ARS, North America 2008

SAND2008-3889P

Reno, Nevada USA

Track Session

Integration of Component, Subsystem, and System Test Data

David G. Robinson
Risk and Reliability Analysis Dept
Sandia National Laboratories

Sandia National Laboratories

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under Contract DE-AC04-94AL85000.

Introduction

As a system evolves

- **suppliers changes**
- **product line undergoes design changes**

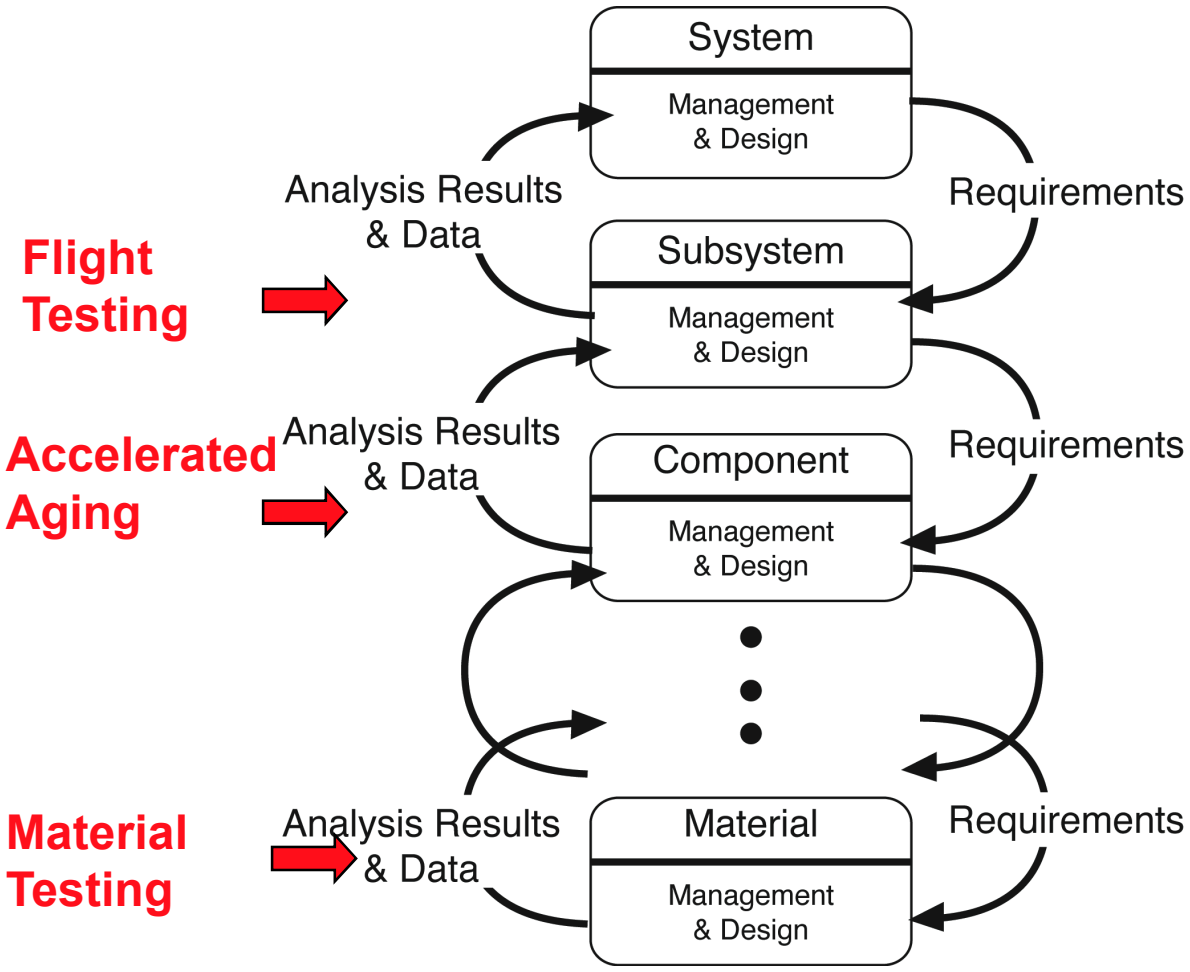
… the population becomes a mixture of various sub-populations

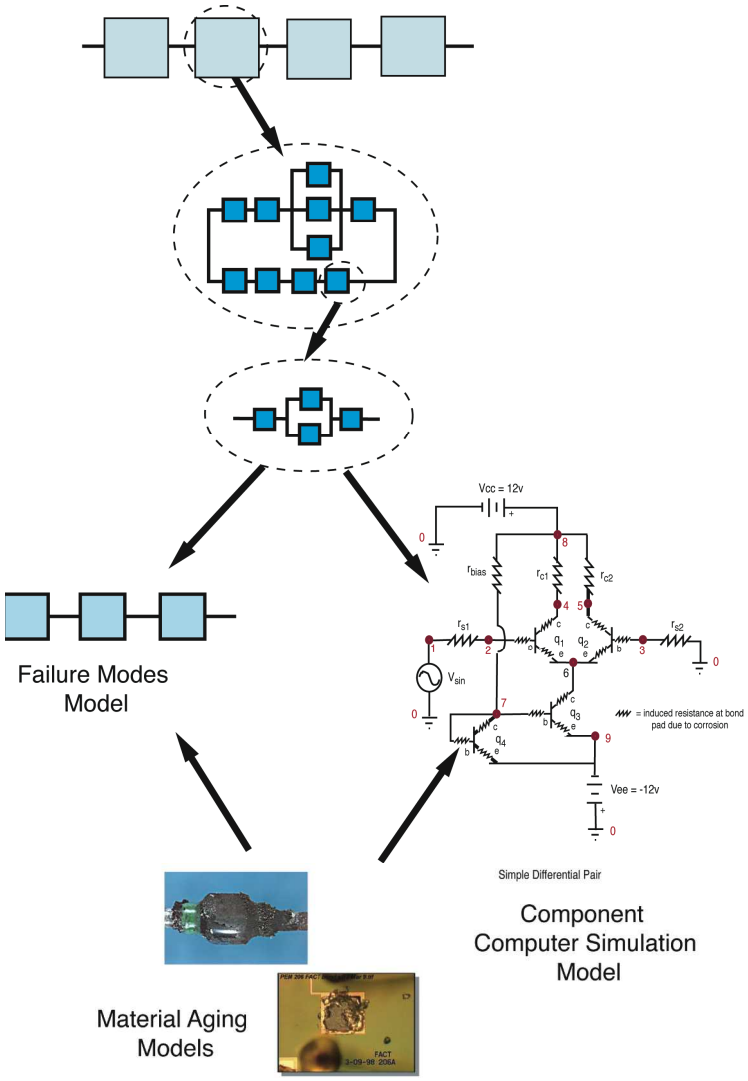
How can you estimate the reliability of entire population and individual sub-populations?

