Bayes' Theorem and its Application in Quantitative Risk Assessment

(c. 1702 – 17 April 1761)

Vincent Ho

21 April 2008

Qualitative Definitions of Risk

$$\text{Risk} = \frac{\text{Hazard}}{\text{Safeguards}}$$

- Risk is never zero by increasing level of safeguards, as long as hazard is present

$$\text{Risk} = \text{Likelihood} \times \text{Consequence}$$

- Classical, but most misleading. More useful in hazard analyses

$$\text{Risk} = \text{Uncertainty} \times \text{Damage}$$

- Without uncertainty or damage, there is no risk
- Anybody can guess extent of damage with different levels of uncertainties

Sources of Uncertainties

- **Impossible to explicitly enumerate all conditions**
- **Inadequate or incorrect information on conditions**
- **Inconsistent interpretation and classification of events**
- **Lack of success data (for number of demands and exposure/mission time)**
- **Limited data sample size; realised risk and unrealised risk**
- **Imperfect mathematical and computer modelling of reality**

Uncertainties

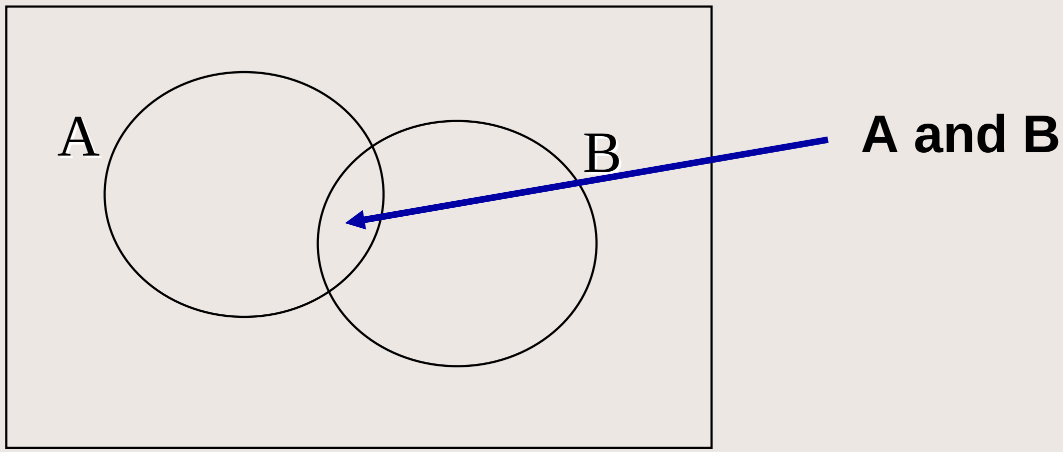
- **In general, there are three types of uncertainties associated with a risk assessment:**
  - **Stochastic uncertainties**
  - **Modelling uncertainties**
  - **Parametric uncertainties**
- **Strictly speaking, A+A≠2xA**
- **It is this explicit consideration of uncertainties distinguishes a risk assessment from a hazard analysis, a PRA from a "QRA"**
- **Uncertainties are measured by level of belief; i.e., probability**

Probability Functions

- The likelihood of an event E is indicated by Probability Function P(E)
- The sum of the probabilities of all elementary outcomes within sample space S, P(S) = 1, with values between 0 and 1
  - P(E) = 1: the event is CERTAIN to occur
  - P(E) = 0: the event is certain NOT to occur
  - Anything in between represents our level of belief of the certainty or uncertainty of an event to occur

The Laws of Probability

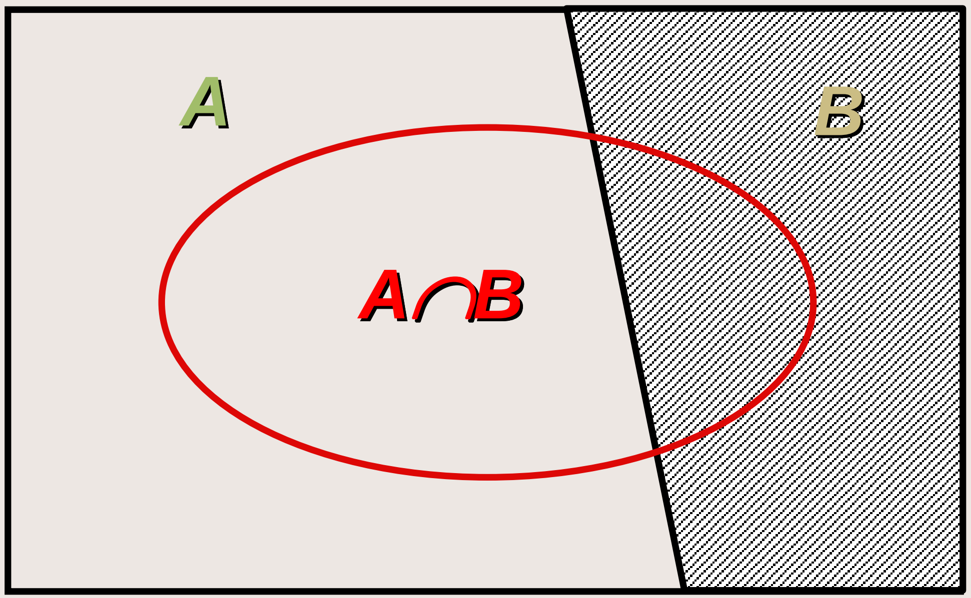
- The probability of any event A is $0 \leq P(A) \leq 1$
- Law of Addition: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$



