The Bayesian Paradigm for Quantifying Uncertainty

- A Tutorial*

By

Nozer D. Singpurwalla,

City University of Hong Kong

*Abstracted from "Reliability and Risk: A Bayesian Perspective" by Nozer D. Singpurwalla, Wiley (UK), 2006. ***Copyrighted***.

1. What is the Essence of the Bayesian Paradigm?

- It is that the only satisfactory way to quantify uncertainty is by probability, and
- That probability is personal to an individual or a group of individuals acting as a team.

2. What is Uncertainty and Why Quantify it?

- Uncertainty is anything that <u>you</u> don't know.

- Thus, like probability, uncertainty is also personal, because your uncertainty could be sure knowledge to another.

- Furthermore, uncertainty is <u>time</u> <u>indexed</u>, because what is uncertain to you now can become known to you later.

- Thus probability should carry two indices, you, denoted by , and time, denoted by τ.

- We quantify uncertainty to invoke the scientific method, and the scientific method mandates measurement.
- Per Lord Kelvin, if you cannot measure it, you cannot talk about it.

3. Notation and Symbols.

- Let X_1 denote an uncertain quantity to , at time τ.

- For example, X_1 could denote the ***failure time*** of a structure, or the ***maximum stress*** experience by the structure over its service life, the tomorrow's ***closing price of a stock***. In other words, X_1 is simply a label.

- Let x_1 denote the possible numerical values that X_1 can take. For example $x_1 = 20$ years, or $x_1 = 15$ lbs/square inch, or $x_1 =$ \$27.82, etc. Thus x_1 *is* ***generic***.

- When X_1 denotes an ***uncertain event*** , like rain or shine tomorrow, or failure or survival by the year's end, or pass or fail, or stock appreciates or depreciates, then X_1 takes only two values $x_1 = 1$ or $x_1 = 0$. Thus $(X_1 = 1)$ will denote the event that it rains tomorrow or the item survives to the year's end.

- In what follows, we will focus on events $(X_1 = x_1)$, where $x_1 = 1$ or 0.

- Our aim is to quantify 's uncertainty about the event $(X_1 = x_1)$ at time τ, using the metric of ***probability***.

- To do so, we need to exploit the background information, or ***history*** H, that has about $(X_1 = x_1)$ at time τ; denote this as H(τ). Bear in mind that H will change with τ, because as time passes on, is liable to know more about $(X_1 = x_1)$ but <u>not</u> for sure if $(X_1 = 1)$ or $(X_1 = 0)$.

- With the above in place, 🚶's uncertainty about the event $(X_1 = x_1)$ at time τ, in the light of H(τ), as quantified by probability, is denoted

$$P^{\tau}_{🚶}(X_1 = x_1; \mathsf{H}(\tau)).$$

- Furthermore, $P^{\tau}_{🚶}(X_1 = x_1; \mathsf{H}(\tau))$ is a number taking all values between 0 and 1(both <u>exclusive</u>, under a personalistic interpretation of probability).

4. Interpretations of Probability.

- ***Relative Frequency***: An objectivistic view according to which probability is the limit of the ratio of number of times that $(X_1 = x_1)$ will occur when the number of possible occurrences of $(X_1 = 1)$ or $(X_1 = 0)$ is infinite, under almost identical circumstances of occurrence.

- Under this interpretation, , H, and τ, do not matter so that

$$P^{\tau}(X_1 = x_1; \mathrm{H}(\tau)) = P(X_1 = x_1),$$

and $P(X_1 = x_1)$ can be assessed only under repeated observation of the event; this view demands hard data on $(X_1 = x_1)$; furthermore, $P(X_1 = x_1)$ is unique.

- The relative frequency notion of probability underlies the frequentist (or sample theoretic) approach to statistical inference with its long run behavior notions of unbiased estimation, Type I & II Errors, Significance Tests, Minimum Variance, Maximum Likelihood, Confidence Limits, Chi-Square and t-Tests, etc.
- This is the approach advocated by Fisher and by Neyman (though unlike Lehman, Neyman was not hostile to the Bayesian argument).

- Bayesian statistical inference <u>rejects</u> the above notions as being irrelvant.

- ***Personalistic Interpretation***: $P^{\tau}(X_1 = x_1; \mathrm{H}(\tau))$ is the amount is prepared to stake, at time τ, in exchange for 1 unit, should $(X_1 = x_1)$ occur, in a 2-sided bet. If $X_1 = x_1$ does not occur (in the future), loses the amount staked. This interpretation assumes a linear utility, (risk aversion) by .

- Here probability is a gamble, and the 2-sided bet ensures that ’s declared probability is a reflection of his(her) true uncertainty. That is, the 2-sided bet ensures honesty, because:

- In a 2-sided bet, if stakes p_1 for the future occurrence of $(X_1 = x_1)$, then should also be prepared to stake $(1 - p_1)$ for $(X_1 = x_1)$ not occurring, and ’s boss gets to choose the side of the bet.

- Under the personalistic (or subjective) interpretation, probability is <u>not</u> <u>unique</u>, it is dynamic with τ, and cannot take the values 0 and 1, i.e. $0 < p_1 < 1$.
- The role played by utility in a 2-sided bet leads one to the claim that personal probability cannot be separated from ‘s utility.

5. The Rules (or Axioms) of Probability.

- Irrespective of how one interprets probability, the following rules are adhered to.
- The rules tell us how to combine several uncertainties (i.e. how the uncertainties ***cohere***).

