Risk, Failure Probability, and Failure Rate

Terminology

1. **Risk** is the:
   a. Potential of losing or gaining something of value (e.g., life, property, performance, schedule, or cost).
   b. Effect of uncertainty on objectives. (Ref. ISO 31000, 2009)
2. In terms of loss, a **risk statement** contains three elements (e.g., as in three columns in a table), namely:
   a. Scenario, what can go wrong?
   b. Likelihood, what is the probability it will happen?
   c. Consequence, what is the impact if it did happen?
3. **Reliability** is the:
   a. Probability
   b. An item (e.g., system, subsystem) will perform its intended function with no failure
   c. For a stated mission time (or number of demands or load)
   d. Under stated environmental conditions.

Risk vs. Reliability

1. For an item of interest, the probability used in:
   a. Risk is the **probability of failure**, denoted $\boldsymbol{Pf}$. Pf is not a **failure rate** (see page 3).
   b. Reliability is the **probability of success**, denoted $\boldsymbol{Ps}$. Ps is not one minus the failure rate.
2. Fundamental math rule: $\boldsymbol{Pf + Ps = 1}$. $Pf = 1 - Ps$ and $Ps = 1 - Pf$ are the **complements**.
3. When one type of probability is known, use the complement to find the other probability.
4. A **risk matrix** that is quantitative (as opposed to qualitative, labels instead of measures) uses the complement of reliability as the likelihood axis and the complement of safety as the consequence axis.

Types of Data and Methods Commonly Used to Make a Probability of Failure or Failure Rate

1. **Demand-based (pass-fail events)** – an item (e.g., starter solenoid) successfully completed its mission upon demand. The life data for this item answers "how many" and is **discrete** data. The *binomial* probability distribution models this item when the events are independent and the fixed **probability of failure** (p_f) is:
   a. $p_f = \frac{failure\ count}{total\ number\ of\ attempts}$ based on classical statistics.
   b. $p_f = \frac{failure\ count + 0.5}{total\ number\ of\ attempts + 1}$ based on one common version of Bayesian statistics (see page 2).

2. **Time-based (hours, cycles, miles)** – an item (e.g., tire) successfully operated for y hours until it failed. The life data for this item answers "how much" and is **continuous** data. The *exponential* (a special case of the Weibull) probability distribution models this item when the **failure rate** (λ) is constant over time and is:
   a. $\lambda = \frac{failure\ count}{total\ run\ time}$ based on classical statistics.
   b. $\lambda = \frac{failure\ count + 0.5}{total\ run\ time}$ based on one version common of Bayesian statistics (see next page 2).

3. **Failure due to variation** – an item failed not as a function of time but due to **static stress**. That is, the item failed because its variable stress (load) exceeded its variable strength (capacity). The *Stress-Strength Interference* method calculates the **probability of failure** (p_f) which can be associated with the overlap (interference, intersection) in the stress and strength distributions. Note: A **safety factor** or the safety margin are not sufficient to address failures due to the variation in the item's stress and the strength.

170505 Tim.Adams@NASA.gov

Risk, Failure Probability, and Failure Rate

Failure Rate Formulas Based on Bayesian Statistics[1]

Data Type	Demand Based (failure on demand)	Time Based (failure while operating)
Failure Probability and Failure Rate Formulas[2,3]	$p_f = \frac{failure\ count + 0.5}{total\ number\ of\ attempts + 1}$	$\lambda = \frac{failure\ count + 0.5}{total\ run\ time}$
Prior Distribution[4]	Beta distribution with α_{prior} = 0.5 and β_{prior} = 0.5 being a Jeffreys Prior	Gamma distribution with α_{prior} = 0.5 and β_{prior} = 0 being a Jeffreys Prior
Likelihood Function	Binomial distribution	Poisson distribution
Posterior Distribution[5]	Beta distribution with parameters $\alpha_{post} = x + \alpha_{prior}$ and $\beta_{post} = n - x + \beta_{prior}$ where x is failure count and n is number of demands. The mean of the beta distribution is $\frac{\alpha}{\alpha+\beta}$.	Gamma distribution with parameters $\alpha_{post} = x + \alpha_{prior}$ and $\beta_{post} = t + \beta_{prior}$ where x is failure count and t is total run time. The mean of the gamma distribution is $\frac{\alpha}{\beta}$.
NASA PRA Procedures Guide[6]	Page C-6 (pdf page 364)	Page C-11 (pdf page 369)
NASA Handbook on Bayesian Inference[7]	Page 34 (pdf page 54)	Page 40 (pdf page 60)

Footnotes:

[1] **Bayesian statistics** quantitatively combines human belief (a subjectively-based probability distribution) with operational or test data (an objectively-based probability distribution).

