



Model Specification and Confidence Intervals for Voice Communication

**Interval Management for Near-Term Operations Validation of Acceptability (IM-NOVA)
Human-in-the-Loop (HITL) Experiment
Air Traffic Management (ATM)
Technology Demonstration – 1 (ATD-1)**

Robert Leonard and David J. Edwards
Virginia Commonwealth University

Sara Wilson
NASA Langley Research Center

April 30, 2013



Outline

- What part of the IM-NOVA project is analyzed?
- Steps of clearance sequence
- Narrow down distribution choices
- Closer look at two particularly good choices
- Details of using one over the other
- Compare results
- Quick look at off-nominal results
- Conclusion with practical sense in mind



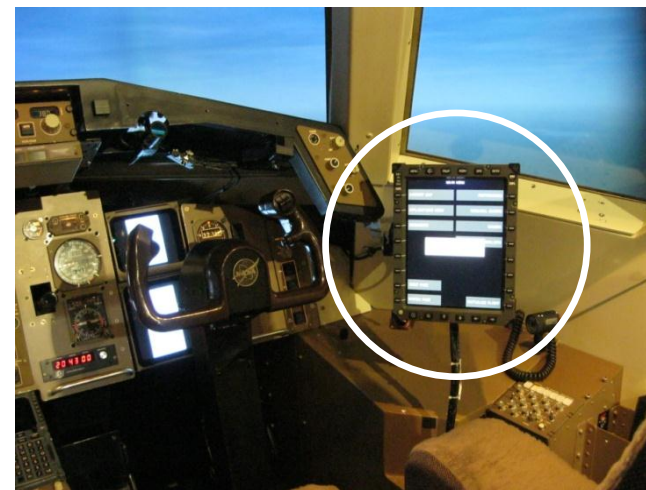
Motivation for this talk

- Increasing air-traffic → increasing radio traffic
- ATD-1 → How can technology help modify the management system to handle this increase in traffic?
- IM-NOVA → Shifting responsibility from ATC to flight crews
- Data link not possible but voice communication can be better understood to improve the issuing of clearances
- This talk focuses on modeling this command/response voice communication to extend these results to other industry related situations

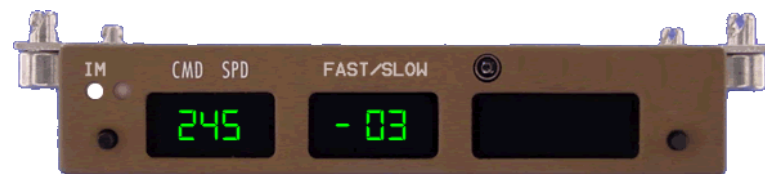


Background Info

- 10, two-person, flight crews served as test pilots within the Integration Flight Deck (IFD), a fixed-base B757 cab
- Confederate controllers (in ATOL):
 - Used on-ground schedule management and route selection tools
 - Issued clearances (standard speed instructions or IM clearances) via voice radio comm.
- The IFD was equipped with a display interface used by flight crews to enter clearances into the onboard system
- Pseudo pilots operated aircraft by responding to controller commands issued via radio comm.



