IMPACTO DA ESPECIALIZAÇÃO DE DADOS DE FALHA EM ANÁLISES PROBABILÍSTICAS DE SEGURANÇA DE PLANTAS DE PROCESSO

Antonio Carlos de Oliveira Ribeiro

TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS PROGRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA NUCLEAR.

Aprovada por:

Prof. Paulo Fernando Ferreira Frutuoso e Melo, D. Sc.

Prof. Antonio Carlos Marques Alvim, Ph.D.

Dr. Marco Antonio Bayout Alvarenga, D. Sc.

RIO DE JANEIRO, RJ – BRASIL
FEVEREIRO DE 2005

RIBEIRO, ANTONIO CARLOS DE OLIVEIRA
Impacto da especialização de dados de falha
em análises probabilísticas de segurança de
plantas de processo, [Rio de Janeiro] 2005
VII, 50 p. 29,7 cm (COPPE/UFRJ, M.Sc.,
Engenharia Nuclear, 2005)
Tese - Universidade Federal do Rio de Janeiro,
COPPE

1. Especialização de dados de falha por
   inferência bayesiana.
2. Estudos de análise de riscos ambientais
3. Análise probabilística de segurança
4. . I COPPE/UFRJ II. Título (série)

AGRADECIMENTOS

A minha esposa e filhos, pelos momentos de apoio e incentivo e principalmente pela paciência durante minhas ausências para me dedicar ao curso.

Ao Prof. Paulo Fernando Ferreira Frutuoso e Melo, agradeço, pelo constante incentivo, pela disponibilidade e principalmente pela amizade cultivada durante nosso convívio.

Aos Engenheiros, Enio Viterbo Júnior , Carlos Headler e Everton Ferreira Fonseca, pelo incentivo, por terem acreditado e me apoiado na realização deste curso.

À empresa BAYER S/A, na pessoa do seu Diretor, Eng. Flavio Abreu, pelo incentivo de aperfeiçoamento de seu quadro técnico do qual faço parte e pela gentil cessão de alguns dados usados neste trabalho .

À empresa SERENO Sistemas Ltda, na pessoa do seu Diretor, Eng. Ricardo Albuquerque, que gentilmente cedeu o software RISKAN®, também usado neste trabalho.

A Eng. Ana Letícia Souza, M.Sc., pelo apoio dado na revisão das simulações computacionais efetuadas.

A todo o corpo docente do Programa de Engenharia Nuclear da COPPE/UFRJ cujo ensinamento foi essencial para a realização desta dissertação.

Aos funcionários da COPPE/Nuclear pela cooperação e amizade criada durante esse nosso convívio.

A todos os colegas da COPPE/Nuclear pelas inúmeras horas conjuntas de trabalho, e estudo dedicados ao curso.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

IMPACTO DA ESPECIALIZAÇÃO DE DADOS DE FALHA EM ANÁLISES PROBABILÍSTICAS DE SEGURANÇA DE PLANTAS DE PROCESSO.

Antonio Carlos de Oliveira Ribeiro

Fevereiro de 2005

Orientador: Paulo Fernando Ferreira Frutuoso e Melo

Programa: Engenharia Nuclear

O objetivo desta tese é apresentar a aplicabilidade da inferência bayesiana em estudos de confiabilidade, usada para especializar dados de falha em análises de segurança, demonstrando o impacto do uso da mesma em estudos de análise de riscos (EAR) ambientais efetuados em plantas industriais de processo, assim como o seu uso em análises probabilísticas de segurança (APS) em instalações nucleares.

Com esta abordagem, encontramos um modo sistemático e auditável de diferenciarmos um EAR de uma instalação industrial que possui uma boa estrutura de projetos e de manutenção de outra que apresente um baixo nível de qualidade nestas áreas. Geralmente, as evidências de taxas de falha e por conseguinte, as freqüências de ocorrência dos cenários de origem dos riscos considerados num EAR, são retiradas de bancos de dados genéricos ao invés da instalação analisada. No caso de APS de instalações nucleares, o uso da especialização dos dados para a instalação em análise já é uma realidade.

Ao longo da tese apresentam-se também algumas limitações e cuidados a serem tomados para que sejam evitados erros no uso da metodologia.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

IMPACT OF FAILURE DATA ESPECIALIZATION ON PROBABILISTIC SAFETY ASSESSMENTS OF PROCESS PLANTS

Antonio Carlos de Oliveira Ribeiro

February 2005

Advisor: Paulo Fernando Ferreira Frutuoso e Melo

Department: Nuclear Engineering

This thesis intend to show the Bayesian inference application to reliability studies, for updating failures rates in safety analysis. The impact of this updating in environmental quantitative risks assessments (QRA) for industrial process plants and also in probabilistic risks assessments for nuclear installations is presented with examples.

The Bayesian approach allows for a structured and auditable way of showing the difference between an industrial installation with a good design and maintenance program from another one that shows a low quality level in these areas. In general the evidence from failures rates and the frequency of occurrence of scenarios considered in QRA, are taken from generic data banks, instead of operational data from the installation under analysis. This is not the case of nuclear installations where the updating of data with the both information is a common practice.

A discussion is also presented on the limitations of this methodology along with the lessons learned from its application.

ÍNDICE

pág.

CAPÍTULO 1

Introdução

1.1-Introdução

A realização de uma Análise Probabilística de Segurança (APS) ou um Estudo de Análise de Riscos (EAR) de uma instalação parte da análise de eventos que isoladamente ou em seqüência, definidas em árvores de falhas [1], levam à degradação do núcleo de um reator nuclear ou a fatalidades em caso de exposição de populações a nuvens tóxicas oriundas de plantas de processo[2] por exemplo.

Devido à insuficiência de informações, existem incertezas associadas a respeito do valor real das taxas de falha dos componentes destes eventos. Essas informações são obtidas então a partir do nosso conhecimento do dispositivo em análise. Por sua vez, este conhecimento envolve as especificações técnicas de projeto e construção, as práticas de manutenção utilizadas, técnicas de medição e definições das taxas de falha, o ambiente operacional ou a opinião de peritos concernentes ao comportamento operacional, o nosso próprio julgamento e experiência, a experiência operacional de dispositivos similares e o histórico operacional do próprio dispositivo em questão.

Além disso, existem limitações matemáticas para o uso dos modelos para estimar taxas de falhas. A primeira delas é que as taxas de falha dos componentes são estimadas pontualmente baseadas em dados disponíveis, sendo válidas para as condições a partir das quais os dados são obtidos e dos conjuntos em que operam. Algumas extrapolações para predição de taxas de falhas são possíveis, mas a natureza empírica inerente a estes modelos pode ser severamente restritiva. Uma restrição a mais na predição de taxas de falha é a sua dependência no correto e consciente uso das informações e aplicação dos modelos pelo usuário.

Os que a aplicam corretamente encontrarão nos modelos uma excelente ferramenta, não se deve ver nos modelos simplesmente um meio de se atender a um número, talvez isso possa até ser atingido mas haverão sérios prejuízos nos sistemas, que inicialmente podem não ser claramente percebidos, mas logo se detecta o impacto desse equívoco nos sistemas.

Podemos encontrar modelos propostos para estimação da confiabilidade inerente de componentes em [2] e [3]. Outro bom modelo observado durante as pesquisas efetuadas na elaboração deste trabalho é o de [4] onde se encontram opções para se decidir sobre estimadores de taxa de falha, modelo este utilizado na estruturação de bancos de dados genéricos tais como [5].

A despeito das diversas dificuldades e incertezas associadas aos modelos de predição de taxas de falha atualmente utilizadas, podemos usar um modelo que permite especializar os dados de falha de uma planta de processo, que é a análise bayesiana de confiabilidade [6],[7] e [8].

Este trabalho tem o objetivo de apresentar as limitações do uso desta metodologia , a partir das informações disponíveis a priori, apresentando as diversas formas de se obter estas informações, comparando ainda as técnicas da inferência estatística pela teoria da amostragem clássica e pela inferência bayesiana, apresentando também um tratamento matemático para as distribuições de probabilidade a priori, verossimilhança e a posteriori. Apresentando também o Teorema de Bayes com o seu uso para distribuições de probabilidades discretas e contínuas e oferece exemplos de especialização de dados com a aplicação da inferência bayesiana, mostrando o seu impacto em uma caso de EAR de uma indústria de processo e em APS de instalações nucleares.

