Notes on binomial prediction intervals and bounds

For SM230 purposes, we distinguish between prediction bounds (for integer random variables $X_{max}$ & $X_{min}$ in $\text{binp}$) and confidence bounds (for parameter $p$ in $\text{binp}$). We make this distinction because we can directly observe the number of successes $X$, but not the overall (or population-wide) probability $p$.

**Given known, constant $p$ & $N$:**

1. To find a one-sided lower prediction bound (LPB) for $X$, vary $X_{min}$ while $X_{max} = N$, or
2. To find a one-sided upper prediction bound (UPB) for $X$, vary $X_{max}$ while $X_{min} = 0$ until the cumulative probability distribution function CDF (= PROB) adds up to or just exceeds some preset value. Typically for one-sided prediction bounds we want $\text{PROB} = 0.9$.

Restated, for the LPB we want to be 90% certain (or a little more) that in $N$ trials we’ll get at least $X_{min}$ successes. For the UPB, we want to be 90% certain (or a little more) that in $N$ trials we’ll get at most $X_{max}$ successes. In Fig. 1 below, parameters $N = 20 & p = 0.65$, and the mean or expected number of successes $\mu = Np = 20*0.65 = 13$.

![Fig. 1](image_url)

To find the 90% LPB in Fig. 1, guess & test using $\text{binp}(0.65,20,20,x)$, & you find that $x=10 \rightarrow 0.94683$. Thus the 90% LPB for $X=10$, and by definition the corresponding upper prediction interval is $10 \leq X \leq 20$. ($\text{solve}(\text{binp}(0.65,20,20,x)=0.9,x)$ also works, but it’s very slow here.)

To find the 90% UPB in Fig. 1, guess & test using $\text{binp}(0.65,20,x,0)$, & you find that $x=16 \rightarrow 0.9556$. Thus the 90% UPB for $X=16$, and by definition the corresponding lower prediction interval is $0 \leq X \leq 16$. Conveniently, here the 90% two-sided prediction bounds are also $X_{min} = 10$, $X_{max} = 16$. In other words, $\text{binp}(0.65,20,16,10)=0.9025$, or we’re 90% certain that in $N$ trials we’ll get 10–16 successes, inclusive. Note that in general the 90% one-sided & 90% two-sided prediction bounds aren’t the same.

**Given known, constant $p$ & an observed number of successes $X$:**

3. To find a one-sided 90% LPB for $N$, set $X_{min} = 0$, $X_{max} = X$, and then increase $N$ from $X_{max}$ until PROB just decreases to 0.9 (or just above 0.9). Then the 90% LPB = $N$, or,
4. To find a one-sided 90% UPB for $N$, set $X_{min} = X$ and increase $X_{max} & N$ together until PROB just increases to 0.9 (or just above 0.9). Then the 90% UPB = $N$.

So if $p = 0.5 & X = 10$ successes, the 90% LPB for $N$ is given by $\text{binp}(0.5,15,10,0) = 0.9408$, or LPB = 15. Interpret this result as “If $N = 15$, there’s a more than 90% chance that I’ll observe at most 10 successes.” Similarly, the 90% UPB for $N$ here is given by $\text{binp}(0.5,26,26,10) = 0.9157$, or UPB = 26. Interpret this result as “If $N = 26$, there’s a more than 90% chance that I’ll observe at least 10 successes.”
Notes on binomial confidence bounds

To determine 90% lower & upper confidence bounds (LCB & UCB, respectively) for binomial \( p \), start by examining Fig. 2. The effect of increasing \( p \) is to shift the PDF’s peak toward higher \( X \). Conversely, decreasing \( p \) shifts the PDF’s peak toward lower \( X \). These shifts make sense, since as in Fig. 1, the mean or expected number of successes \( \mu = Np \) (\( N = 16 \) in Fig. 2). Thus decrease \( p \) and you decrease \( \mu \).

Now the goal is not to vary \( X_{\text{max}} \) or \( X_{\text{min}} \) while taking \( p \) as known. Instead, we use \( N, X_{\text{min}}, \) and \( X_{\text{max}} \) as statistical data to test whether some particular assumed value for \( p \) is likely.

If \texttt{binp} returns a small value (typically, \( \leq 0.1 \)) for an assumed \( p \) value, then this \( p \) was inconsistent with our data and we look for another, more consistent \( p \) value. Suppose we observe 12 successes in 16 trials. Figure 2 does not show this observed fact, but instead illustrates several possible binomial PDFs that result from assuming different \( p \) values for \( N = 16 \). These PDFs may or may not be consistent with our observations.

Given \( X \) successes and known, constant \( N \):

1. to find a one-sided 90% lower confidence bound (LCB), set \( X_{\text{max}} = N, X_{\text{min}} = X, \) & decrease \( p \) from 1 until PROB decreases to 0.1, or
2. to find a one-sided 90% upper confidence bound (UCB), set \( X_{\text{max}} = X, X_{\text{min}} = 0, \) & increase \( p \) from 0 until PROB increases to 0.1.

So to find the 90% LCB in Problem 3.7.11 (p. 3.24), calculate \texttt{solve(binp(p,16,16,12)=0.1,p)}

\[ \rightarrow p = 0.5611 \]

Thus the 90% LCB for the coach’s overall \( p = 0.5611 \). Interpret this result as follows: “If \( p \) is as small as 0.5611, there is only a 10% chance that the coach can have 12 or more successes (wins) in 16 trials (games). If \( p \) were smaller, his probability of having 12 or more wins would be unacceptably small (PROB<0.1), so I conclude that the 90% LCB for \( p = 0.5611 \). In other words, I’m 90% confident he’ll win fewer than 12 of 16 games\(^\dagger\) if the unknown \( p = 0.5611 \).”

Similarly, to find the 90% UCB in Problem 3.7.11, calculate \texttt{solve(binp(p,16,12,0)=0.1,p)}

\[ \rightarrow p = 0.8862 \]

Thus the 90% UCB for the coach’s overall \( p = 0.8862 \). Interpret this result as follows: “If \( p \) is as large as 0.8862, there is only a 10% chance that the coach can have 12 or fewer wins in 16 games. If \( p \) were larger, his probability of having 12 or fewer wins would be unacceptably small (PROB<0.1), so I conclude that the 90% UCB for \( p = 0.8862 \). In other words, I’m 90% confident that he’ll win more than 12 of 16 games if the unknown \( p = 0.8862 \).”

\( \dagger \) Recall that \( \texttt{binp(p,N,N,X+1)+binp(p,N,X,0)=1.0} \). Similarly, \( \texttt{poisson(}\lambda*\text{t,}\infty,X+1)+\texttt{poisson(}\lambda*\text{t,X,0})=1.0 \). Thus the best, most accurate substitute for \( \texttt{poisson(}\lambda*\text{t,}\infty,X+1) \) is \( 1-\texttt{poisson(}\lambda*\text{t,X,0}) \) (e.g., \( \texttt{poisson(6,}\infty,10) = 1-\texttt{poisson(6,9,0)} = 0.08392 \)).