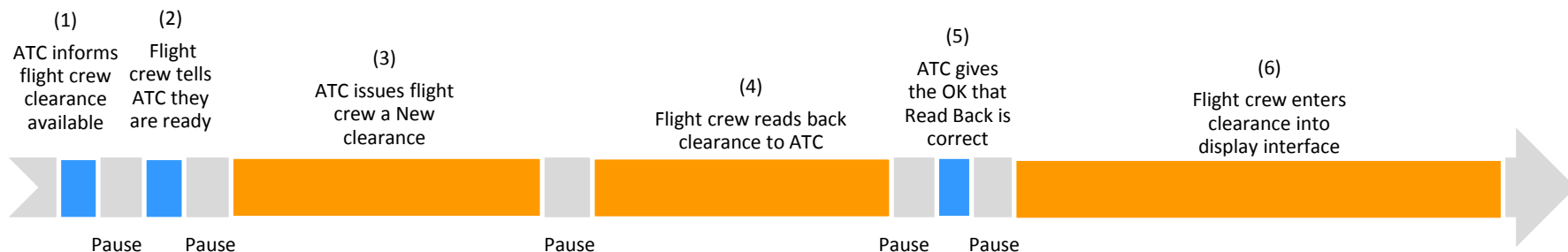
Flight-deck display interface



Target and error speed values shown on guidance displays



Nominal Time Line for Clearance



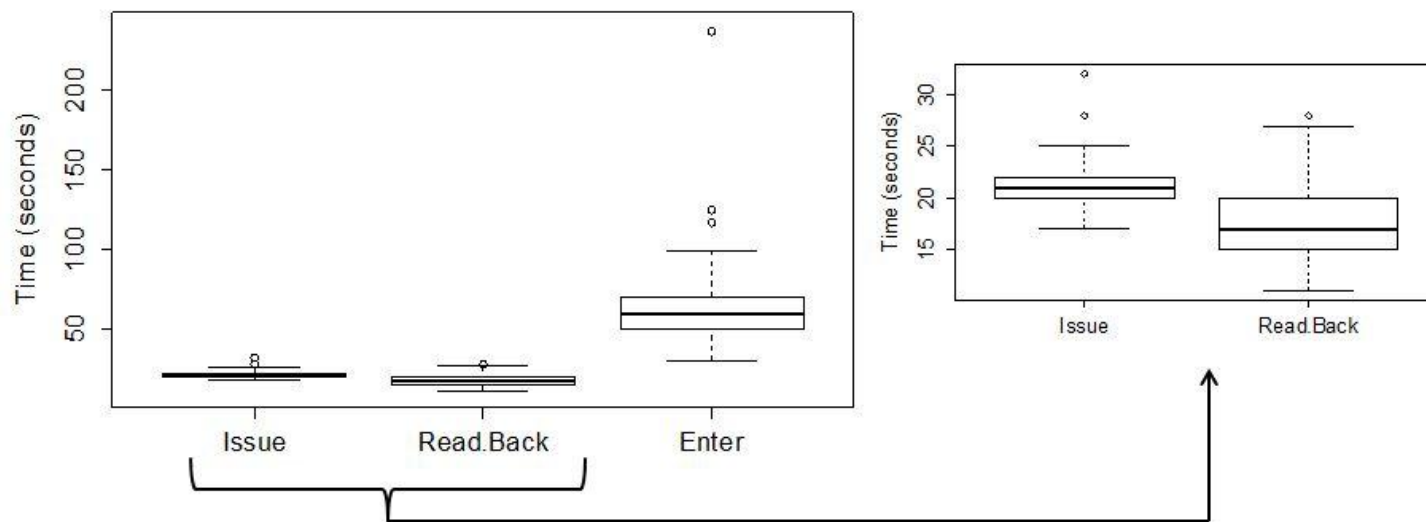
1. ATC begins radio communication by informing the flight crew that a clearance is ready.
2. After a pause, the flight crew responds to indicate to ATC that they are ready.
3. Then, ATC issues the clearance and the flight crew writes down the clearance.
4. Next the flight crew reads back the clearance to the ATC.
5. ATC informs the flight crew that the read back of the clearance is correct
6. Finally, the flight crew enters the clearance into the display interface.

IM Clearance Example

*NASA 65, for Interval Spacing, cross **JIFFY** at **1859 plus 13 seconds**, when able space **113 seconds** behind **Alpha Alpha Lima 544** on **Cedar Creek 6 arrival, Alexandria Transition**. Report commencing interval spacing.*



Sample Statistics



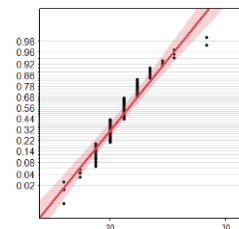
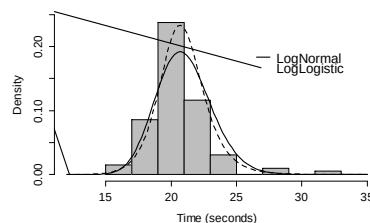
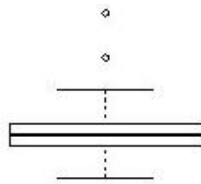
Step	Sample Size	Mean (seconds)	SD (seconds)	Outliers (seconds)
Issue	99	21.0	2.2	(28, 28, 32)
Read Back	99	17.8	3.6	(28)
Enter	99	62.8	24.1	(117, 125, 237)



Outliers

Total of 7 outliers in this study

- Outliers are **numerically** distant from the rest of the data
- They may also be visible



- Measurement error or real life occurrences?
 - 3 for ATC issuing a clearance
 - ATC stumbles, pauses, and/or self-corrects
 - 1 when flight crew was reading back a clearance
 - Flight crew couldn't write down the clearance
 - 3 for entering a clearance into the display interface
 - 2 times the clearance was completely entered but display interface wasn't activated
 - The other was a continuation from the read back outlier