Data is available from:

- **different types of tests: flight, laboratory, etc.**
- **different levels of indenture: materials aging, piece-part, component, system**

Must recognize value of data from all sources





"Value" of data is measured by the impact of the data on the credibility of the reliability estimates

Agenda

- Provide some historical background and discuss positive and negatives of possible analysis approaches
- Briefly review the latest approach currently employed for aggregating data from:
  - multiple testing sources and
  - multiple levels of indenture
- Introduce a series of simple, real reliability analysis examples
- Summary

Historical Background …

- Allegory of Fortune, Dossi (1486-1542)
- Lady on right is the Goddess Fortuna (Lady Luck)
- Man on left is Chance
- Fortuna
  - holding a basket of fruit symbolizing good luck
  - Only has one sandal which means she can also bring bad luck
  - Sitting on bubble - luck won't last
- Chance
  - Holding lottery tickets
  - Dossi was court painter near Venice and Venice just introduced a state lottery to raise money

QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.

Painting is in J. Paul Getty Museum and notes from Dossi exhibit

Historical Background

- 1713 - Bernoulli suggests that there is a difference between the logic used for
  - Games of chance (deductive)
  - Decision making (inductive)
- 1763 - Thomas Bayes (in posthumous article published by friend)
  - Proposes solution to Bernoulli's question
  - Develops foundation for what we call 'Bayes' theorem, but the formulae we use is primarily due to …
- 1787-1812 - Laplace
  - "The theory of probabilities is at bottom nothing but common sense reduced to calculus.~Laplace, *Theorie analytique des probability"*, 1820
  - Uses probability to prove stability of solar system
  - Use Bayes theorem to estimate the mass of Saturn given observational data and laws of physics
  - Stated: P{mass of Saturn is between m1 and m2 | data, model}:
    - "… it is a bet of 11,000 to 1 that the error in this result is not 1/100th of it's value."
  - 150 years of data have not changed that estimate by 0.63%
- 1922 - Fisher
  - Develops maximum likelihood methods and establishes Frequentist perspective of statistics
  - Uses statistics to prove that smoking does not cause lung cancer

Conceptual Differences

- Frequentist statistics do not allow the use of probability theory in estimating the mass of Saturn:
  - since the mass of Saturn is not a random variable
  - To use Frequentist statistics an ensemble of universes would have have to be imagined in which everything is constant except the mass of Saturn.
- "What is the probability that OJ is guilty?"
  - Frequentist think that such questions can not be addressed by statistics since there is no underlying concept of a random process
- Frequentist approach cannot be applied for things that don't exist yet, i.e. cannot be applied during design

Conceptual Differences

- Confidence Intervals versus Credibility Intervals
  - If an ensemble of samples is generated from a population, and a confidence interval calculated for each sample, then a certain percentage (the confidence level) of all the intervals will include the unknown population parameter. The upper and lower values for the confidence interval are referred to as the confidence limits.
  - However, confidence limits are not related to the probability that a parameter value or response falls within certain limits. Risk analysts often refer to credibility limits for characterizing the uncertainty in the risk associated with a particular accident. Credibility limits reflect the probability that a parameter or system response falls within certain limits.
  - VIP distinction: you can add credibility intervals but you cannot add confidence intervals

Confidence vs Credibility Intervals

- Confidence and credibility intervals are NOT the same.
- A confidence interval says something about a parameter AND a random variable (or statistic) based on it.
- A credibility interval describes the knowledge about the parameter; it must always be based on a specification of the knowledge *before* making the observations, as well as the observations
- In many cases, computed confidence intervals correspond to credibility intervals with a certain *prior knowledge* assumed.