- Consider two uncertain events at time τ, say $(X_1 = x_1)$ and $(X_2 = x_2)$, $x_i = 1$ or 0, $i = 1, 2$, and an individual  with history $\mathrm{H}(\tau)$. Then:

- i) ***Convexity***:

$$P^{\tau}(X_i = x_i;\, \mathrm{H}(\tau)) = p_i,$$

with $0 < p_i < 1$.

- ii) ***Addition***:

$$P^{\tau}(X_1 = x_1 \ \underline{\mathrm{or}}\ X_2 = x_2;\, \mathrm{H}(\tau)) = p_1 + p_2$$

but <u>only</u> when $X_1 = x_1$ and $X_2 = x_2$ are mutually exclusive.

- iii) ***Multiplication***:

$$P^{\tau}(X_1 = x_1 \text{ and } X_2 = x_2;\ \mathrm{H}(\tau))$$
$$= P^{\tau}(X_1 = x_1 \mid X_2 = x_2;\ \mathrm{H}(\tau)) \cdot P^{\tau}(X_2 = x_2;\ \mathrm{H}(\tau)).$$

- The middle term is called the conditional probability of the event $(X_1 = x_1)$ supposing that $(X_2 = x_2)$ were to be true.
- It is very important to note that conditional probabilities are in the ***subjunctive mood***.

6. Operationalizing Conditional Probability.

- In the relative frequency theory, conditional probability is a definition; it is the ratio of two probabilities. Thus

$$P(X_1 = x_1 \mid X_2 = x_2) = \frac{P(X_1 = x_1 \text{ and } X_2 = x_2)}{P(X_2 = x_2)},$$

- if $P(X2 = x2) \neq 0$.

- In the personalistic theory, if:
  $P^{\tau}(X_1 = x_1 \mid X_2 = x_2; \mathrm{H}(\tau)) = \pi$, say, $0 < \pi < 1$, then π is the amount staked by at time τ, in the light of $\mathrm{H}(\tau)$, on event $(X_1 = x_1)$ in a 2- sided bet, but under the stipulation that the bet will be settled only if $(X_2 = x_2)$ turns out to be true.
- All bets are off if $(X_2 = x_2)$ does not turn out to be true.

- Note that at time τ, the disposition of both X_1 and X_2 is not known to . Thus it is the subjunctive mood that is germane to conditional probability.

- Important Convention:

- All quantities known to at time τ with certainty, are written after the semi-colon; e.g. H(τ). All quantities unknown to at time τ, but conjectured by at τ, like $(X_2 = x_2)$ are written after the vertical slash. Thus we have:

$$P^{\tau}(X_1 = x_1 | X_2 = x_2; \mathrm{H}(\tau)).$$

7. Independence, Dependence, & Causality.

- $(X_1 = x_1)$ and $(X_2 = x_2)$ are said to be independent events if

$$P^{\tau}(X_1 = x_1 \mid X_2 = x_2; \mathrm{H}(\tau)) = P^{\tau}(X_1 = x_1; \mathrm{H}(\tau)),$$

for all values $x_{1,}\ x_2$; or else, they are dependent.

- Thus independence means that your disposition to bet on say $(X_1 = x_1)$ will not change under the (supposed) added knowledge of the disposition of $(X_2 = x_2)$.
- Consequently, mutually exclusive events are necessarily dependent.

- Since $(X_1 = x_1)$ independent of $(X_2 = x_2)$ implies that $(X_2 = x_2)$ is independent of $(X_1 = x_1)$, and $(X_1 = x_1)$ dependent of $(X_2 = x_2)$ implies that $(X_2 = x_2)$ is dependent of $(X_1 = x_1)$, the notion of dependence does not encapsulate causality.

- The notion of causality involves a time ordering in the occurrence of $(X_1 = x_1)$ and $(X_2 = x_2)$, if any, whereas the notions of independence and dependence refer to the disposition of 's mind towards bets on $(X_1 = x_1)$ and $(X_2 = x_2)$ at time τ, irrespective of how and when X_1 and X_2 reveal themselves.

- To summarize, $(X_1 = x_1)$ dependent of $(X_2 = x_2)$ does <u>not</u> imply that $(X_2 = x_2)$ causes $(X_1 = x_1)$ or that it does not cause $(X_1 = x_1)$.

- <u>Note</u>: The notions of independence and dependence reflect the <u>judgment</u> of  at τ.
- Whereas a causal relationship between the two events in question may lead to the judgment of dependence, an absence of causality between two events does <u>not</u> necessarily imply independence of the events.

8. Generalizing the Rules of Probability.

- For convenience, we skip writing , τ, and H(τ) but recognize their presence (in the personalistic context).

- Then for k uncertain events $(X_1 = x_1)$, … , $(X_k = x_k)$,

- i) $P(X_1 = x_1 \text{ or } X_2 = x_2 \text{ or } \ldots \text{ or } X_k = x_k)$
$$= \sum_{i=1}^{k} P(X_i = x_i),$$
if all the k events are mutually exclusive, and

- ii) $P(X_1 = x_1 \text{ and } X_2 = x_2 \text{ and } \ldots \text{ and } X_k = x_k)$
$$= P(X_1 = x_1 \mid X_2 = x_2 \text{ and } \ldots \text{ and } X_k = x_k)$$
$$\cdot P(X_2 = x_2 \mid X_3 = x_3 \text{ and } \ldots \text{ and } X_k = x_k) \cdot \ldots \cdot$$
$$\cdot P(X_{k\text{-}1} = x_{k\text{-}1} \mid X_k = x_k) \cdot P(X_k = x_k)$$
$$= \prod_{i=1}^{k} P(X_i = x_i),$$
if all the k events are judged independent.