Fitting Distributions

- Distributions fit
 - normal, lognormal, gamma, Weibull, loglogistic, logistic, exponential (all continuous distributions)
- Maximum likelihood estimation (MLE) was used to estimate the parameters
 - Closed/non-closed forms depending on distribution
- Akaike's corrected information criterion (AICc) used to determine what models are “best” and therefore “most useful”
- Competing choices further compared



Ways to Compare Distribution Fits

- **Information Criterion**
 - Likelihoods, AIC, AICc, BIC, etc.
 - Quantified way to compare distribution
 - Why distributions are different is not so clear. Values alone mean nothing.
- **Overlay pdfs on histograms**
 - Visually inspect single curve or compare different curves to same histogram
 - Can possibly give more insight than comparing criterion values
 - Discriminating between distribution choices can still be difficult
- **Probability Plots**
 - Compare single or multiple distributions to same sample
 - Can be more precise than overlaying pdfs onto histograms → more insight
- **Various Test Statistics**
 - Chi-square, Kolmogorov-Smirnov, Shapiro-Wilk, etc.
 - Suppose to answer the question “Does this sample come from this distribution?”
 - Problem - Not always reliable, sometimes distribution specific, etc.
 - Can also be used to discriminate between competing distribution choices



AICc for Clearance Steps

$$AICc = -2\log(L) + 2k + \frac{2k(k+1)}{n-k-1}$$

Distribution	Issuing	Reading Back	Entering
Normal	444	537	914
Lognormal	430	<u>529</u>	850
Gamma	434	531	865
Weibull	485	546	911
Loglogistic	<u>419</u>	533	<u>841</u>
Logistic	426	539	866
Exponential	803	771	1018

- Choose the “best” model by looking for the smallest AICc value
- Loglogistic has the smallest AICc for issuing and entering a clearance
- Lognormal is the smallest for reading back a clearance
- Problem – We can’t really see what makes our choices the best. AICc values alone mean nothing. Only when comparing the values can we make an informed decision.

