1 near $p = 0$.

Average (like minimum) coverage probability is just another metric.

We could adjust the Clopper-Pearson bounds (lowering their minimum coverage) to have same average coverage behavior.

Trying to raise the score test bounds to have their minimum coverage equal that of the Clopper-Pearson bounds would be more problematic.

Both types of bounds show roughly the same type of zigzag swings.

Coverage Probability of Score Test 95% Upper Bounds with Continuity Correction

nominal confidence level = 0.95

sample size n = 100
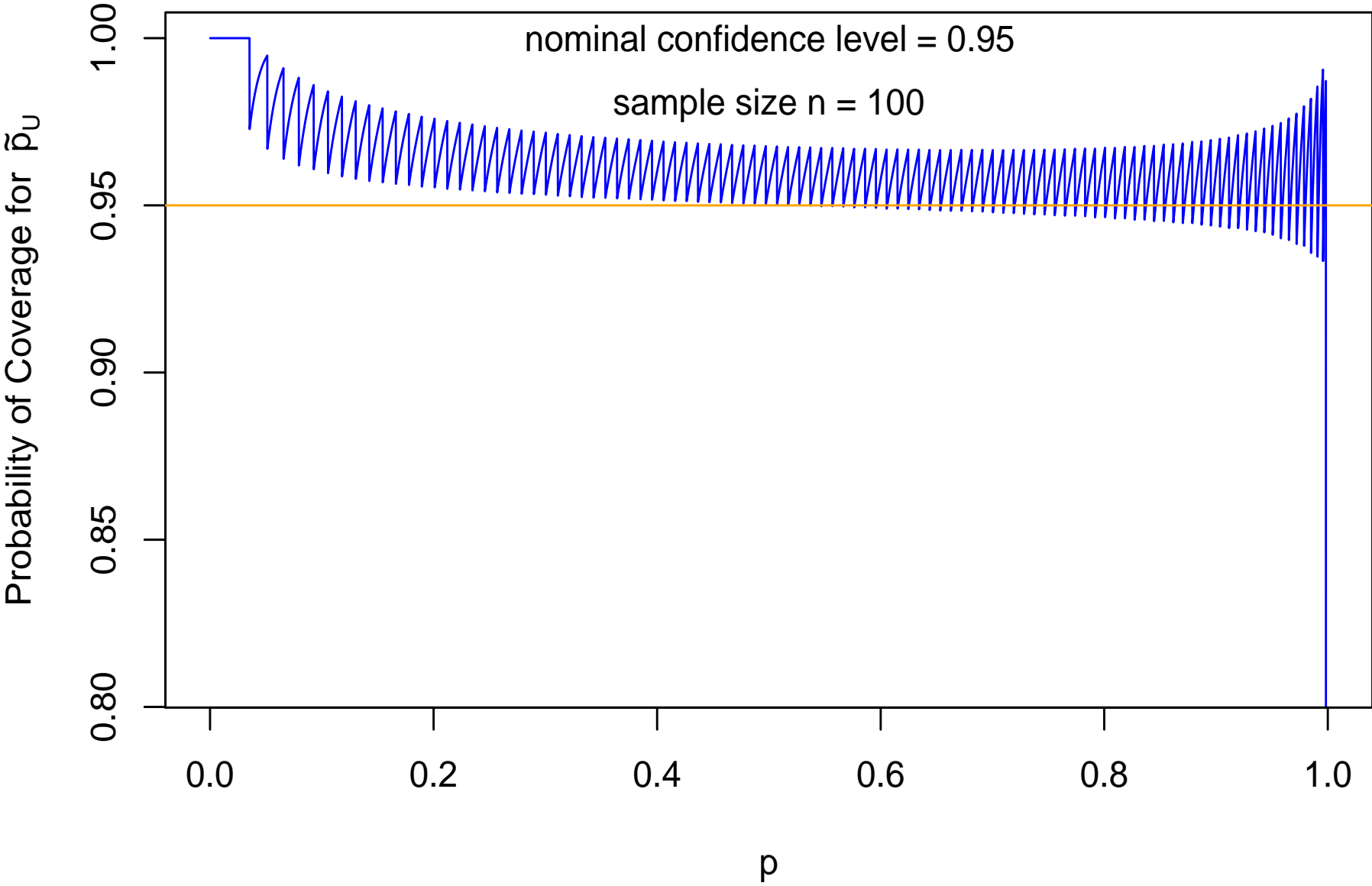
Coverage Probability of Score Test 95% Confidence Intervals

nominal confidence level = 0.95

sample size n = 100
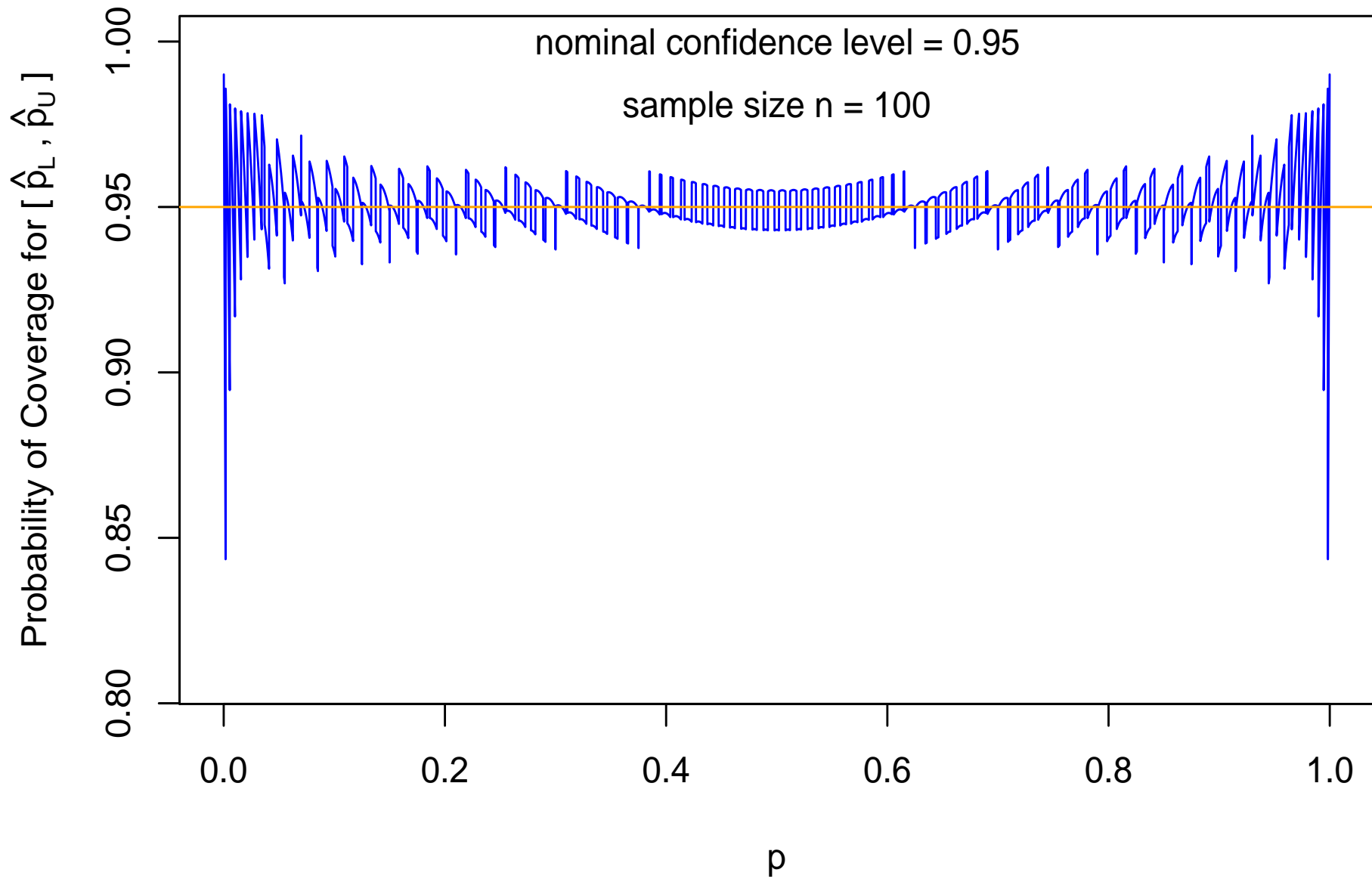
Corrosion Inspection

Airplanes are inspected for corrosion at a 10 year inspection interval.

All n inspected aircraft made it through corrosion inspection without any findings.

Customer wants 95% lower confidence bounds on the probability of an airplane passing its corrosion inspection without any findings.

The customer did not tell me n, only that 2.5% of the fleet had been inspected. How do we deal with this, not knowing the fleet size?

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.

$X =$ number of aircraft without any corrosion found at their respective inspections is a binomial random variable with parameters n and success probability p.

$$X = n \implies \hat{p}_L(.95, n, n) = (1 - .95)^{1/n}.$$

Plot $\hat{p}_L(.95, n, n)$ against n and the customer can take it from there.

Lower 95% Confidence Bounds as Function of n

y-axis: 95 % lower confidence bound on p=P(no corrosion in 10 years)

x-axis: number n of planes inspected, all without corrosion findings

48
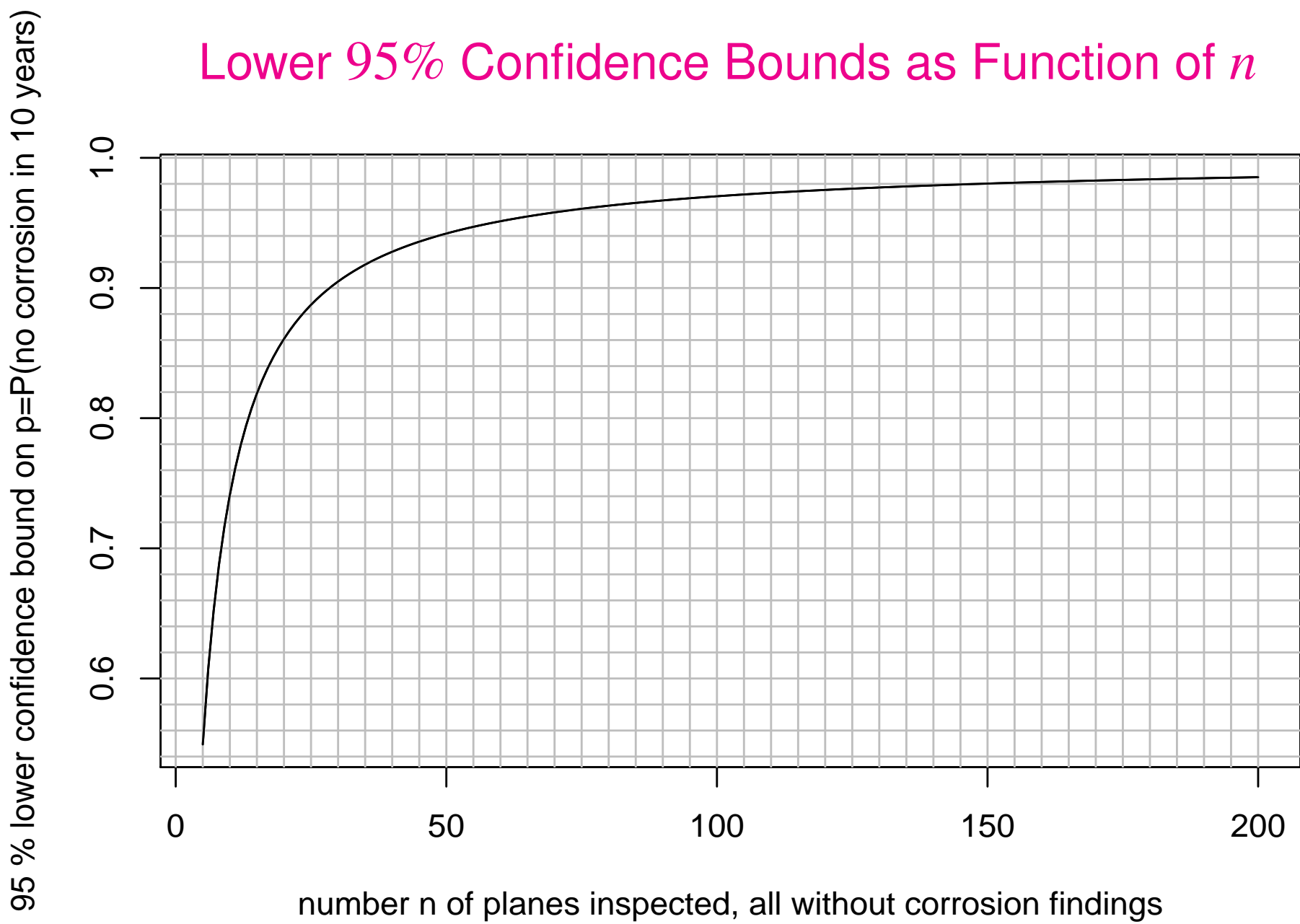
Some Commentary

Each such inspection involves extensive stripping of aircraft interior.

An airline may have asked about the necessity of this onerous and costly task after finding no corrosion in n inspection.

Maybe it is felt that the inspection interval 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 and cheaper to fix than when it has progressed beyond extensive or impossible repair.

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.

Space Shuttle Stud Hang-Ups

Blast
Container

SRB
Aft Skirt

Holddown
Stud

Frangible
Nut

Aft Skirt
Shoe

MLP
Support
Post

Ejection into post
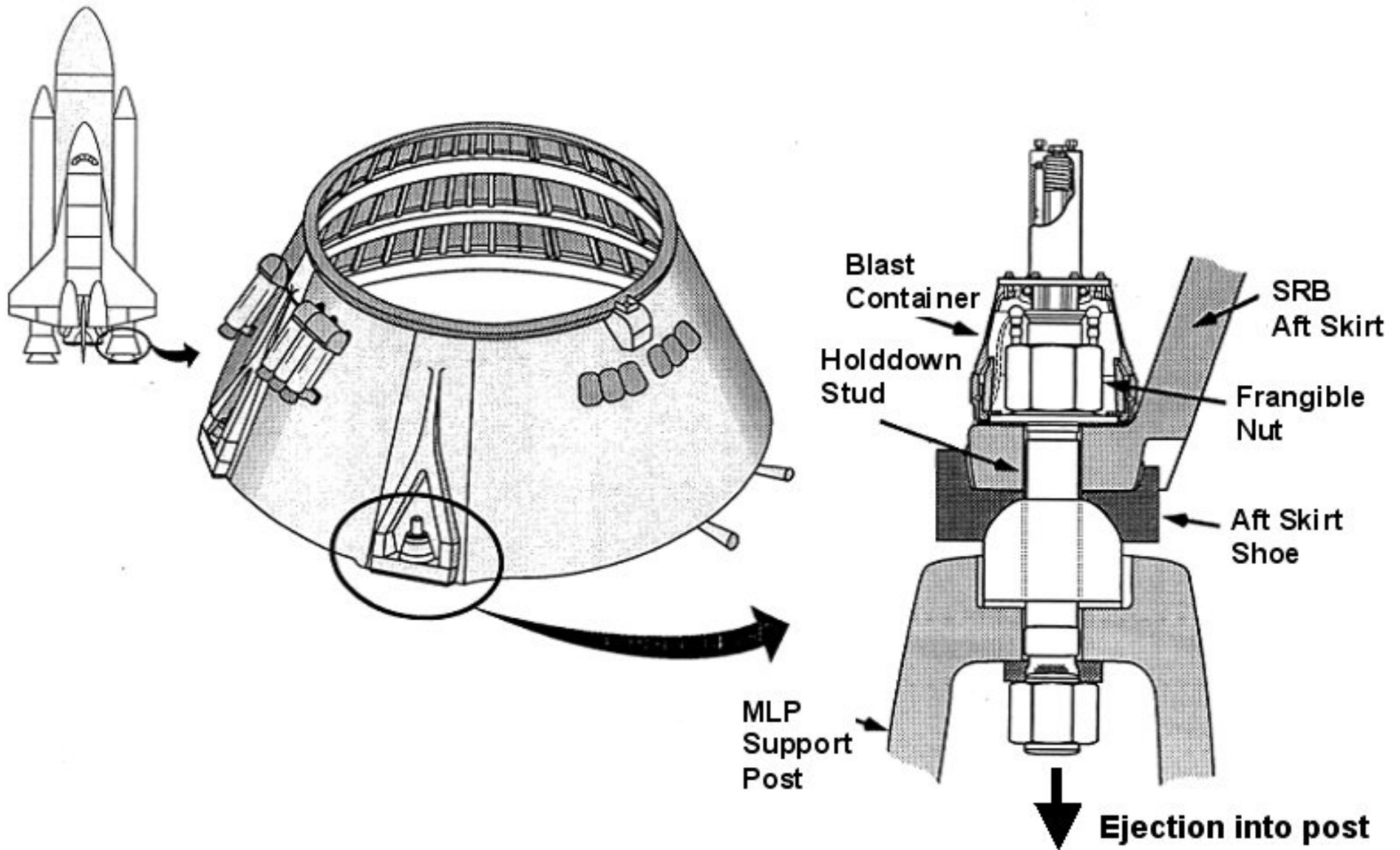
The Issues and Facts

The studs have $3.5''$ diameter. If they do not drop through cleanly they hang up.