- If the events $(X_1 = x_1)$ and $(X_2 = x_2)$ are <u>not</u> mutually exclusive, then i) above leads us to the result that

$$P(X_1 = x_1 \text{ or } X_2 = x_2)$$

$$= P(X_1 = x_1) + P(X_2 = x_2) - P(X_1 = x_1 \text{ and } X_2 = x_2)$$

$$= P(X_1 = x_1) + P(X_2 = x_2) - P(X_1 = x_1 | X_2 = x_2) \cdot P(X_2 = x_2)$$

$$= P(X_1 = x_1) + P(X_2 = x_2) - P(X_1 = x_1) \cdot P(X_2 = x_2),$$

if X_1 and X_2 <u>also</u> happen to be independent (in addition to being not mutually exclusive).

9. Why these Rules?

- There are two arguments, one pragmatic, the other mathematical/logical, which lead to the conclusion that not following these rules leads to ***incoherence*** (i.e. a sure loss no matter what the outcome; e.g heads I win, tails you lose).

- i) The first argument is based on ***scoring rules*** and is due to de Finetti and generalized by Lindley.

- ii) The second argument is based on certain axioms of "rational behavior", called ***behavioristic axioms***, and is due to Ramsey and Savage.

- To Kolmogorov, the axioms of probability are a given (like commandments) and are the starting point for the theory of probability.
- Cardano – the Italian polymath – discovered the rules (axioms) of probability as a way to gamble without a sure loss.
- Some psychologists, like Khaneman and Tversky, and some economists like Allais and Ellsberg, claim individuals do not like to be scored, nor do they behave according to the axioms of rational behavior, and thus cast pallor on the axioms of probability.

- The above argument has opened the door to alternatives to probability, like ***possibility theory***, ***upper and lower probabilities***, and ***fuzzy logic***.

- Lindley and Savage have rejected such alternatives to probability on grounds that the behavioristic axioms underlying the axioms of probability are normative. They prescribe rational behavior, just like how the Paeno Axioms prescribe the rules of arithmetic.

10. Extending the Rules of Probability.

- Some simple manipulations of the convexity, the addition, and the multiplication rules enable us to derive two new and very important consequences of the above rules. These are:

- i) ***The Law of Total Probability*** (or ***Extension of Conversation***) – due to La Place:

$$P(X_1 = x_1) = P(X_1 = x_1 \mid X_2 = 0) \cdot P(X_2 = 0) + P(X_1 = x_1 \mid X_2 = 1) \cdot P(X_2 = 1)$$

$$= \sum_{i=1}^{2} P(X_1 = x_1 \mid X_2 = x_i) \cdot P(X_2 = x_i).$$

- Here, assessing the uncertainty about $(X_1 = x_1)$ is facilitated by contemplating the dispositions of X_2.

- ii) ***Bayes' Law*** (or the ***Law of Inverse Probability***) – due to Bayes and La Place:

$$P(X_1 = x_1 \mid X_2 = x_2) = \frac{P(X_2 = x_2 \mid X_1 = x_1)P(X_1 = x_1)}{P(X_2 = x_2)}$$

$$= \frac{P(X_2 = x_2 \mid X_1 = x_1)P(X_1 = x_1)}{\sum_{i=1}^{2} P(X_2 = x_2 \mid X_1 = x_i)P(X_1 = x_i)}$$

(by the Law of Total Probability), so that

Posterior Probability

Prior Probability

- $$P(X_1 = x_1 \mid X_2 = x_2) \propto P(X_2 = x_2 \mid X_1 = x_1) P(X_1 = x_1)$$

Conditional Probability

since the role of the denominator is to simply ensure that the left hand side is a probability.

- Note: Bayes' Law being a part of the theory of probability, only deals with uncertain events, or contemplated conditioning events.
- Observe the inversion of arguments in the posterior and the conditional probabilities.

11. Bayes' Law and Observed Data – Notion of Likelihood.

- Recall, Bayes' Law:

$$P(X_1 = x_1 \mid X_2 = x_2) \propto P(X_2 = x_2 \mid X_1 = x_1) P(X_1 = x_1).$$

- Suppose now that $(X_2 = x_2)$ has actually been observed by , i.e. it is no more contemplated.

- Then, the left hand side should be written as $P(X_1 = x_1; X_2 = x_2)$ because now $(X_2 = x_2)$ has become a part of 's history at τ, namely H(τ).

- The term $P(X_2 = x_2 \mid X_1 = x_1)$ is no more a probability – because probability is germane only for the unknowns.
- The above expression is therefore written as L$(X_1 = x_1; X_2 = x_2)$, and it is now called the ***likelihood*** of $(X_1 = x_1)$, in the light of the actually observed $(X_2 = x_2)$.

12. More on the Likelihood.

- It is <u>not</u> a probability and therefore need not obey the rules of probability.

- It is a <u>relative</u> <u>weight</u> which assigns to the unknown events $(X_1 = 1)$ and $(X_1 = 0)$ in the light of the observed $(X_2 = x_2)$.

- It is generally assigned by , upon flipping the arguments in $P(X_2 = x_2 \mid X_1 = x_1)$. Thus the often expressed view that the likelihood is a probability.

13. Bayes' Law with Likelihood.

Posterior

Prior

- $$P^{\tau}(X_1 = x_1; X_2 = x_2, \mathrm{H}(\tau))$$
$$\propto \mathrm{L}^{\tau}(X_1 = x_1; X_2 = x_2, \mathrm{H}(\tau)) \cdot P^{\tau}(X_1 = x_1; \mathrm{H}(\tau)).$$

Likelihood

- This law prescribes a mathematical process by which changes his(her) mind in the light of new information (data).
- However, in doing so we encounter caveats.