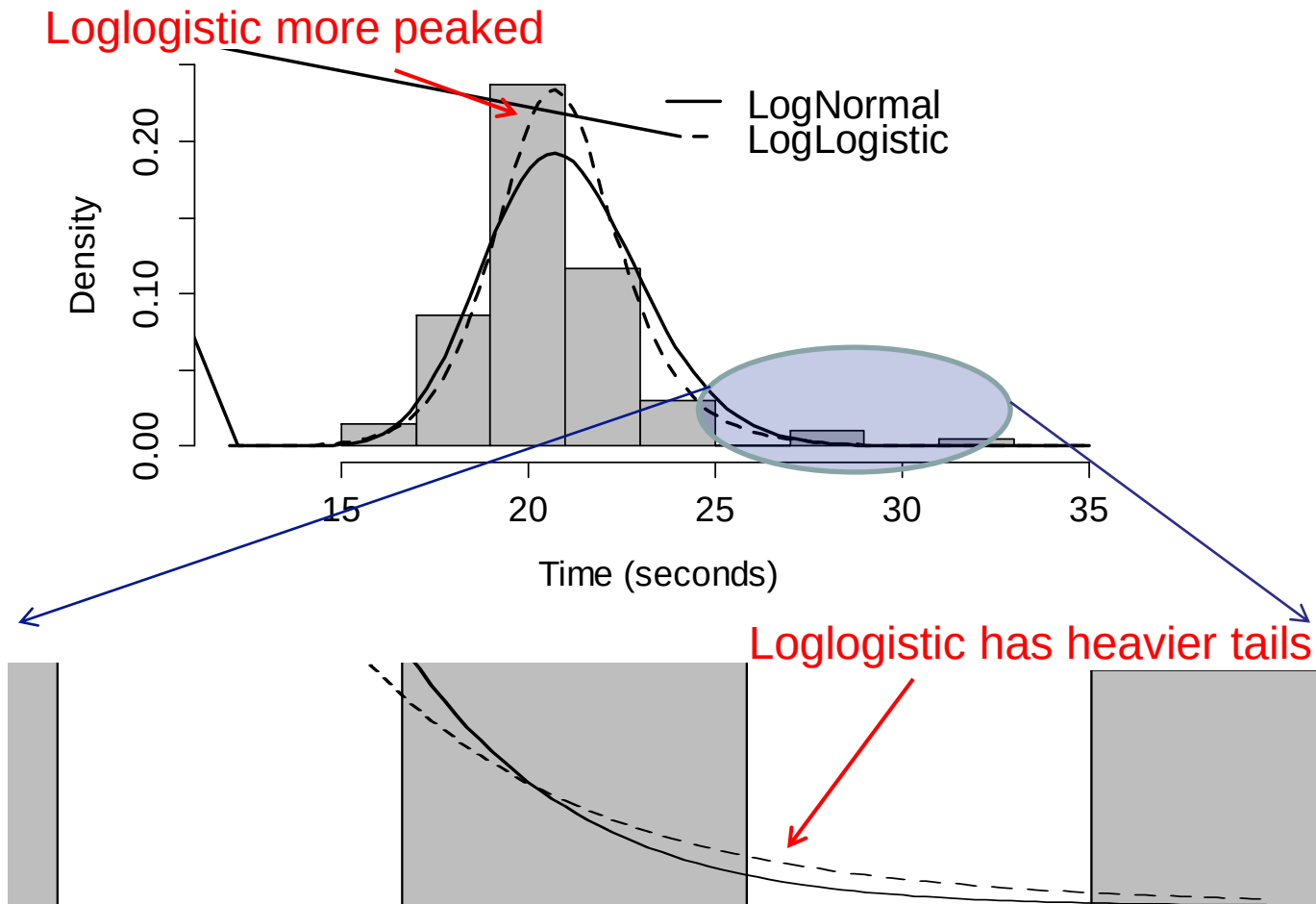


Comparing the Two Narrowed Down Choices

Lognormal vs. Loglogistic	
Skewed distribution with thinner tails , thus giving more probability to nominal results	Skewed distribution with thicker tails , thus giving more probability to off-nominal results
Closed form parameter estimates $\hat{\mu} = \frac{\sum_k \ln x_k}{n} \quad \hat{\sigma}^2 = \frac{\sum_k (\ln x_k - \hat{\mu})^2}{n}$	• Iterative methods necessary to estimate parameter estimates • JMP, R, Minitab and other statistical software can be used
CI calculations on the mean can be calculated by plugging in closed form parameter estimate values	CI calculations on the distribution mean require simulations such as bootstrapping, normal approximation, etc.



Differences between Lognormal and Loglogistic





Parameter Estimates for Both Distributions

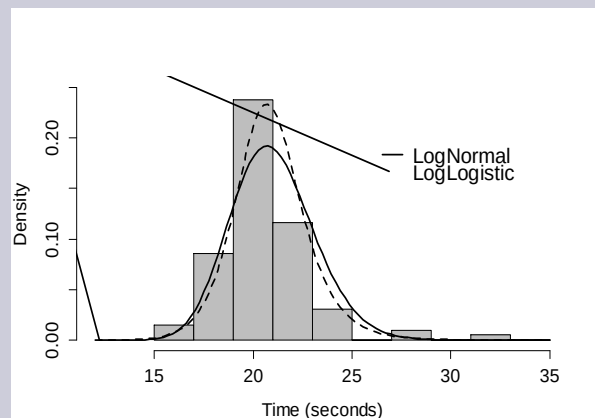
Clearance Step	Distribution	Location Parameter $\hat{\mu}$ (log seconds)	Scale Parameter $\hat{\sigma}$ (log seconds)
Issue	Lognormal	3.04	0.10
	Loglogistic	3.04	0.05
Read Back	Lognormal	2.86	0.20
	Loglogistic	2.86	0.11
Enter	Lognormal	4.09	0.29
	Loglogistic	4.08	0.15

- Maximum Likelihood Estimation used $\max L(\vec{\theta}; \vec{x}) = \max \prod_{i=1}^n f(x_i; \theta_1, \dots, \theta_k)$.
- These estimates maximize the probability that the proposed distribution produced the given sample
- Notice that we can still glean information about the sample variance from the scale parameter



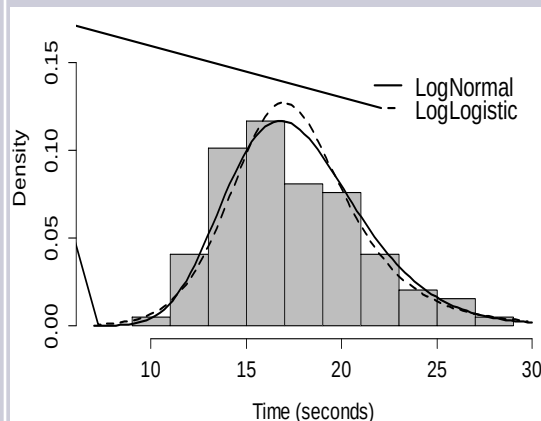
1st Comparison - Overlay pdfs on Histograms