[2] When the *failure count is zero*, these two Bayesian-based formulas are commonly used.

[3] When the *failure count is zero* and the data type is time-based, one method in classical statistics calculates the failure rate using: $\lambda = \frac{1/3}{total\ run\ time}$.

[4] A **Jeffreys Prior** is used when there is insufficient information to form an informed prior distribution. Thus, the Jefferys Prior is referred to as a noninformative prior and is intended to convey little prior belief or information. A **noninformative prior** allows the data (described by the likelihood function) to speak for themselves.

[5] A Bayesian-based failure-rate formula is the mean (average) of its posterior distribution. This mean is commonly called the point Bayes' estimate. A **posterior distribution** is derived from Bayes' Theorem (Bayes-Laplace Theorem). This Theorem uses a **prior distribution** (to represent the value of the failure rate as a belief or best estimate prior to collecting field data) and a **likelihood function** (the failure distribution for field data that was collected after the stated belief). The posterior distribution is shifted in the direction of the likelihood function that was used.

[6] Source: https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20120001369.pdf

[7] Source: https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20090023159.pdf

170505 Tim.Adams@NASA.gov

Risk, Failure Probability, and Failure Rate

Illustration: Failure Rate vs. Failure Probability

Question:

- What is the probability of a flood(s) occurring in a ten-year period for an area of land that is classified by the National Flood Insurance Program (NFIP) as being in a 100-year floodplain?

Method 1 – Duration is *continuous* data (i.e., clock time, t, is a non-negative real number):

- Assume the "100-year floodplain" means: The hazard rate or failure rate (λ) is one flood every 100 years, this rate remains constant over time (t), and t is any non-negative real number $\{t \in \mathbb{R} \mid t \geq 0\}$.
- Since $\lambda = 1/100$ is a constant or fixed rate over time, the **exponential** distribution, a continuous probability distribution, can be used as the math model. This model has no memory of previous failures (floods).
- The probability of success or reliability form of the exponential distribution is $R(t) = e^{-(\frac{t}{\theta})}$, where θ is the average or mean time between failure (MTBF) and the reciprocal of λ. Since $\lambda = \frac{1}{100}$, then $\theta = 100$.
- The probability of success (no flood event) during a 10-year period is $R(10) = e^{-(\frac{10}{100})} = 0.904837$.
- The probability of failure (at least one flood event) during a 10-year period is 1 - 0.904837 = 0.095163 **≈ 9.5%**.
- In Excel, the two previous steps can be worked as one using the complement of success space or the cumulative distribution function (failure space): ***=1-EXP(-1/100*10)*** or ***=EXPON.DIST(10,1/100,TRUE)***.
- A related math model is the complement of the cumulative **Poisson**. Let the count of failure events (x) be zero and the mean be the product of time and the failure rate. Use ***=1-POISSON.DIST(0,10*1/100,TRUE)***.

Method 2 – Duration is *discrete* data (i.e., number of successes, x, and trials, n, are non-negative integers):

- Assume the "100-year floodplain" means: There is a probability (p) of one flood every 100 years, this probability is the same from year to year, and year counts (no floods in x years for the duration of n years) are non-negative integers where $x \leq n$. In addition, call this the probability of failure (p_f), the probability of a one flood in one year.
- Since $p_f = 1/100$ is the same each year, each year is independent of one another, the year count is fixed and not infinite, and there are exactly two mutually exclusive outcomes (success and failure) for each year, the binomial distribution can be used to obtain the probability of observing *x* successes in *n* independent trials.
- The probability of success or reliability form of the **binomial** distribution for obtaining exactly x number of successes (no-flood years) in n trials (years) with a given probability of success where $p_s = 1 - p_f$ is:

$$b(x, n, p_s) = \left(\frac{n!}{x!\,(n-x)!}\right)(p_s)^x(1-p_s)^{n-x}$$

- The overall probability of failure being the probability of one or more flood events (years) in 10 trials (years) uses the complement of the above the formula where $x = 10, n = 10, and\ p_s = 0.99$.
- In Excel, the previous step can be worked as ***=1-BINOM.DIST(10,10,0.99,FALSE)*** resulting in 0.095618 **≈ 9.6%**. An alternative method with the binomial distribution in success space is the cumulative form. In this form, let $x = 9$ (for at most 9 flood-free years out of 10 years) being ***=BINOM.DIST(9,10,0.99,TRUE)***.