14. What is a Probability Model?

- Where do specifications such as the Bernoulli, the exponential, the Weibull, the Gaussian, the bivariate exponential, etc. come from?

- Consider $P(X_1 = x_1)$, and invoke the law of total probability by extending the conversation to some unknown (perhaps unobservable) quantity, say θ, where $0 < \theta < 1$. Then

- $$P(X_1 = x_1) = \int_0^1 P(X_1 = x_1 \mid \theta) P(\theta) \mathrm{d}\theta$$

replacing the summation by the integral, since θ is assumed continuous.

- $P(X_1 = x_1|\theta)$ is called a ***probability model*** for X_1;

- if $P(X_1 = 1|\theta) = \theta \Leftrightarrow P(X_1 = 0|\theta) = 1 - \theta$, then the probability model is called a ***Bernoulli Model***.

- Thus to summarize, under a Bernoulli Model

$$
\begin{aligned}
P^{\tau}(X_1 = 1; \mathrm{H}(\tau)) &= \int_0^1 P^{\tau}(X_1 = 1 \mid \theta; \mathrm{H}(\tau)) \cdot P^{\tau}(\theta; \mathrm{H}(\tau))\,\mathrm{d}\theta \\
&= \int_0^1 \theta \cdot P^{\tau}(\theta; \mathrm{H}(\tau))\,\mathrm{d}\theta.
\end{aligned}
$$

Predictive of X

Prior on θ

- The essence of the above is that under a Bernoulli Model, were we to know θ, then $P(X_1 = 1|\theta) = \theta$, but since we know θ only probabilistically, we average over all the values of θ to obtain $P(X_1 = 1)$, which is now devoid of θ.

15. Meaning of θ: To…

- de Finetti– θ is just a Greek symbol which makes $(X_1 = 1)$ independent of H(τ).

- Popper - θ is a chance or a propensity (i.e. a tendency for $X_1 = 1$).
- For *induction* under the Bernoulli model go to slide 54.
- For *hypotheses testing* go to slide 66.

16. The Exponential & Weibull Models.

- Let T denote an unknown (at time τ) time to failure of a structure, with T taking a value t, for some $t \geq 0$.

- Let H(τ) denote the background knowledge possessed by  about the structure at τ.

-  needs to assess the survivability of T, for a mission time, $t^* > 0$. Thus, we need

$$P^{\tau}(T > t^*; \mathrm{H}(\tau)) = \int_0^{\infty} P^{\tau}(T > t^* \mid \lambda; \mathrm{H}(\tau)) \cdot P^{\tau}(\lambda; \mathrm{H}(\tau)) \mathrm{d}\lambda$$

by extending the conversation to $\lambda > 0$.

- Suppose that chooses an ***exponential distribution*** as a probability model for T. Then for ,
- $$P^{\tau}(T > t^{*} \mid \lambda; \mathsf{H}(\tau)) = \exp(-\lambda t^{*})$$

and now

$$P^{\tau}(T > t^{*}; \mathsf{H}(\tau)) = \int_{0}^{\infty} \mathrm{e}^{-\lambda t^{*}} \cdot P^{\tau}(\lambda; \mathsf{H}(\tau))\,\mathrm{d}\lambda.$$

Predictive of T ***Exponential Failure Model*** ***Prior on λ***

- If  prefers to choose a Weibull with shape $\beta > 0$ and scale 1 as a probability model for T, then

$$P^{\tau}(T > t^{*}; \mathrm{H}(\tau)) = \int_{0}^{\infty} \mathrm{e}^{-(t^{*})^{\beta}} \cdot P^{\tau}(\beta; \mathrm{H}(\tau))\,\mathrm{d}\beta.$$

$$\underset{\textit{Predictive of } T}{\uparrow} \qquad \underset{\textit{Weibull Failure Model}}{\uparrow} \qquad \underset{\textit{Prior on } \beta}{\uparrow}$$

- In either case, the predictive distribution entails integration for which either numerical methods or MCMC is of use.

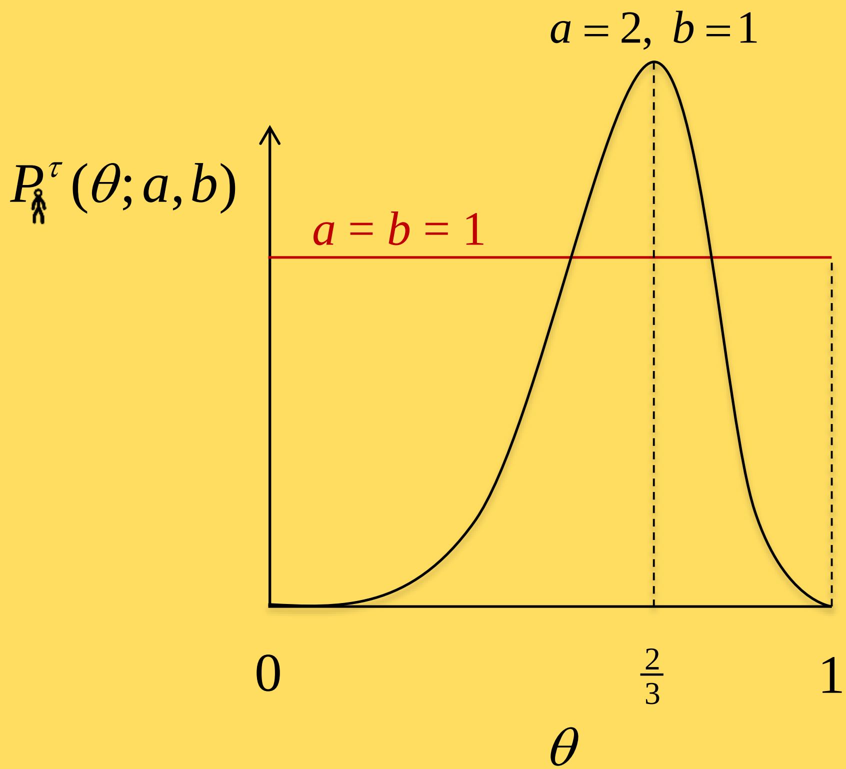
17. Choice of a Prior on Chance (Propensity) θ.