Issue



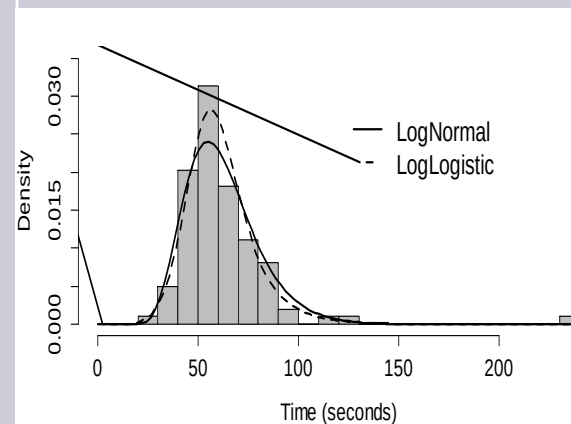
Loglogistic appears to better capture the activity of the sample

Read Back



Both are comparable fits but the lognormal appears to be a better fit

Enter

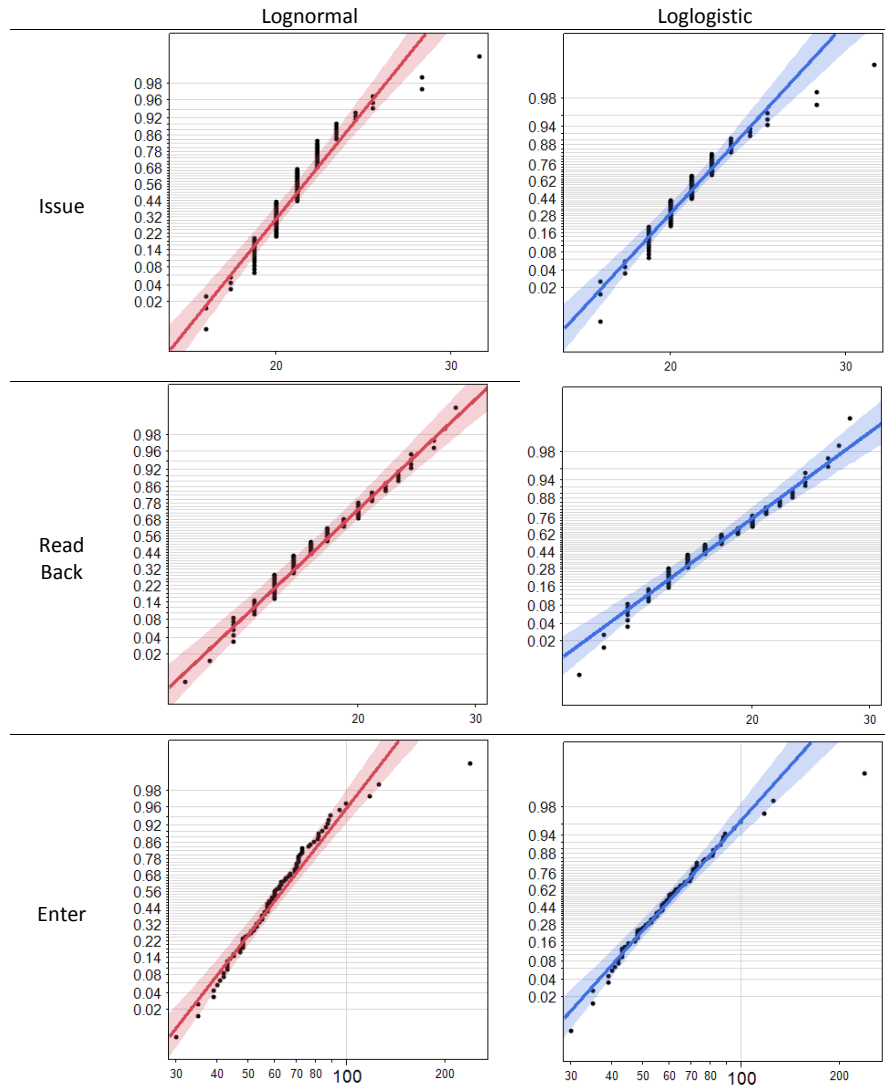


Similar to issuing a clearance, the loglogistic appears to better capture the activity of the sample

Visually we are seeing what the AICc values were indicating –
Loglogistic is a better fit for issuing and entering a clearance while
lognormal is a better fit for the reading back a clearance.



2nd Comparison – Probability Plots



What are we looking for?

- Linear conformance – departures from the straight line indicate that a distribution is having issues capturing the activity of the sample
- Confidence bands also help us
- Behavior of our data. Stacking?