This can even be felt in the shuttle cabin.

It impacts some performance parameters (indicators): how severely?

In an extreme case the studs just snap under the tremendous takeoff force.

In 114 liftoffs there were 21 with one hang-up and 2 with two hang-ups.

No liftoff with more than 2 hang-ups.

There are 4 studs per Solid Rocket Booster (SRB), 8 total.

http://www.eng.uab.edu/ME/ETLab/HSC04/abstracts/HSC147.pdf

and   http://www.nasa.gov/offices/nesc/home/Feature_1.html

Modeling and a Check

If we treat each liftoff as 8 Bernoulli trials (one for each stud) we have experienced $114 \times 8 = 912$ trials, with $21 \times 1 + 2 \times 2 = 25$ "successes."
We get $\hat{p} = 25/912 = .02741$ as estimate of our success probability.

If indeed the 8 studs per liftoff act as independent Bernoulli trials we would estimate the probability of seeing i hang-ups in a liftoff as

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

From this one gets $\hat{p}_0 = 0.8006$, $\hat{p}_1 = 0.1805$, $\hat{p}_2 = 0.0178$, and $\hat{p}_{3+} = 0.0010$.

Based on this one would have expected $E_i = \hat{p}_i \times 114$ liftoffs with i stud hang-ups.
$\implies E_0 = 91.27$, $E_1 = 20.58$, $E_2 = 2.03$ and $E_{3+} = 0.12$

These 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.

Upper Confidence Bounds

The Clopper-Pearson 95% upper confidence bound for 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 $f_4(p)$, since

$$P_p\left(f_4(p) \leq f_4(\hat{p}_U(.95, X, 912))\right) = P_p\left(p \leq \hat{p}_U(.95, X, 912)\right) \geq .95$$

Because of the small risk upper bound of 0.00013 it was decided to focus further concerns to $X \leq 3$.

Simulations of Indicators

NASA has intricate programs that can simulate liftoffs with many randomly varying environmental inputs, material properties, etc. It takes considerable time to run these simulations.

Before each mission (with different payloads) liftoff simulations are carried out: 800 each, under 0 hang-ups, 1 hang-up, 2 hang-ups, and 3 hang-ups.

For each simulation many indicator variables (stresses, deviation angles, etc.) are monitored. Each such indicator Y has a safe range or a safe threshold, say y_0.

The task was to determine the 3σ risk for each of those indicators.

In the engineering world 3σ refers to a normal distribution and implies a risk of .00135 of a normal random variable deviating by more than 3σ above (or below) its mean. The 2-sided risk is .0027.

Law of Total Probability

Recall the following law of total probability which facilitates probability calculations for events, that are complex initially but become easier to manage when split into several mutually exclusive cases.

Suppose $\quad B_1 \cup \ldots \cup B_k = \Omega$ (full probability space) $\quad$ with $B_i \cap B_j = \emptyset$ for $i \neq j$

$$
\begin{aligned}
P(A) &= P(A \cap \{B_1 \cup \ldots \cup B_k\}) = P(\{A \cap B_1\} \cup \ldots \cup \{A \cap B_k\}) \\
&= P(A \cap B_1) + \ldots + P(A \cap B_k) \quad \text{since } \{A \cap B_i\} \cap \{A \cap B_j\} = \emptyset \text{ for } i \neq j \\
&= P(A|B_1)P(B_1) + \ldots + P(A|B_k)P(B_k)
\end{aligned}
$$

since $P(A|B_i) = P(A \cap B_i)/P(B_i)$ by definition.

The Risk of Exceeding a Threshold y_0

It is desired 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 $\bar{F}_p(y_0)$

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

For many (but not for all) of the response variables one can plausibly assume that $P(Y > y_0 | X = x)$ increases with x.

Using $\psi(x) = P(Y > y_0 | X = x)$ for $x = 0, 1, 2, 3$ and $\psi(x) = 1$ for $x > 3$ we can write

$$\bar{G}_p(y_0) = E_p(\psi(X))$$

Based on the general binomial monotonicity property 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.

Simulated CDF of Indicator Variable

$\hat{F}(x) =$ proportion of sample $X_i \leq x$

0 hang–ups
1 hang–up
2 hang–ups
3 hang–ups
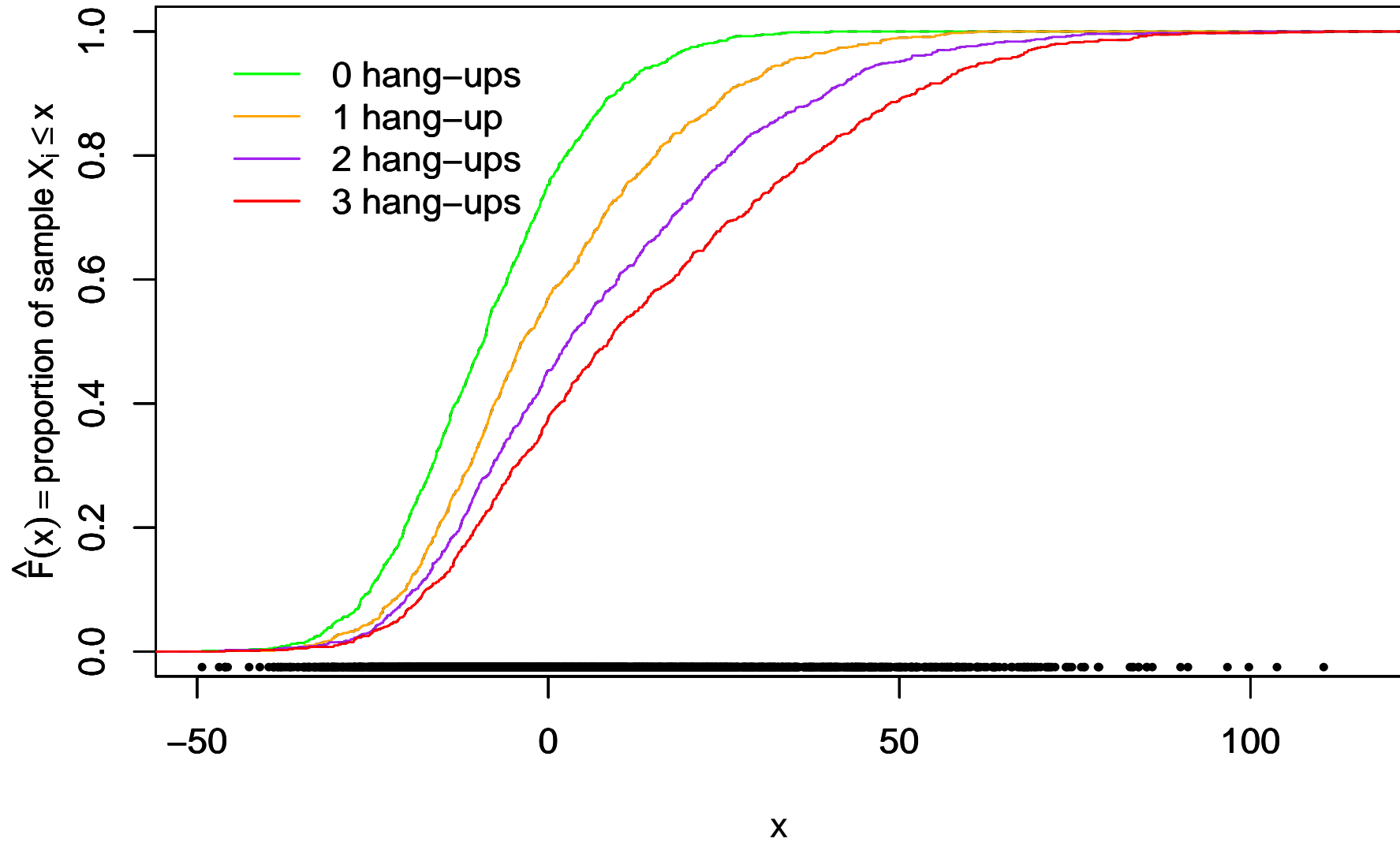
Magnified CDF of Indicator Variable
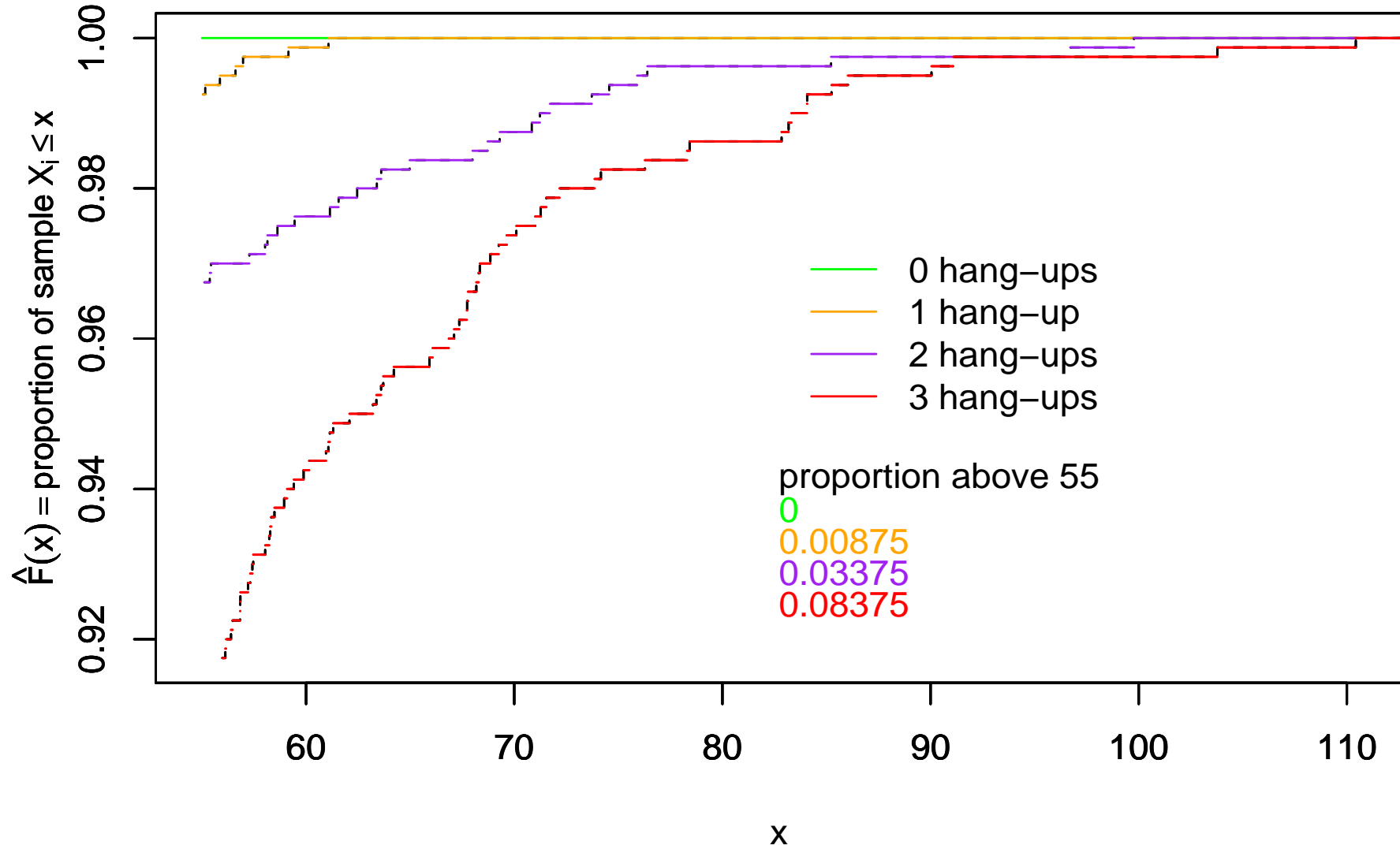
Bounding $\bar{G}_p(55) \geq P(Y > 55)$

Based on the numbers $0, .00875, .03375$, and $.08375$ from the previous slide and the upper bound $\hat{p}_U(.95, 25, 912) = .03807645 = \hat{p}_U$ for p we calculate the following 95% upper confidence bound for $\bar{G}_p(55) \geq P(Y > 55)$

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

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

This is larger than the $.00135$ risk associated with the 3σ requirement.

Even the estimate of $\bar{G}_p(55)$ (using $\hat{p} = 25/912$ instead of $\hat{p}_U$) only yields 0.002300867.

Some Comments

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.

These kinds of analyses can whittle down the number of cases where serious engineering review is warranted.

Perspectives on Risk Levels

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: # of flight critical systems ≤ 100.

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 $100 \times 10^{-9} = 10^{-7}$.

This is still well below the current rate of hull losses and most of these are not due to any aircraft design deficiencies. See

$$\mathrm{http}://\mathrm{www.boeing.com/news/techissues/pdf/statsum.pdf}$$

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

Perspectives on Risk Levels

Establishing a 10^{-9} risk directly from system exposure data is difficult
if not impossible, and certainly impracticable during the design phase.

The low exponent (-9) of such risks is built up by multiplying verifiable risks of higher order, e.g., $10^{-4} \times 10^{-5} = 10^{-9}$

Such multiplication is based on statistical independence and redundancy in the design, like having at least two engines on a plane, or doing engine maintenance at different times or by different mechanics.

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.

The Poisson Distribution and Monotonicity

For a Poisson random variable X with mean λ ($\lambda > 0$) we have

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

On intuitive grounds, a larger mean λ should entail a decrease in $P_\lambda(X \leq k)$.