- Law of Multiplication: $P(A \cap B) = P(A) \times P(B)$

Law of Multiplication

- Actually, $P(A \cap B) = P(B) \times P(A|B)$
  where $P(A|B)$ is probability of A given B has occurred
- If A and B are statistically independent,
  - $P(B|A) = P(B)$, then
  - $P(A \cap B) = P(A) \times P(B|A) = P(A)\, P(B)$



Bayes' Theorem

- Bayes' Theorem is a trivial consequence of the definition of conditional probability, but it is very useful in that it allows us to use one conditional probability to compute another
- Given that A and B are events in sample space S, and P(B) $\neq$ 0, conditional probability is defined as:
  - $P(A \cap B) = P(A|B)\,P(B)$
  - $P(A \cap B) = P(B|A)\,P(A)$
  - $P(B|A)\,P(A) = P(A|B)\,P(B)$

$$P(B \mid A) = \frac{P(A|B)P(B)}{P(A)} = \frac{P(A \cap B)}{P(A)}$$

Probability vs Frequency

- Frequency is a measure of the rate of occurrence. E.g., failure rate of a pump is 6.2×10^{-3}/hr
- Probability is a measure of the level of belief, a fraction, or failure per demand. It is dimensionless. E.g., the failure rate of the pump is

Frequency	Probability
1.0×10^{-4}/hr	0.2
2.0×10^{-3}/hr	0.5
3.2×10^{-3}/hr	0.2
4.5×10^{-2}/hr	0.1

with a mean of 6.2×10^{-3}/hr

- The parameter failure rate is denoted as λ, and the probability of the failure rate is $P(\lambda = 1.0 \times 10^{-4}/hr) = 0.2$

Uncertainties Based on Evidence

- In general, two approaches to estimate parameters:
  - Frequentist
    - Based only on observed data and an adopted model
    - Characterized by scientific objectivity
  - Bayesian
    - Appropriately combining prior intuition or knowledge with information from observed data
    - Characterized by subjective nature of prior opinion
- Each approach is valid when applied under specific circumstances
- Neither approach uniformly dominates the other

Frequentist Statistics

- **Relative frequency λ: proportion of times an outcome occurs**

$$\lambda = \frac{\text{Number of successful trial}}{\text{Total number of trials (N)}}$$

- **For $N \rightarrow \infty$, the relative frequency tends to stabilize around some number: probability estimates**
- **Frequentist statistics will completely break down if no data or no experience history (N =0)**
  - **New technology**
  - **Rare events**
  - **No failure record**

Bayesian Statistics

- **Bayesian statistics measures degrees of belief by using intuition knowledge (prior belief), updating it by evidence (likelihood) to obtain a posterior belief P(B|A) = P(A|B)*P(B) / P(A)**
- **To process knowledge**

P(Cause|Effect) = P(Effect|Cause) P(Cause) / P(Effect)

Posterior $\propto$ likelihood x prior

$$p(\theta_j | \text{data}) = p(\text{data} | \theta_j)\, p(\theta_j) / \Sigma\, p(\text{data} | \theta_i)\, p(\theta_i)$$

basically a normalizing constant

Bayesian Statistics

- **In a QRA, we know what the damage effects and their contributing factors are, <u>we want to know the likelihood of the contributing factors</u>**