Interpretation

- The upper tails are where we are having the most departure
- Loglogistic is better for issuing and entering a clearance, but the lognormal appears to capture the activity as good as or better than the loglogistic for any step



3rd Comparison – Goodness-of-Fit Test

- Many hypothesis tests available
- Many not reliable due to being sensitive to assumptions
- Chose the Kolmogorov-Smirnov test statistic because it is easy to interpret and understand with the behavior of our data



Kolmogorov-Smirnov Test

- Testing the null hypothesis that the data come from a particular distribution
- $P\text{-value} < 0.05$ rejects the null hypothesis
- Used for continuous distributions
- P-values can be compared for discriminating between competing distribution choices



KS Test Results

Issue	Lognormal (3.04,0.10)	Loglogistic (3.04, 0.05)
	$p_{LN} = 0.0086$	$p_{LL} = 0.0598$
Read Back	Lognormal (2.86,0.20)	Loglogistic (2.86, 0.11)
	$p_{LN} = 0.2450$	$p_{LL} = 0.2316$
Enter	Lognormal (4.09,0.29)	Loglogistic (4.08, 0.15)
	$p_{LN} = 0.4486$	$p_{LL} = 0.9600$

- Null hypothesis **rejected** that the sample for issuing a clearance came from a lognormal distribution but not a loglogistic distribution
- Read Back more likely to come from a lognormal
- Issue and Enter more likely to come from a loglogistic
- Results affected by our data behaving discretely...So do we trust these results?



Confidence Interval Calculations

Different Procedures

Lognormal

- Closed form parameter estimates make things easier
- Naïve vs. Cox Methods

Compare these two methods looking at variance and sample size

Loglogistic

- Requires more work
- We implemented bootstrapping (resample with replacement) but normal approximation or other methods could have been used as well