When to use …

- Frequentist methods
  - Significant amounts of observational/computational data
  - Construction of an ensemble is possible, i.e. repeatable observations
  - Data comes from a single source
  - Measurements are repeatable
  - No uncertainty in data, only in the ability to measure or observe
  - No uncertainty in the underlying physical model
- Bayesian methods
  - Limited amounts of observational/computational data
  - Not possible to conceptualize what an 'ensemble' would look like
  - Data comes from multiple sources or multiple levels of indenture from within the system
  - Want to characterize the uncertainty in system characteristics, e.g. reliability, mass, mean failure time …
    - Cannot calculate confidence intervals for the reliability/risk of a system when the data comes from multiple sources or levels of indenture
  - Explicitly need to include expert judgment

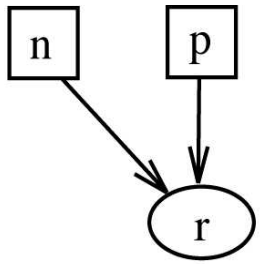
Alternative Approach

- A Bayesian approach provides an organized and objective means of estimating reliability of heterogeneous populations.
- Hierarchical Bayesian analysis formally captures the conditional relationship between parameters and the data available
- HB is easy to extend to account for underlying time-dependent processes, e.g. aging degradation.
- Following discussion goes through the approach using a simple Bernoulli testing situation (pass/fail)

Mathematical Approach: Simple Example

- Consider a simple problem of characterizing the probability of *r* successes in *n* tests:

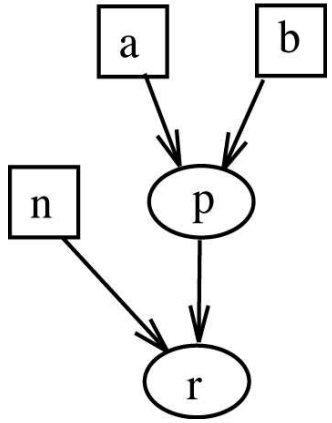
$$f(X=r\mid p)=\frac{n!}{r!(n-r)!}(1-p)^{n-r}(p)^{r}$$



- Reliability is explicitly recognized as a random variable

$$g(p\mid r)=\frac{g(p)f(X=r\mid p)}{\int_{o}^{1}g(p)f(X=r\mid p)dp}$$

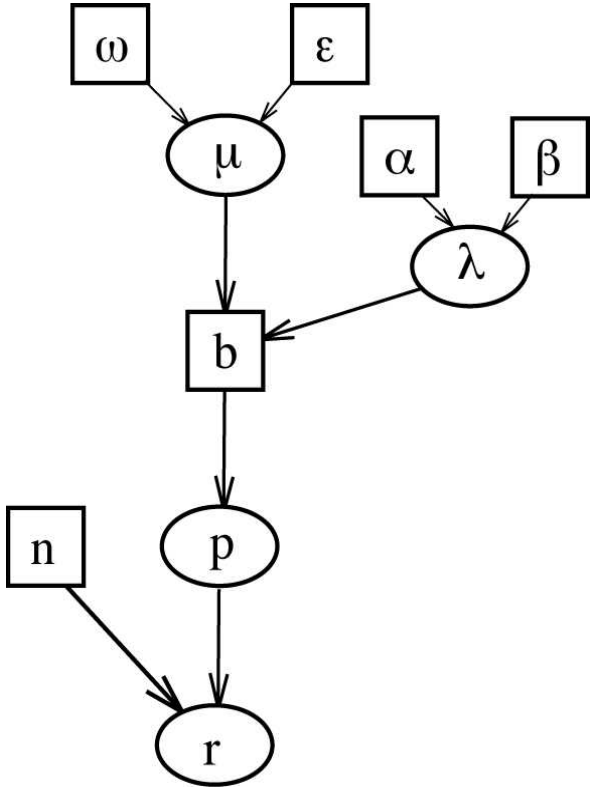
$$=\frac{g(p)f(r\mid p)}{f(r)}$$



Extension of Basic Model

- For mathematical reasons, transform beta distributed probability into a random variable distributed across the real line:

$$\text{logit}[p] = \log\left[\frac{p}{1-p}\right] = \mu + \lambda$$



- The parameter λ represents the error induced by the testing not representing true (transformed) operational reliability.
- Since λ is a gamma distributed random variablew/ parameters α and β, the error will always be non-negative
- For data collected from *operational* testing λ=0

Extension of Basic Model

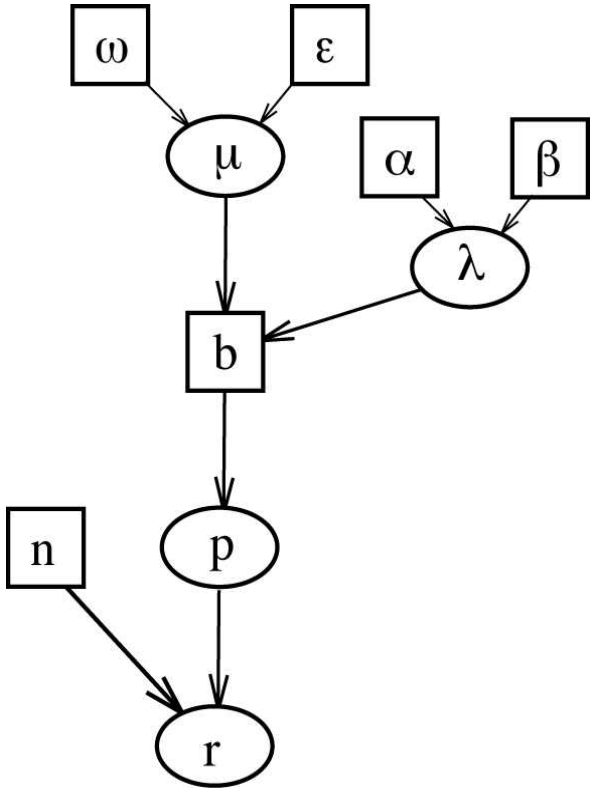
- Recall:

$$\text{logit}[p] = \log\left[\frac{p}{1-p}\right] = \mu + \lambda$$

- While *p* was a beta distributed random variable, μ is a Gaussian random variable with mean and variance:

