Chapter 13
Bayesian Reliability Analysis

Marvin Rausand
marvin.rausand@ntnu.no

RAMS Group
Department of Production and Quality Engineering
NTNU

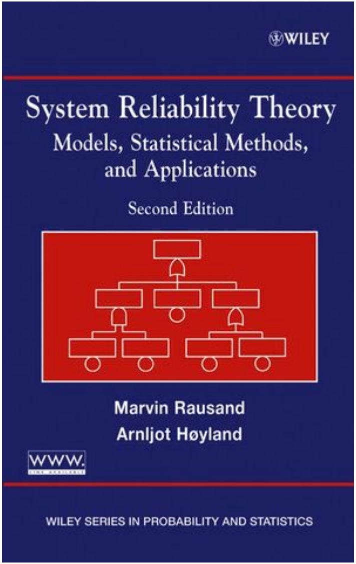
(Version 0.1)

NTNU – Trondheim
Norwegian University of
Science and Technology

Slides related to the book

System Reliability Theory
Models, Statistical Methods,
and Applications

Wiley, 2004



Homepage of the book:
http://www.ntnu.edu/ross/books/srt

Purpose

Frequentist probability - 1

- A *frequentist* is a person who interpret *probability* as the limit of a frequency.
- A *random variable* can be observed and measured. Examples of random variables are time-to-failure (T), and number of failures (N).
- A random variable can be described by a model or *probability distribution* that can contain one or more *parameters*. The distribution of the time-to-failure T may, for example, be exponential with density $f(t) = \lambda e^{-\lambda t}$, where t is an observed value of T and λ is a parameter.
- The parameter (e.g., λ) is considered to have a constant, but unknown value.

Frequentist probability - 2

- The value of the parameter (e.g., λ) cannot be observed or measured. We can only get information about the value of the parameter by performing experiments (i.e., repeated observations of T) and based on this data set estimate the parameter.
- To consider the parameter as a random variable with a distribution has no meaning for a frequentist.

Subjective probability

- The Bayesian approach involves a very different way of thinking about probability compared to the frequentist approach
- The probability of an event or a statement measures a person's *degree of belief* about the event or statement.
- In the Bayesian approach, we can also talk about the probability of a non-observable event or statement, e.g., the probability that God exists.
- Wherever there is uncertainty, this can be described by a probability (distribution)
- This covers both aleatory and epistemic uncertainty

Thomas Bayes

Thomas Bayes (1701-1761) was an English statistician, philosopher, and priest and is acknowledged as the father of *Bayes theorem*. This theorem is fundamental for a special branch of probability theory based on *subjective probability*.

We talk about the *Bayesian approach* to probability and call people who are using this approach *Bayesians*.

REV. T. BAYES

Differences with frequentists and Bayesians

- To a frequentist, data are repeatable, parameters are not:

$$\Pr(\text{Data} \mid \text{Parameters})$$

- To a Bayesian, the parameters are uncertain, the observed data are not

$$\Pr(\text{Parameters} \mid \text{Data})$$

  The data is often referred to as the available *evidence*.

In the Bayesian view, a probability is assigned to a statement/hypothesis, whereas under the frequentist view, a hypothesis is typically tested without being assigned any probability.

Bayes theorem

Consider two events A and B From the definition of conditional probability we know that

$$\Pr(A \cap B_k) = \Pr(A \mid B) \Pr(B) = \Pr(B \mid A) \Pr(A)$$

This expression yields the well-known *Bayes theorem*:

$$\boxed{\Pr(B \mid A) = \frac{\Pr(A \mid B) \Pr(B)}{\Pr(A)}}$$

Bayes theorem is simple – but is still very powerful!

Bayes equation – example

Example – Bayes Theorem

A patient takes a lab test and the result is A (positive). A positive result (A) indicates that the patient has a special type of cancer (B). It is known that the test returns a correct positive result with probability $\Pr(A \mid B) = 0.99$ and a correct negative result with probability $\Pr(A^* \mid B^*) = 0.95$. Furthermore, evidence indicates that 3% of the population in this age group has this type of cancer, such that our prior belief is $\Pr(B) = 0.03$. Our question is now:

- What is the probability that this patient (with a positive test) has the special type of cancer?

According to Bayes theorem

$$\Pr(B \mid A) = \frac{\Pr(A \mid B)\Pr(B)}{\Pr(A)} = \frac{\Pr(A \mid B)\Pr(B)}{\Pr(A \mid B)\Pr(B) + \Pr(A \mid B^*)\Pr(B^*)} = 0.38$$

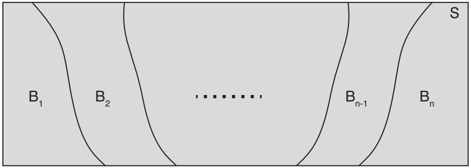
We conclude that a patient with a positive lab test has cancer with 38% probability.