Compare all 3 results for CIs

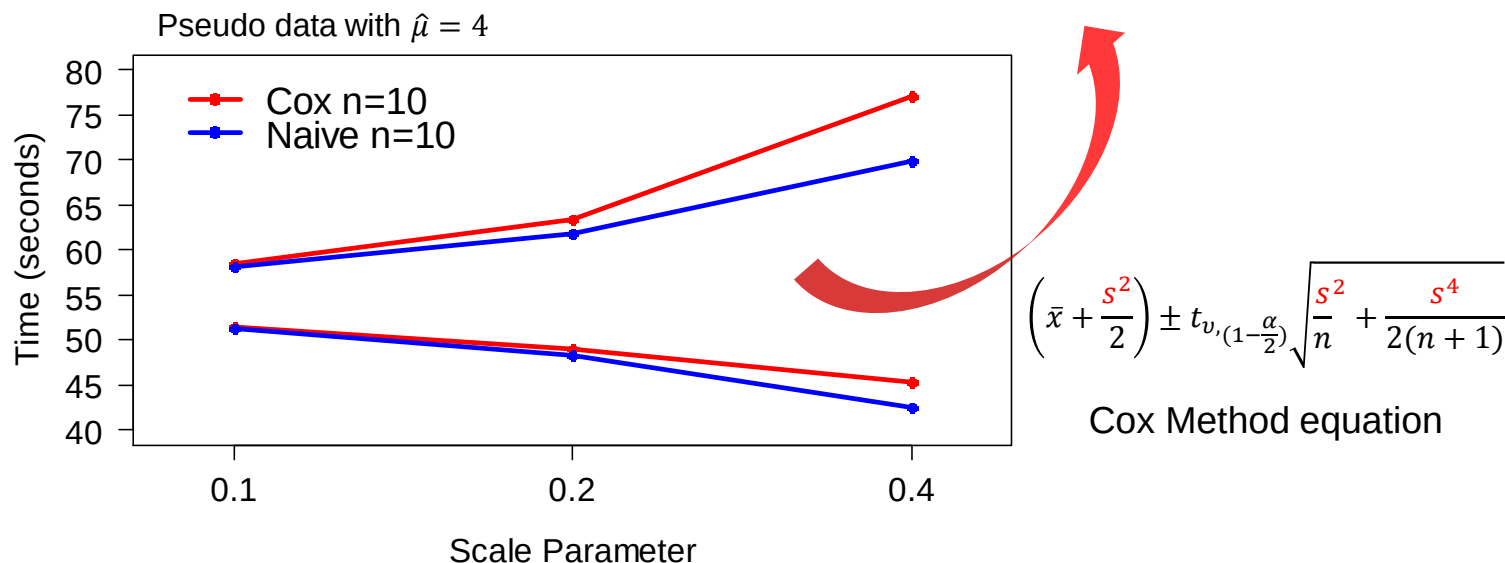


Lognormal CI - Naïve vs. Cox Methods

Naïve Method	Cox Method
1. Normalize the data by using a log-transformation 2. Using normal theory, construct confidence interval limits on μ, the mean of the transformed data, as	1. Normalize the data by using a log-transformation 2. Using normal theory, construct confidence interval limits on μ, the mean of the transformed data, as
$\bar{x} \pm t_{v, (1-\frac{\alpha}{2})} \frac{s}{\sqrt{n}}$	$\left(\bar{x} + \frac{s^2}{2} \right) \pm t_{v, (1-\frac{\alpha}{2})} \sqrt{\frac{s^2}{n} + \frac{s^4}{2(n+1)}}$
3. Transform the limits calculated in step 2 back to their original scale by taking the exponential of the two values	3. Transform the limits calculated in step 2 back to their original scale by taking the exponential of the two values
• Both methods use the sample mean and variance of the transformed data • The naïve method uses $\bar{x}$ and $\frac{s}{\sqrt{n}}$, which are both biased • The Cox method uses $\left(\bar{x} + \frac{s^2}{2} \right)$ and $\sqrt{\frac{s^2}{n} + \frac{s^4}{2(n+1)}}$, which are both minimum variance unbiased estimators (MVUEs) of the expected value and standard error • Note the influence of the sample variance (s^2) in both the expected value and standard error when using the Cox method	



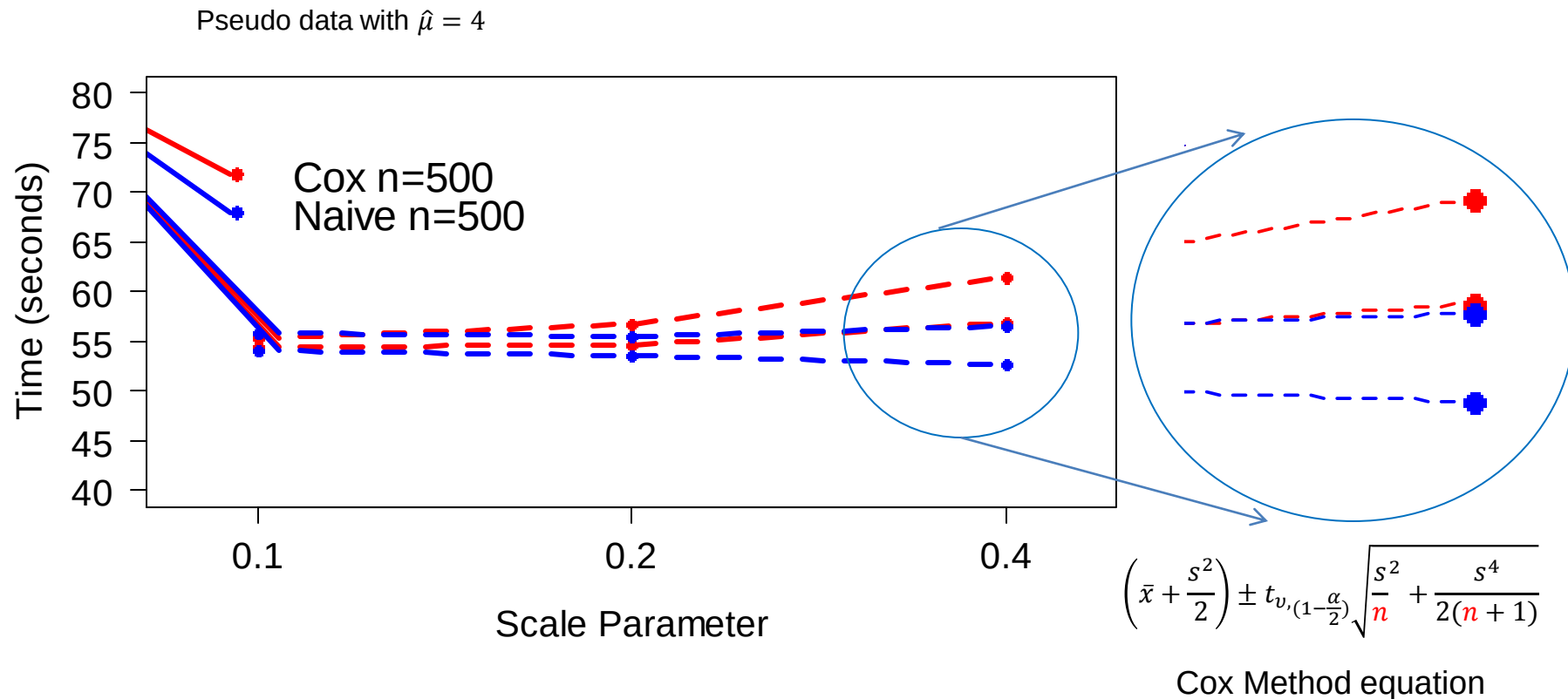
Comparing Naïve and Cox Methods



- The divergence between the two methods becomes more prominent as the variance increases
- For a **small sample size**, the CIs for the two methods **still overlap**
- This divergence can become more influential in decision making when the sample size increases