- The simplest possibility is a uniform on $(0, 1)$; that is

$$P^{\tau}(\theta; \mathrm{H}(\tau)) = 1, \quad 0 < \theta < 1.$$

- If the propensity of $(X_1 = 1)$ is higher than that of $(X_1 = 0)$, a beta with parameters (a, b) makes sense:
- $$P^{\tau}(\theta; \mathrm{H}(\tau)) = P^{\tau}(\theta; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1}(1-\theta)^{b-1},$$

where $\Gamma(x) = \int_0^1 \mathrm{e}^{-u} u^{x-1} \mathrm{d}u = (x-1)!.$

- 

- As a gets bigger than b, the mode shifts to the right and vice-versa.

18. Model & Predictive Failure Rates.

- Consider the exponential failure model

$$P(T > t^* \mid \lambda) = \exp(-\lambda t^*).$$

- Then it can be seen that the conditional probability

$$P(t^* < T < t^* + \mathrm{d}t^* \mid T > t^*, \lambda) = \frac{-\frac{\mathrm{d}}{\mathrm{d}t^*} P(T > t^* \mid \lambda)}{P(T > t^* \mid \lambda)} = \lambda.$$

Model Failure Rate of T

- The quantity

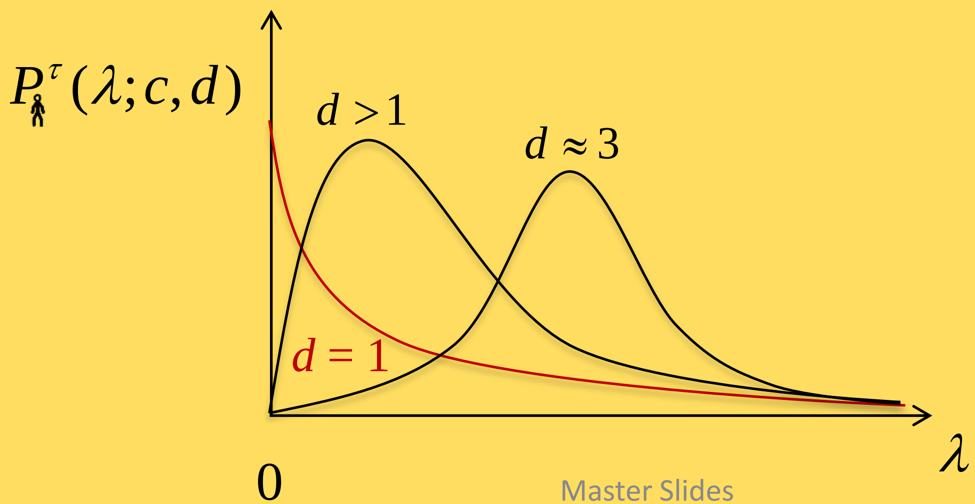
$$P(t^* < T < t^* + \mathrm{d}t^* \mid T > t^*; \mathrm{H}(\tau))$$

$$= -\frac{\frac{\mathrm{d}}{\mathrm{d}t^*} P(t^* < T < t^* + \mathrm{d}t^* \mid T > t^*; \mathrm{H}(\tau))}{P(t^* < T < t^* + \mathrm{d}t^* \mid T > t^*; \mathrm{H}(\tau))}$$

is called the ***predictive failure rate*** of T.

19. Prior on Exponential Model Failure Rate λ.

- Since $\lambda > 0$, a meaningful prior on λ is a gamma distribution with parameters (***scale***) c and (***shape***) d; that is
- $$P^{\tau}(\lambda; \mathrm{H}(\tau)) = P^{\tau}(\lambda; c, d) = \frac{e^{-\lambda c}(\lambda c)^{d-1} c}{(d-1)!}.$$



- The ***mean time to failure*** is $1/\lambda$, denoted MTTF.

- Note: MTTF $=$ (Model Failure Rate)$^{-1}$

  but only for the exponential failure model.

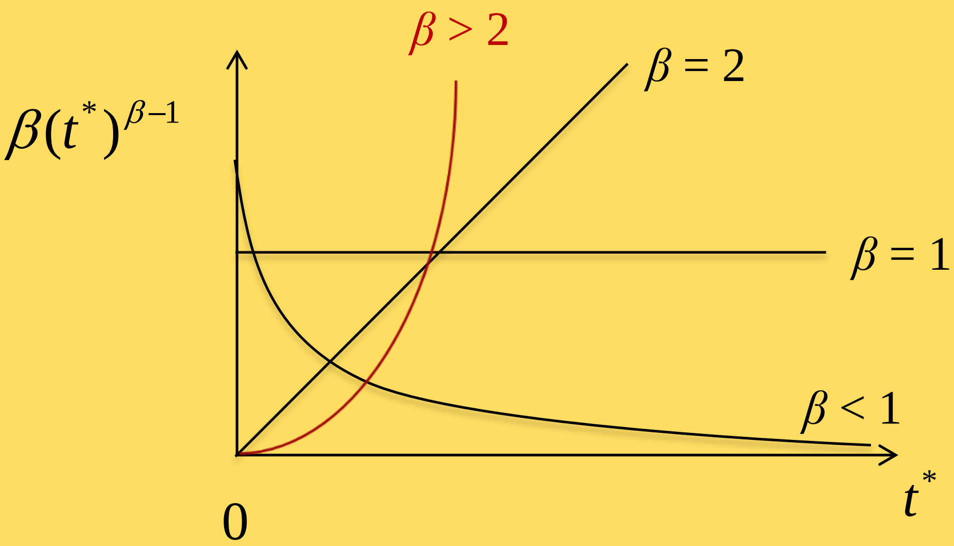
20. Model Failure Rate of the Weibull Failure Model.