Partition of the sample space

The computation of Bayes theorem requires that we partition the sample space (in B and B^* in the above example).

The set of events $B_1, B_2, \ldots$ is said to be a *partition* of the sample space S when

1. $\Pr\left(\bigcup_{i=1}^{\infty} B_i\right) = 1$, (i.e., the partition covers the whole sample space)
2. $B_i \cap B_j = \emptyset \quad \text{for } i \neq j$, (i.e., the events are non-overlapping)
3. $\Pr(B_i) > 0 \quad \text{for each } i$



Total probability

Let $B_1, B_2, \ldots$ be a *partition* of the sample space S, and consider an event A in S. The probability of A can now be written as

$$\Pr(A) = \sum_{i=1}^{\infty} \Pr(A \cap B_i) = \sum_{i=1}^{\infty} \Pr(A \mid B_i) \Pr(B_i)$$

This formula is called the *Law of total probability*.

Bayes theorem for events

Let A be an event in S, such that $\Pr(A) > 0$ and let $B_1, B_2, \ldots$ be a partition of the sample space.

The conditional probability $\Pr(B_k \mid A)$ is given by Bayes theorem

$$\Pr(B_k \mid A) = \frac{\Pr(A \mid B_k)\Pr(B_k)}{\Pr(A)}$$

By using the law of total probability, we may write Bayes theorem as

$$\boxed{\Pr(B_k \mid A) = \frac{\Pr(A \mid B_k)\Pr(B_k)}{\sum_{i=1}^{\infty} \Pr(A \mid B_i)\Pr(B_i)}}$$

Bayes theorem for distributions

Example – discrete distributions

Two types of products are available. The first type has a specific defect with probability θ_1 and the second type has the same defect with probability θ_2. We receive a sample of products without information about which type it is. We test n products and record X, the number of defects among the n products. The random variable X is binomially distributed with n trials and probability either θ_1 or θ_2 (depending on the type of product). To determine the probability that the product is of type 1, we use Bayes theorem.

$$p(\theta_1 \mid x) = \frac{p(x \mid \theta_1)p(\theta_1)}{p(x)} = \frac{p(x \mid \theta_1)p(\theta_1)}{p(x \mid \theta_1)p(\theta_1) + p(x \mid \theta_2)p(\theta_2)}$$

where $p(\theta_1)$ and $p(\theta_2)$ is the prior probability distribution for the type of product we have got.

Bayes theorem for continuous distributions

A component has time-to-failure T that is exponentially distributed with parameter λ, such that $f_{T|\Lambda}(t \mid \lambda) = \lambda e^{-\lambda t}$.

Our prior belief about the value of the parameter is expressed by assuming that the failure rate is a random variable Λ with probability distribution $f_\Lambda(\lambda)$.

When the time t is observed, the probability probability distribution for Λ is given by Bayes theorem

$$\boxed{f_{\Lambda|T}(\lambda \mid t) = \frac{f_{T|\Lambda}(t \mid \lambda) f_\Lambda(\lambda)}{\int_\Omega f_{T|\Lambda}(t \mid \lambda) f_\Lambda(\lambda)\, d\lambda}}$$

This distribution is called the *posterior distribution* of Λ when the evidence t is given.

Prior distribution

- Our prior belief about the uncertain parameters is a fundamental part of Bayesian probability.
- Our prior belief about an uncertain parameter Θ is expressed by a probability distribution, for example, given by the *prior density* $f_\Theta(\theta)$.[1]
- The prior density may be flat or peaked depending on our prior knowledge about the parameter value.
- When the data available is scarce, our prior belief dominates our inference about the parameter.

[1] The parameter may have a continuous or discrete distribution.

Weak prior information

- If we accept the subjective nature of Bayesian statistics and are not comfortable using subjective priors, then many have argued that we should try to specify prior distributions that represent no prior information.
- These prior distributions are called *non-informative*, or *weak* priors.
- The idea is to have a completely flat prior distribution over all possible values of the parameter.