$$\mu \sim N\left(\omega, 1/\varepsilon^2\right)$$

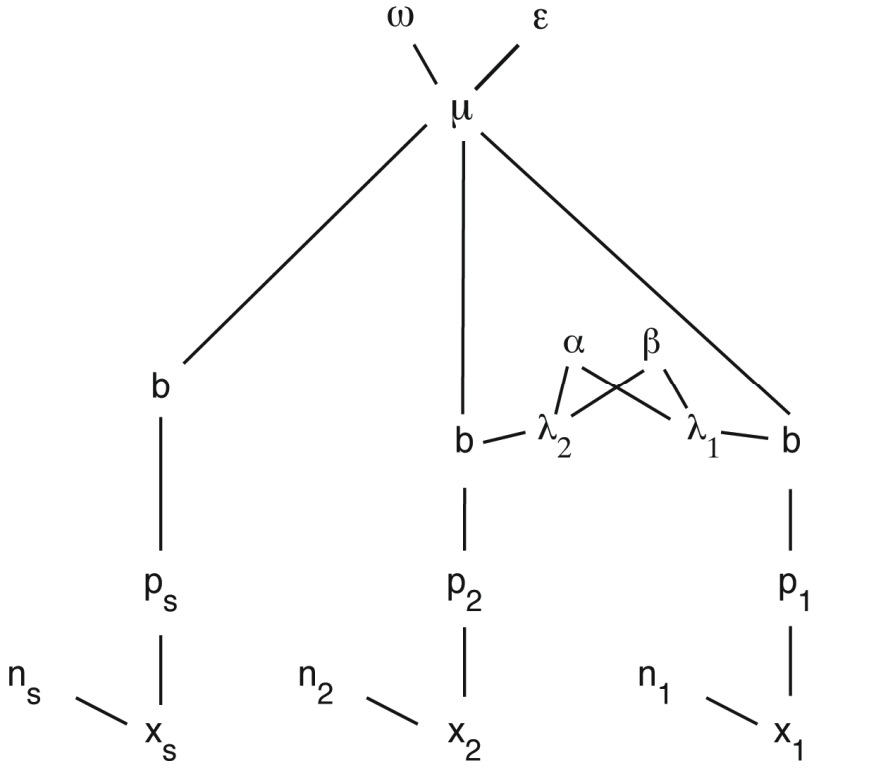
- Previous experience and engineering judgment allows us to put at least vague priors on λ and μ



Topic 1: Multiple Test Regimes

- Consider a more complicated situation where data may come from more than one test regime, e.g. flight testing (s), laboratory tests (1) , accelerated tests (2)

- Since there is some similarity between the tests, the parameter μ is shared across all test regimes.