- When $P^{\tau}(T > t^{*} \mid \beta) = \mathrm{e}^{-(t^{*})^{\beta}}$,

$$P(t^{*} < T < t^{*} + \mathrm{d}t^{*} \mid T > t^{*}, \beta) = \beta (t^{*})^{\beta - 1}$$

is the model failure rate.

- Depending on the choice of β it encapsulates aging ($\beta > 1$), non-aging ($\beta = 1$), or things like improvement with age ($\beta < 1$) – work hardening.



- Since $\beta > 0$, the gamma distribution would be a suitable prior for β.

21. Predictive Probabilities Under Bernoulli Models.

- Let $(X_i = 1)$ if the i-th unit survives to some time t^*, $i = 1, 2$; $(X_i = 0)$ otherwise.

- Let H(τ) be the background information of 🚶 at time $\tau > 0$, $\tau < t^*$.

- What are $P^{\tau}_{\text{🚶}}(X_i = 1;\ \mathrm{H}(\tau))$, and
$P^{\tau}_{\text{🚶}}(X_1 = 1 \text{ and } X_2 = 1;\ \mathrm{H}(\tau))$?

- Focus on $i = 1$, and ignoring  and τ, consider $P(X_1 = 1; \mathrm{H}(\tau))$

$$= \int_0^1 P(X_i = 1 \mid \theta) P(\theta; a, b) \mathrm{d}\theta$$

$$= \int_0^1 \theta \cdot 1 \mathrm{d}\theta = \tfrac{1}{2} \text{ if } a = b = 1; \quad \text{similarly } P(X_2 = 1; \mathrm{H}(\tau)).$$

- Now consider $P(X_1 = 1 \text{ and } X_2 = 1; \mathrm{H}(\tau))$

$$= \int_0^1 P(X_1 = 1 \text{ and } X_2 = 1 \mid \theta; \mathrm{H}(\tau)) P(\theta; \mathrm{H}(\tau)) \mathrm{d}\theta$$

$$= \int_0^1 P(X_1 = 1 \mid \theta) P(X_2 = 1 \mid \theta) P(\theta; a, b) \mathrm{d}\theta$$

$$= \int_0^1 \theta \cdot \theta \cdot 1 \mathrm{d}\theta = \int_0^1 \theta^2 \mathrm{d}\theta = \tfrac{1}{3}, \text{ if } a = b = 1.$$

- Similarly, $P^{\tau}(X_1 = 1 \text{ and } X_2 = 0; \mathrm{H}(\tau))$, when $a = b = 1$ is

$$= \int_0^1 \theta(1-\theta)\cdot 1 \mathrm{d}\theta = \int_0^1 \theta \mathrm{d}\theta - \int_0^1 \theta^2 \mathrm{d}\theta$$

$$= \tfrac{1}{2} - \tfrac{1}{3} = \tfrac{1}{6}.$$

- Observe that

$$P(X_1 = 1; \mathrm{H}(\tau)) > P(X_1 = 1 \text{ and } X_2 = 1; \mathrm{H}(\tau))$$
$$> P(X_1 = 1 \text{ and } X_2 = 0; \mathrm{H}(\tau))$$
$$= P(X_2 = 1, X_1 = 0; \mathrm{H}(\tau)).$$

- Thus $P(X_1 = 0 \text{ and } X_2 = 0) = 1 - \tfrac{1}{3} - \tfrac{1}{6} - \tfrac{1}{6} = \tfrac{1}{3}$.

Inductive Prediction Under Bernoulli Models.

- The scenario of predicting $X_2 = 1$, when we know for sure that $X_1 = 1$ – at time τ – that is the case of induction, involves some subtle, if not tricky, arguments.

- We need to assess $P^{\tau}(X_2 = 1;\, X_1 = 1,\, \mathrm{H}(\tau))$.
- We start by ignoring the fact that $X_1 = 1$ is known, and suppressing and $\mathrm{H}(\tau)$, consider the proposition:

- $P(X_2 = 1 \mid X_1 = 1) = \int_0^1 P(X_2 = 1 \mid \theta, X_1 = 1)\, P(\theta \mid X_1 = 1)\,\mathrm{d}\theta$

$$= \int_0^1 P(X_2 = 1 \mid \theta)\, P(\theta \mid X_1 = 1)\,\mathrm{d}\theta$$

$$= \int_0^1 \theta\, P(\theta \mid X_1 = 1)\,\mathrm{d}\theta.$$

- But by Bayes' Law,

$$P(\theta \mid X_1 = 1) \propto P(X_1 = 1 \mid \theta) P(\theta; \mathrm{H}(\tau)), \text{ or}$$

$$P(\theta; X_1 = 1) \propto \mathrm{L}(\theta; X_1 = 1) P(\theta; a, b),$$

since $(X_1 = 1)$ is actually observed.