Informative priors

- An informative prior is an accurate representation of our prior beliefs.
- We are not interested in the prior being part of some conjugate family.
- An informative prior is essential when we have few or no data for the parameter of interest.
- Elicitation, in this context, is the process of translating someone's beliefs into a distribution

Elicitation

- It is unrealistic to expect someone to be able to fully specify their beliefs in terms of a probability distribution.
- Often, they are only able to report a few summaries of the distribution.
- We usually work with medians, modes and percentiles.
- Sometimes they are able to report means and variances, but there are more doubts about these values.
- Once we have some information about their beliefs, we fit some parametric distribution to them.
- These distribution almost never fit the judgements precisely.
- There are nonparametric techniques that can bypass this.
- Feedback is essential in the elicitation process

Likelihood function

The probability density of the variable T given the parameter θ is $f_{T|\Theta}(t \mid \theta)$. In a more general setting, we observe some Data and the conditional probability density of this Data, given the parameter θ is $f_{\text{Data}|\Theta}(\text{Data} \mid \theta)$.

This function tells us how probable it is to obtain some specific Data when we know the value of the parameter θ.

This may be inverted and we may be interested in determining how likely it is that the parameter has the value θ, when the specific Data has been observed. This function is called the *likelihood function* and is written as

$$L(\theta \mid \text{Data}) = f_{\text{Data}|\Theta}(\text{Data} \mid \theta)$$

where the $f(\cdot)$ function is considered as a function of θ.

Bayes theorem

$$f(\theta \mid \text{Data}) = \frac{L(\text{Data} \mid \theta) f(\theta)}{\int L(\text{Data} \mid \theta) f(\theta)\, d\theta}$$

Modern computational methods for Bayesian analysis use simulation to generate a sample from $f(\theta \mid \text{Data})$

When can we use Bayesian methods?

- You have a probability distribution over the states of a variable of interest, Θ that describes your prior beliefs
- You learn something new, for example, that some other random variable T has a particular value t
- You would like to update you beliefs about Θ, to incorporate this new evidence (i.e., that $T = t$).

- $f_\Theta(\theta)$ is the probability distribution over the variable of interest Θ *prior* to the addition of your new observation.
- $f_{\Theta|T}(\theta \mid t)$ is the (conditional) probability distribution of the variable of interest *posterior* to your observation, i.e., after that you have observed that $T = t$.

Conjugate priors

- When we move away from non-informative priors, we might use priors that are in a convenient form.
- That is a form where combining them with the likelihood produces a distribution from the same family.
- In our example, the beta distribution is a conjugate prior for a binomial likelihood.

What do you need to know to use Bayesian methods?

- You need to express your prior beliefs about Θ as a probability distribution $f_{\Theta}(\theta)$
- You must be able to relate your new evidence to your variable of interest in terms of it's likelihood $L(\theta \mid t)$
- You must be able to multiply

Sources of prior information

- Previous experience with the same failure mechanism and test or environment
- Knowledge from physics of failure
- Expert opinion

Bayesian

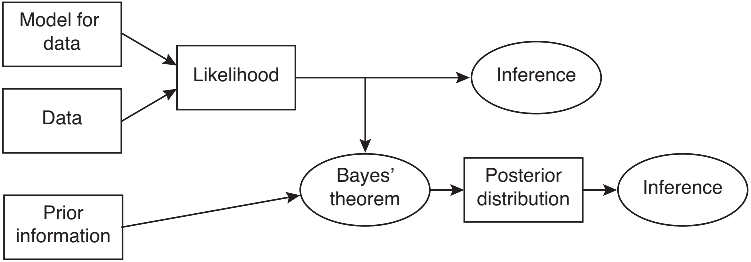


Figure "borrowed" from William Q. Meeker

Computation

- Markov chain Monte Carlo (MCMC) techniques allow us to access our posterior distributions even in complex models

Sensitivity analysis

- It is clear that the elicitation of prior distributions is far from being a precise science.
- A good Bayesian analysis will check that the conclusions are sufficiently robust to changes in the prior.
- If they are not, we need more data or more agreement on the prior structure.

Advantages

- Bayesian statistics offers a framework to deal with all the uncertainty.
- Bayesians make use of more information – not just the data in their particular experiment.
- The Bayesian paradigm is very flexible and it is able to tackle problems that frequentist techniques could not.
- In selecting priors and likelihoods, Bayesians are showing their hands – they can't get away with making arbitrary choices when it comes to inference.
- Can incorporate prior information to supplement limited data, often providing important improvements in precision (or cost savings).
- Can handle,with relative ease,complicated data-model combinations for which no ML software exists (e.g., combinations of random effects and censored data).

Disadvantages

- Bayesian methods are often more complex than frequentist methods
- There is not much software to give scientists off-the-shelf analyses
- Subjectivity: all the inferences are based on somebody's beliefs

Concluding remarks

- Bayesian methods provide a means to incorporate engineering information or other prior information into a formal statistical analysis.
- Incorrect prior information can lead to seriously misleading inferences. Caution is required.
- Use of Bayesian methods in reliability data analysis will certainly increase in the future.