prior probability, i.e., before seeing the data

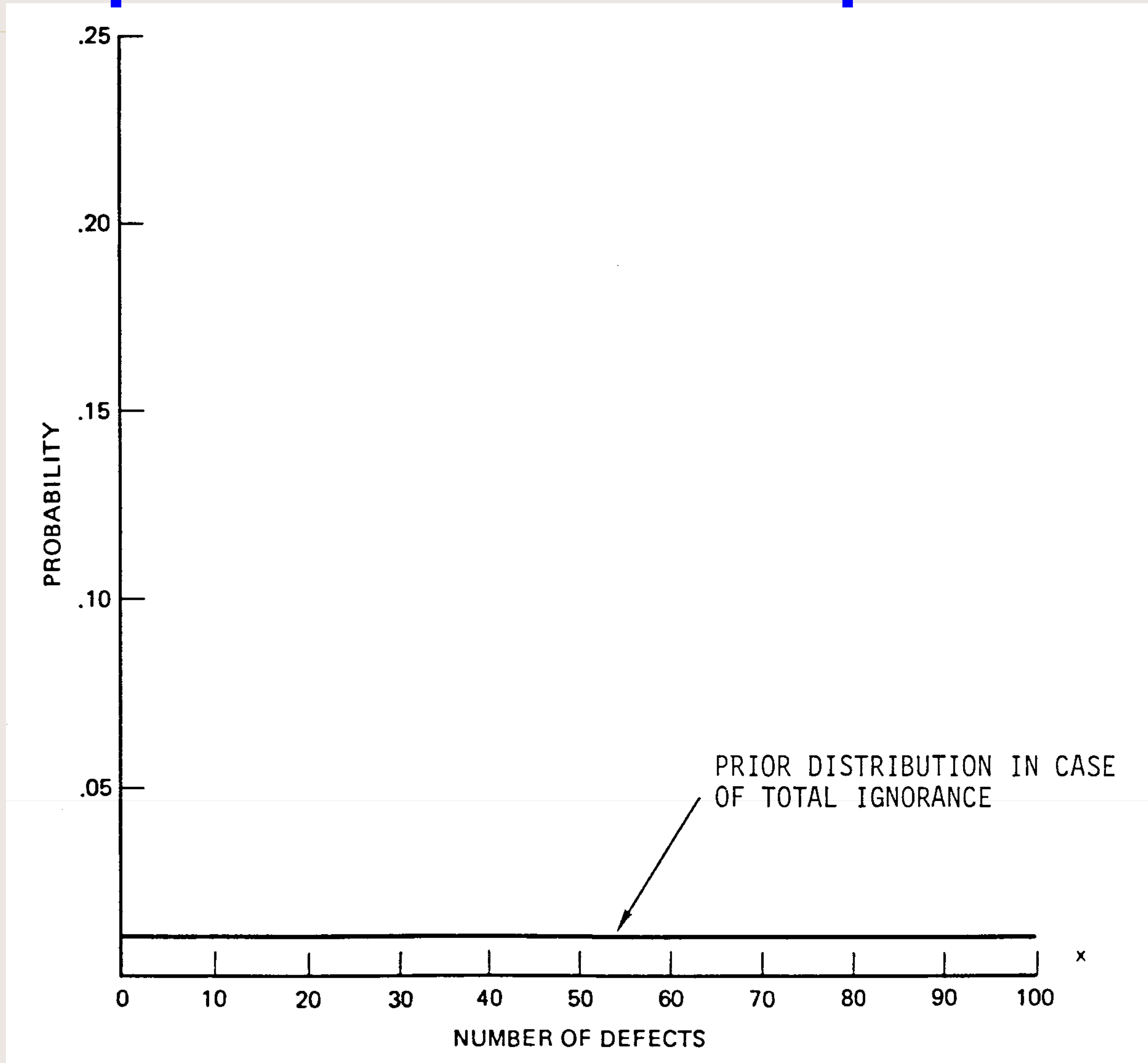
The likelihood of seeing evidence given prior

$$\pi'(\lambda \mid E) = \frac{\pi(\lambda) L(E \mid \lambda)}{\int_0^\infty d\lambda\, \pi(\lambda) L(E \mid \lambda)}$$

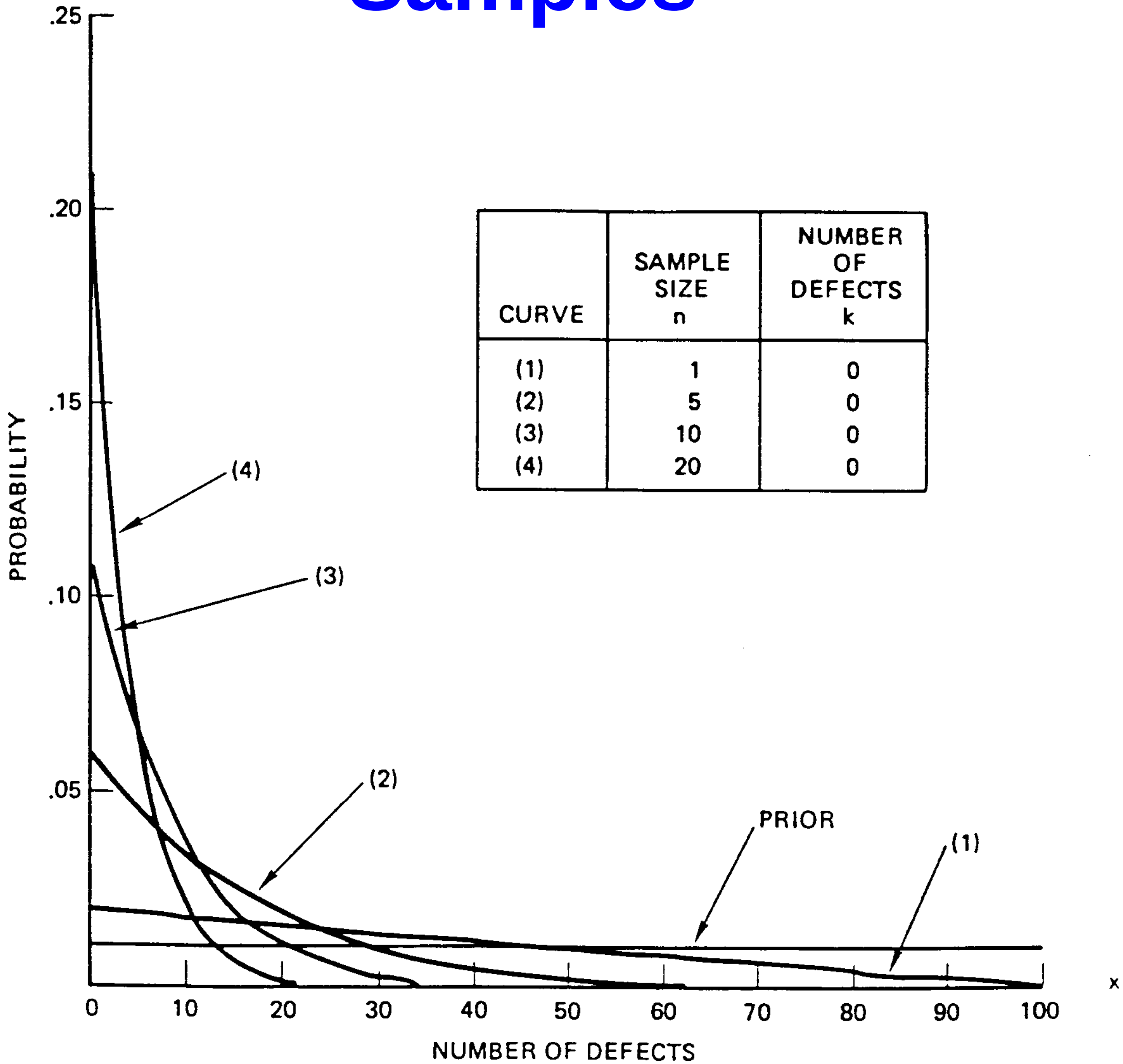
posterior probability, i.e., after seeing the data