Para o caso de EAR, utilizou-se um cenário típico de uma planta de processo, isto é, uma vazamento de amônia líquida, causando uma nuvem tóxica, que expõe pessoas a fatalidades a longas distâncias a partir do ponto do vazamento.

Efetuou-se uma modelagem com simulação computacional da dispersão da nuvem tóxica e obteve-se o risco individual de uma pessoa fora da instalação (considerando uma lesão fatal) e comparando este risco antes e depois da especialização dos dados.

Para aplicação em uma APS de uma instalação nuclear, exemplificamos com algumas evidências utilizadas em APS efetuada para um usina nuclear similar a Angra I , onde se compara a freqüência de ocorrência de eventos iniciadores considerados na análise, com outras instalações similares. Apresentam-se as considerações efetuadas a respeito das distribuições de probabilidade utilizadas na especialização dos dados e os resultados obtidos.

Conclui-se resumindo as vantagens do emprego da metodologia pelos orgãos reguladores de EAR e APS, apresentando também cuidados e recomendações a serem tomadas para o uso satisfatório da especialização dos dados.

CAPÍTULO 2

Metodologia utilizada para a especialização de dados

2.1- Qual é o problema ?

Um caso muito comum em estudos de confiabilidade é o de especializar dados de falha a serem utilizados para uma planta em particular . O ponto de partida (informação a priori) , nesses casos, é o emprego de bancos de dados genéricos, e então incorporar a eles os dados de falha específicos da planta (função de verossimilhança). Desta maneira, teremos como resultado uma informação a posteriori, que incorpora a evidência colhida na planta. Poderíamos perguntar, neste caso, por que não usar os dados da planta somente. Contudo, os dados da planta podem significar, por exemplo, nenhuma falha em um dado período de operação, ou seja, se um componente não falhou em um dado período, a que conclusão podemos chegar acerca, por exemplo, da sua taxa de falha? Ainda assim, esta evidência é valiosa e precisamos aproveitá-la.

Ainda acerca dos dados iniciais para a inferência bayesiana, podemos obtê-los a partir de bancos de dados genéricos, dados da própria planta ou ainda da opinião de especialistas e podemos inferir os seguintes comentários ou considerações sobre as vantagens e desvantagens destas origens:

Dados genéricos:

-São número baseados num grande de ocorrências de falhas;

-Contêm uma grande variedade de modelos, fabricantes e tipos de componentes;

-Raramente são conservativos por natureza, isto é, vários dados genéricos foram criados usando fontes que restringem um número de falhas reportáveis e que subestimam as verdadeiras características de falha dos componentes;

-Resultam em estimações baseadas em grandes populações de componentes com diferentes políticas de manutenção, diferentes intervalos de testes, diferentes processos e condições ambientais, o que acarreta uma maior incerteza;

-Os dados extraídos de bancos de dados genéricos são aceitos na comunidade internacional, por falta de informações mais precisas.

Dados específicos da planta:

-Representam melhor o comportamento de falha dos componentes considerados;

-A inclusão de dados específicos da planta em EAR e APS aumenta a credibilidade destes estudos. Além disso permite, a comparação de performance das plantas;

-Nem sempre é possível colher dados para todos os componentes dos eventos iniciadores apontados pelos modelos de confiabilidade. Componentes altamente confiáveis, tais como instrumentação, circuitos de controle e outros componentes com longo tempo médio entre falha, podem nunca ter falhado na história da planta ou no espaço de tempo considerado. A falta de histórico de falha torna difícil a estimação do valor verdadeiro da taxa de falha. Além disso, é praticamente impossível obter um espaço amostral significativo de exposição de certos componentes (por exemplo o número de demandas atribuídas a um relé), usando dados somente da planta.

Opinião de especialistas.

-Também conhecido como julgamento de engenharia. Na ausência de informações de bancos de dados externos e da própria planta, podemos usar a opinião de especialistas, entretanto, isto depende da qualidade dos especialistas entrevistados e do método de obtenção das informações.

Para efeito de tratamento do problema em questão, como já dito, iremos combinar e especializar dados a partir de bancos de dados de falha genéricos incorporando os dados de falha específicos da planta.

2.2-Descrição da metodologia .

2.2.1- Considerações gerais.

Não serão detalhados os modelos estatísticos das distribuições utilizados em estudos de confiabilidade. Com relação à parametrização destas distribuições, que compõe a base metodológica matemática do teorema de Bayes, recomendamos a consulta a Amaral Netto [6] e Santos [7].

Entretanto, antes de aplicarmos a metodologia bayesiana a especialização de dados de falha, faremos algumas abordagens conceituais sobre as funções de densidade de probabilidade (fdp) que estruturam os modelos para estimação de taxas de falha.

2.2.2-Experimentos, variáveis aleatórias e espaços amostrais.

Considerando que experimentos são processos repetitivos para os quais os resultados são incertos, não se espera que todas as tentativas de um experimento atinjam os mesmos resultados, devido às incertezas associadas aos processos. Então, identificamos o grupo de saídas possíveis do espaço amostral, denominado de S, do experimento. O espaço amostral do experimento pode incluir pontos de uma forma contínua ou pontos discretos distintos. Por exemplo, o tempo de vida de um bulbo de lâmpada pode ser qualquer numero real positivo, entretanto o interruptor tem um comportamento do tipo discreto, podendo estar ligado ou desligado. Podemos definir o conceito de uma variável aleatória (v.a.) como qualquer valor de um a função definida no espaço amostral do

experimento. O grupo de valores assumido pela v. a. é chamado de espaço amostral da v.a.. Uma v.a. é dita discreta se o seu espaço amostral é finito ou infinito enumerável (é o caso do interruptor). Uma v.a. que assume valores num intervalo (ou intervalos) de valores é uma v.a. contínua (é o caso do bulbo da lâmpada) [8].

2.2.3 – Distribuições estatísticas disponíveis para uso em modelagem para estimação de taxas de falha.

Existem diversas distribuições estatísticas bem conhecidas que podem modelar tanto um espaço amostral contínuo quanto um espaço amostral discreto. Uma boa referência para consulta sobre estas distribuições além de [6] e [7] já citadas anteriormente, é o apêndice G do AIChE-Guideline [9]. Alguns exemplos das distribuições estatísticas utilizadas são relacionados na Tabela 1.

Tabela 1 : Resumo com descrição e aplicabilidade de algumas distribuições estatísticas.

Distribuição	Descrição	Aplicabilidade
Modelos discretos		
Binomial	Distribuição de probabilidade do número de falhas em *n* demandas independentes nas quais em cada demanda existem dois resultados possíveis – falha ou sucesso	Apropriada para situações nas quais um equipamento opera somente na demanda, e se atributos constantes e independentes são adequados.
Poisson	Distribuição de probabilidade de um evento isolado, ocorrendo em um numero específico de vezes em um dado período quando a taxa de ocorrência é fixa . A ocorrência deve ser afetada somente pela chance. Exemplo disto é o número de toques telefônicos num período ou o número de defeitos num sistema.	Apropriada para situações onde um evento pode ocorrer em qualquer momento.
Modelos Contínuos		
Exponencial	Algumas vezes referida como distribuição 'exponencial negativa' . A distribuição é caracterizada por um simples parâmetro, λ, a taxa de falha é assumida como constante ao longo do tempo.	Usualmente assumida na ausência de outras informações; por isso é a distribuição mais usada em trabalhos de confiabilidade. Não é apropriada para modelagens de início ou fim de vida útil de equipamentos
Normal	A distribuição normal é a distribuição simétrica mais conhecida, tem dois parâmetros, a média, μ, e o desvio padrão, σ. É a distribuição geralmente usada para caracterizar dados que giram em torno de um valor médio e desviam deste valor em quantidades absolutas.	Largamente usada em engenharia da confiabilidade, particularmente para tratar certos tipos de falhas ligadas a partes de processos automatizados, fenômenos físicos naturais, e equipamentos que aumentam a taxa de falha no tempo.
Lognormal	É a distribuição naturalmente usada quando desvios de um modelo são mais influenciados por fatores, proporções ou percentagens, do que por valores absolutos tais como na distribuição normal. Uma boa abordagem matemática para esta distribuição é encontrada em [18]	É usada para tratar certos tipos de falhas que crescem no tempo, tais como fadiga de um metal , vida útil de isolamento de componentes elétricos, e falhas de processos contínuos tais como processos químicos. É muito usada em sistemas reparáveis e também para descrever a faixa possível de falhas de um equipamento quando há uma certa incerteza sobre estes valores. Foi a distribuição usada no 'Reactor Safety Study' (Rasmussen, 1975)

2.2.4 - O teorema de Bayes.