Comments on the two methods:

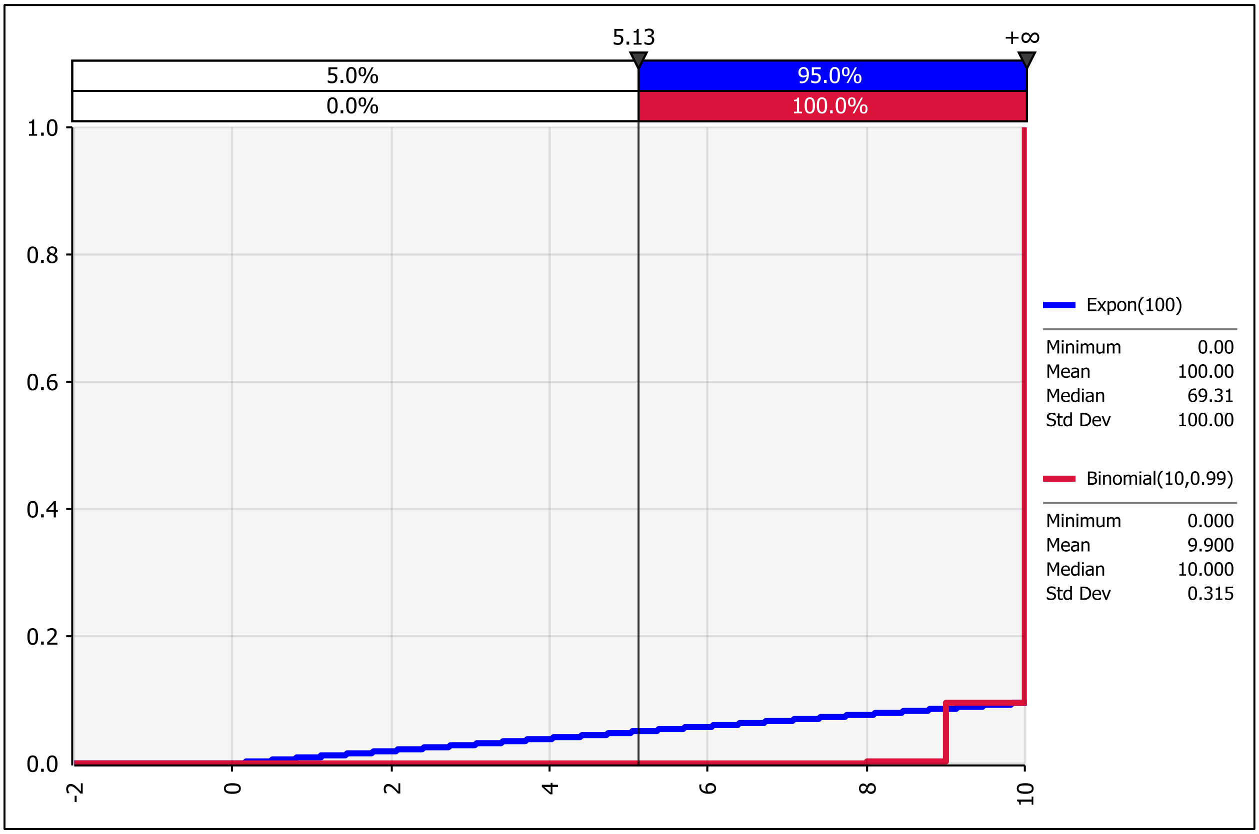
- These methods do not exactly agree since the Poisson and binomial distributions have an asymptotic relationship. The *Poisson distribution approximates the binomial distribution* when n is large and p is small. The exponential distribution is a special case of the Poisson when the number of events in the interval associated with a process equals zero. The next page graphically compares the two above methods.
- Note: The exact case of the binomial in failure space simplifies to $(1 - p_f)^n$ when $x = 0$ (i.e., no failure or flood in every year or trial). In this case, $\boldsymbol{e^{-\lambda t}} \approx \boldsymbol{(1 - p_f)^n}$, when $\lambda = p_f\ and\ t = n$.

170505 Tim.Adams@NASA.gov

Risk, Failure Probability, and Failure Rate

Purpose: Plot the cumulative distribution functions (CDFs) for the exponential distribution (a continuous distribution, with a mean = 100 years, and time truncated at 10 years) and the binomial distribution (a discrete probability distribution using p_s=0.99 with n=10 independent trials or years).

Interpretation: When the exponential's t = 10 and the binomial's n= 10, these two math models intersect at essentially the same value on the vertical axis (0.095) which means: "In a 100-year floodplain, there is a 9.5% probability of failure (at least one flood) in any 10-year period."



170505 Tim.Adams@NASA.gov