normalization involves summing over all possible hypotheses

How Many Defects in a Population of 100 Components?

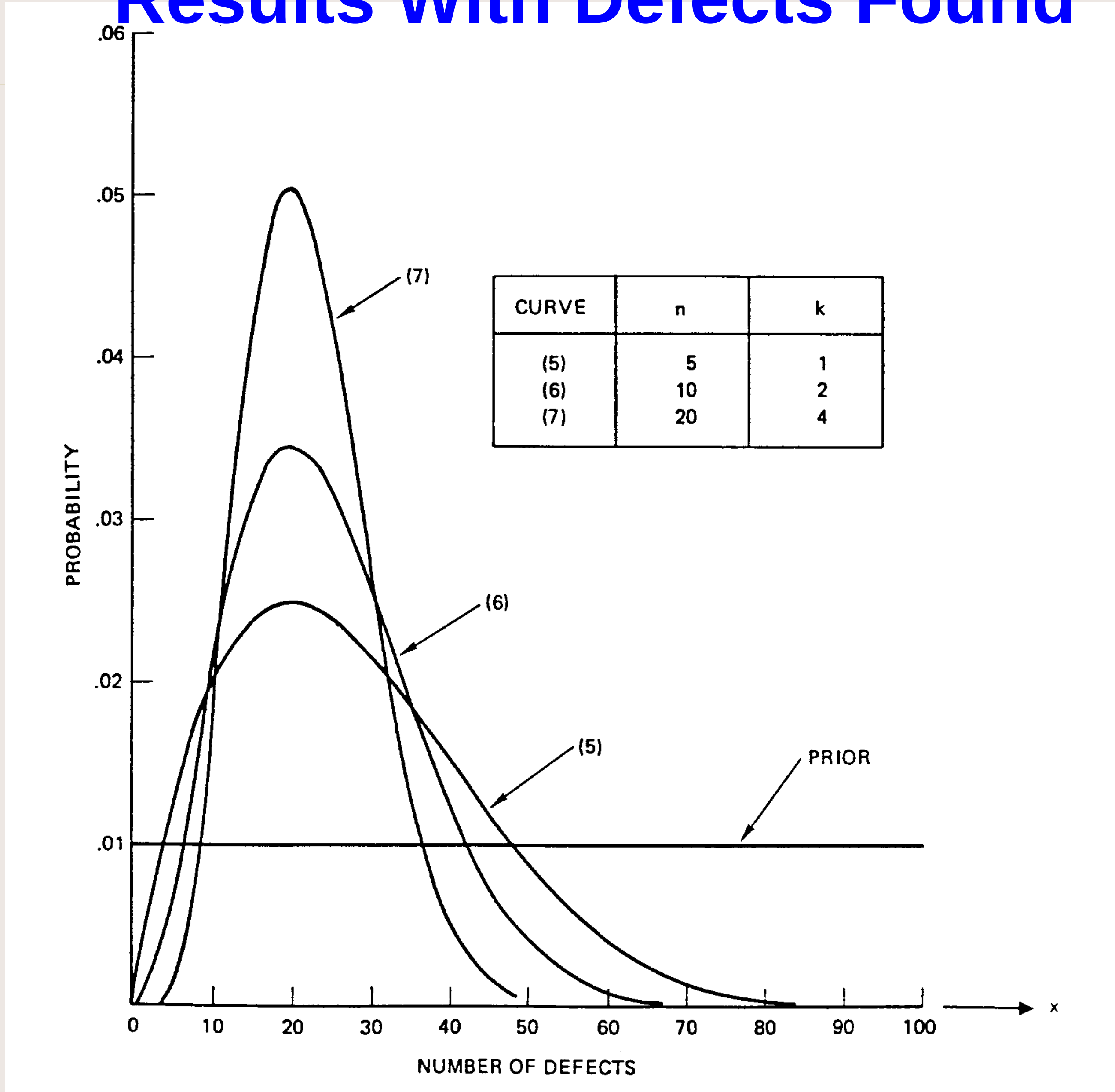


State of Knowledge after Various Samples

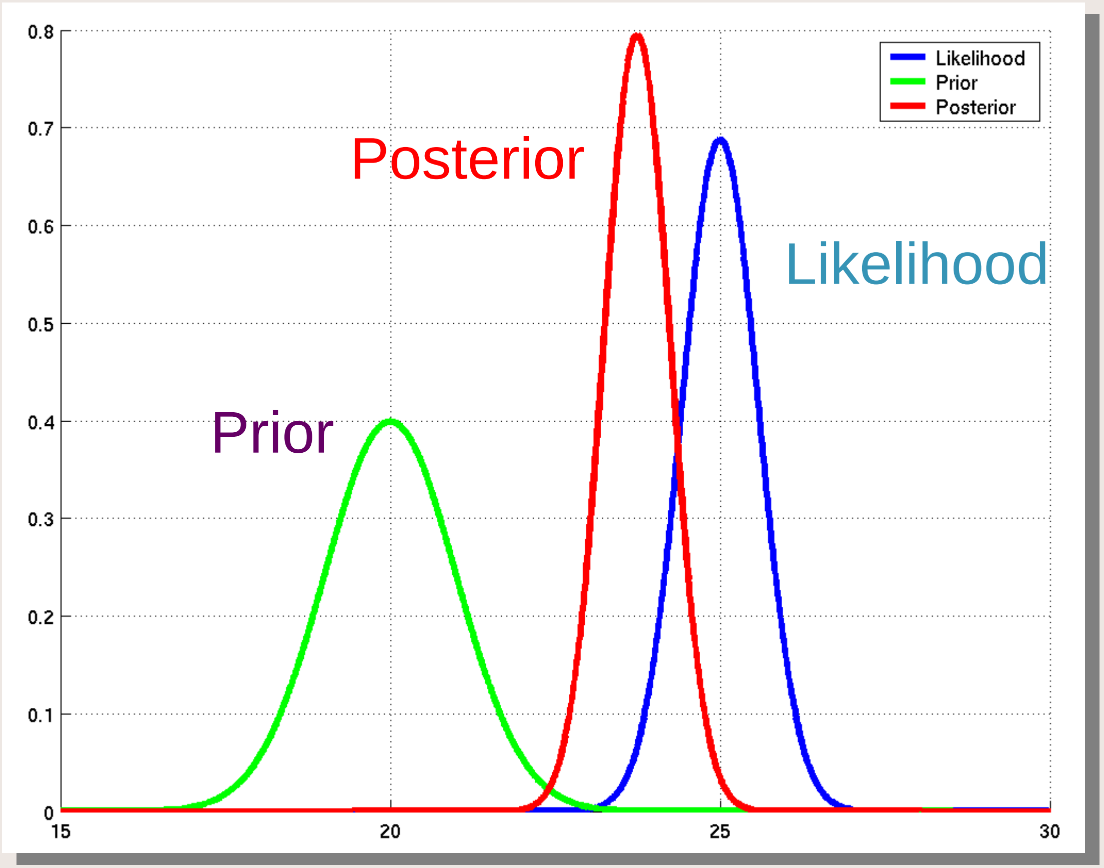


CURVE	SAMPLE SIZE n	NUMBER OF DEFECTS k
(1)	1	0
(2)	5	0
(3)	10	0
(4)	20	0

State of Knowledge after Various Sample Results With Defects Found



Typical Shapes of Probability Functions for Prior, Likelihood and Posterior



Application Example- Fire Frequency of a Complex Engineering System

- Model the likelihood function by the Poisson distribution and the prior by a gamma distribution (conjugate of Poisson)
- Then the posterior distribution is also a gamma distribution
- A gamma distribution (α, β) is