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

Connection to the Gamma Distribution

By the Fundamental Theorem of Calculus we have

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

$$\implies P_\lambda(X \leq k) = \sum_{i=0}^k \frac{\exp(-\lambda)\lambda^i}{i!} = \int_\lambda^\infty \frac{\exp(-x)x^{k+1-1}}{\Gamma(k+1)}\,dx \qquad (6)$$

since both the left and right side converge to 1 as $\lambda \to 0$.

Here the righthand side of (6) is the right tail probability $P(V \geq \lambda)$

of a Gamma random variable V with scale 1 and shape parameter $k+1$.

$$G_{k+1}(\lambda) = P(V \leq \lambda) = \int_0^\lambda \frac{\exp(-x)x^{k+1-1}}{\Gamma(k+1)}\,dx$$

is also called incomplete Gamma function with parameter $k+1$.

Uses of the Poisson Distribution

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. Thus the often used description: Random failures.

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 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.

Binomial Distribution $\approx$ Poisson Distribution for Small p

For small p: $\text{Binomial}(n, p) \approx \text{Poisson}(\lambda = np)$.

In fact, assuming $np_n = \lambda$ the limiting behavior of the binomial probabilities

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

$$\longrightarrow \frac{\exp(-\lambda)\lambda^x}{x!} \quad as \quad n \to \infty$$

is very instrumental in motivating the Poisson distribution.

Even though the above argument involves a limiting operation ($n \to \infty$) 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.

Poisson-Binomial Approximation

Let $I_1, \ldots, I_n$ be independent Bernoulli r.v.s 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 is quite complicated 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 $\lambda = \sum_{i=1}^{n} p_i = n\bar{p}$.

If $X \sim \mathsf{P}(\lambda)$ we can approximate $P(W \in A)$ by $P(X \in A)$ for any set A.

Approximation Error Bound

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 d_{TV}

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

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

Poisson-Binomial Approximation
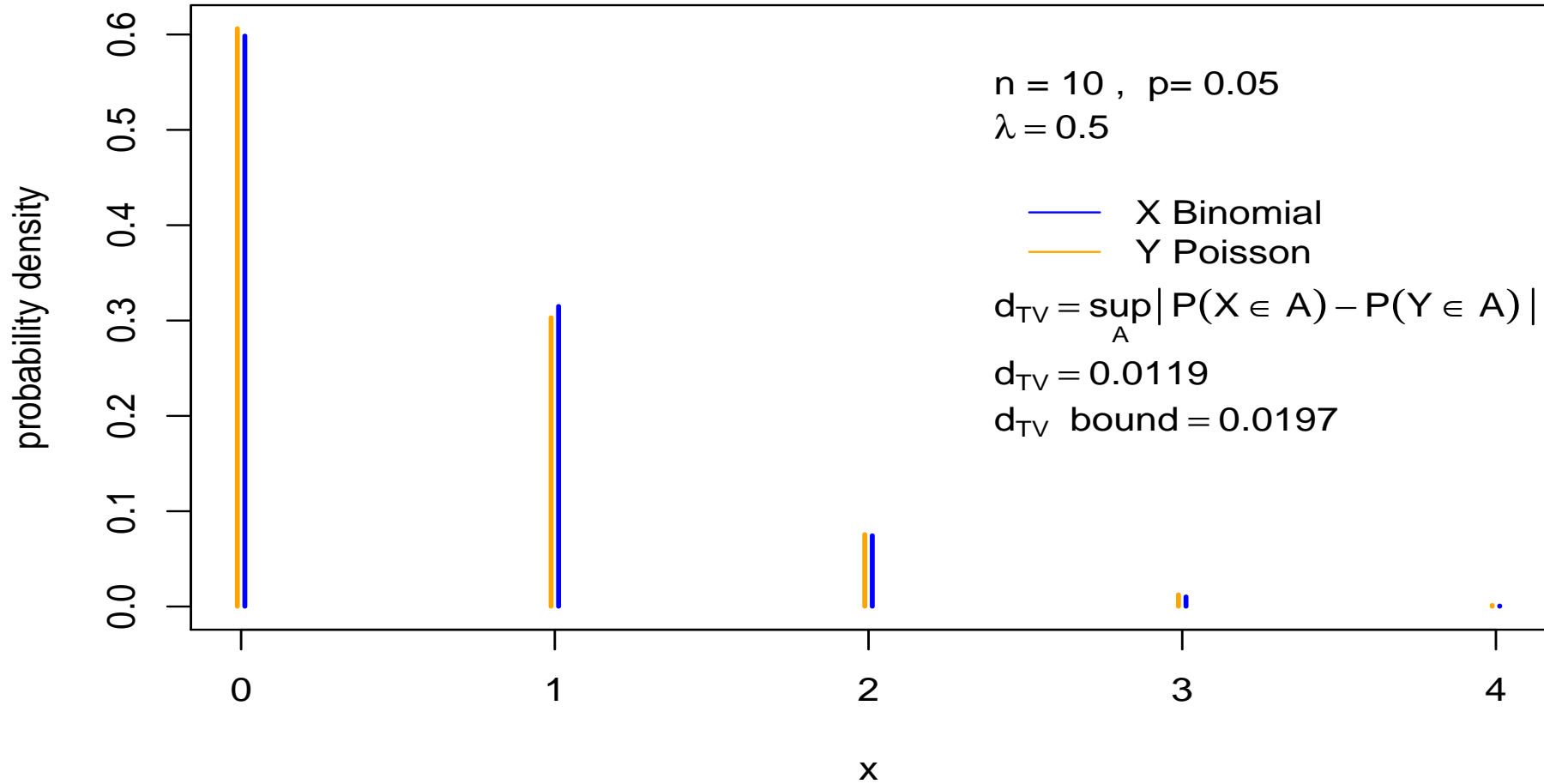
n = 10 , p= 0.05
$\lambda = 0.5$

—— X Binomial
—— Y Poisson

$d_{TV} = \sup_{A} \left| P(X \in A) - P(Y \in A) \right|$

$d_{TV} = 0.0119$
d_{TV} bound $= 0.0197$

Poisson-Binomial Approximation

probability density

0.15
0.10
0.05
0.00

n = 100 , p= 0.05
$\lambda = 5$

—— X Binomial
—— Y Poisson

$d_{TV} = \sup_{A} \left| P(X \in A) - P(Y \in A) \right|$

$d_{TV} = 0.0126$
d_{TV} bound $= 0.0497$

0   5   10   15

x

Poisson-Binomial Approximation

n = 1000 , p= 0.05
$\lambda = 50$

— X Binomial
— Y Poisson
$d_{TV} = \sup_A |\, P(X \in A) - P(Y \in A)\,|$
$d_{TV} = 0.0124$
d_{TV} bound $= 0.05$

Poisson-Binomial Approximation

n = 1000 , p= 0.005
$\lambda = 5$

— X Binomial
— Y Poisson

$d_{TV} = \sup_{A} \left| P(X \in A) - P(Y \in A) \right|$

$d_{TV} = 0.00123$

d_{TV} bound $= 0.00497$
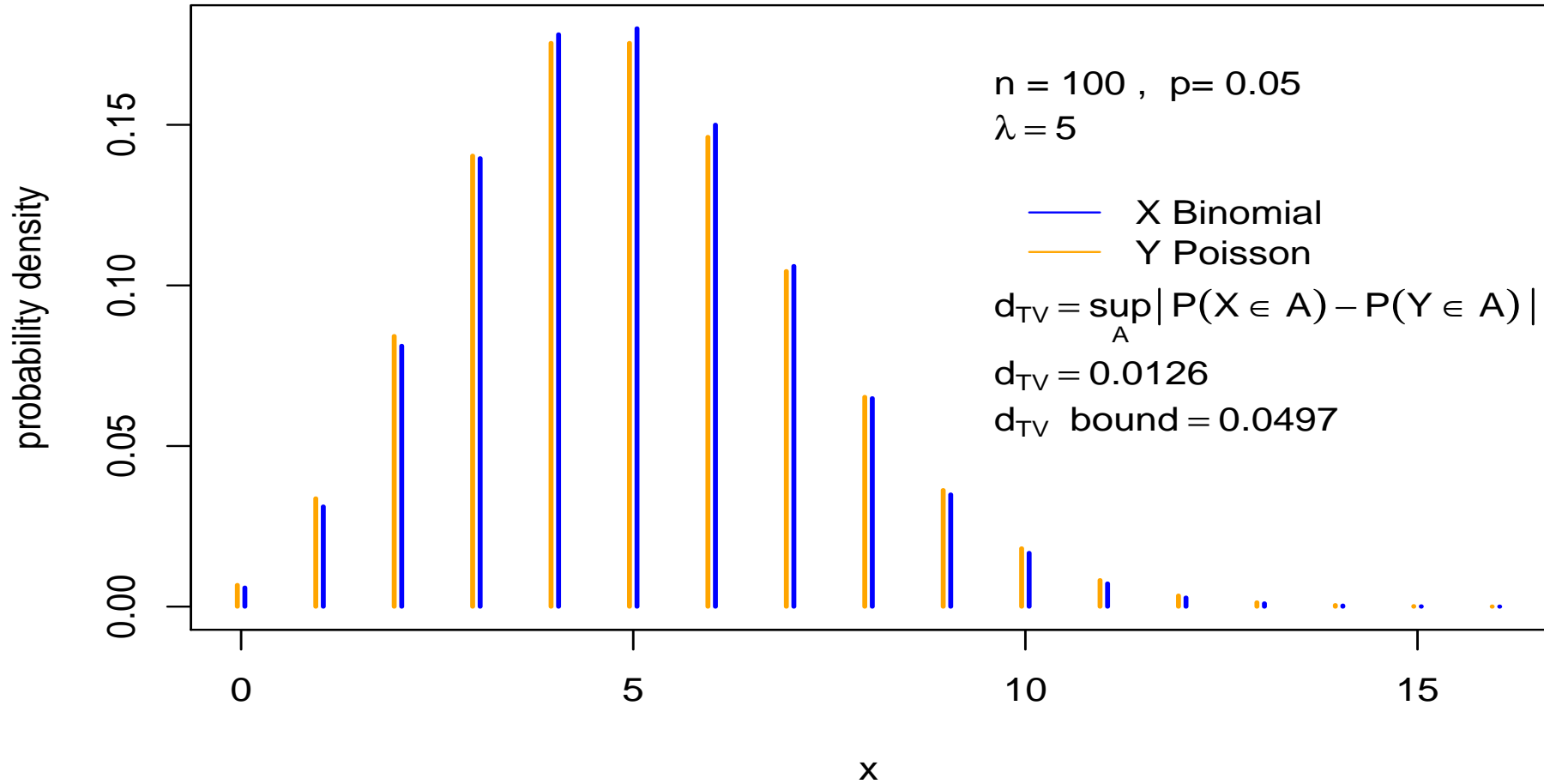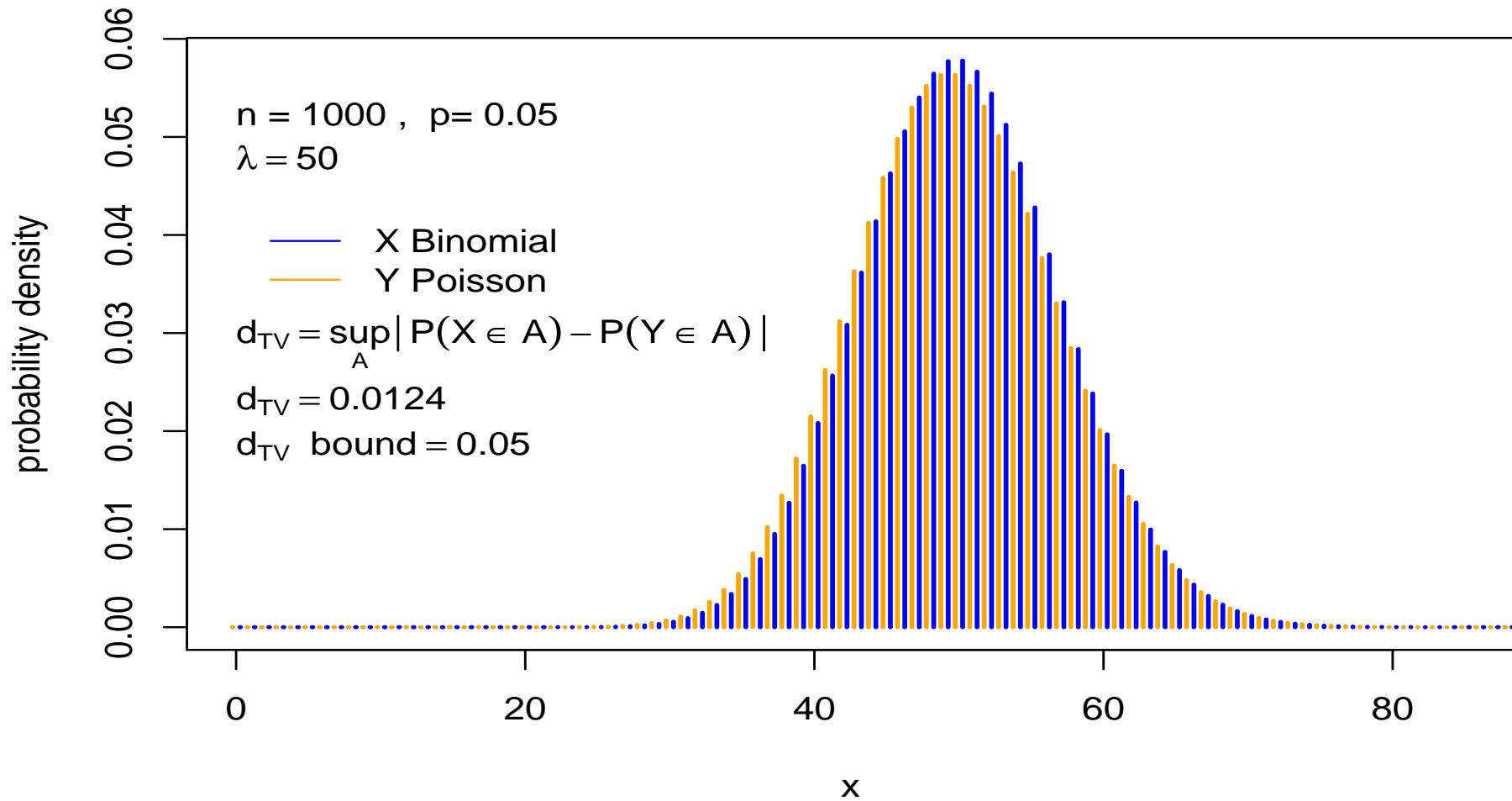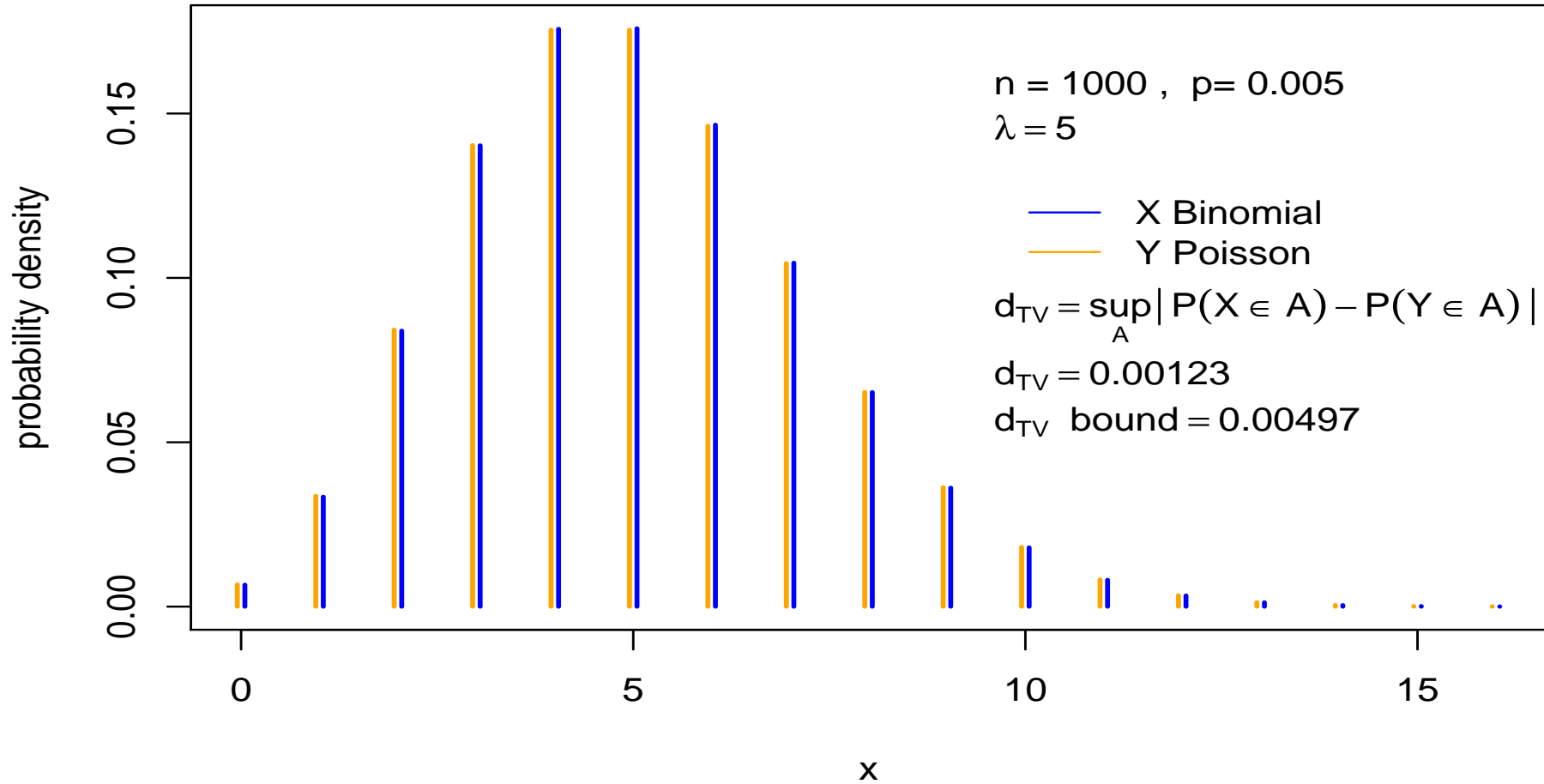
Commentary on Approximations