O ponto de partida para se chegar ao teorema é a definição de probabilidade condicional. Dada uma coleção de eventos A_i, i=1,2,...,N e um outro evento B, teremos:

$$P(A_i \cap B) = P(A_i|B)P(B) \quad , \tag{1.1}$$

e da mesma forma:

$$P(B \cap A_i) = P(B|A_i)P(A_i) \tag{1.2}$$

Das equações (1.1) e (1.2), resulta:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} \quad , \tag{1.3}$$

Podemos interpretar o resultado expresso pela eq. (1.3) da seguinte forma: $P(A_i)$ é a probabilidade de ocorrência de um evento Ai e P(B) representa o resultado de um experimento. A probabilidade $P(A_i|B)$ é a nossa estimativa atualizada, a qual leva em conta o resultado do experimento mencionado. Para que obtenhamos este resultado, devemos ser capazes de estimar a probabilidade de um resultado experimental B, dado A_i.

Para determinar P(B), vamos considerar que A só pode ser representado pelos eventos A_1, A_2,..., An. Como A só pode ser representado por um evento A_i de cada vez, os eventos A_i são mutuamente excludentes, ou seja:

$$\sum_{i=1}^{N,} P(A_i) = 1 \tag{1.4}$$

e, da mesma forma:

$$\sum_{i=1}^{N} P(A_i|B) = 1 \tag{1.5}$$

Somando a eq. (1.3) membro a membro sobre todos os eventos A_i e substituindo nela a eq. (1.5), obteremos:

$$P(B) = \sum_{i=1}^{N} P(B|A_i)P(A_i), \tag{1.6}$$

que é conhecida como probabilidade total.

Dessa maneira, a eq. (1.3) pode finalmente ser escrita como:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^{N} P(B|A_i)P(A_i)}, \tag{1.7}$$

que é o teorema de Bayes para variáveis aleatórias discretas.

Da mesma forma, podemos aplicar o teorema a variáveis aleatórias contínuas. Neste caso utilizaremos funções de densidade de probabilidade condicionais. Para tal necessitamos de uma versão apropriada da eq. (1.7) na qual A e B sejam contínuas.

Para trabalhar com variáveis contínuas vamos utilizar a notação X para o evento A e a notação Y para o evento B. Neste caso, necessitaremos da densidade conjunta de X e Y e também das densidades condicionais, ou seja:

$$f(x,y) = f_X(x|y_n)f_Y(y_n) \tag{1.8}$$

e

$$f(x,y) = f_Y(y_n|x)f_X(x) \tag{1.9}$$

a partir das quais podemos escrever:

$$f_X(x|y_n) = \frac{f_Y(y_n|x)f_X(x)}{f_Y(y_n)} \tag{2.0}$$

Como $f_X(x|y_n)$ é uma função de densidade de probabilidade, temos:

$$\int_{-\infty}^{\infty} f_X(x|y_n)dx = 1 \tag{2.1}$$

Substituindo a eq. (2.0) na eq. (2.1), teremos:

$$\int_{-\infty}^{\infty} f_X(x|y_n)dx = \frac{1}{f_Y(y_n)} \int_{-\infty}^{\infty} f_Y(y_n|x)f_X(x)dx = 1 \tag{2.2}$$

de onde concluímos que:

$$f_Y(y_n) = \int_{-\infty}^{\infty} f_Y(y_n|x)f_X(x)dx \tag{2.3}$$

Substituindo a eq. (2.3) na eq. (2.0) teremos finalmente:

$$f_X(x|y_n) = \frac{f_Y(y_n|x)f_X(x)}{\int_{-\infty}^{\infty} f_Y(y_n|x)f_X(x)dx} \tag{2.4}$$

que é a apresentação do teorema para variáveis aleatórias contínuas.

Vamos discutir a eq. (2.4) com um exemplo, tomando como base uma situação concreta.

Desejamos estimar um parâmetro *x* de uma distribuição de probabilidade, por exemplo: uma taxa de falha, um tempo médio para falhar (MTTF), uma vida característica, a resistência média de uma estrutura, um parâmetro de forma, etc. Como não temos certeza do valor desse parâmetro x, vamos considerá-lo como uma variável aleatória X, cuja função de densidade $f_X(x)$ expressa a nossa incerteza a esse respeito.

Especificamente, vamos considerar a incerteza acerca de uma taxa de falha constante (modelo exponencial). Admitindo que a nossa incerteza pode ser expressa por meio de uma distribuição normal e representando a variável aleatória por Λ, teremos:

$$f_{\Lambda}(\lambda) = \frac{1}{\sqrt{2\pi}\sigma_o} \exp\left[-\frac{1}{2\sigma_o^2}(\lambda - \lambda_o)^2 \right] \tag{2.5}$$

onde λ_o é a nossa melhor estimativa pontual. A variância indica o nosso grau de incerteza. A eq. (2.5) representa a distribuição a priori.

A densidade $f_y(y_n/x)$ é a probabilidade de que um experimento que pode resultar em diversos valores distintos de Y apresentará um resultado particular y_n, dado que o parâmetro tomou um valor *x*.

Admitiremos, como exemplo, que N componentes são ensaiados durante um intervalo de tempo t_o e que, ao final do ensaio, *n* tenham falhado. Vamos considerar que y_n é a variável discreta *n*. Admitindo ainda que os ensaios sejam independentes, a função de densidade necessária é:

$$f_N(n|\lambda) = \binom{N}{n} p(\lambda)^n [1 - p(\lambda)]^{N-n} \tag{2.6}$$

onde p(λ) é a probabilidade de falha de um componente que opera durante um intervalo de tempo t_o, que é a função de verossimilhança, ou seja, no caso do modelo exponencial de falha:

$$p(\lambda) = 1 - e^{-\lambda t_o}. \tag{2.7}$$

Deste modo, teremos:

$$f_N(n|\lambda) = \binom{N}{n}(1 - e^{-\lambda t_o})^n \exp[-(N-n)\lambda t_o] \tag{2.8}$$

Introduzindo as alterações de variáveis comentadas, a eq. (2.4) pode ser escrita como:

$$f_\Lambda(\lambda|n) = \frac{\left(1-e^{-\lambda t_o}\right)^n \exp[-(N-n)\lambda t_o] f_\Lambda(\lambda)}{\int_{-\infty}^{\infty} \left(1-e^{-\lambda' t_o}\right)^n \exp\left[-(N-n)\lambda' t_o\right] f_\Lambda(\lambda')\, d\lambda'} \tag{2.9}$$

onde $f_\Lambda(\lambda)$ é dada pela eq. (2.5) para o exemplo sob discussão. É importante frisar que a incerteza pode ser expressa por qualquer distribuição de probabilidade. A escolha da melhor distribuição, pode ser ditada por um teste de aderência.

A função de densidade $f_\Lambda(\lambda|n)$ é a nossa estimativa da incerteza de λ, modificada para levar em conta os dados colhidos no ensaio, em termos do número de falhas. Podemos também trabalhar com os próprios tempos de falha para fazer a atualização da função de densidade, como veremos mais adiante.

Uma vez obtida a função de densidade a posteriori, eq.(2.9), a qual incorpora as evidências do ensaio, então podemos atualizar a nossa estimativa pontual do valor médio da taxa de falha:

$$\lambda_1 = \int_{-\infty}^{\infty} \lambda f_\Lambda(\lambda|n) d\lambda \tag{3.0}$$

De maneira análoga, podemos também calcular a variância da distribuição a posteriori e verificar, dessa forma, se obtivemos uma maior precisão com os resultados do ensaio:

$$\sigma_1^2 = \int_{-\infty}^{\infty} \lambda^2 f_\Lambda(\lambda|n)d\lambda - \lambda_1^2, \tag{3.1}$$

lembrando que

$$\sigma^2 = E(X^2) - [E(X)]^2,$$

onde $E(X^n) = \int_{-\infty}^{\infty} x^n f(x)dx$ é o momento de ordem n da variável aleatória X.

Antes de apresentarmos a aplicação do teorema com casos práticos de EAR e APS, faremos uma abordagem conceitual para solidificar o entendimento da metodologia, mostrando as nuances e limitações da mesma.

2.2.5- Inferência Estatística.