$$x_i \sim Binomial(n_i, p_i)$$

$$\text{logit}[p_s] = \log\left[\frac{p_s}{1-p_s}\right] = \mu$$

$$\text{logit}[p_i] = \log\left[\frac{p_i}{1-p_i}\right] = \mu + \lambda$$

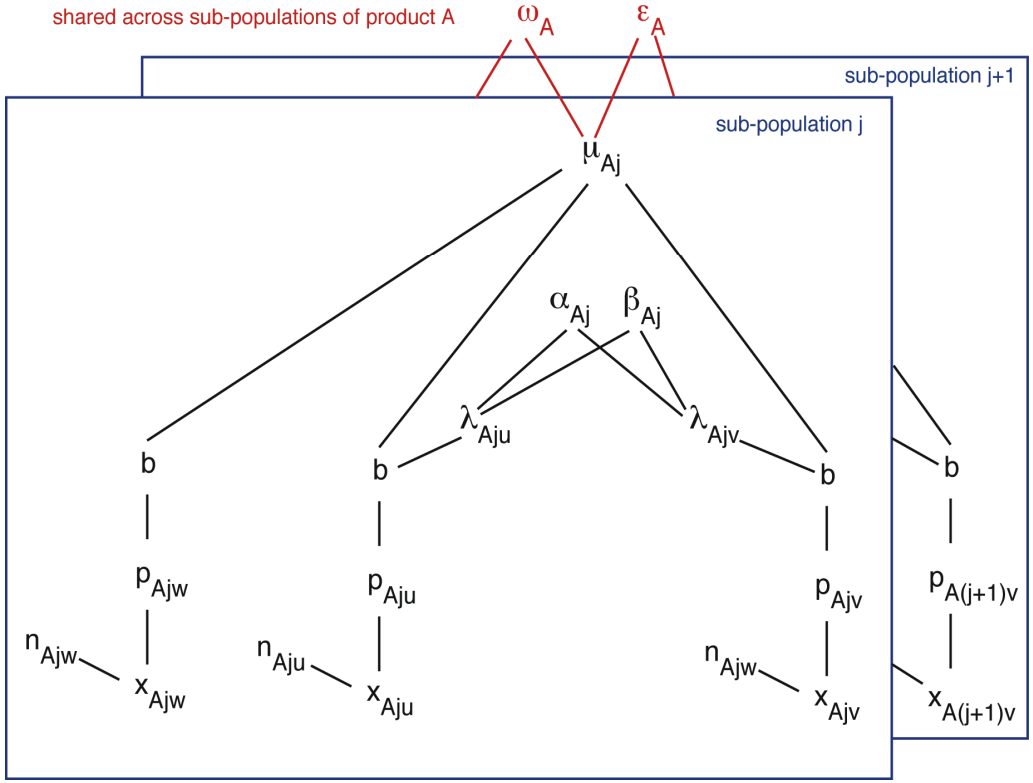
$$\lambda_i \sim Gamma(\alpha, \beta)$$

$$\alpha \sim Gamma(\upsilon, \zeta)$$

$$\beta \sim Gamma(\eta, \xi)$$

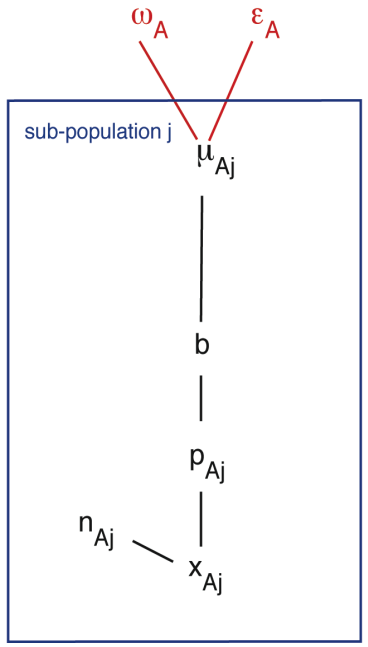
Multiple Sub-populations

- Consider a product where:
  - new supplier becomes available
  - new manufacturing process is introduced
  - unique defect found in segment of population
- All products are similar, but do represent unique sub-populations



Single test regime

shared across sub-populations of product A



Multiple test regimes and multiple sub-populations

Example

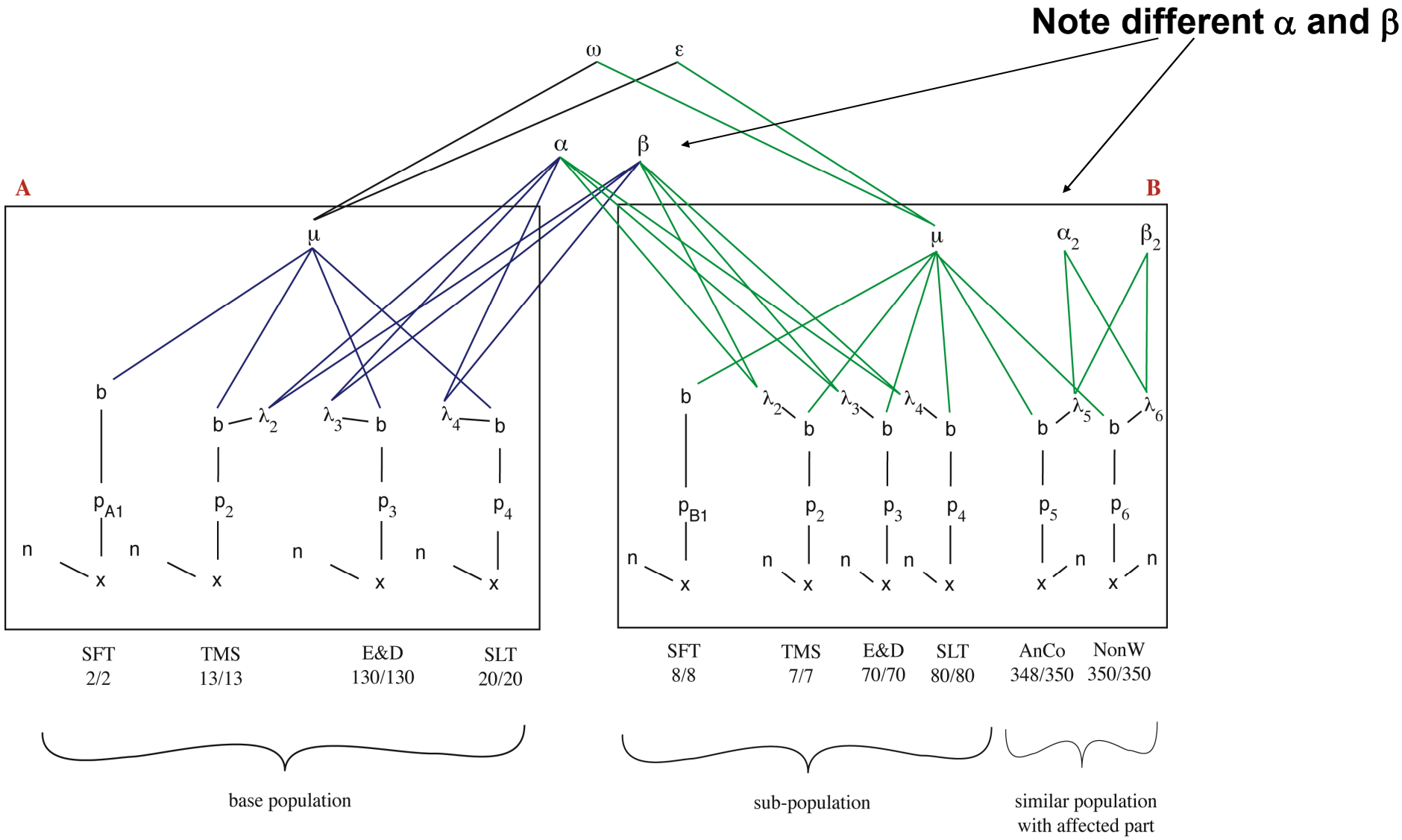
- Data is available from three different test regimes in addition to operational flight testing.

Test Regime	base population (Success/Test)	sub-population (Success/Tests)
Production Tests (TMS)	13/13	7/7
Environmental (E&D)	130/130	70/70
Laboratory Tests (SLT)	20/20	80/80
Flight Tests (SFT)	2/2	8/8

- Data is also available on two products that are similar (but not identical) to the original component.