The previous 4 slides 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 on d_{TV}, i.e., $[(1 - \exp(-\lambda))/\lambda] \times np^2 = (1 - \exp(-np)) \times p$.

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}.

In the first plot we have $\lambda = .5$. The bound on d_{TV} is different from p and is not

quite as conservative compared to the actually achieved value of d_{TV}.

R Program for Calculating Poisson-Binomial Probabilities

A deceptively simple but recursive R program for calculating $P(W = k)$

```
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])
   }
}
```

Comments on R Program for Calculating $P(W = k)$

Note that there are about 2^n recursions, $n = $ length of $(p_1, \ldots, p_n)$

$$P(W = k) = P\left(\sum_{i=1}^{n} I_i = k\right) = p_1 \times P\left(\sum_{i=2}^{n} I_i = k - 1\right) + (1 - p_1) \times P\left(\sum_{i=2}^{n} I_i = k\right)$$

and so on ...

Thus the time and memory requirements grow beyond practical limits very fast.
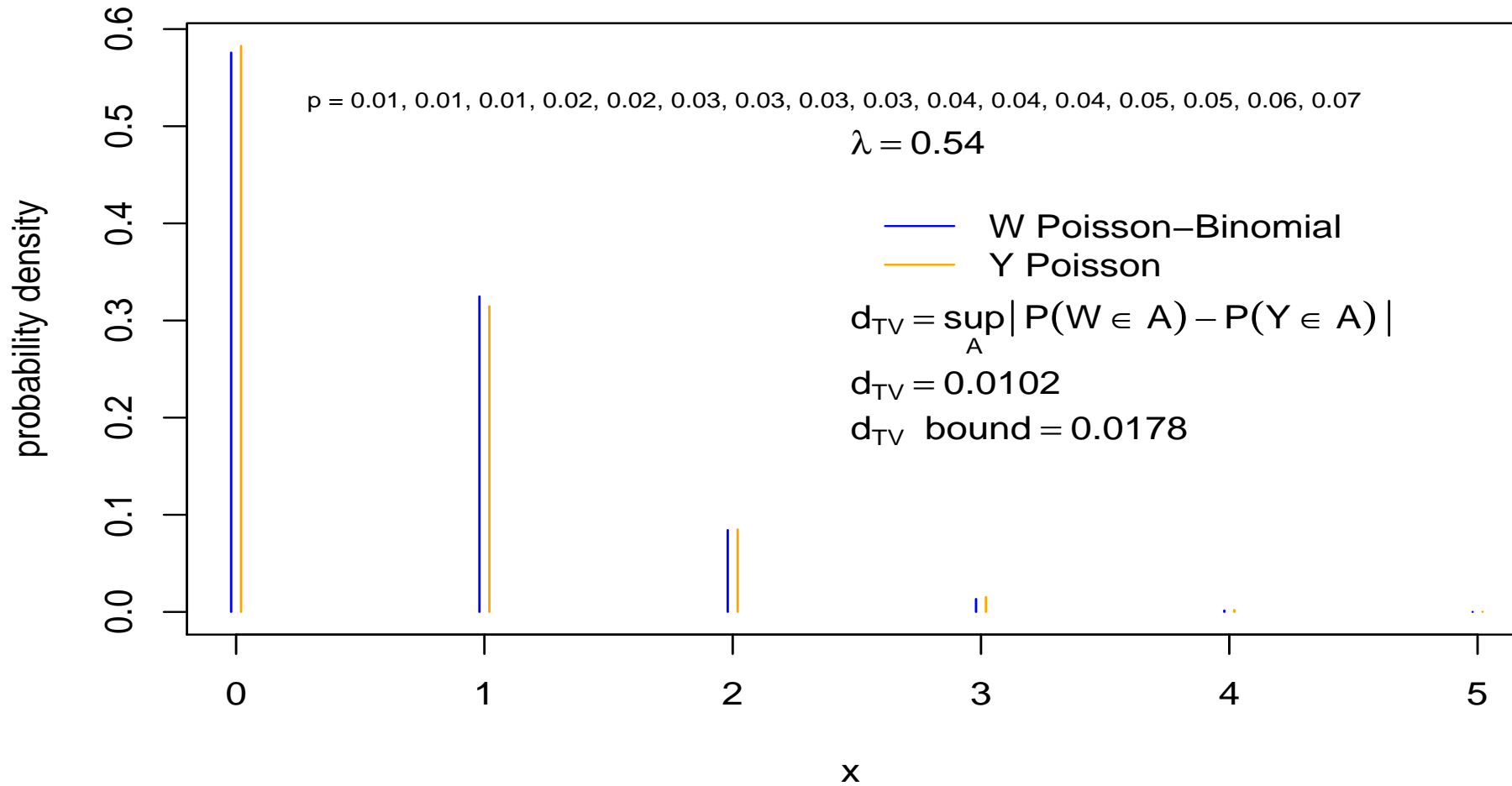
$2^{20} = 1048576!$

The next slide shows the results for $n = 16$ with the following probabilities:

$$p = .01, .01, .01, .02, .02, .03, .03, .03, .03, .04, .04, .04, .05, .05, .06, .07 \ .$$

$2^{16} = 65536$: 30 sec, $2^{17} = 131072$: 62 sec,

$2^{18} = 262144$: 131 sec, $2^{19} = 524288$: 275 sec

Poisson-Binomial Approximation

p = 0.01, 0.01, 0.01, 0.02, 0.02, 0.03, 0.03, 0.03, 0.03, 0.04, 0.04, 0.04, 0.05, 0.05, 0.06, 0.07

$\lambda = 0.54$

—— W Poisson–Binomial
—— Y Poisson

$d_{TV} = \sup_A | P(W \in A) - P(Y \in A) |$

$d_{TV} = 0.0102$

d_{TV} bound $= 0.0178$

Upper Confidence Bounds for λ

Test 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 $H(\lambda_0)$.

Reject $H(\lambda_0)$ at significance level α when the p-value $p(x, \lambda_0) = P_{\lambda_0}(X \leq x) \leq \alpha$.

The duality principle gives $C(x)$ as the set of all acceptable $H(\lambda_0)$, i.e.,

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

For the 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.

Upper Confidence Bound Calculation

$P_\lambda(X \le x)$ is continuous in λ and strictly decreasing from 1 to 0.

$\implies$ $C(x) = [0, \hat{\lambda}_U(\gamma, x))$ where $\hat{\lambda}_U(\gamma, x)$ is the unique value of λ that solves

$$P_\lambda(X \le x) = \sum_{i=0}^{x} \frac{\exp(-\lambda)\lambda^i}{i!} = \alpha = 1 - \gamma = 1 - \int_0^\lambda \frac{\exp(-t)t^{x+1-1}}{\Gamma(x+1)} \, dt \ .$$

Get this upper bound $\hat{\lambda}_U(\gamma, x)$ as the γ-quantile of the Gamma distribution with scale 1 and shape parameter $x + 1$

$$\hat{\lambda}_U(\gamma, x) = \texttt{GAMMAINV}(\gamma, \texttt{x} + 1, 1) = \texttt{qgamma}(\gamma, \texttt{x} + 1)$$

Thus we can treat $\hat{\lambda}_U(\gamma, X)$ as a $100\gamma\%$ upper confidence bound for λ.

Example and Special Case

Example: $x = 2$ with $\gamma = .95 \implies \hat{\lambda}_U(.95, 2) = 6.295794$.

For the special case $x = 0$ there is an explicit formula for the upper bound,

$$\hat{\lambda}_U(\gamma, 0) = -\log(1 - \gamma).$$

$\gamma = .95 \implies \hat{\lambda}_U(.95, 0) = 2.995732 \approx 3,$

another instance of the Rule of Three.

Lower Confidence Bounds for λ

Test $H(\lambda_0) : \lambda = \lambda_0$ against the alternatives $A(\lambda_0) : \lambda > \lambda_0$.

Large values of X would serve as evidence against the hypothesis $H(\lambda_0)$.

Observing $X = x$ 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.,

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

The random set $C(X)$ has the desired coverage probability

$$P_{\lambda_0}(\lambda_0 \in C(X)) = 1 - P_{\lambda_0}(\lambda_0 \notin C(X)) \geq 1 - \alpha = \gamma \qquad \text{for any } \lambda_0 > 0.$$

Here $\geq$ becomes $=$ for some λ_0. Thus the confidence coefficient $\bar{\gamma}$ of $C(X)$ is γ.

Calculation of Lower Confidence Bounds for λ

$P_{\lambda_0}(X \geq 0) = 1$ for all $\lambda_0 > 0 \implies C(0) = (0, \infty)$.

$P_\lambda(X \geq x)$ is continuous in λ and strictly increasing from 0 to 1 for $x > 0$.

$\implies C(x) = (\hat{\lambda}_L(\gamma, x), \infty)$ where $\hat{\lambda}_L(\gamma, x)$ is the unique value of λ which solves

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

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.

For $x > 0$ get $\hat{\lambda}_L(\gamma, x) = \texttt{GAMMAINV}(1 - \gamma, \texttt{x}, 1) = \texttt{qgamma}(1 - \gamma, \texttt{x})$

For $x = 1$ get explicitly $\hat{\lambda}_L(\gamma, 1) = -\log(\gamma)$, since

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

$$\hat{\lambda}_L(1-\alpha/2,x) < \hat{\lambda}_U(1-\alpha/2,x)$$

$\hat{\lambda}_L(1-\alpha/2,x) < \hat{\lambda}_U(1-\alpha/2,x)$ for $x = 0,1,2,\ldots$ and $0 < \alpha < 1$.

Suppose to the contrary that $\quad \hat{\lambda}_L(1-\alpha/2,x) \geq \hat{\lambda}_U(1-\alpha/2,x) \quad$ for some x.

Then there exists a λ_0 and x with $\quad \hat{\lambda}_L(1-\alpha/2,x) \geq \lambda_0 \geq \hat{\lambda}_U(1-\alpha/2,x)$.

For that λ_0 and x we reject $H(\lambda_0)$ when testing against $A(\lambda_0) : \lambda > \lambda_0$

or when testing it against $\tilde{A}(\lambda_0) : \lambda < \lambda_0$.

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

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

which leaves us with a contradiction. Hence our assumption can't be true.

Confidence Intervals for λ

$$\hat{\lambda}_L(1-\alpha/2,x) < \hat{\lambda}_U(1-\alpha/2,x) \quad \text{for all } x = 0,1,2,\ldots \text{ and } 0 < \alpha < 1$$

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

Thus $(\hat{\lambda}_L(1-\alpha/2,X),\ \hat{\lambda}_U(1-\alpha/2,X))$ is a $100\gamma\%$ level confidence interval for λ.

Its confidence coefficient usually is $\bar{\gamma} > \gamma$, for same reasons given previously.