Em Spojtvoll [4] apresenta-se uma metodologia interessante para se decidir sobre qual o melhor estimador de um parâmetro a partir da quantidade de dados disponíveis. Podemos encontrar também em 'Martz, Waller'[8] no capítulo 3, uma boa abordagem sobre inferência estatística, onde se apresenta a técnica de estimação pontual, a função de verossimilhança e o estimador de máxima verossimilhança, estimadores não tendenciosos , assim como a definição de intervalos de confiança para esses estimadores. O que vale ressaltar é o conceito de intervalo de confiança para um estimador de uma taxa de falha assim como o conceito de intervalo de probabilidade, que é o conceito para o teorema de Bayes , uma vez que neste o estimador é tratado como uma variável aleatória (v.a.).

2.2.6 – A inferência bayesiana em confiabilidade.

.O famoso artigo do Reverendo Thomas Bayes (1702-1761) foi publicado postumamente em 1763 (após sua morte) e fundamenta a base da metodologia denominada ' Inferência estatística bayesiana'. Devido à sua importância, o artigo foi novamente publicado em 1958. Em função da sensibilidade e dificuldades com alguns métodos de

inferências estatísticas, os fatores seguintes têm contribuído para o recente ressurgimento desta metodologia: algumas soluções de inferências estatísticas baseiam-se em suposições que carecem de convenientes soluções matemáticas, é difícil desenvolver um plano de amostragem que forneça um bom resultado mediante um alto nível específico de qualidade dos dados se na prática não se consegue obter este nível. A inclusão de dados subjetivos pode ser efetuada pela metodologia bayesiana [8], pelo fato da mesma poder explicitar o modo como ela incorpora estes dados. Para tal podemos lançar mão ainda do surgimento de computadores cada vez mais velozes para cálculos, ou modelagens matemáticas que facilitem o tratamento dos dados.

2.2.7- Inferência pela teoria da amostragem versus inferência bayesiana.

Existem diferenças entre a teoria da amostragem e métodos de inferência bayesiana. Tomemos como exemplo um estudo sobre a vida útil de componentes eletrônicos em determinadas condições de operação. Para este estudo, vamos assumir que as observações são independentes e exponencialmente distribuídas com uma vida média θ. A distribuição conjunta de probabilidade de um experimento amostral de n observações $Y' = (y_1, \ldots, y_n)$ será então:

$$f(y/\theta) = \frac{1}{\theta^n} \exp\left[-\frac{1}{\theta} \sum_{i=1}^{n} y_i\right], \qquad 0 < y_i < \infty, \tag{3.2}$$

e estamos interessados em fazer inferências sobre θ dados os n valores disponíveis.

2.2.7.1-Inferências baseadas na teoria da amostragem

Na abordagem usada na teoria da amostragem a vida média desconhecida θ é assumida como uma constante fixa. Um estimador pontual $\boldsymbol{\Theta}$ (Y), que é uma função de um grupo de dados Y , é escolhido por algum método, tal como: Máxima Verossimilhança,

mínimos quadrados, ou métodos dos momentos. Por exemplo, os estimadores de máxima verossimilhança e pelo método dos momentos de $\boldsymbol{\Theta}$ são:.

$$\boldsymbol{\Theta} = \boldsymbol{\Theta}(\mathrm{Y}) = \sum_{i=1}^{n} \frac{Y_i}{n}$$

Ao imaginarmos a amplitude de todas as possibilidades hipotéticas dos vetores de dados $Y_1, Y_2, \ldots$ que poderiam ser gerados pela eq. (3.2), para um dado valor de $\boldsymbol{\Theta}$, obteríamos a distribuição amostral correspondente de $\boldsymbol{\Theta}(\mathrm{Y})$. Por exemplo, a distribuição amostral de $U=2n\,\boldsymbol{\Theta}/\,\theta$ é uma distribuição do qui quadrado de parâmetro 2n com uma f.d.p dada por:

$$f(u) = \frac{1}{2^n \Gamma(n)} u^{n-1} \exp\left(-\frac{u}{2}\right), \qquad 0 < u < \infty. \tag{3.3}$$

Um intervalo de confiança para o estimador de θ poderia ser calculado para se obter uma idéia de como o valor $\boldsymbol{\Theta}(\mathrm{Y})$ esta afastado do valor real θ , que gerou as n observações. Por exemplo o Intervalo de Confiança Total (*ICT*), 100(1-γ) %, do estimador para θ é dado por

$$100(1-\gamma)\%ICT \quad para \quad \theta : \quad \left[\frac{2n\Theta}{\chi^2_{1-\gamma/2}(2n)}; \frac{2n\Theta}{\chi^2_{\gamma/2}(2n)} \right] \tag{3.4}$$

No caso de amostragens repetitivas, onde os intervalos de confiança são computados para cada amostra, os intervalos de confiança computados poderiam incluir o valor real de θ em 100(1- γ)% das amostras. É importante lembrar que o intervalo de confiança não pode ser interpretado como uma afirmação probabilística sobre θ , posto que θ não é uma v.a.. Por exemplo, podemos obter um intervalo de confiança do tempo médio entre falhas (MTTF) de um componente de um evento iniciador de uma APS de uma usina nuclear a partir da combinação dos intervalos de confiança dos MTTF's de seus componentes . Uma

vez que a confiança não é uma probabilidade relativa a θ , os métodos para combiná-los não são tão bem definidos como no caso de probabilidades.

A teoria da inferência por amostragem conduz a exemplos indutivos (Fig.1). , por exemplo, o valor final superior do *ICT* da eq.(3.3) é o maior valor que θ pode assumir sem a observação do estimador $\mathbf{\Theta}$(Y) com uma probabilidade menor que $\gamma/2$ disto acontecer, de acordo com a eq.(3.4). Uma afirmação similar, pode ser feita para o valor final inferior do *ICT*.

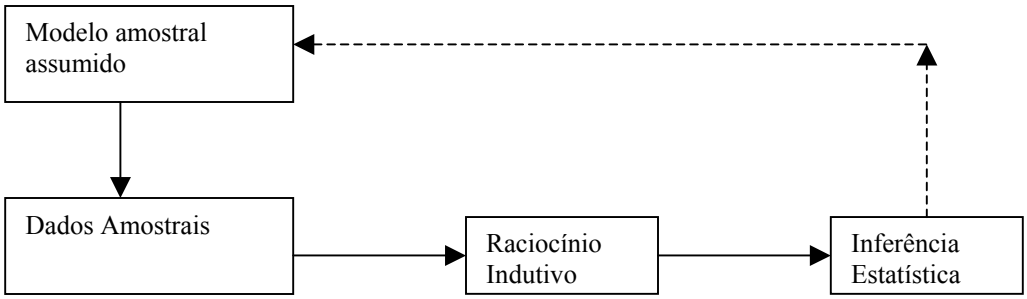


Figura 1: Inferência baseada na teoria da amostragem

2.2.7.2- A inferência Bayesiana.

O raciocínio da metodologia bayesiana é muito mais direto, é dedutivo. Para atingir esta abordagem direta, a vida média de θ é assumida como uma v.a. com uma f.d.p. a priori $g(\theta)$. Esta distribuição expressa o estado de conhecimento ou ignorância sobre θ antes da amostra ser analisada.

Dada a distribuição a priori, um modelo de probabilidade $f(y \mid \theta)$ e o valor de y, o teorema de Bayes é usado para calcular a f.d.p posterior $g(\theta \mid y)$ de $\mathbf{\Theta}$, dado o valor de y.