Data Source	Sub-population (Success/Test)
Similar Product in Similar Application (AnCo)	**348/350**
Similar Product in Different Application (NonW)	350/350

Reliability Digraph Model



Results: Posterior Density Parameters

Note lack of difference in λ between base and sub-population

Base population

Sub-population

Parameter	Mean	Std Dev	MC error	10% LCL	50% (median)	90% UCL
λ_{A2}	0.850852	1.09814	0.00462	0.019353	0.458959	2.19021
λ_{A3}	0.748619	0.942924	0.00408	0.017949	0.415528	1.91268
λ_{A4}	0.836194	1.0741	0.004564	0.019652	0.454081	2.15452
μ_A	9.4713	2.61004	0.004698	6.38112	9.14324	13.0196
p_{A1}	0.99938	0.001456	2.75E-06	0.99831	0.999893	0.999998
p_{A2}	0.998176	0.00631	1.06E-05	0.995837	0.999751	0.999995
p_{A3}	0.998864	0.002363	3.38E-06	0.996786	0.999762	0.999995
p_{A4}	0.998344	0.005032	7.82E-06	0.995979	0.999752	0.999995
λ_{B2}	0.842587	1.06255	0.004603	0.020702	0.461799	2.17675
λ_{B3}	0.726124	0.883682	0.004	0.018166	0.416616	1.84889
λ_{B4}	0.717833	0.86816	0.003911	0.018409	0.415042	1.82368
λ_{B5}	2.34891	2.1123	0.013962	0.100225	1.85668	5.2946
λ_{B6}	0.920242	1.15197	0.005906	0.014833	0.503232	2.3989
μ_B	8.2338	1.90895	0.012371	6.21277	7.82711	10.8297
p_{B1}	0.999239	9.70E-04	4.38E-06	0.998	0.999601	0.99998
p_{B2}	0.997053	0.010612	2.51E-05	0.994572	0.999171	0.999957
p_{B3}	0.998274	0.002933	9.32E-06	0.995771	0.999219	0.999958
p_{B4}	0.998324	0.002731	8.58E-06	0.995865	0.999221	0.999958
p_{B5}	0.99609	0.003252	7.53E-06	0.991822	0.997003	0.999148
p_{B6}	0.998622	0.001517	4.44E-06	0.996724	0.999104	0.999908
α	1.12955	0.847077	0.004617	0.329461	0.897164	2.23849
α_2	1.19643	0.962578	0.003861	0.273203	0.934264	2.46495
β	1.63138	1.13749	0.003678	0.48437	1.36187	3.12657
β_2	1.0357	0.914519	0.002919	0.20982	0.77389	2.19335

Note difference in λ even when they share the same distribution characteristics α_2 and β_2.

Topic 2: Multiple Levels of Indenture

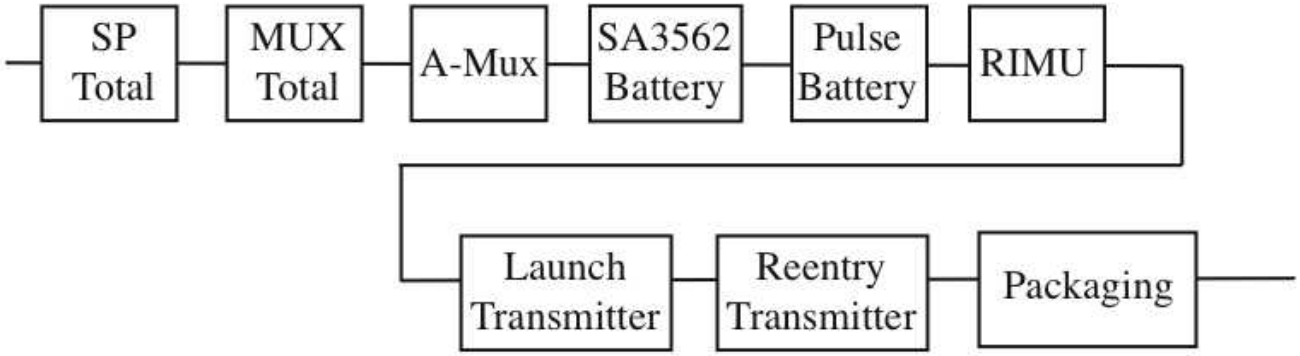
- Aggregation of data from component and system testing is difficult and can lead to problems - care must be taken!

- Using Frequentist approach, no credibility limits can be generated so we don't know the value of additional testing, i.e. how is our confidence/credibility improved by spending money on additional experiments or flight tests.

- Alternatively, hierarchical Bayesian methods provide an objective methodology for aggregating data and provide insight into value of information via credibility interval construction.