Poisson Approximation to Binomial

For small p the binomial distribution of X is 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.
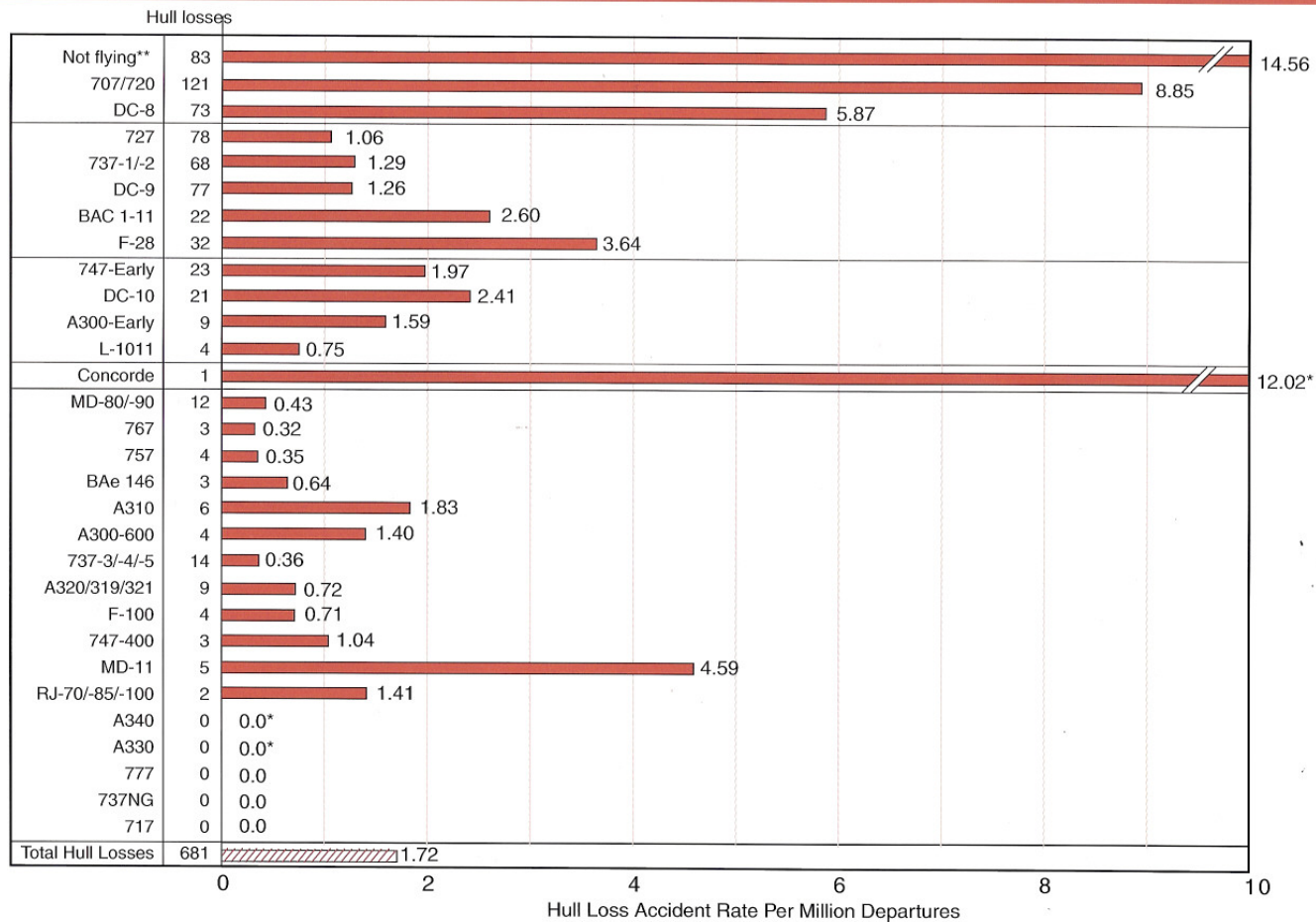
Here p would denote the probability of such an incident during a particular flight.

Typically p is very small and n, as accumulated over the whole fleet, is very large.

Aircraft Accidents

Accident Rates by Airplane Type
Hull Loss Accidents — Worldwide Commercial Jet Fleet — 1959 Through 2001

	Hull losses	Hull Loss Accident Rate Per Million Departures
Not flying**	83	14.56
707/720	121	8.85
DC-8	73	5.87
727	78	1.06
737-1/-2	68	1.29
DC-9	77	1.26
BAC 1-11	22	2.60
F-28	32	3.64
747-Early	23	1.97
DC-10	21	2.41
A300-Early	9	1.59
L-1011	4	0.75
Concorde	1	12.02*
MD-80/-90	12	0.43
767	3	0.32
757	4	0.35
BAe 146	3	0.64
A310	6	1.83
A300-600	4	1.40
737-3/-4/-5	14	0.36
A320/319/321	9	0.72
F-100	4	0.71
747-400	3	1.04
MD-11	5	4.59
RJ-70/-85/-100	2	1.41
A340	0	0.0*
A330	0	0.0*
777	0	0.0
737NG	0	0.0
717	0	0.0
Total Hull Losses	681	1.72

** The Comet, CV-880/-990, Caravelle, Mercure, Trident & VC-10 are no longer in commercial service, and are combined in the "Not Flying" bar.

 * These types have accumulated fewer than 1 million departures.

14

2001 STATISTICAL SUMMARY, JUNE 2002

BOEING

86

Statistical Summary of Commercial Jet Airplane Accidents

Accident data on commercial jet aircraft are updated yearly at

$$\mathtt{http://www.boeing.com/news/techissues/pdf/statsum.pdf}$$

This summary provides ample opportunities for applying the confidence bound methods developed so far.

These reports do not show such bounds.

The previous slide shows a page on hull loss accidents from the 2001 report, still showing the Concorde line (no longer in 2005 report).

The next slide 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.

Confidence Bounds on Aircraft Accident Rates

Hull Loss Accidents –– Worldwide Commercial Jet Fleet 1959–2001

accident rates/million departures & 95% upper confidence bounds
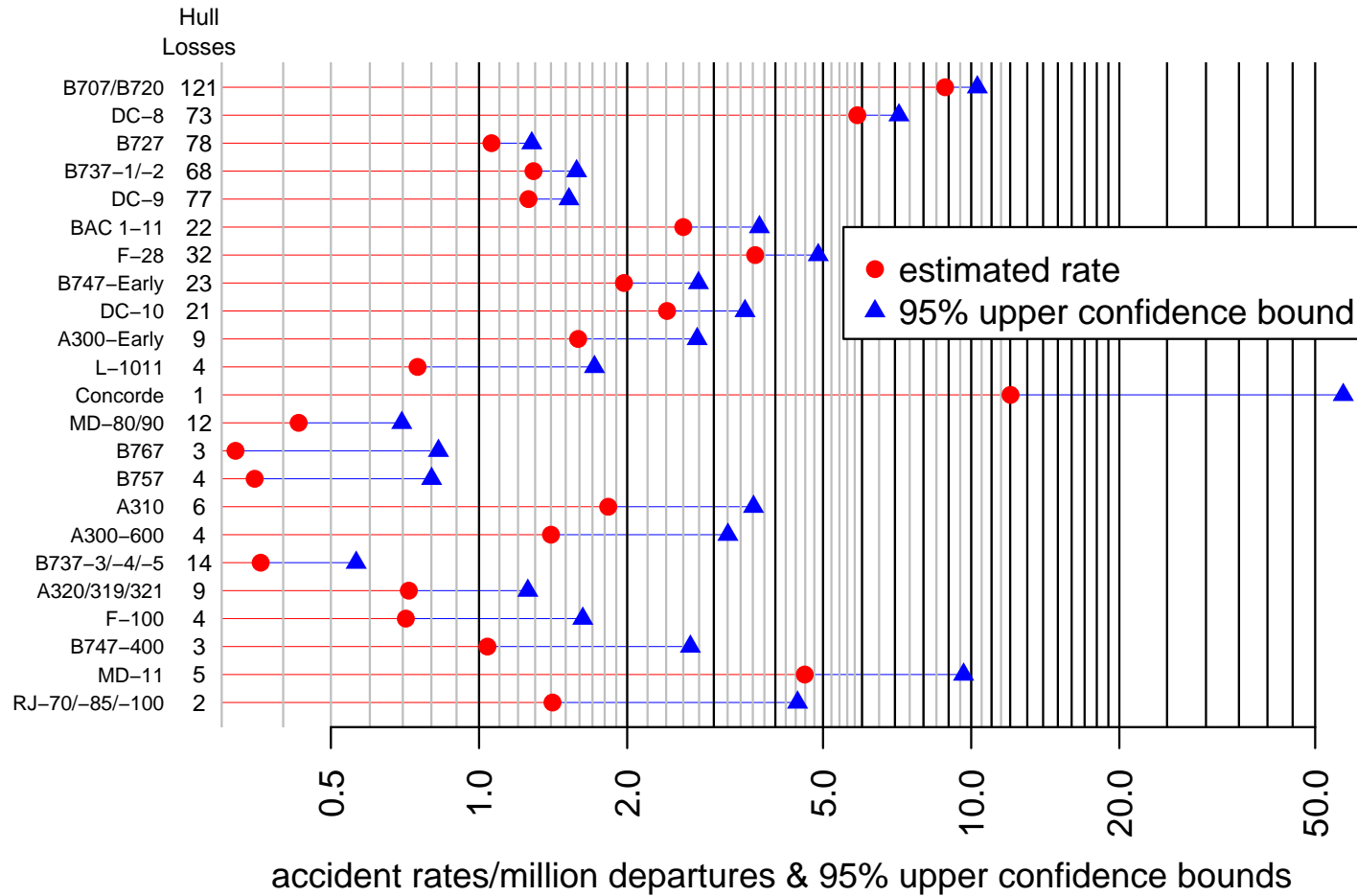
Details on Upper Confidence Bounds Calculation

95% upper confidence bounds were calculated based on the Poisson distribution and based on the binomial distribution.

$x = $ # of hull losses and rate $\rho = $ # of losses per 1 million departures

$\longrightarrow n = $ # of departures during which each aircraft model was tracked:

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

For the Concorde: $x = 1$, $\rho = 12.02$, hence $n = 1000000/12.02 \approx 83195$.

$p = $ probability of a hull loss/departure. Poisson approximation model with $\lambda = np$

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

which converts to an upper bound

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

If we use a binomial based approach directly we get

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

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.

Thus it suffices to show just one upper bound column in the following Table.

Table of 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

Hypergeometric Distribution

Randomly sample 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 X as

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

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

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

while $F_D(x) = 0$ for $x < D_m$ and $F_D(x) = 1$ for $x \geq D_M$.

Monotonicity Property

We expect that $F_D(x) = P_D(X \le x) \searrow$ or $P_D(X \ge x) \nearrow$ as $D \nearrow$.

Think of an urn containing D red, 1 pink, and $N - D - 1$ white balls.

Let $X = $ # of red balls in the grab when the pink ball is viewed as white.

Let $Y = $ # of red balls in the grab when the pink ball is viewed as red.

Show $P_D(X \ge x) < P_D(Y \ge x) = P_{D+1}(X \ge x)$ provided $D_m + 1 \le x \le D_M$.

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.

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

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

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

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 $\{\omega\}^c$.

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

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

Upper Confidence Sets

We view small values 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 $p(x, D_0) = P_{D_0}(X \leq x) \leq \alpha = 1 - \gamma$.

The corresponding confidence set consists of all acceptable $H(D_0)$ at level α, i.e.,

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

We have the following coverage probability property for the random set $C(X)$

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

The parameter D is discrete $\Rightarrow$ minimum coverage probability is usually $> \gamma$, except for some select set of γ.

Upper Confidence Bounds

For $x < n$ this p-value $p(x, D_0) = P_{D_0}(X \leq x)$ is decreasing from 1 to 0 with increasing D_0, and it is strictly decreasing while $0 < p(x, D_0) < 1$.

$\implies$ there exists a unique largest D_0 with $p(x, D_0) = P_{D_0}(X \leq x) > \alpha$.

We denote this value by $\hat{D}_U(\gamma, x)$.

For $x = n$ we have $p(n, D_0) = P_{D_0}(X \leq n) = 1 > \alpha$ for all D_0.

Thus $C(n) = [0, N]$ and we define $\hat{D}_U(\gamma, n) = N$ in that case.

$\implies$ $C(x) = [0, \hat{D}_U(\gamma, x)]$. Note the closed form of the interval here.

Replace the lower bound 0 by x since we saw that many in the sample.