Para exemplificarmos, vamos supor que a distribuição a priori de **Θ** seja a uniforme num intervalo θ_1 até θ_2 , onde $0< \theta_1< \theta_2 < \infty$:

$$g(\theta) = \frac{1}{\theta_1 - \theta_2} \quad , 0< \theta_1 \leq \theta \leq \theta_2 < \infty. \tag{3.5}$$

Acredita-se a priori que θ é um valor do intervalo especificado (θ_1 , θ_2). Usando o teorema de Bayes , a partir da eq. (3.2) e da eq. (3.5), a distribuição posterior resultante é calculada como:

$$g(\theta/y) = \frac{\left(\sum y_i\right)^{n-1} \exp\left(\sum y_i/\theta\right)}{\theta^n\left[\Gamma\left(n-1,\sum y_i/\theta 1\right)-\Gamma\left(n-1,\sum y_i/\theta 2\right)\right]}, \qquad \theta_1 \leq \theta \leq \theta_2 , \quad n>1, \tag{3.6}$$

onde o somatório de y_i vai de 1 a n e Γ (a,z) é a função gama incompleta definida como

$$\Gamma(a,z) = \int_a^z y^{a-1} \exp(-y)dy, \qquad a > 0 \tag{3.7}$$

Argumentos dedutivos são usados com a distribuição a posteriori para se fazer uma inferência bayesiana sobre θ. Como exemplo, uma estimador pontual para θ é a média da distribuição posterior da eq. (3.7) dada por :

$$E(\mathbf{\Theta}/y) = \sum y_i \left[\frac{\Gamma(n-2,\sum y_i/\theta_1)-\Gamma(n-2,\sum y_i/\theta_2)}{\Gamma(n-1,\sum y_i/\theta_1)-\Gamma(n-1,\sum y_i/\theta_2)} \right] \qquad n>2.$$

Os intervalos dos estimadores para θ são também calculados a partir da distribuição posterior

$$\int_{-\infty}^{\theta^*} g(\theta/y)d\theta = \frac{\gamma}{2} \tag{3.8}$$

e

$$\int_{\theta_*}^{-\infty} g(\theta/y)d\theta = \frac{\gamma}{2} \tag{3.9}$$

A f.d.p. a priori na análise baeysiana geralmente incorpora noções subjetivas, uma vez que a freqüência dos valores de θ raramente é conhecida. É a distribuição que representa o grau de conhecimento de θ antes que os dados y sejam obtidos.

Uma distinção da inferência bayesiana é o fato de que ela torna explícito o uso da informação a priori na análise. Isto contrasta com a abordagem da teoria da amostragem na qual esta informação é considerada somente de maneira informal.

A Figura 2 retrata a metodologia da inferência bayesiana. O processo começa com um modelo amostral validado. Uma f.d.p. a priori é postulada para estes parâmetros desconhecidos no modelo amostral e a inferência bayesiana é aplicada, os dados da amostra e a f.d.p. a priori são então combinados pelo teorema, o raciocínio dedutivo é usado em conjunto com o resultado da distribuição posteriori para produzir a inferência desejada sobre os parâmetros assumidos no modelo amostral.

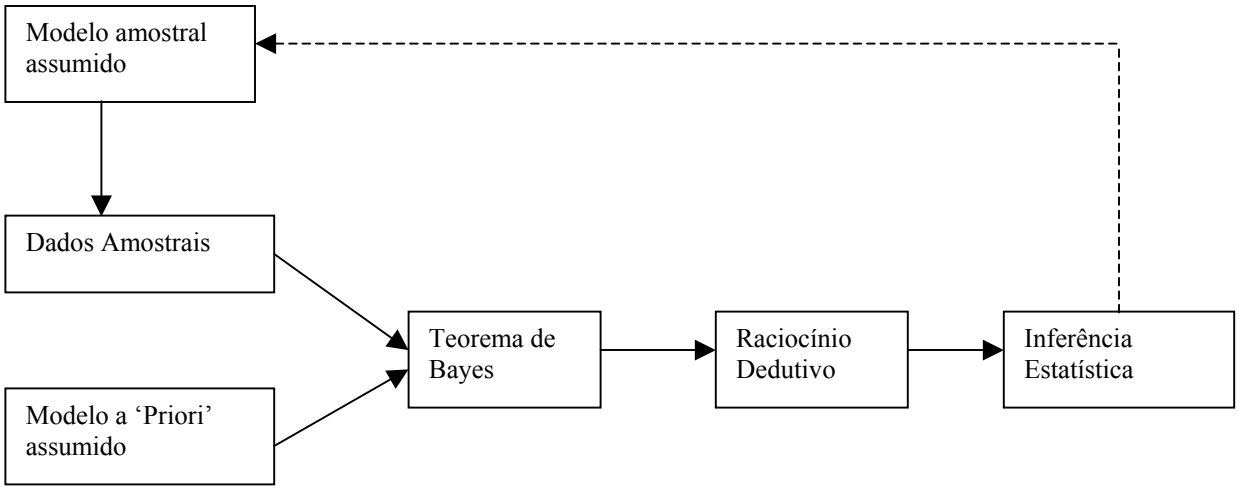


Fig. 2 – Inferência Bayesiana

Podemos apresentar mais duas diferenças entre a teoria da amostragem e a abordagem bayesiana. A primeira é que a inferência estatística baseada na teoria da amostragem é usualmente mais restritiva pelo fato do uso exclusivo dos dados da amostra. O uso da experiência passada quantificada pela distribuição a priori produz uma inferência mais informativa nos casos em que a mesma realmente reflete a variação do parâmetro. Isto quer dizer que um alto grau de acerto desta inferência depende da qualidade dos dados incorporados pela distribuição a priori. A segunda é que a metodologia bayesiana normalmente requer um espaço amostral menor para atingir a mesma qualidade dos resultados obtidos pela teoria da amostragem. Em muitos casos esta é a grande motivação prática para o uso da metodologia bayesiana e conseqüentemente da informação a priori.

2.2.8- A distribuição a Priori.

Considere-se por exemplo um problema de inferência estatística no qual as observações são tomadas de um f.d.p., $f(x|\theta)$, onde θ é um parâmetro com um valor desconhecido. Então assumimos que o valor desconhecido de θ deve pertencer a um espaço amostral $\mathbf{\Theta}$. O problema é tentar determinar, baseado em informações da f.d.p. de $f(x|\theta)$, onde está o provável valor de θ no espaço amostral paramétrico $\mathbf{\Theta}$. Em vários casos, depois de várias informações disponíveis provenientes de $f(x|\theta)$, o experimento ou estatística será capaz de resumir sua prévia informação ou conhecimento sobre onde em $\mathbf{\Theta}$ o valor de θ é provável de ocorrer a partir da construção de uma distribuição de probabilidade de θ em $\mathbf{\Theta}$**.** Em outras palavras, antes que qualquer dado experimental seja coletado ou observado, o experimento ou experiência passada do analista o levará a acreditar que θ é mais provável de estar em determinadas regiões de $\mathbf{\Theta}$ do que em outras. Podemos assumir que a verossimilhança relativa das diferentes regiões pode ser expressa

em termos de uma distribuição de probabilidade em $\boldsymbol{\Theta}$. Esta é chamada de distribuição a priori de θ , porque ela representa a verossimilhança relativa de que o verdadeiro valor de θ se encontra em cada região de $\boldsymbol{\Theta}$ antes de serem observados quaisquer valores de $f(x|\theta)$.

Esta pode ser uma limitação da aplicação do teorema de Bayes. Da escolha correta desta f.d.p. dependerá o resultado da inferência estatística bayesiana. A distribuição a priori reflete uma informação importante a respeito dos valores possíveis da taxa de falha antes de termos obtido a nossa própria evidência experimental.

Para a maior parte dos dispositivos usados na área nuclear para elaboração de uma APS ou na indústria de processo para elaboração de um EAR, não existe uma fonte de dados num formato ideal para representar a distribuição a priori no Teorema de Bayes. Em geral, as fontes de dados disponíveis não especificam as circunstâncias em que a falha ocorreu, nem o ambiente para qual os dados por elas fornecidos se aplicam. Portanto, a maioria destas fontes adota uma família de distribuições para representar a distribuição de taxa de falha e fornece, para cada modo de falha, os percentuais que definem um intervalo de confiança de 90% sobre esta distribuição (*ICT*). Em outras palavras, estabelecem valores máximo e mínimo para a variação de taxa de falhas e ajustam uma distribuição entre valores extremos. Tais distribuições são interpretadas como curvas representativas da população, ou seja, curvas da variabilidade populacional [10]. Esta variabilidade de taxa de falhas é devida, dentre outros fatores, a diferentes fabricantes, diferentes modelos, diferentes condições de operação, de manutenção, de ambiente operacional e às diferenças devidas à flutuação randômica inerente entre componentes idênticos.

Se todo o nosso conhecimento acerca de um dispositivo resumir-se à informação de que ele é um membro da população, então adotamos a curva de variabilidade populacional como distribuição a priori. Porém, na maioria dos casos dispomos de algumas informações, baseadas no projeto do dispositivo, em experiências em aplicações similares, no julgamento de peritos e no nosso próprio conhecimento a respeito do dispositivo. Estas informações adicionais, usadas conjuntamente com a variabilidade populacional (fontes de dados genéricos), auxiliam na escolha de uma distribuição a priori adequada.