Comparing Naïve and Cox Methods



- Tighter interval due to **larger sample size**
- As the variance increases these tighter intervals no longer overlap - **significant difference**



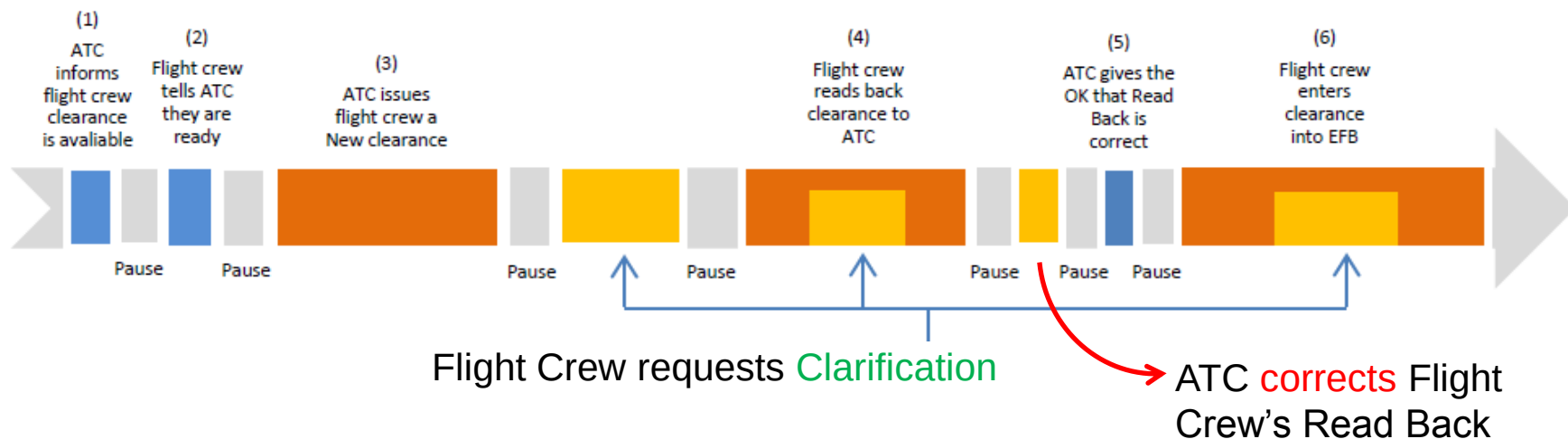
Comparing Lognormal and Loglogistic CIs

Clearance Step	Distribution Fit	Method	Expected Value	95% C.I.
Issue	Lognormal (3.04, 0.10)	Naïve	20.9	(20.5, 21.3)
		Cox	21.0	(20.6, 21.4)
	Loglogistic (3.04, 0.05)	Bootstrap	20.9	(20.6, 21.3)
Read Back	Lognormal (2.86, 0.20)	Naïve	17.5	(16.8, 18.2)
		Cox	17.8	(17.1, 18.5)
	Loglogistic (2.86, 0.11)	Bootstrap	17.8	(17.0, 18.6)
Enter	Lognormal (4.09, 0.29)	Naïve	59.7	(56.4, 63.3)
		Cox	62.3	(58.8, 66.1)
	Loglogistic (4.08, 0.15)	Bootstrap	61.4	(58.1, 65.0)

- Lognormal and Loglogistic **not operationally different**
- We have relatively low variance



Off-Nominal Results



Total of 14 off-nominal runs

- Off-nominal results not necessarily outliers
- These results “pack-up” the body-to-tail area of a distribution
- 6 due to flight crew **corrected** by ATC after read back of clearance
- 8 due to the flight crew requesting **clarification**
 - 2 before read back of clearance
 - 3 during read back
 - 3 after read back, while entering the clearance into the EFB



Conclusion

- Loglogistic is a better fit for issuing and entering a clearance
- Lognormal is a better fit for reading back a clearance
- Variance, outliers and off-nominal results play an important role in understanding the “best” and therefore “most useful” model for issuing, reading back and entering a clearance
- Characteristics of the lognormal make it easier to conduct inference on the data
- There are no operationally different results in choosing the loglogistic over the lognormal
- Therefore, the lognormal distribution and the calculations resulting from its choice are acceptable for the practical purposes of analyzing the data sets presented for this experiment
- Questions/Comments?