$\implies$ $C(x) = [x, \hat{D}_U(\gamma, x)]$

$\hat{D}_U$ via the Function `hypergeo.conf`

Example: $N = 2500$, $n = 50$, $x = 11$ and $\gamma = .95$

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


$confidence
  nominal    minimum
0.9500000 0.9500011
```

`hypergeo.conf` $\rightarrow$ confidence coefficient $\bar{\gamma}$ (minimum coverage probability) when the logic flag `cc.flag=T` is set. Here $\bar{\gamma} = 0.9500011 \approx .95$.

$\hat{D}_U \rightarrow 95\%$ upper bound of $\hat{D}_U/N = 0.3364$ for the population proportion $p = D/N$. Based on binomial distribution get $\hat{p}_U = 0.3378$, not much different.

Lower Confidence Sets

Test $H(D_0) : D = D_0$ against the alternative $A(D_0) : D > D_0$

High values of X can be viewed as evidence against the hypothesis $H(D_0)$.

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

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

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

We usually have $> \gamma$ for all D_0, because of the discrete nature of the parameter D_0.

Lower Confidence Bounds

$x > 0$: the p-value $P_{D_0}(X \geq x)$ increases from 0 to 1 as D_0 increases from 0 to N,

it increases strictly as long as $0 < p(x, D_0) = P_{D_0}(X \geq x) < 1$

$\implies$ there is a smallest value of D_0 for which $P_{D_0}(X \geq x) > \alpha$. Denote it by $\hat{D}_L(\gamma, x)$.

$x = 0$: $P_{D_0}(X \geq 0) = 1$ for all D_0, $\implies C(0) = [0, N]$ and define $\hat{D}_L(\gamma, 0) = 0$.

$\implies C(x) = [\hat{D}_L(\gamma, x), N]$, 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.

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.

$\implies C(x) = [\hat{D}_L(\gamma, x), N - (n - x)]$.

$\hat{D}_L$ via the Function `hypergeo.conf`

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., we get $\hat{D}_L = \hat{D}_L(\gamma, 11) = 324$ as 95% lower confidence bound for D.

Again the confidence coefficient was requested and is given as 0.9500011.

$\rightarrow \hat{D}_L(\gamma, 11)/N = 324/2500 = .1296$ as 95% lower bound for $p = D/N$.

The corresponding binomial based lower confidence bound is $\hat{p}_L = 0.1286$.

$$\hat{D}_L(1-\alpha/2,x) \leq \hat{D}_U(1-\alpha/2,x)$$

$D_0 = \hat{D}_U(1-\alpha/2,x)$ is by definition the largest D such that $P_D(X \leq x) > \alpha/2$

$D_1 = \hat{D}_L(1-\alpha/2,x)$ is the smallest D such that $P_D(X \geq x) > \alpha/2$ or

such that $P_D(X \leq x-1) < 1-\alpha/2$.

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

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, see next slide.

$P_{D_0}(X \leq x-1) < 1-\alpha/2 \implies D_1 \leq D_0$, using the alternate definition of D_1.

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

The above inequality holds quite generally for all x and D.

Use again the urn 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.
B consists of all grabs of n balls that don't contain the pink ball.

X counts the number of red balls in the grab, interpreting pink as white.
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 $\omega \in A$.

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

Monotonicity of Confidence Bounds

A consequence of the previous inequality is that $\hat{D}_U(\gamma, x)$ and $\hat{D}_L(\gamma, x)$ are strictly

increasing in x in the allowable range.

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, i.e., $P_{D+1}(X \leq x+1) \geq P_D(X \leq x)$, we see that

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

Confidence Intervals

Since $\hat{D}_L(1-\alpha/2,x) \leq \hat{D}_U(1-\alpha/2,x)$ for any x and $0 < \alpha < 1$

we may use $[\hat{D}_L(1-\alpha/2,X), \hat{D}_U(1-\alpha/2,X)]$ as a confidence interval

with coverage probability $\geq \gamma$, since

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

The confidence coefficient $\bar{\gamma}$ is typically $> \gamma$, due to the discrete nature of D
and because the one-sided bounds may not have minimum coverage probability
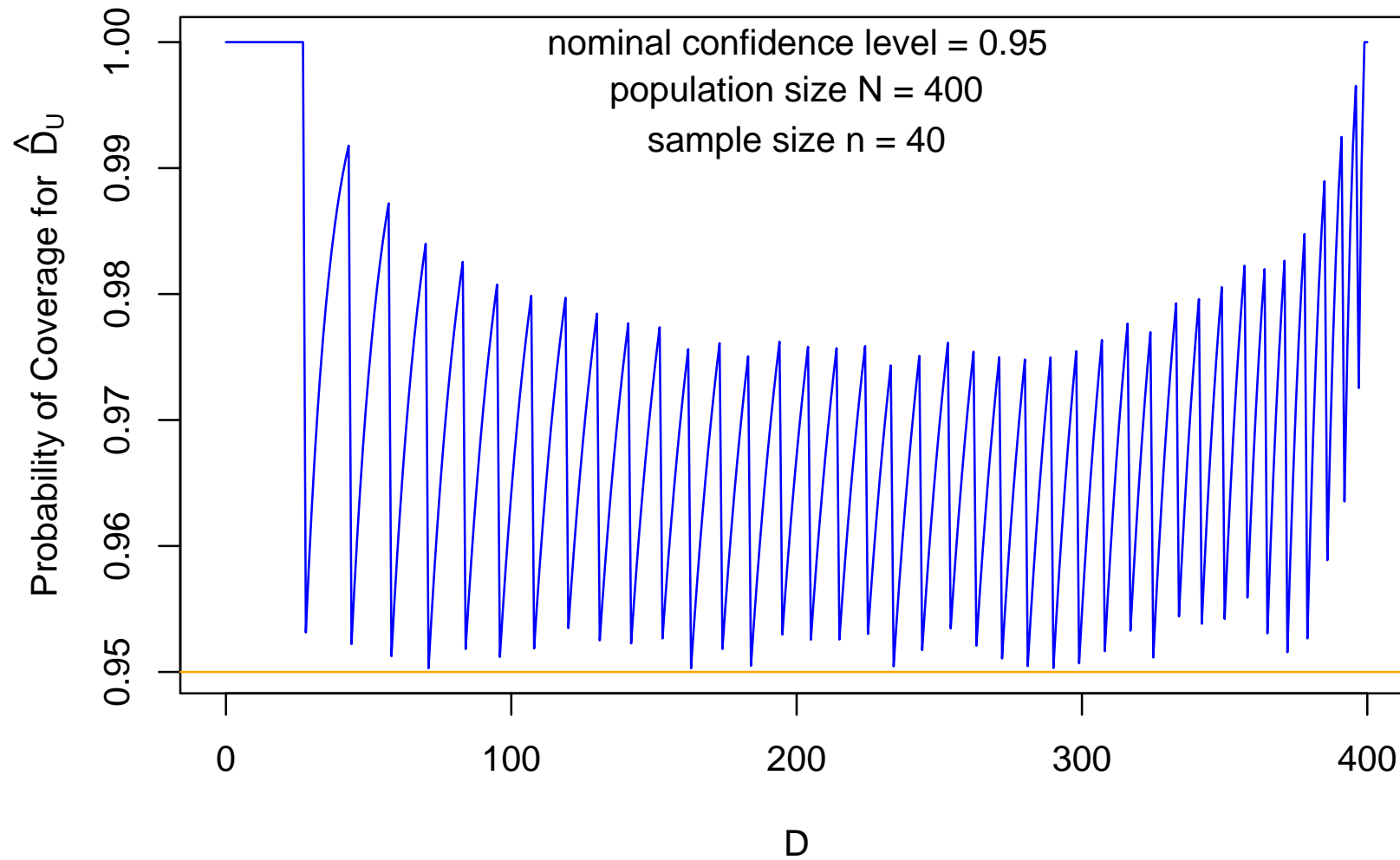at the same D.

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 $\gamma = .95$.

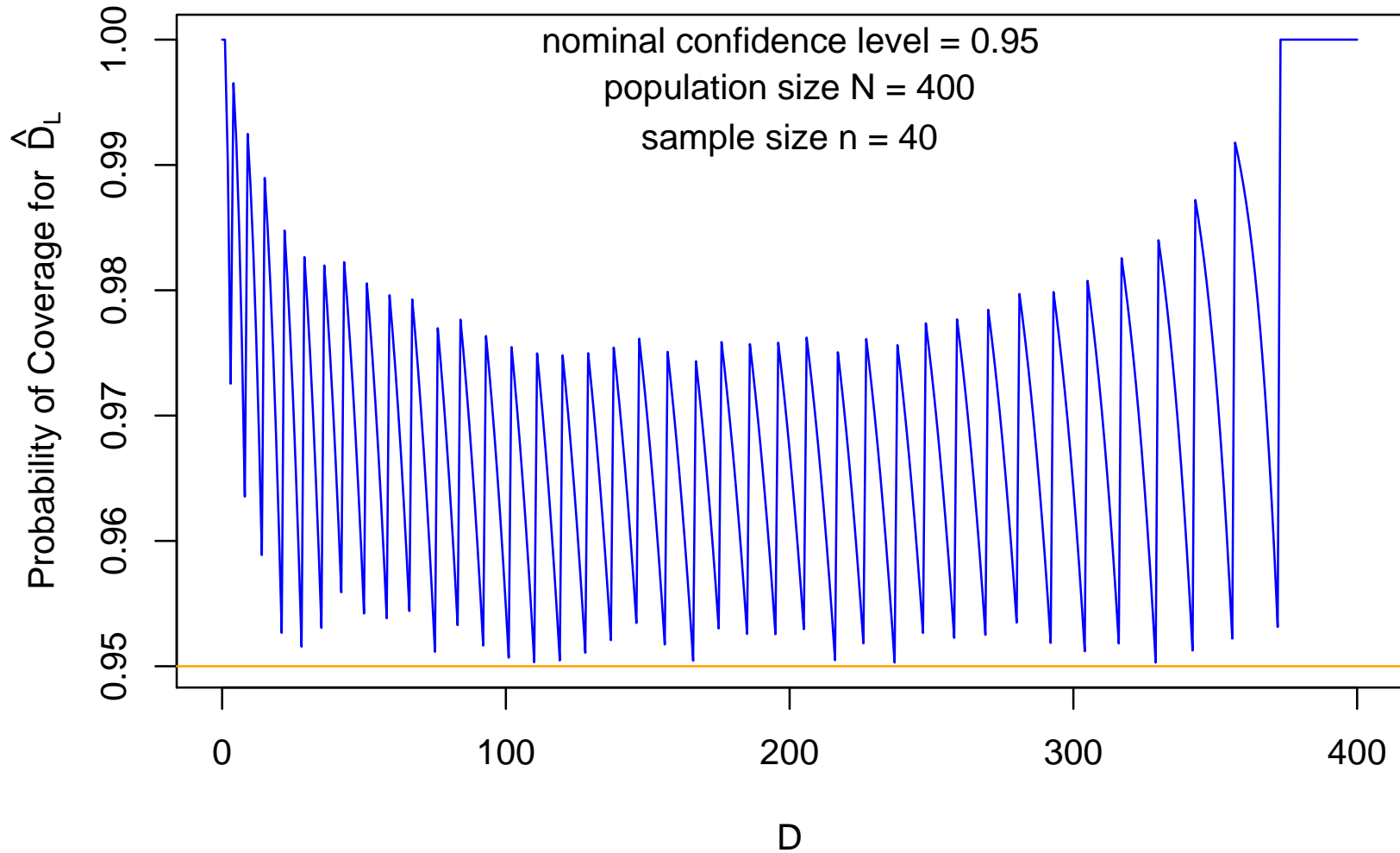The results are shown on the following slides.

The functions used to produce these plots are `D.minimum.coverage.upper`, `D.minimum.coverage.lower`, and `D.minimum.coverage.int`.

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 $\bar{\gamma} \geq \gamma$.
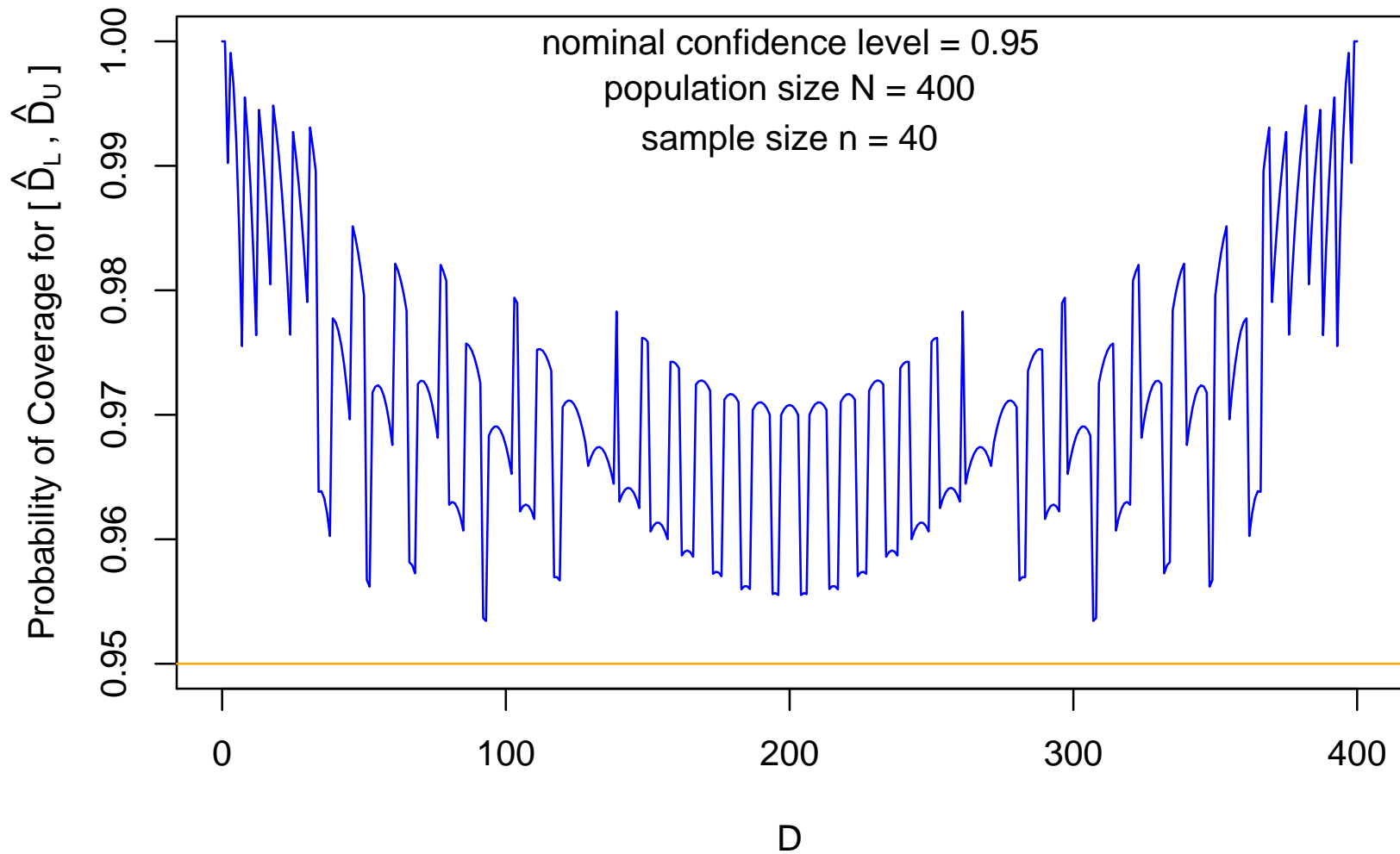
Coverage Probabilities for $\hat{D}_U(.95, x)$

nominal confidence level = 0.95
population size N = 400
sample size n = 40

Probability of Coverage for $\hat{D}_U$

D

Coverage Probabilities for $\hat{D}_L(.95, x)$

nominal confidence level = 0.95
population size N = 400
sample size n = 40

Coverage Probabilities for $[\hat{D}_L(.975,x), \hat{D}_U(.975,x)]$

$\bar{\gamma} = .9503$ for one-sided bounds, $\bar{\gamma} = 0.9534$ for the intervals.

In the case of one-sided bounds it can be shown that there is a distinct and quite rich 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 $\gamma(D^\star, x^\star) = \gamma^\star = 1 - \alpha^\star$.

We then have $\hat{D}_U(\gamma^\star, x^\star) = D^\star$ and $\bar{\gamma}^\star = \gamma^\star$.

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

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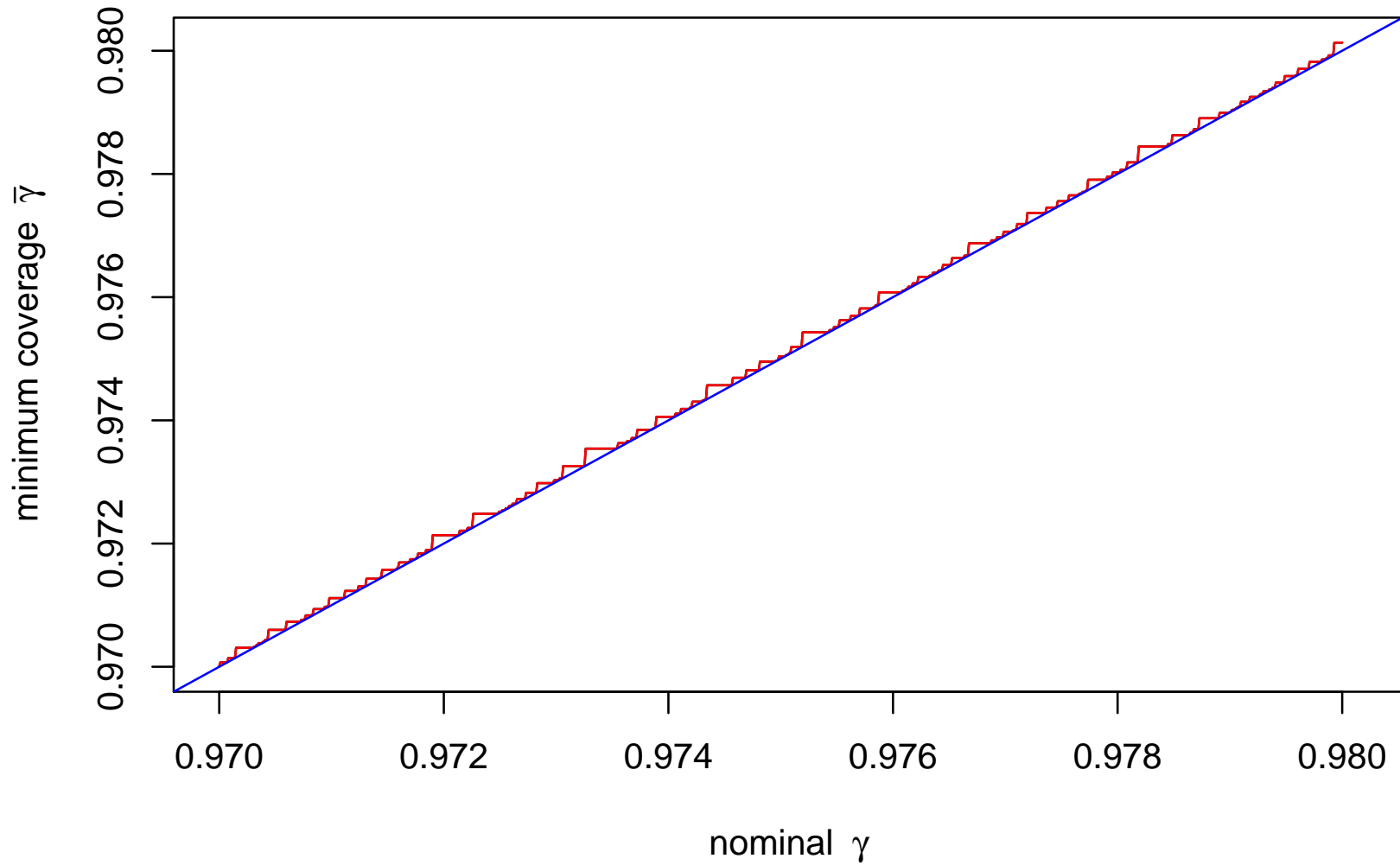
Since $\hat{D}_U(\gamma^\star, X)$ is strictly increasing in X and since $\hat{D}_U(\gamma^\star, x^\star) = D^\star$ we have

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

and hence

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

Confidence Coefficient $\bar{\gamma}$ vs. Nominal γ for Upper/Lower Bounds

Confidence Coefficient γ̄ vs. Nominal γ for Intervals
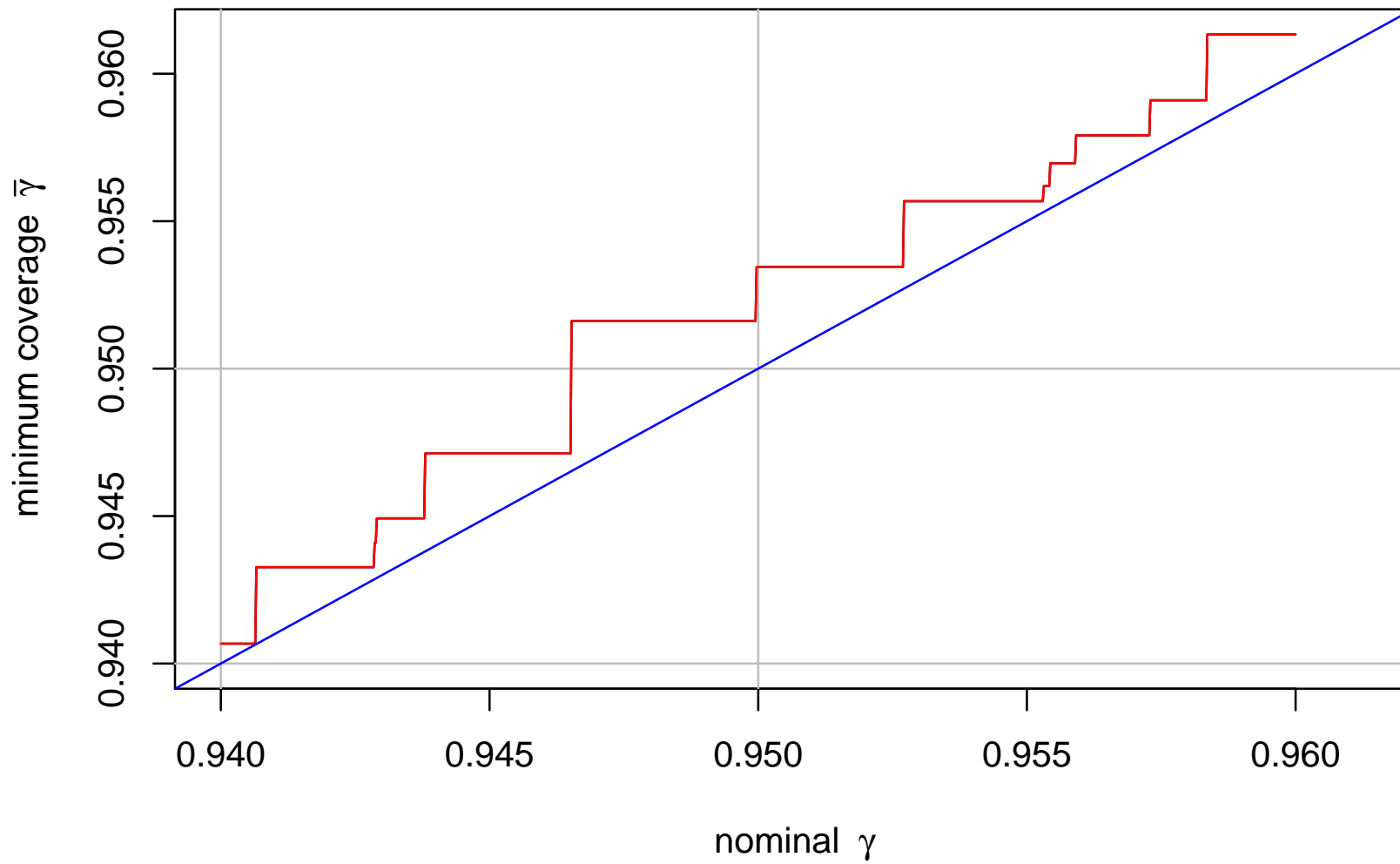
Some Comments

In the case of upper/lower bounds the discrepancies between γ and $\bar{\gamma}$ appear to be small, with $\gamma = \bar{\gamma}$ for many γ over the short displayed range $[.97, .98]$.

Thus the discreteness of D does not appear to be a serious issue, based on the examined case $n = 40$, $N = 400$.

For intervals the minimum coverage $\bar{\gamma}$ shows much stronger deviations from γ.

This is due to the much wider zigzag swings within the coverage probabilities of one-sided bounds and the fact that minima may not be attained at the same D for either one-sided bound.

Keep in mind that we need to distinguish between nominal γ, actual coverage probability that changes with D and minimum coverage probability $\bar{\gamma}$.

$\bar{\gamma} = \gamma$ or $\bar{\gamma} \approx \gamma$ for only a few values of γ in the displayed range $[.94, .96]$.

The Classical Birthday Problem

The classical birthday problem is often presented in elementary probability courses.

The chance that n random persons have birthdays on n different days in the year is

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

Thus the probability of at least one match is

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

This is assumes a non leap year and equal likelihoods of all 365 possible birthdays for all n persons. Unequal birthday probabilities increase $P(M)$.

A student once asked the following question:

What is the chance of seeing at least one pair of matching or adjacent birthdays?

Probability of Matching or Adjacent Birthdays

Let A be the event of getting n birthdays at least one day apart. Then we have

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

365 ways to pick a birthday for person 1.   There are $365-n$ non-birthdays (NB).