- Suppose that we choose L(θ; X_1 = 1) = θ, and set a = b = 1; then
- $P(\theta; X_1 = 1) \propto \theta = c\theta$, where c is a constant.
- To find the constant of proportionality c, we integrate

$$\int_0^1 P(\theta; X_1 = 1)\mathrm{d}\theta = 1 = \int_0^1 c\theta\mathrm{d}\theta \Rightarrow c = 2.$$

- Thus $P(\theta; X_1 = 1) = 2\theta \sim Beta(a = 2, b = 1)$.

- Thus to summarize,

$$P(X_2 = 1; X_1 = 1, a = b = 1) = \int_0^1 2\theta \mathrm{d}\theta = \tfrac{2}{3}.$$

- Consequently,

$$\begin{aligned} P(X_1 = 1, X_2 = 1; \mathrm{H}(\tau)) &= \tfrac{1}{3} \\ &< P(X_2 = 1; \mathrm{H}(\tau)) \\ &< P(X_2 = 1; X_1 = 1, \mathrm{H}(\tau)) = \tfrac{2}{3}. \end{aligned}$$

- The ability to do integrations is crucial. Thus a need for MCMC methods.
- Go to slide 66 for <u>*hypotheses testing*</u>.

23. Markov Chain Monté Carlo – The Gibbs Sampler.

- There are several MCMC methods, one of which is the ***Metropolis-Hastings Algorithm***, a special case of which is the ***Gibbs-Sampler***.

- Gibbs sampling is a technique for generating random variables from a distribution without knowing its density.

- Suppose there exists a joint density $f(x,y)$, and we are interested in knowing characteristics of the marginal $f(x) = \int_y f(x, y)dy.$

- Then Gibbs sampling enables us to obtain a sample from $f(x)$ without requiring an explicit specification of $f(x)$, but requiring a specification of $f(x|y)$ and $f(y|x)$.

- The technique proceeds as follows:
  i) Choose y_0^1, and generate x_0^1 from $f(x \mid y_0^1)$.
  ii) Now generate y_1^1 from $f(y \mid x_0^1)$.
  iii) Next generate x_1^1 from $f(x \mid y_1^1)$.
  iv) Repeat steps ii) and iii) k times to obtain $(y_0^1, x_0^1), (y_1^1, x_1^1), \ldots, (y_k^1, x_k^1) -$ the ***Gibbs Sequence***.

- Result: When k is large, the distribution of x_k^1 is $f(x) \Rightarrow x_k^1$ is a sample point from $f(x)$. This result is from the theory of Markov Chains.

- To get a sample of size m from $f(x)$, repeat steps i) through iv) m times using m different starting values $y_0^1, y_0^2, \ldots, y_0^m$.

- The ***Hammersley-Clifford*** theorem asserts that a knowledge of the conditionals asserts a knowledge of the joint.

24. Gibbs Sampling Under a Bernoulli.

- Recall the scenario of a Bernoulli(θ) probability model with a uniform distribution for θ. The predictive distribution is:

$$P^{\tau}(X_1 = 1; \mathrm{H}(\tau)) = \int_0^1 P(X_1 = 1 \mid \theta) \cdot 1 \mathrm{d}\theta$$

$$= \int_0^1 \theta \mathrm{d}\theta = ?,$$

if you have forgotten your integration.

- Using Bayes' Law, we have seen that

$$P(\theta \mid X_1 = 1) = 2\theta.$$

- Thus knowing $P(X_1 = 1 | \theta) = \theta$ and $P(\theta | X_1 = 1) = 2\theta$, we can generate for some large k a Gibbs Sequence

$$(X_0^1, \theta_0^1), (X_1^1, \theta_1^1), \ldots, (X_k^1, \theta_k^1),$$

and thence a sample of size *m*, $(X_k^1, X_k^2, \ldots, X_k^m)$, from which $P(X_1 = 1; \mathrm{H}(\tau))$ can be obtained as

$$\sum_1^m X_k^i / m.$$

25. Application: Gibbs Sampling from an Exponential.

- Recall the scenario of an exponential(λ) failure (probability) model with a gamma (scale c, shape d) distribution for λ. The predictive is

$$P^{\tau}(T > t; \mathrm{H}(\tau)) = \int_0^{\infty} \mathrm{e}^{-\lambda t} P^{\tau}(\lambda; c, d)\,\mathrm{d}\lambda$$

$$= \int_0^{\infty} \mathrm{e}^{-\lambda t} \frac{\mathrm{e}^{-\lambda c}(\lambda c)^{d-1} c}{\Gamma(d)} \mathrm{d}\lambda.$$

- Using Bayes' Law, it can be shown that $P(\lambda|t)$ is also a gamma (scale $c + t$, shape $d + 1$).

- Thus, knowing $P(t \mid \lambda) = \lambda e^{-\lambda t}$ and $P(\lambda \mid t)$, we can generate, for some large k (= 1000, say) a Gibbs Sequence $(\lambda_0^1, t_0^1), (\lambda_1^1, t_1^1), \ldots, (\lambda_k^1, t_k^1)$ and thence a sample of size m, $(t_k^1, t_k^2, \ldots, t_k^m)$ from which $P(T > t; \mathrm{H}(\tau))$ can be obtained as

$$\tfrac{1}{m}[\# t_k^i > t].$$

BAYESIAN HYPOTHESIS TESTING

Bayesian Hypothesis Testing

- The testing of hypothesis is done to support a theory or a claim in the light of available evidence.

- It is useful in astronomy, particle physics, forensic science, drug testing, intelligence, medical diagnosis, and acceptance sampling in quality control and reliability.