$$\pi(\lambda) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$$

Application Example- Fire Frequency of a Complex Engineering System

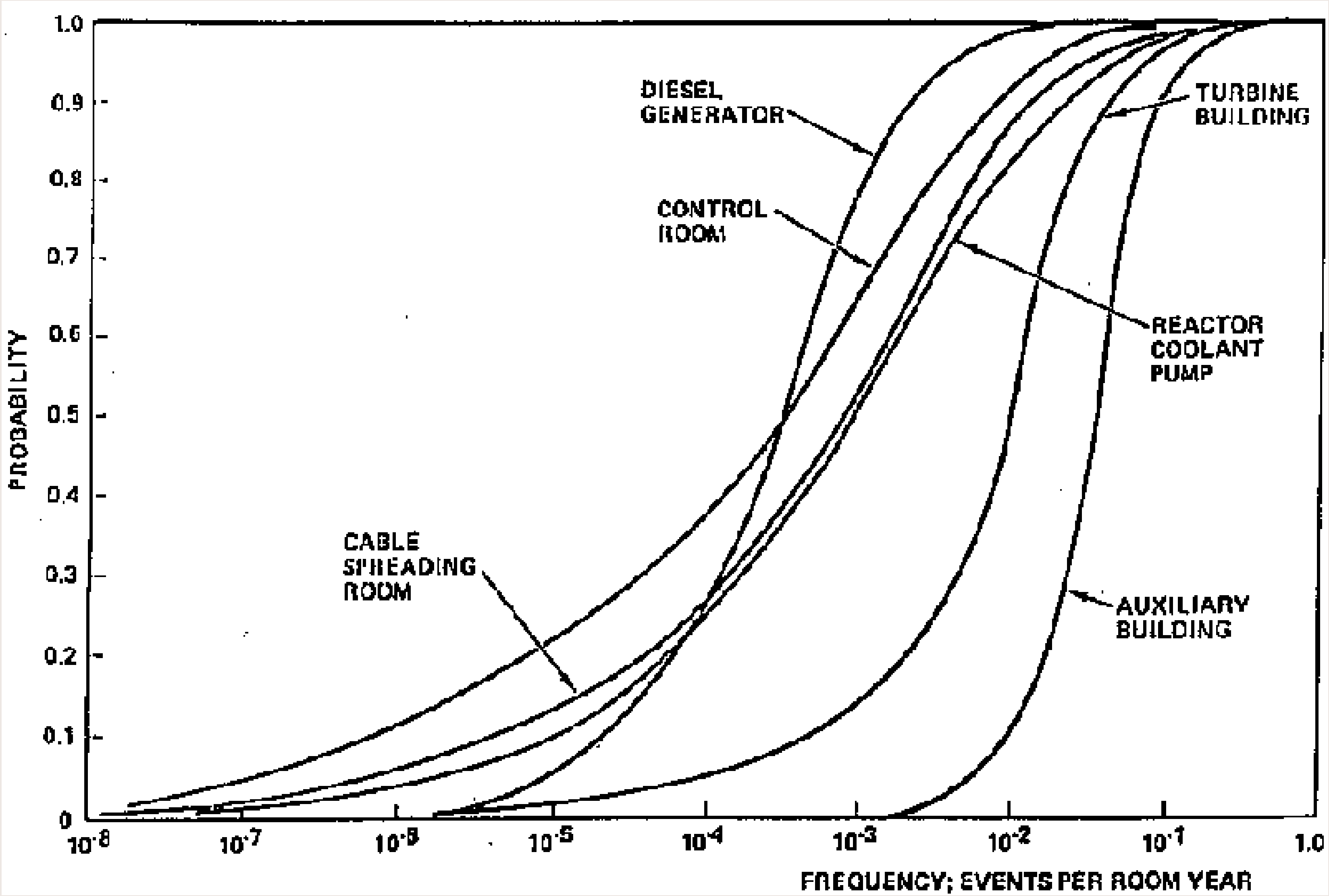
- Evidence is Poisson (r, T) where

  r = number of fires in the system being analysed

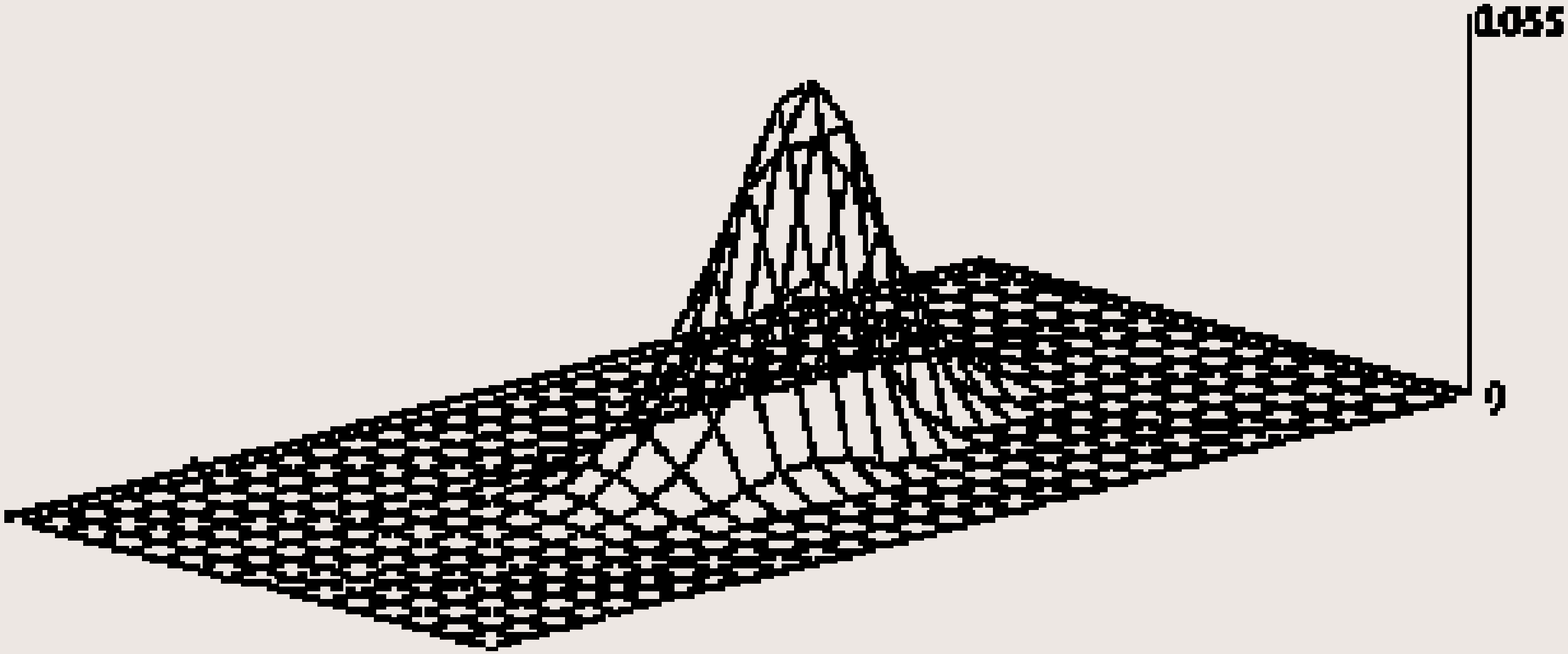
  T = number of years covered by the system