Use the remaining $n-1$ birthdays (B) to each fill one of the remaining $365-n-1$ gaps between the non-birthdays, $\binom{365-n-1}{n-1}$ ways. That fixes the NB–B pattern.

$(n-1)!$ ways to assigns these birthdays to the remaining $(n-1)$ persons.

The next slide provides a simplifying visualization diagram for the above reasoning, assuming a year of 20 days and $n=7$ birthdays.

Visualization Diagram

year of N = 20 days and n = 7 birthdays
separated by at least one non–birthday (NB)

NB

NB

NB
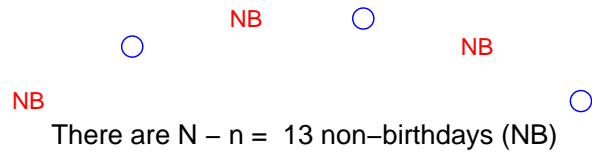
There are N − n = 13 non–birthdays (NB)

NB

NB

There are also N − n = 13 potential birthday positions
(blue circles) separated by NB's.

NB

n = 7 of the potential birthday positions need to be chosen
to give us a valid pattern where each birthday is separated
from the next by at least one NB.

NB

This will give us a pattern of (N − n) + n = 13 + 7
non–birthdays and birthdays for all N = 20 days of the year.

However, the same circular pattern can arise by rotation
in a multiplicity of ways.

NB

The birthday date for person 1 fixes the pattern
with respect to calendar time.

NB

NB

NB

NB

NB

NB

NB
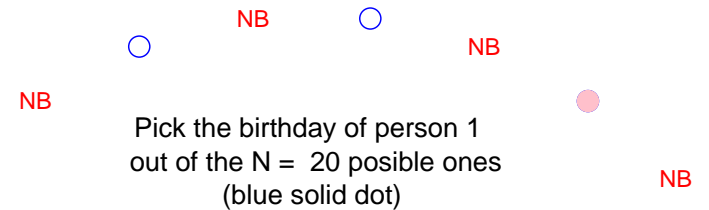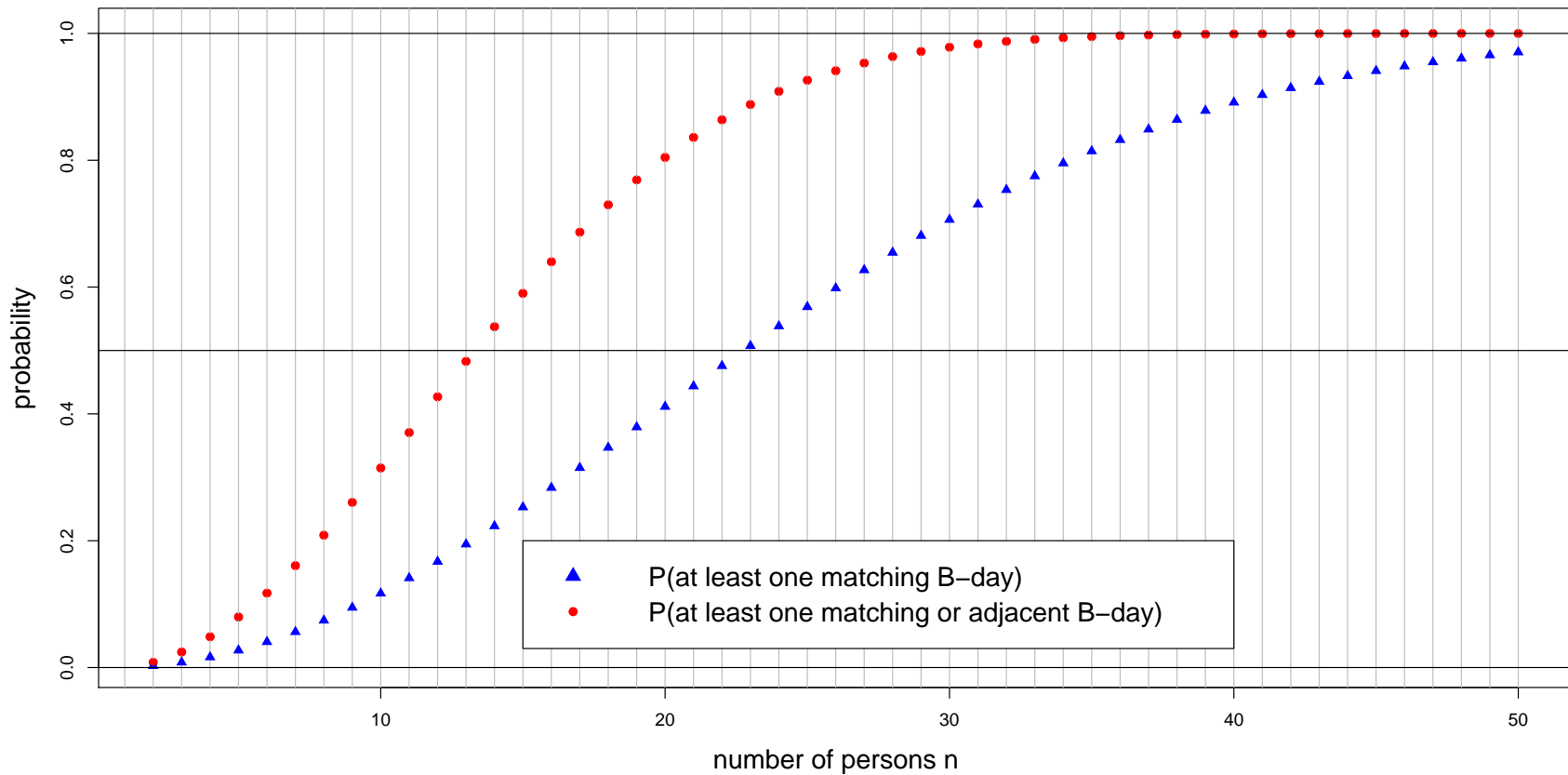
Pick the birthday of person 1
out of the N = 20 posible ones
(blue solid dot)

NB

NB

NB

Out of the remaining N − n − 1 = 12 potential birthday positions
pick the remaining n − 1 = 6 birthdays, without assigning
them to specific persons. (pink solid dots)

NB

NB

NB

NB

NB

NB

NB
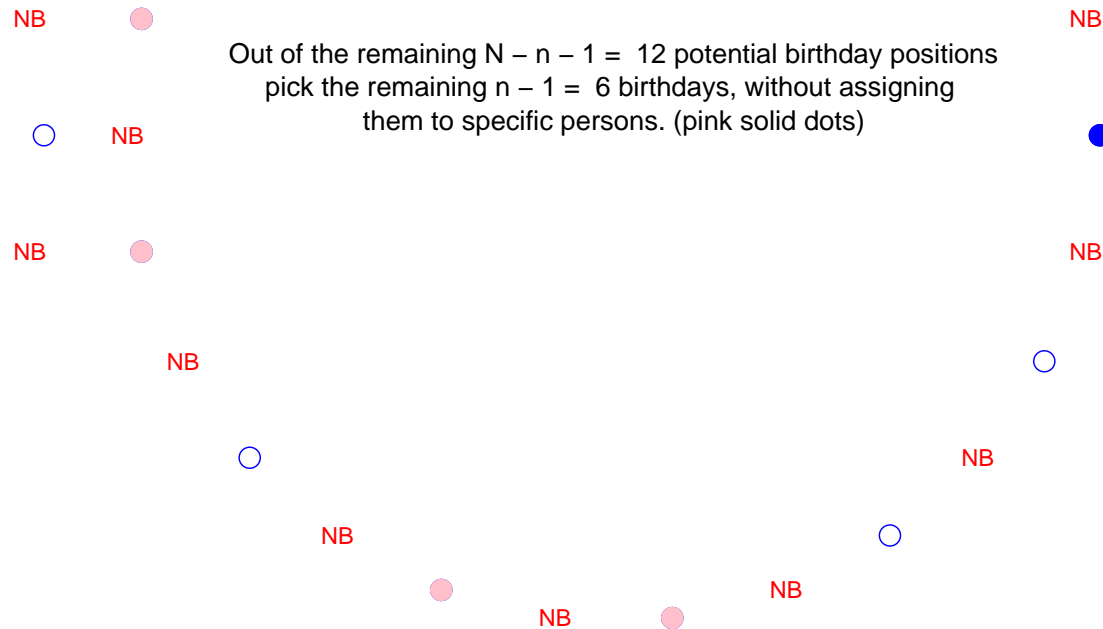
$P(M)$ and $P(A^c)$

$n = 14$ gives the smallest n for which $P(A^c) \geq .5$, in fact $P(A^c) = .5375$.

Hypergeometric Confidence Bounds with a Birthday Problem Twist

The following problem arose in connection with the Inertial Upper Stage of the Military Space Program.

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

$n = 12$ out of $N = 200$ parts (rocket motors) are to be tested in some form.

It is expected that all tested items will be non-defective. Find $\hat{D}_U(\gamma, 0)$.

A ship-set of $k = 16$ parts is randomly chosen from the population of N

to serve as stage separation motors.

Find lower bound $\hat{p}_1$ for the probability that this ship-set contains ≤ 1 defective.

After addressing this problem it was pointed out that the 16 parts will be arranged

in a circular pattern. All that matters is that no two defective parts be placed

next to each other. Find lower confidence bound $\hat{p}_2$ for the probability that no two

defective parts are placed next to each other.

Assumptions and Notation

This does not quite match the birthday problem since we cannot have two defective parts in the same position.

However, we can base probability calculations on a similar counting strategy.

We assume that the $n = 12$ parts are selected randomly from the population.

The same is assumed for the ship-set of $k = 16$, where we assume that the $n = 12$ tested ones were previously returned to the population.

The $k = 16$ parts are arranged in a random circular order.

$D = \#$ of defective parts in the population of N parts.

$X = \#$ of defective items in the tested sample of $n = 12$.   We expect $X = 0$.

Approach

Based on $X = 0$ what kind of upper confidence bounds $\hat{D}_U(\gamma, 0)$ can we get?

Recall that $\hat{D}_U(\gamma, 0) = D \iff P_D(X = 0) > \alpha = 1 - \gamma$ and $P_{D+1}(X = 0) \leq \alpha$.

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, we can associate the highest possible confidence level with each $\hat{D}(\gamma, 0) = D$.

Note that any $\gamma > \gamma_D$ would result in $P_{D+1}(X \leq 0) > 1 - \gamma$ and thus in $\hat{D}_U(\gamma, 0) \geq D + 1$.

Also note that this choice of γ_D ensures that $\bar{\gamma} = \gamma_D$, as was shown previously.

Upper bound on $D \implies$ lower bounds $\hat{p}_1$ and $\hat{p}_2$.

Confidence Level Options

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
20	0.746	0.904	0.973	0.993	0.998	1.000	0.508	0.867
25	0.821	0.947	0.989	0.998	1.000	1.000	0.377	0.800
30	0.876	0.971	0.996	0.999	1.000	1.000	0.272	0.727
35	0.914	0.985	0.998	1.000	1.000	1.000	0.190	0.649
40	0.942	0.992	0.999	1.000	1.000	1.000	0.130	0.571
42	0.950	0.994	1.000	1.000	1.000	1.000	0.111	0.540

Interpretation

We can get $\hat{D}(\gamma, 0) = 10$ with varying confidence levels, depending on the strength of testing evidence, i.e., for how big a sample of size n we observed 0 failures.

By choosing n higher we can attach higher confidence to the stated bounds $\hat{p}_1$ and $\hat{p}_2$, provided we still stay with 0 failures among the larger tested sample of size n.

We cannot force 0 failures and the confidence level γ only has meaning in the context of a random upper bound $\hat{D}_U(\gamma, X)$.

Using this approach we pretend to be clairvoyant in saying $X = 0$ will be observed.

This method is useful as a planning tool when weighing confidence against the cost of testing, in view of expected results $(X = 0)$.

Lower Bound $\hat{p}_1$

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

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

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 $\hat{D}_U(\gamma, 0)$.

Lower Bound $\hat{p}_2$

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

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

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

with $\psi(y) \searrow$ in y and

$$
Q^\star(D) = E_D\left(\psi(Y)\right) \searrow \text{ in } D \implies \hat{p}_2 = Q^\star(\hat{D}_U(\gamma, X))
$$

$\psi(y) = P(A_y)$

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. How many of these patterns have no two F's next to each other?

G_1 can be in any of the k places. We can arrange the remaining $m - 1$ G's in $(m - 1)!$ ways to form some circular sequence order of G's.

Next we place the y F's between the G's so that only one F goes into any one gap.

There are $\binom{m}{y}$ ways of designating the gaps to receive just one F each.

There are $y!$ ways of arranging the order of the y F's.

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

Note that $P(A_0) = P(A_1) = 1$ and $P(A_y) = 0$ for $y > k - y$.

$$\psi(y) = P(A_y) \searrow$$

$$\frac{P(A_{y+1})}{P(A_y)} = \frac{\binom{k-y-1}{y+1} \big/ \binom{k-1}{y+1}}{\binom{k-y}{y} \big/ \binom{k-1}{y}}$$

$$= \frac{(k-y-1)!}{(y+1)!(k-2(y+1))!} \frac{(y+1)!(k-y-2)!}{(k-1)!} \frac{y!(k-2y)!}{(k-y)!} \frac{(k-1)!}{y!(k-y-1)!}$$

$$= \frac{(k-2y)(k-2y-1)}{(k-y)(k-y-1)} < 1$$

$$\Longleftrightarrow (k-2y)(k-2y-1) < (k-y)(k-y-1)$$

$$\Longleftrightarrow -2k+3y+1 < 0$$

which is true for $y = 1, \ldots, k/2$.

$$Q^\star(D+1) = E_{D+1}(\psi(Y)) \leq E_D(\psi(Y)) = Q^\star(D)$$

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

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

$A < k(D+1)/(N+1) < B.$ $\qquad \psi(y) \searrow \implies a = \inf_A \psi(y) \geq \sup_B \psi(y) = b \implies$

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

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

Interpretation of Tabulated Confidence Levels

How do we interpret the results in the previous Table in relation to $\hat{p}_1$ and $\hat{p}_2$?

For the originally proposed sample size $n = 12$ for testing, we can be 82.1% confident that $D \leq 25$ and that the probability of having less than two defective items in the ship-set of $k = 16$ is at least .377, 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 82.1% confident that the probability of seeing no such adjacency is at least .80, which may still not be satisfactory.

128

p1p2.hat for Computing $\hat{p}_1$ & $\hat{p}_2$