A Simple Architecture

- Let X be an unknown quantity
- Let $P(X|\theta)$ be a probability model for X, where θ is a parameter.
- Suppose that θ can take only two values, $\theta = \theta_0$ or $\theta = \theta_1$ (the case of a simple versus a simple hypothesis).
- Let $\mathrm{P}(\theta = \theta_0) = \Pi_0 \Rightarrow P(\theta = \theta_1) = \Pi_1 = 1 - \Pi_0$

A Simple Architecture (continued)

- Suppose that X has revealed itself as x.
- Can we now say conclusively and emphatically, that (H_0 –the **null** hypothesis) or (H_1 –the **alternate** hypothesis) is true?

- Very rarely a yes, but most often a no.
- The Bayesian paradigm does not permit an acceptance or rejection of a hypothesis (without an involvement of the underlying utilities).
- All that one can do under the Bayesian paradigm claim that a knowledge of x enhances our opinion of either H_0 or of H_1 .

A Simple Architecture (continued)

- The quantity Π_0/Π_1 is termed (our) *prior odds* on H_0 against H_1.
- $\frac{\Pi_0}{\Pi_1} = 1 \Rightarrow$ H_0 and H_1 are equally likely true, a priori, and in our opinion.
- $\frac{\Pi_0}{\Pi_1} > 1 \Rightarrow \mathrm{H}_0$ is (a priori) more likely to be true than H_1, in our opinion.

How Should Evidence x be Incorporated?

- By Bayes' Law

$$P_0 \stackrel{\text{def}}{=} P(\theta_0; x, H) \propto L(\theta_0; x)\Pi_0$$

the posterior of θ_0 under x and

$$P_1 \stackrel{\text{def}}{=} P(\theta_1; x, H) \propto L(\theta_1; x)\Pi_1$$

- Let $\frac{P_0}{P_1} = \boldsymbol{posterior\ odds}$ on H_0 against H_1
- Then $\frac{P_0}{P_1} = \frac{\mathrm{L}(\theta_0;x)}{L(\theta_1;x)} \cdot \frac{\Pi_0}{\Pi_1}$, by simple algebra.
- Thus

  Posterior Odds = Ratio of Likelihood · Prior Odds.

- The ratio of likelihoods is called the *Bayes' Factor* **B** in favor of H_0 against H_1.

- Thus Posterior Odds = (Bayes Factor).
  (Prior Odds).

- Equivalently, Bayes' Factor **B** $= \dfrac{Posterior\ Odds}{Prior\ Odds}$

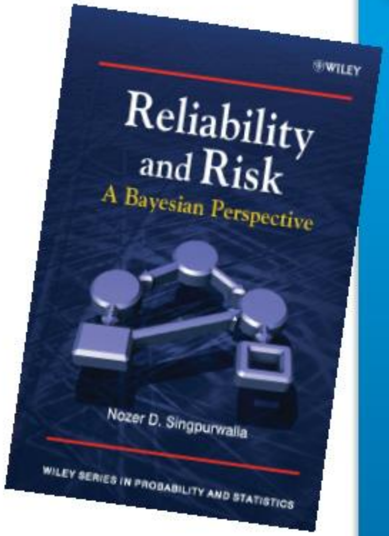
- The logarithm of B is called the *Weight of Evidence*.

20% OFF Quote Code VB205

MATHEMATICS & STATISTICS

Reliability and Risk - A Bayesian Perspective

By Nozer D. Singpurwalla



We all like to know how reliable and how risky certain situations are, and our increasing reliance on technology has led to the need for more precise assessments than ever before. Such precision has resulted in efforts both to sharpen the notions of risk and reliability, and to quantify them. Quantification is required for normative decision-making, especially decisions pertaining to our safety and wellbeing. Increasingly in recent years Bayesian methods have become key to such quantifications.

Reliability and Risk provides a comprehensive overview of the mathematical and statistical aspects of risk and reliability analysis, from a Bayesian perspective. This book sets out to change the way in which we think about reliability and survival analysis by casting them in the broader context of decision-making. This is achieved by:

- Providing a broad coverage of the diverse aspects of reliability, including: multivariate failure models, dynamic reliability, event history analysis, non-parametric Bayes, competing risks, co-operative and competing systems, and signature analysis
- Covering the essentials of Bayesian statistics and exchangeability, enabling readers who are unfamiliar with Bayesian inference to benefit from the book
- Introducing the notion of "composite reliability", or the collective reliability of a population of items
- Discussing the relationship between notions of reliability and survival analysis and econometrics and financial risk

Reliability and Risk can most profitably be used by practitioners and research workers in reliability and survivability as a source of information, reference, and open problems. It can also form the basis of a graduate level course in reliability and risk analysis for students in statistics, biostatistics, engineering (industrial, nuclear, systems), operations research, and other mathematically oriented scientists, wherein the instructor could supplement the material with examples and problems.

Hardback 396 pages 2006 ISBN 978-0-470-85502-7
USD $130.00 £74.95 €89.90

WILEY-BLACKWELL

Connect with us:

Wiley-Blackwell Statistics and Mathematics

www.twitter.com/Wiley_Stats

Statistics by Wiley

How to order:

EUROPE, MIDDLE EAST, ASIA & AFRICA
John Wiley & Sons Ltd
Tel: +44 (0)1243 843294 Fax: +44 (0)1243 843296
E-mail: cs-books@wiley.co.uk
www.wiley.com

NORTH, CENTRAL & SOUTH AMERICA
John Wiley & Sons Inc
Tel: 877 762 2974 Fax: 800 597 3299
E-mail: custserv@wiley.com
www.wiley.com

GERMANY, SWITZERLAND & AUSTRIA
Wiley-VCH Verlag GmbH
Tel: +49 6201 606 400 Fax: +49 6201 606 184
E-mail: service@wiley-vch.de
www.wiley-vch.de