Conforme [2], os dados podem ser genéricos, específicos , objetivos e subjetivos. Quando o dado é representativo da população, o classificamos como genérico, se o dado é proveniente do dispositivo que estamos analisando, constitui um dado específico. Quando é obtido a partir experiências ou de expressões analíticas comprovadas experimentalmente, é classificado como objetivo, caso o dado seja proveniente da opinião de especialistas, ele é classificado como subjetivo.

2.2.8.1- Fontes de dados genéricos:

a) Reactor Satety Study – WASH 1400, Apêndice III [11]

Este relatório contém dados de falhas de dispositivos mecânicos, estimativas da freqüência de ocorrência de eventos iniciadores e dados referentes a falhas de modo comum. Grande parte destes dados foi obtida a partir da opinião de especialistas, que forneceram valores máximos e mínimos recomendados para a taxa de falha. Neste estudo admitiu-se que as incertezas, referentes às taxas de falha dos dispositivos estão distribuídas lognormalmente e os valores mínimos, recomendados e os máximos foram adotados como sendo, respectivamente, os percentis 5%, 50% e 95% de uma distribuição lognormal.

b) NREP- National Reliability Evaluation Program Procedure Guide- Apêndice C [12]:

Os dados de falhas foram obtidos a partir de especialistas e do guia IREP (NUREP/CR-2728) [12]. Para cada modo de falha foi atribuído um valor nominal da taxa de falha e um erro relacionado aos valores superior e inferior de um intervalo de 80%.

O valor nominal foi multiplicado e dividido pelo fator de erro, obtendo-se, respectivamente, os extremos superior e inferior para o intervalo de variação da taxa de falha. Uma distribuição lognormal foi ajustada entre os valores extremos sendo calculado, então, o valor médio desta distribuição [12].

c) OREDA- Offshore Reliability Data Bank [5].

Resume informações de taxas de falhas de dispositivos, de plataformas marítimas do Mar de Norte e do Mar Adriático, utilizados em instalações de processos de exploração e produção de petróleo. A metodologia usada descarta informações do início e do final de vida útil, evidenciando períodos onde as taxas de falha são relativamente constantes. Os modos de falha possuem um intervalo de confiança total de 90% usando procedimentos adequados de estimação das taxas de falha [4]. Utiliza uma f.d.p. do tipo gama para os parâmetros com taxa de falha de comportamento contínuo, não fazendo referências específicas de f.d.p. sobre estimadores de taxa de falha do tipo falha na demanda.

d) AIChE-CCPS Data Bank [9]

O banco de dados foi formado a partir de dados genéricos disponíveis de confiabilidade e de taxas de falha de componentes, incluindo estudos de confiabilidade, trabalhos de pesquisa publicados e bancos de dados de confiabilidade e relatórios governamentais que contêm informações ligadas a processos químicos, nucleares, plataformas marítimas de petróleo e indústrias de petróleo ao redor do mundo.

Utiliza uma f.d.p. lognormal com intervalo de confiança total de 90% para o registro dos dados, recomendando o uso de uma f.d.p exponencial para tratamentos futuros do modo de falha dos componentes.

2.2.8.2- Uso da opinião de especialistas. (Distribuições a priori não informativas)

No caso de não dispormos de informações genéricas, suficientes para escolhermos a distribuição a priori, podemos usar a opinião de especialistas. Deste modo, a quantidade de dados disponíveis pode ser ampliada através da quantificação destas opiniões. Esta quantificação deve ser efetuada através de procedimentos de avaliação apropriados.

A avaliação subjetiva exige que o especialista atribua um valor de probabilidade para o seu grau de crença a respeito do evento analisado. Isto, por sua vez, exige que o especialista tenha a habilidade de formular sua própria opinião e atribuir um valor de probabilidade a este grau de crença associado com a sua opinião. A avaliação subjetiva depende então do grau de conhecimento do especialista e do procedimento usado para avaliar tal conhecimento. O conteúdo depende de quem sabe o que e o quanto sabe, ou seja, depende da experiência técnica do especialista em fazer julgamentos não tendenciosos. O procedimento usado para efetuar a avaliação subjetiva pode afetar significativamente a qualidade da informação obtida [13] e [14].

Os especialistas tendem, freqüentemente, a superestimar o grau de certeza de suas estimativas, sendo que fornecem um julgamento melhor se a definição do evento considerado incluir a descrição precisa dos modos e das circunstâncias da falha. Portanto, deve-se usar um procedimento estruturado de entrevistas e técnicas adequadas para reduzir as tendências na quantificação do julgamento subjetivo.

Uma metodologia interessante para obtenção e uso de informações oriundas de especialistas é o método Delphi [14], onde se combina o conhecimento de um grupo de especialistas a partir de: questionários formulados e comentados pelos próprios especialistas, a garantia do anonimato das informações e em alguns casos dos próprios especialistas, possibilidade de ajustar um ponto de vista baseado no resumo do resultado do grupo e a realização de dois a cinco ciclos iterativos.

Geralmente o uso de informações de especialistas leva a uma limitada informação a priori do parâmetro de interesse, ou seja a informação não é substancial em comparação com a informação que se esperava a partir da amostra de dados. Isto leva a acreditar que há um conjunto de valores cujos elementos têm igual probabilidade de representar o verdadeiro valor do parâmetro em questão.

O método mais empregado, em tal caso, é selecionar como a priori uma distribuição que é localmente uniforme, isto é, aproximadamente distribuída uniformemente no intervalo de interesse. Contudo, há outros modos de se definir distribuições a priori não informativas e uma definição mais geral, inclusive com exemplos práticos, pode ser encontrada em [8].

2.2.8.3- Distribuições a Priori conjugadas

Algumas distribuições de probabilidade quando são usadas como a priori são mais convenientes que outras, quando a amostra é obtida de determinadas funções de verossimilhança. Em tais casos, tem-se o que se chama de a priori conjugadas, isto é, a distribuição posterior é da mesma família a priori. Isto simplifica a análise, pois permite obtermos a distribuição a posteriori analiticamente.

Como exemplo podemos afirmar que caso a evidência experimental obtida seja do tipo temporal, com função verossimilhança calculada através da distribuição de Poisson, e a distribuição a priori pertença à família de distribuições gama, a distribuição a posteriori será também uma distribuição gama [15] e [16]. Da mesma forma podemos afirmar que para uma verossimilhança binomial e uma a priori beta temos uma a posteriori beta. Um outro grupo de conjugadas é a família da distribuição exponencial que são as distribuições: normal, lognormal, binomial, Poisson, e a própria exponencial [7].

2.2.9- A função posteriori

Suponhamos agora que as n varáveis aleatórias $X_{1,.....,}X_n$ de uma amostragem aleatória de uma distribuição para qual a função de densidade de probabilidade (f.d.p) ou a função de probabilidade(f.p.) é $f(x|\theta)$. Suponhamos também que o valor do parâmetro θ é desconhecido e que a f.d.p ou f.p. a priori de θ é $\xi(\theta)$. Para simplificar, devemos assumir que o espaço paramétrico $\mathbf{\Theta}$ é qualquer intervalo de um conjunto de números reais, e que $\xi(\theta)$ é a f.d.p a priori em $\mathbf{\Theta}$. Do mesmo modo assumiremos $f(x|\theta)$ como uma f.d.p em $\mathbf{\Theta}$.

Desde que as varáveis aleatórias $X_{1,.....,}$ X_n de uma amostragem aleatória de uma distribuição para qual a f.d.p é $f(x|\theta)$ a sua f.d.p conjunta $f(X_{1,.....,}\ X_n \mid \theta)$ será dada pela equação :

$$f(X_{1,.....,}\ X_n \mid \theta) = f(x_1 \mid \theta).....f(X_n \mid \theta). \qquad (4.0)$$

Se usarmos a notação vetorial $x = (x_1,.....,X_n)$ então a distribuição conjunta na eq. (4.0) pode ser escrita simplesmente como $f_n(x|\theta)$.