- Frequentist will show fire frequency $\lambda = r/T$, but the process will break down if r =0
- Prior, gamma (α, β), is based on generic experience
- Then the posterior distribution is gamma (α', β') where $\alpha' = \alpha + r$; $\beta' = \beta + T$
- Bayesian update can ensure a meaningful set of data being used with $\lambda = \alpha' / \beta'$, even r=0
- Prior is usually obtained by another set of Bayesian operations

Probability Curves for Fire Frequency (Example)



Posterior Distribution of Fire Suppression Time (Example)



Properties of Bayesian Updating

- **Need to estimate Prior and obtain the appropriate evidence**
  - **With weak evidence, prior dominates results**
  - **With strong evidence, results insensitive to prior (dominated by evidence)**
  - **Successive updating gives same result as one-step updating with consistent evidence**
- **Provide a robust method in assessing initializing event frequency, failure rates, event tree split fractions**
  - **Drawn on generic experience**
  - **Plant specific data (including no failure)**

Probability of Impact….

2008

Conference Announcement and Call for Participation

PSAM 9

International Conference on Probabilistic Safety Assessment and Management

An IAPSAM Conference

18 - 23 May 2008 • Hong Kong • China

www.psam9.org

End

Supplementary Notes
(Advance Level)

Bayes' Theorem Example

Suppose we have a new untested system. We estimate, based on prior experience, that 80% of chance the reliability (probability of successful run) R_1 = 0.95, and 20% of chance that R_2 = 0.75. We run a test and find that it operate successfully. What is the probability that the reliability level is R_1.

Prior Distribution	
R	P
0.95	0.8
0.75	0.2
Mean = 0.	

S_i = event "System test results in a success"

$$P(R_1|S_1) = \frac{P(R_1)P(S_1|R_1)}{P(R_1)P(S_1|R_1) + P(R_2)P(S_1|R_2)}$$

=0.8*0.95/(0.8*0.95 + 0.2*0.75) = 0.835 (updated from 0.8)

Posterior Distribution	
R	P
0.95	0.835
0.75	1-0.835 = 0.165
Mean = 0.917	

Now, let's assume that a second test is also successful

$$P(R_1|S_2) = \frac{P(R_1)P(S_2|R_1)}{P(R_1)P(S_2|R_1) + P(R_2)P(S_2|R_2)}$$

=0.835*0.95/(0.835*0.95 + (1-0.835)*0.75)

=0.79325/(0.79325+0.12375)=0.865

2nd Posterior Distribution	
R	P
0.95	0.865
0.75	1-0.865 = 0.135
Mean = 0.923	

What is $P(R_1|S_3)$ if the S_3 is a failure?

Bayes' Theorem Example

Prior Distribution	
R	P
0.95	0.8
0.75	0.2
Mean = 0.95*0.8+0.75*0.2 = 0.91	

Posterior Distribution | S1	
R	P
0.95	0.835
0.75	1-0.835 = 0.165
Mean = 0.917	

Posterior Distribution |S2	
R	P
0.95	0.865
0.75	1-0.865 = 0.135
Mean = 0.923	

S_3 = event "System test results in a failure"

$$P(1-R_1)\,|\,S_3) = \frac{P(1-R_1)\,P(S_3\,|\,1-R_1)}{P(1-R_1)\,P(S_3\,|\,1-R_1) + P(1-R_2)\,P(S_3\,|\,1-R_2)}$$

Posterior Distribution |S2	
1-R	P
0.05	0.865
0.25	1-0.865 = 0.135
Mean = 0.077	

$P(1\text{-}R1|S_3)$= 0.865*0.05/(0.865*0.05 + (1-0.865)*0.25)