```
>  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
```

Adjusting γ

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 $\hat{D}_U(0,\gamma) = \texttt{D.U} = 12$.

What if $X = x > 0$?

In that case the upper bound $\hat{D}_U(\gamma, x)$ will increase.

$\hat{p}_1 = Q(\hat{D}_U(\gamma, x))$ and $\hat{p}_2 = Q^\star(\hat{D}_U(\gamma, x))$ will decrease.

However, we should not return defective items into the population.

$\implies N \longrightarrow N' = N - x$ and $D \longrightarrow D' = D - x$

Should use $\hat{D}'_U(\gamma, x) = \hat{D}_U(\gamma, x) - x$ when plugging into $Q(D)$ and $Q^\star(D)$

i.e., use $\hat{p}_1 = Q(\hat{D}'_U(\gamma, x))$ and $\hat{p}_2 = Q^\star(\hat{D}'_U(\gamma, x))$

Example with $X = 1$

```
>  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
```

Final Warning

$X = 1$ results in lower confidence bounds for $Q(D)$ and $Q^\star(D)$.

Do not 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 no empty warning. I have experienced this kind of ingenuity from engineers.

Webster's: "engineer" and "ingenuity" have the same Latin origin: ingenium.

Comparing Two Poisson Means

Assume $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{Poisson}(\mu)$ are independent.

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 $\kappa = \pi_1/\pi_2$, 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 the total observed count $X + Y$, i.e., consider the conditional distribution of Y given $T = X + Y = t$

Conditional Distribution of Y Given $X+Y=t$

$$
\begin{aligned}
P(Y=k|X+Y=t) &= \frac{P(Y=k,X+Y=t)}{P(X+Y=t)} = \frac{P(Y=k,X=t-k)}{P(X+Y=t)} \\[2mm]
&= \frac{P(Y=k)P(X=t-k)}{P(X+Y=t)} \\[2mm]
&= \frac{[\exp(-\mu)\mu^k]/k! \times [\exp(-\lambda)\lambda^{t-k}]/(t-k)!}{\exp(-[\mu+\lambda])[\mu+\lambda]^t/t!} \\[2mm]
&= \binom{t}{k} p^k (1-p)^{t-k},
\end{aligned}
$$

where $p = \mu/(\lambda+\mu) = 1/(1+\rho)$.

Here we used the fact that $X+Y$ has a Poisson distribution with mean $\lambda+\mu$.

$X + Y \sim \text{Poisson}(\lambda + \mu)$

$$P(X+Y=t) \;=\; \sum_{i=0}^{t} P(X=i \;\;\cap\;\; Y=t-i)$$

$$=\; \sum_{i=0}^{t} P(X=i) \times P(Y=t-i)$$

$$=\; \sum_{i=0}^{t} \frac{\exp(-\lambda)\lambda^{i}}{i!} \times \frac{\exp(-\mu)\mu^{t-i}}{(t-i)!}$$

$$=\; \frac{\exp(-[\lambda+\mu])(\lambda+\mu)^{t}}{t!} \; \sum_{i=0}^{t} \binom{t}{i} \left(\frac{\lambda}{\lambda+\mu}\right)^{i} \left(\frac{\mu}{\lambda+\mu}\right)^{t-i}$$

$$=\; \frac{\exp(-[\lambda+\mu])(\lambda+\mu)^{t}}{t!}$$

Confidence Bounds for ρ

The conditional distribution of Y given $X + Y = t$ is binomial, $n = t$ and $p = \frac{1}{1+\rho}$

We write $\quad \mathcal{D}(Y \mid X + Y = t) \sim \text{Binomial}(t, p = 1/(1+\rho))$

$\rho = 1/p - 1$ is strictly decreasing in p.

Confidence bounds and intervals translate easily to bounds and intervals for ρ.

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

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

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

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

Confidence Bounds for π_1/π_2 (small π_i)

Assume $X \sim \text{Binomial}(n_1, \pi_1)$, $Y \sim \text{Binomial}(n_2, \pi_2)$ are independent with $\pi_i \approx 0$.

Using Poisson approximations for both binomial distributions

$$X \sim \text{Poisson}(\lambda = n_1\pi_1) \quad \text{and} \quad Y \sim \text{Poisson}(\mu = n_2\pi_2)$$

and noting $\quad \rho = (n_1\pi_1)/(n_2\pi_2) = \kappa \times n_1/n_2 \quad \text{or} \quad \kappa = \pi_1/\pi_2 = (n_2/n_1) \times \rho$

confidence bounds and intervals for ρ convert to corresponding ones for $\kappa = \pi_1/\pi_2$

by multiplying the bounds or interval end points for ρ by n_2/n_1.

A Fictitious Example

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.

We have $Y = 5$ and $T = X + Y = t = 0 + 5$, $n_1 = 11.128 \times 10^6$ and $n_2 = 55.6 \times 10^6$.

$$\implies \hat{p}_U(.95, 5, 5) = 1 \quad \text{and thus} \quad \hat{\kappa}_L(.95, 5, 5) = \frac{n_2}{n_1} \times \left(\frac{1}{1} - 1 \right) = 0$$

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

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

Can view 4.099871 as a 95% upper confidence bound for π_1 / π_2. 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 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.

Comparing Hull Loss Rates across Airplane Models

Was the Concorde crash significant compared to experiences of other airplanes?

We made two comparisons, comparing it with the MD11 and the B767.

The MD11 had $X = 5$ hull losses and a rate of 4.59 per 10^6 departures

the Concorde had $Y = 1$ such loss and a rate of 12.02 per 10^6 departures.

$\implies n_1 = 1089325$ and $n_2 = 83195$ departures for MD11 and Concorde, resp.

$\implies (.04273, 18.06)$ 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.

Comparing Hull Loss Rates across Airplane Models

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 still > 1,

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 on the next slide the coverage behavior of this lower bound and indicate by vertical lines positions for various values of π_1/π_2.

Coverage Probabilities
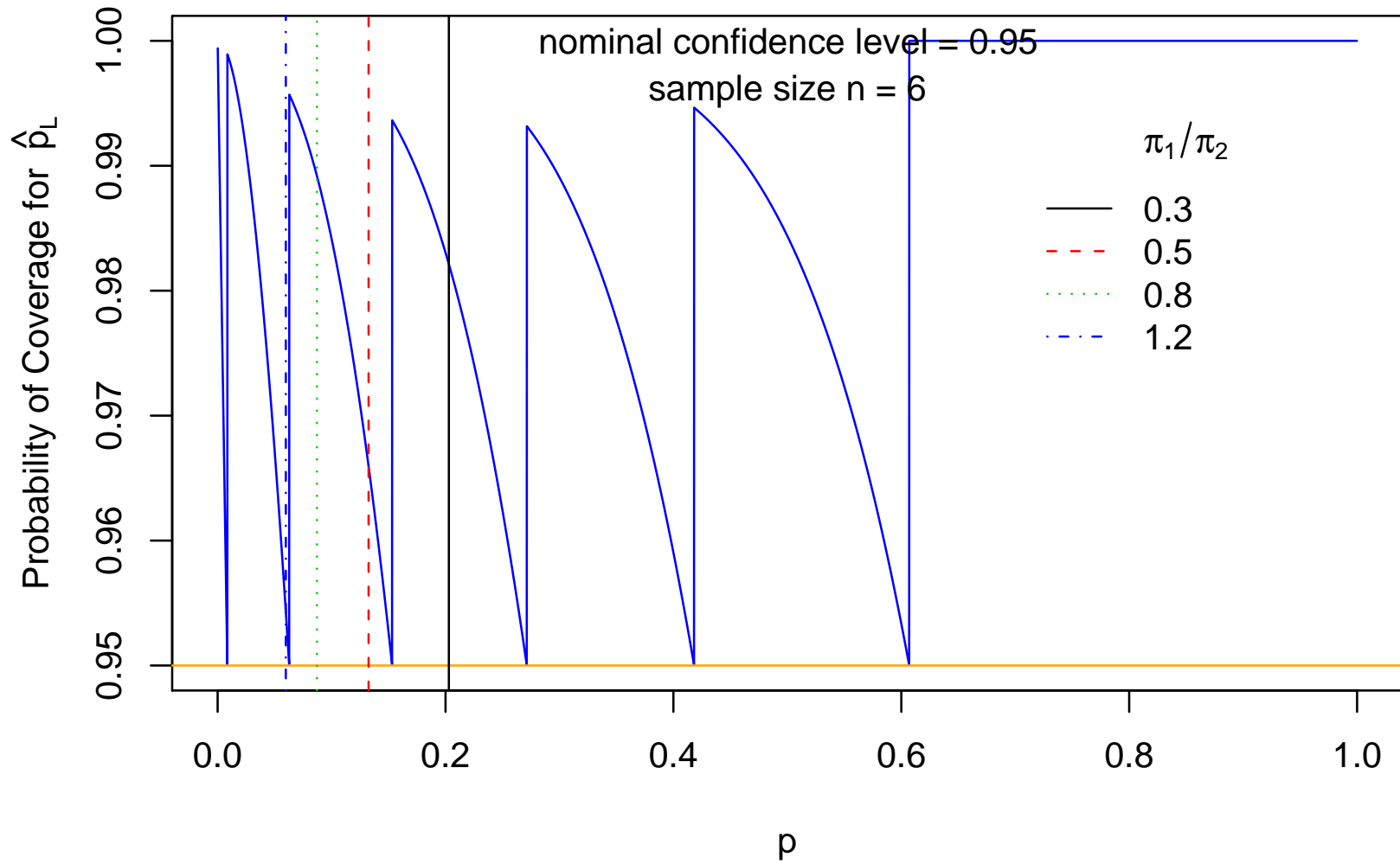
Comparing Hull Loss Rates across Airplane Models

The B767 had $X = 3$ hull losses and a rate of $.32$ per 10^6 departures.

This translates into $n_1 = 9375000$ departures for the B767.

Here our 95% upper bound for π_1 / π_2 is $.6876$

This is clearly below 1 and thus indicates a difference in experience

between the B767 and the Concorde, in favor of the B767.

Concorde Takeoff at JFK