Desde que o parâmetro θ por si só é agora considerado como pertencente a uma distribuição para a qual a f.d.p. é $\xi(\theta)$, a conjunta $f_n(x|\theta)$ pode ser considerada a f.d.p condicional de $X_{1,.....,}$ X_n para uma dado valor de θ. Se multiplicarmos esta f.d.p. condicional pela f.d.p $\xi(\theta)$, obteremos uma f.d.p dimensional (n+1) de $X_{1,.....,}$ X_n e θ na

forma de $f_n(x|\theta)\ \xi(\theta)$. A f.d.p. marginal conjunta de $X_{1,\ldots\ldots,}\ X_n$ pode agora ser obtida pela integração desta f.d.p conjunta em todos os valores de θ. Entretanto, a f.d.p. condicional n-dimensional conjunta $g_n(x)$ de $X_{1,\ldots\ldots,}\ X_n$ pode ser escrita na forma:

(4.1)

Além disso, a f.d.p condicional. de θ fornece $X_1 = x_{1,\ldots\ldots,} X_n = x_{n,}$ a qual podemos denotar por $\xi(\theta|x)$ que deverá ser igual à f.d.p conjunta de $X_{1,\ldots\ldots,}\ X_n$. Por isso temos:

$$\xi(\theta|x) = \frac{f_n(x|\theta)\xi(\theta)}{g_n(x)} \qquad \text{para } \theta \in \mathbf{\Theta} \tag{4.2}$$

A distribuição de probabilidade em $\mathbf{\Theta}$ representada pela f.d.p. condicional da eq (4.2) é então denominada de distribuição a posteriori de θ, porque esta é a distribuição de θ depois dos valores $X_{1,\ldots\ldots,}\ X_n$ serem observados. Similarmente, a f.d.p. condicional de θ na eq. (4.2) é chamada de f.d.p. posterior de θ. Podemos dizer que a f.d.p anterior $\xi(\theta)$ representa a verossimilhança relativa , antes dos valores $X_{1,\ldots\ldots,}\ X_n$ terem sido observados, que o valor real de θ se encontra em cada região variável de $\mathbf{\Theta}$. Isto quer dizer que a f.d.p. $\xi(\theta|x)$ representa esta verossimilhança relativa depois que os valores $X_1 = x_{1,\ldots\ldots,} X_n = x_{n,}$ foram observados.

2.2.10- A Função Verossimilhança.

A função verossimilhança de um parâmetro θ é a função que associa a cada θ pertencente ao espaço $\mathbf{\Theta}$ o valor $p(x/\theta)$. Assim .

$$l(\theta,x) : \Theta \rightarrow R^+ \qquad \theta \rightarrow l(\theta;x) = p(x|\theta)$$

A função verossimilhança associa para um valor fixo de x a probabilidade de ser observado x a cada valor de θ . Assim , quanto maior o valor de l , maiores as chances de encontrarmos o verdadeiro valor de θ relativo ao evento fixado. Portanto ao fixarmos um valor de x e variarmos os valores de θ observamos a plausibilidade (verossimilhança) de cada um dos valores de θ.

Para melhor entendimento, supondo que dispomos de uma amostra aleatória $x = (x_{1,\ldots,}x_n)$ de uma f.d.p. de Poisson com parâmetro θ. Então, a função de verossimilhança de θ é a densidade conjunta de todas as observações, dado o parâmetro θ, ou seja,

$$l(\theta;x) = \prod_{i=1}^{n} f(x_i|\theta) \qquad = \prod_{i=1}^{n} \frac{\theta^{x_i} e^{-\theta}}{x_i!} = \frac{\theta^{n\overline{x}} e^{-n\theta}}{\prod_{i=1}^{n} x_i!} \qquad \text{onde:} \qquad \overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

A partir da definição anteriormente efetuada a respeito da função posterior, podemos afirmar que o denominador do membro direito da eq. (4.2) é simplesmente a integral do numerador sobre todos os possíveis valores de θ. No caso em que todos os possíveis valores $x_{1,\ldots..}, x_{n.}$ não dependam de θ, podemos resumi-los como uma constante, quando o membro direito da eq. (4.2) é considerado uma f.d.p de θ. Podemos entretanto substituir a eq. 4.2 com a seguinte relação:

$$\xi(\theta|x) \propto f_n(x|\theta)\xi(\theta) \tag{4.3}$$

O símbolo de proporcionalidade $\propto$ é usado aqui para indicar que o membro esquerdo só é igual ao direito a partir de uma constante, cujo valor pode depender dos valores observados $x_{1,\ldots..}, x_{n.}$, mas não depende de θ. A constante apropriada para a qual será possível a igualdade dos dois lados na relação da eq. 4.3, pode ser determinada usando o fato de que

$\int_{\Omega} \xi(\theta|x)d\theta = 1$, porque $\xi(\theta|x)$ é uma f.d.p. de θ.

Quando a f.d.p. conjunta das observações de uma amostra aleatória é considerada como uma função de θ, dados os valores $x_{1,\ldots..}, x_{n}$ ela é denominada de função de verossimilhança . Nesta terminologia, a relação da eq. 4.3 estabelece que a f.d.p posterior de θ é proporcional ao produto da função verossimilhança pela f.d.p a priori de θ.

Pelo uso da relação de proporcionalidade da eq. 4.3, é sempre possível se determinar a f.d.p. a posteriori de θ sem ser necessário explicitamente calcular a integral da eq. 4.1. Se reconhecermos que o membro direito da eq. 4.3 igual a uma das f.d.p's anteriormente relacionadas na seção 2.2.3, a menos de um fator constante, então podemos facilmente determinar a constante apropriada que irá converter o membro direito da eq. 4.3 em uma f.d.p adequada de θ. Para isto, já existem vários programas computacionais ou modelos de cálculo que incorporam esta informação como mencionado em [6] e [7].

O que nos motiva a aplicarmos a inferência bayesiana como já vimos é a existência de uma evidência ou experiência que temos sobre um determinado parâmetro, uma taxa de falha por exemplo. Precisamos dar um tratamento estatístico, ou seja, uma f.d.p. para esta evidência e então fazermos uma inferência com a informação a priori disponível sobre este parâmetro e então especializarmos esta informação a qual é a função verossimilhança.

Alguns cuidados devem ser tomados com a escolha da função verossimilhança, assim como o experimento ou amostra a qual a mesma irá representar. Na maioria das plantas químicas existem dados de taxa de falha. Entretanto, eles não são organizados e arquivados de um modo que permita o uso imediato dos mesmos nos EAR. Esta é uma deficiência já comentada em relação a f.d.p a priori. Devemos ressaltar que a coleta de dados e conversão dos mesmos não é uma tarefa trivial, são informações específicas da planta e requerem muitas vezes esforços substanciais para obtê-los. Nenhum tratamento estatístico pode ser eficiente para dados incompletos ou tendenciosos. A chave do sucesso para a formação dos bancos de dados próprios está no uso de mão de obra treinada para obtê-los, procedimentos detalhados para a coleta, tratamento e arquivamento adequado dos mesmos de modo que possuam uma boa rastreabilidade e sejam facilmente auditáveis.

Finalmente, os bancos de dados devem ser estruturados de modo que forneçam uma taxonomia bem estruturada da taxa de falha.

O uso de campos destinados a registros de falhas em "software" de controle de custos e estoque de sobressalentes de manutenção, já existentes na maioria das empresas, resulta em uma boa transparência na rastreabilidade dos dados. Um exemplo de formulário para coleta dos dados de falha pode ser encontrado em [9].

Para falhas sob demanda relatadas, intervalos de confiança podem ser calculados por aproximações usando-se as f.d.p: qui-quadrada, normal, lognormal ou Weibull, dependendo da relação entre N (número de falhas), e D (número de demandas) [9]. A aproximação da qui-quadrada é aplicada por exemplo para grandes valores de D e pequenos valores de N, e ainda quando $D/N > 10$. As aproximações pelas demais distribuições são recomendadas quando os valores de N e D são grandes e ambos N e D/N são maiores do que 10. Para relatos de falha no tempo, a estimação e intervalos de confiança são especificados a partir de distribuições do tipo: exponencial, normal, lognormal, Weibull e gama. As referências [16] e [17] contêm uma abordagem compreensiva de estimação de parâmetros, confiabilidade, taxas de falhas e incertezas associadas a quantidades e à forma de tratamentos dos dados.

Podemos lançar mão ainda de procedimentos de teste que podem verificar os dados coletados, para verificar se estes dados empíricos da amostra selecionada estão de acordo com os dados teóricos adequados para se assumir determinado modelo de f.d.p.. Em [8] podemos encontrar detalhes dos métodos de aderência, do qui-quadrado e o teste de Kolmogorov- Smirnoff.

CAPITULO 3

Resultados

3.1- Exemplos de aplicação

Apresentaremos a seguir dois exemplos para demonstrar a aplicação da metodologia. Inicialmente, mostraremos um exemplo de especialização de taxa de falha para EAR, onde se apresenta a curva de isorisco de uma planta de processo usando a taxa de falha de banco de dados genérico (externo a planta) e a curva após a atualização bayesiana da taxa de falha.