Simple Example for Comparison

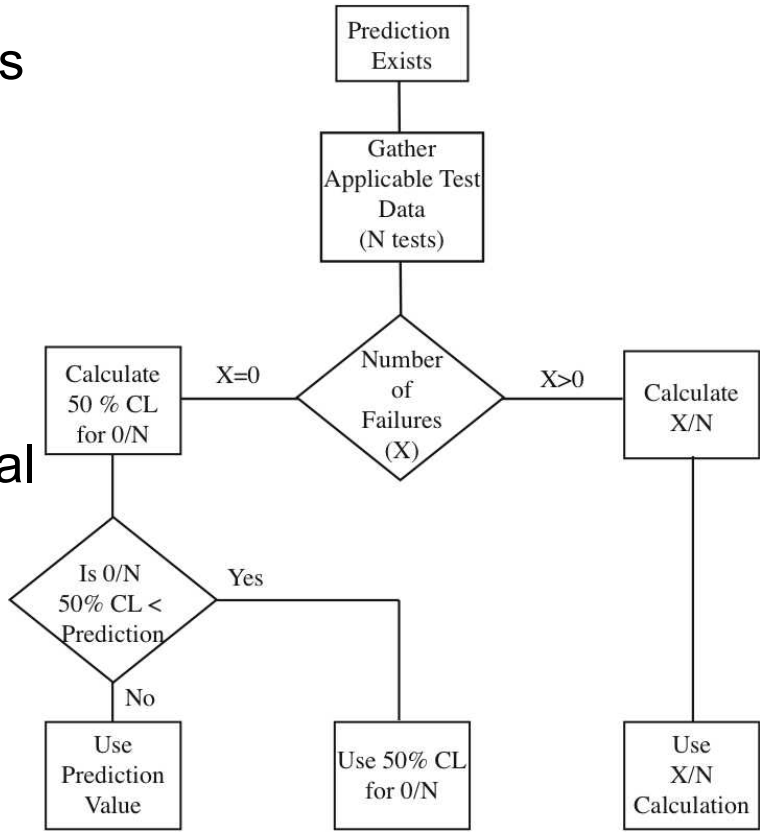
- Consider a system of 9 components, some of which must operate sequentially:



- Scenario:
  - 15 operational flight tests
  - Component failure on flight number 13 prevented complete operation of all 9 components (some components successfully operated until connector failed and prevented other components from being powered on at critical point in flight)
  - Summary: 14 successful flight tests, and one test where only one component was tested (and failed); the other 8 components were never energized

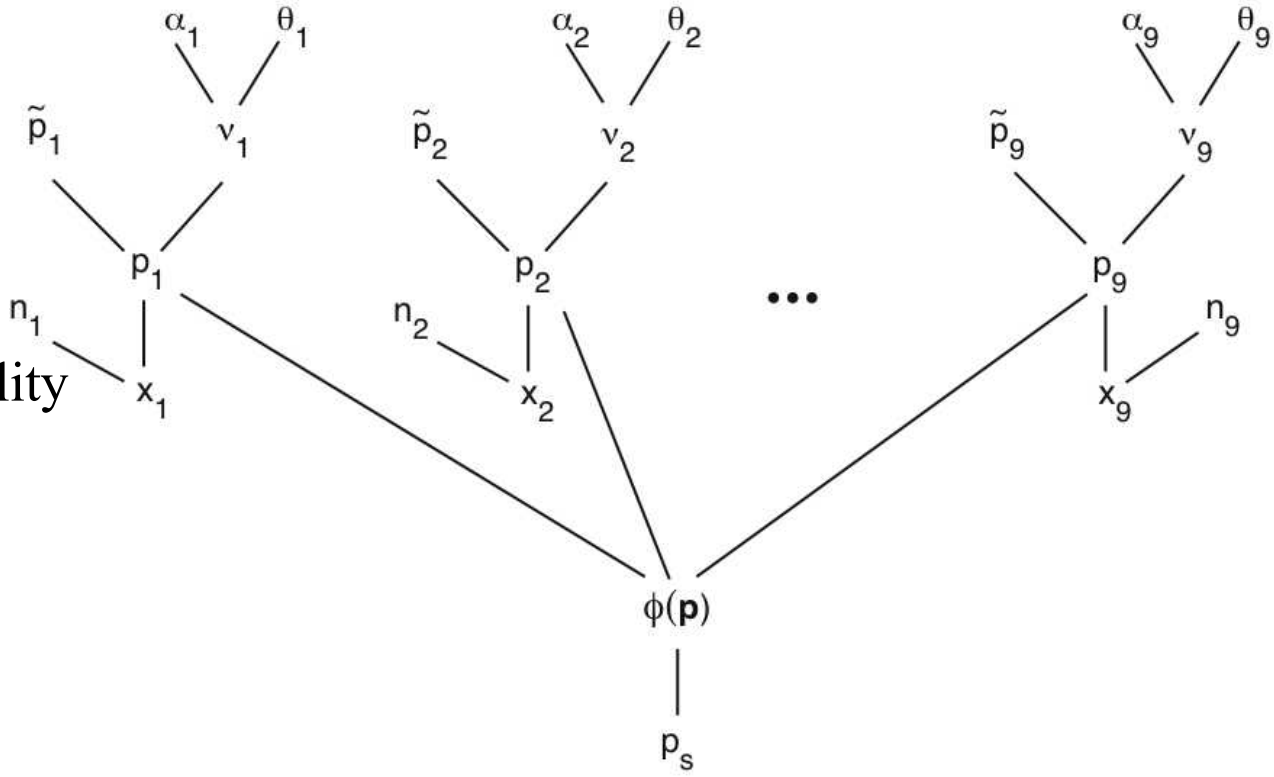
System Reliability Estimate

- Historically, SNL statisticians used the traditional, approved approach:
  - Obtain initial component reliability estimates based on expert opinion and MIL-217
  -
  - After each test (with no failure), estimate the 50th percentile based on 0 failures and N successes
  - Use the lesser of the 50th percentile or initial reliability estimate based on expert opinion
  - When a failure occurs, then use successes/failures to estimate system reliability



Alternative Reliability Estimate

- A hierarchical Bayesian approach follow the traditional Sandia results closely, but Bayesian approach transitions from expert opinion to reliance on data is done gradually, as data is collected, and not at discrete points.