=0.04325/(0.04325+0.03375)=0.562

Posterior Distribution |S3	
R	P
0.95	0.562
0.75	1-0.562 = 0.438
Mean = 0.8624	

Bayes's theorem

- Let us recall Bayes's theorem:

$$f(\theta | x) = \frac{f(\theta) f(x | \theta)}{\int_{-\infty}^{\infty} f(\theta) f(x | \theta) d\theta}$$

- Where f(θ) is density of the prior probability distribution for parameter(s) of interest, f(x|θ) is density of conditional probability distribution for x given θ, f(θ|x) is posterior density of the distribution of the parameter of interest - θ. Integral is the normalisation coefficient that ensures that integral of posterior is equal to 1.
- Bayesian estimation is fundamentally different from the maximum likelihood estimation. In maximum likelihood estimation parameters we want to estimate are not random variables. In Bayesian statistics they are.
- Prior, likelihood and posterior have the following interpretations:
- Prior: It reflects the state of our knowledge about the parameter(s) before we have seen the data. E.g. if this distribution is sharp then we have fairly good idea about the parameter of interest.
- Likelihood: How likely it is to observe current observation if parameter of interest would have current value.
- Posterior: It reflects the state of our knowledge about the parameter(s) after we have observed (and treated) the data.

Bayes's theorem and learning

- Bayes's theorem in some sense reflects dynamics of learning and accumulation of the knowledge. Prior distribution encapsulates the state of our current knowledge. When we observe some data then they can change our knowledge. Posterior distribution reflects it. When we observe another data then our current posterior distribution becomes prior for this new experiment. Thus every time using our current knowledge we design experiment, observe data and store gained information in the form of new prior knowledge. Sequential nature of Bayes's theorem elegantly reflects it. Let us assume that we have prior knowledge written as f(θ) and we observe the data - x. Then our posterior distribution will be f(θ|x). Now let us assume that we have observed new independent data y. Then we can write Bayes's theorem as follows:

$$f(\theta|x,y)=\frac{f(\theta)f(x,y|\theta)}{\int_{-\infty}^{\infty}f(\theta)f(x,y|\theta)d\theta}=\frac{f(\theta)f(x|\theta)f(y|\theta)}{\int_{-\infty}^{\infty}f(\theta)f(x,y|\theta)d\theta}=\frac{f(\theta|x)f(y|\theta)}{\int_{-\infty}^{\infty}f(\theta|x)f(y|\theta)d\theta}$$

- Last term shows that posterior distribution after observing and incorporating information from x is now prior for treatment of the data y. That is one reason why in many Bayesian statistics book priors are written as f(θ|I), where I reflects the information we had before the current observation. If data are not independent then likelihood becomes conditional on parameter and on the previous data.
- One more important point is that prior is different from a priori. Prior is knowledge available before this experiment (or observation) a priori is before any experiment. In science we do not deal with the problem of knowledge before any experiment.

Prior, likelihood and posterior

- Before using Bayes's theorem as an estimation tool we should have the forms of prior, likelihood and posterior.
- Likelihood is usually derived or approximated using physical properties of the system under study. Usual technique used for derivation of the form of the likelihood is central limit theorem.
- Prior distribution should reflect state of knowledge. Converting knowledge into distribution could be a challenging task. One of the techniques used to derive prior probability distribution is maximum entropy approach. In this approach entropy of distribution is maximised under constraint defined by the available knowledge. Some of the knowledge we have, can easily be incorporated into maximum entropy formalism. Problem with this approach might be that not all available knowledge can easily be used, Another approach is to study the problem, ask experts and build physically sensible prior. One more approach is to find such prior that when used in conjunction with the likelihood they give easy and elegant forms for posterior distributions. These type of priors are called conjugate priors. They depend on the form of likelihood. Here is the list of some of conjugate priors used for one dimensional cases:

Likelihood	Parameter	Prior/Posterior
Normal	mean (μ)	Normal
Normal	variance (σ2)	Inverse gamma
Poisson	λ	Beta
Binomial	π	Gamma

Importance of prior distributions

- One of the difficult (and controversial) parts of the Bayesian statistics is finding prior and calculating posterior distributions. Convenient priors can easily be incorporated into calculations but they are not ideal and they may result in incorrect results and interpretation. If prior knowledge says that some parameters are impossible then no experiment can change it. For example if prior is defined so that values of the parameter of interest are positive then no observation can result in non 0 probability for negative values. If some values of the parameter have extremely small (prior) probability then one might need many, many experimental data to see that these values are genuinely possible.
- Bayesian statistics assumes that probability distribution is known and it in turn involves integration to get the normalisation coefficient. This integration might be tricky and in many cases there is no analytical solution. That was main reason why conjugate prior were so popular. With advent of computers and various integration techniques this problem can partially be overcome. In many application of Bayesian statistics prior is tabulated and then sophisticated numerical integration techniques are used to derive posterior distributions.
- Popular approximate integration techniques used in Bayesian statistics involve: Gaussian integration, Laplace approximation, numerical integration based on stochastic approaches (Monte-Carlo, Gibbs sampling, Markov Chain Monte Carlo).