$\tilde{p}$ = initial estimate of reliability

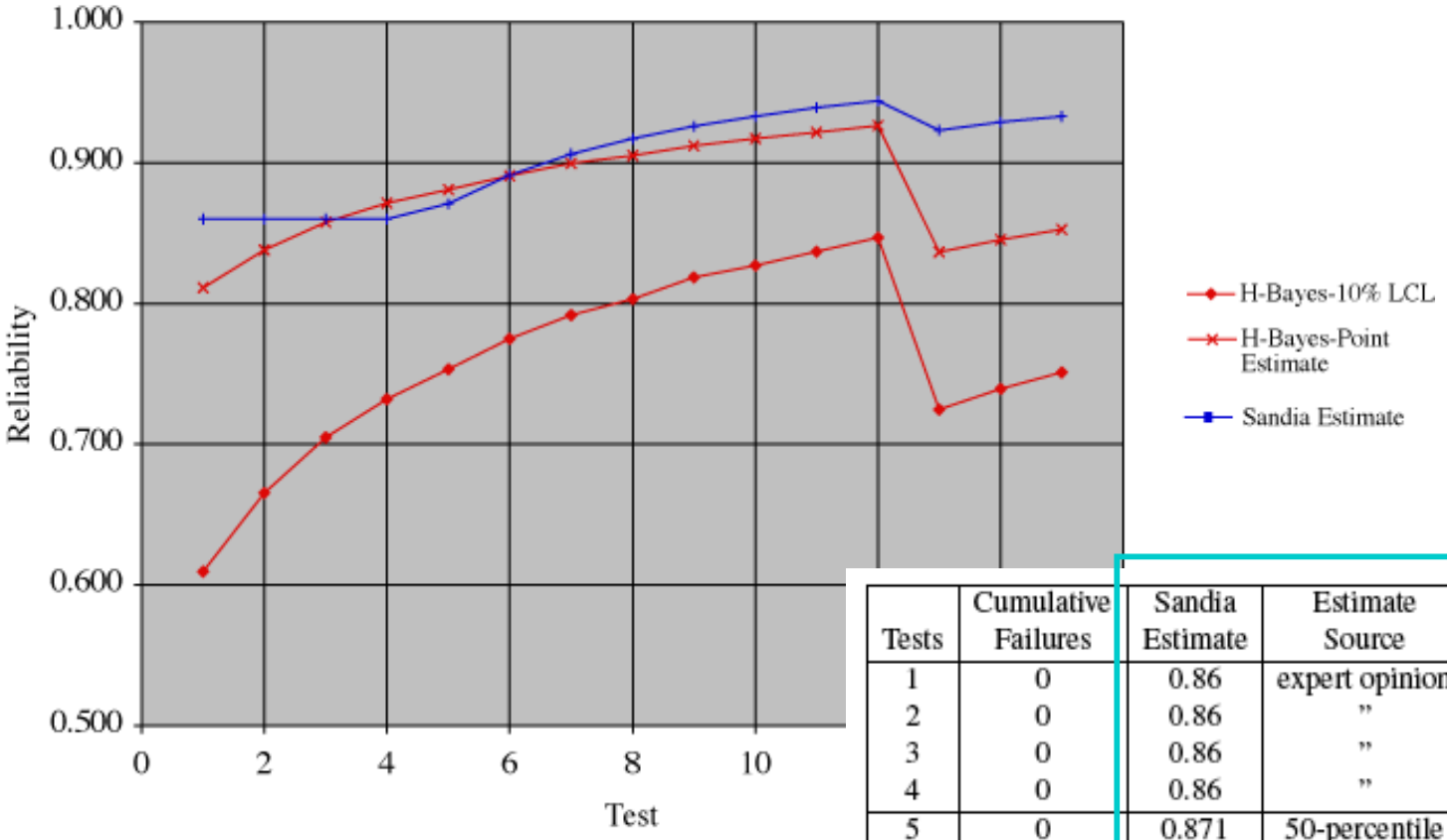
ν= value of initial estimate of reliability $\tilde{p}$

n_i = number of tests for component i.

x_i = number of successes for component i.

$\phi(p) = p_{system} = p_1 \bullet p_2 \bullet p_3 \bullet \ldots \bullet p_9$

Comparison of Results



Tests	Cumulative Failures	Sandia Estimate	Estimate Source	Bayesian Estimate (2)	10% LCL	50% (median)
1	0	0.86	expert opinion	0.8627	0.7975	0.8679
2	0	0.86	"	0.8656	0.8021	0.8704
3	0	0.86	"	0.8684	0.806	0.8731
4	0	0.86	"	0.8708	0.8091	0.8755
5	0	0.871	50-percentile	0.873	0.813	0.8777
6	0	0.891	"	0.8755	0.8159	0.8801
7	0	0.906	"	0.8777	0.8195	0.882
8	0	0.917	"	0.8821	0.8254	0.8865
9	0	0.926	"	0.8821	0.8254	0.8865
10	0	0.933	"	0.8838	0.8281	0.8883
11	0	0.939	"	0.886	0.8315	0.8902
12	0	0.944	"	0.8878	0.8339	0.8922
13	1	0.923	failure data	0.8742	0.8178	0.878
14	1	0.929	"	0.8762	0.8208	0.8805
15	1	0.933	"	0.8779	0.8231	0.882

Summary

- A number of issues related to system reliability assessment were explored:
  - Integration of data from multi-levels of analysis
  - Integration of data from various, possibly disparate sources
- Observations:
  - HB methods provide a well-proven framework for aggregating data from a variety of sources
  - Aggregating component and system level testing has been well demonstrated
  - HB methods account for the value of additional component testing, e.g. 1 success/1 test is not the same as 1000 successes/1000 tests
- Secondary issues can be addressed in a synergistic manner:
  - Decision making related to allocation of resources, in particular the development of optimal sampling plans focused on combining component and system level testing
  - Potential for integration of materials aging model

Questions?

Miracle Mile Fishing Hole