3.2- Aplicação para um EAR

Imaginemos um cenário de uma tubulação de amônia líquida, rompida, numa ponte de tubulações a 5,5m de altura no interior de um complexo industrial, com diâmetro nominal de 100 mm (4"), comprimento de 100 m, interligada a reservatório contendo 30.000 m^3. Entre a tubulação e o vaso, temos uma válvula que fecha automaticamente, atuada por sensor de pressão baixa, que detecta o vazamento, limitando o inventário de produto vazado em aproximadamente 4.130 m^3 . Este volume de produto vazado foi estimado considerando as condições físicas no interior da tubulação (pressão interna de 13,0 bar, temperatura interna de -33^oC) e tempo de atuação do operador para fechar manualmente a válvula, de 10 minutos .

Para calcularmos o risco de um indivíduo da população externa a um complexo industrial, onde ocorre este cenário apresentado, usaremos a freqüência de ocorrência anual do evento. A taxa de falha do conjunto é a taxa de falha da tubulação e da válvula, ou seja

,(λt×λv.). Para a tubulação usando dados do AICHE [9] temos, λt = 2,2 x 10E-07/m x 100m, λt = 2,2 x 10E-05

Para a válvula, na instalação em análise, (dados fornecidos pela Bayer S/A), obtivemos a informação de 43758 demandas em três anos, ou seja 14758 demandas por ano, teríamos então 32,4676 falhas por ano.

A freqüência anual do evento fica então λt×λv = 2,2 x 10E-05 x 32,4676 = 7,14x10E-04.

Por se tratar de uma planta de processo químico, vamos especializar a taxa de falha da válvula pela inferência bayesiana, usando a informação a priori a partir de um banco de dados genérico, como por exemplo o AIChE [9]. Temos então uma taxa de falha de 2,2 x 10E-03 falhas por demanda, cuja f.d.p. é uma lognormal com intervalos de confiança: inferior de 5% equivalente a 0.306x 10E-03 , ou seja , ($\lambda_{0,05}$=0.000306) e o limite superior de 95% equivalente a 6.22 x 10-03 , ou seja, ($\lambda_{0.95}$=0.00662).

A função densidade de probabilidade de uma lognormal [18] é dada por :

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\lambda\omega} \exp\left\{-\frac{1}{2\omega^2}\left[\ln\left[\frac{\lambda}{\lambda_{50}}\right]\right]^2\right\} \tag{4.4}$$

sendo: $\lambda_{50}=\sqrt{(\lambda_{0.05} \times \lambda_{0.95})}$ $\lambda_{50}=\sqrt{(0.000306 \times 0.00662)} : \lambda_{50} = 1.4233 \times 10^{-3}$

$$n = \frac{\lambda_{0.95}}{\lambda_{50}} \qquad n = \frac{0.00662}{1.4233 \times 10^{-3}} : n = 2.1567,$$

$$\omega = \frac{1}{1.645} \times \ln n \qquad \omega = \frac{1}{1.645} \times \ln 2.1567 : \omega = 0.46722,$$

temos: $$f_X(x) = \frac{1}{\sqrt{2\pi} \times \lambda \times 0.46722} \exp\left(-\frac{1}{2 \times 0.46722^2}\left(\ln\left(\frac{\lambda}{1.4233 \times 10^{-3}}\right)\right)^2\right) \tag{4.5}$$

Agora vamos aplicar uma experiência de campo a partir de dados retirados de notas de reparo de um sistema integrado de manutenção (Dados fornecidos pela Bayer S/A – a partir de dados do sistema de gerenciamento de manutenção SAP® módulo PM).

Neste experimento, após o acompanhamento dos modos de falha de válvulas com acionamento automático do modelo CAMFLEX II ® e sua malha de acionamento, obtivemos que após 43758 solicitações num período de observação de três anos, foram evidenciadas apenas 6 falhas.

Adotaremos para esta evidencia de campo, que nos permite especializar os dados, uma função de verossimilhança, como uma f.d.p. binomial, dada a característica do modo de falha da válvula, com N=43758 (número de demandas) ; n=6 (número de falhas) ; então temos:

$$f_Y(y_n|x) = \binom{N}{n} p(\lambda)^n [1 - p(\lambda)]^{N-n} \tag{4.6}$$

Para um baixo valor de λ podemos usar uma aproximação da distribuição binomial da distribuição de Poisson [10], assumindo uma evidencia de *n* falhas num tempo de operação *T* :.

$$f_Y(y_n|x) = \exp(-\lambda \times T)\frac{(\lambda \times T)^n}{n!} \tag{4.7}$$

assumindo λ×T = N×λ [12], fica então :

$$f_Y(y_n|x) = \exp(-\lambda \times N)\frac{(\lambda \times N)^n}{n!} \tag{4.8}$$

substituindo os valores de *n* e *T* em (4.8) temos:

$$f_Y(y_n|x) = \exp\left[(-\lambda \times 43758) \times \frac{(\lambda \times 43758)^6}{6!}\right] \tag{4.9}$$

O novo valor de λ após a evidencia de campo será, $f_X(x|y_n)$, tal que:

$$f_X(x|y_n) = \frac{f_Y(y_n|x) f_X(x)}{\int_{-\infty}^{\infty} f_Y(y_n|x) f_X(x) dx} \tag{5.0}$$

efetuando as substituições de (4.5) e (4.9) em (5.0) temos:

$$f_X(x|y_n) = \frac{\frac{1}{\sqrt{2\pi}\times\lambda\times 0.46722}\exp\left(-\frac{1}{2\times 0.46722^2}\left(\ln\left(\frac{\lambda}{1.4233\times 10^{-3}}\right)\right)^2\right)\times\exp\left((-\lambda\times 43758)\times\frac{(\lambda\times 43758)^6}{6!}\right)}{\int_{0.0}^{\infty}\frac{1}{\sqrt{2\pi}\times\lambda\times 0.46722}\exp\left(-\frac{1}{2\times 0.46722^2}\left(\ln\left(\frac{\lambda}{1.4233\times 10^{-3}}\right)\right)^2\right)\times\exp\left((-\lambda\times 43758)\times\frac{(\lambda\times 43758)^6}{6!}\right)d\lambda} \qquad (5.1)$$

resolvendo a integral do denominador da eq.(5.1) :

$$\int_0^{\infty}\exp(-\lambda\times 43758)\times\frac{(\lambda\times 43758)^6}{6!}\times\frac{1}{\sqrt{2\pi}\times\lambda\times 0.46722}\exp(-\frac{1}{2\times 0.46722^2}(\ln(\frac{\lambda}{1.4233\times 10^{-3}}))^2)d\lambda \quad =$$

= 2,7932 x 10E-05

como $f_X(x|y_n)$ é uma função de densidade de probabilidade, temos que :

$\int_0^{\infty} f_X(x|y_n)$ = 1, (condição de normalização), aplicando a eq.(5.1) temos:

$$\int_0^{\infty}\frac{1}{\sqrt{2\pi}\times\lambda\times 0.46722\times 2.7932\times 10^{-5}}\exp(-\frac{1}{2\times 0.46722^2}(\ln(\frac{\lambda}{1.4233\times 10^{-3}}))^2)\times\exp(-\lambda\times 43758)\times\frac{(\lambda\times 43758)^6}{6!}d\lambda$$

= 0.99999 , ou seja , estamos atendendo à condição de normalização.

A nova taxa de falha vem então da média da f.d.p. a posteriori de λ:

$$\lambda_1 = \int_0^{\infty}\lambda f_X(x|y_n)d\lambda \qquad (5.2)$$

substituindo nossos valores em (5.2) teremos:

$$\int_0^{\infty}\lambda\frac{1}{\sqrt{2\pi}\times\lambda\times 0.46722\times 2.7932\times 10^{-5}}\exp(-\frac{1}{2\times 0.46722^2}(\ln(\frac{\lambda}{1.4233\times 10^{-3}}))^2)\times\exp(-\lambda\times 43758)\times\frac{(\lambda\times 43758)^6}{6!}d\lambda$$

= 3,0226x10E-04 , ou seja, λ_{v1} = 3,0226x10E-04 que comparado com o valor anterior genérico λ = 2,2 x 10E-03, reflete a realidade do complexo industrial com melhores programas de manutenção, padrões de projeto e condições operacionais.

A Figura 3 mostra a comparação gráfica das curvas das f.d.p. a priori e posterior do nosso exemplo .

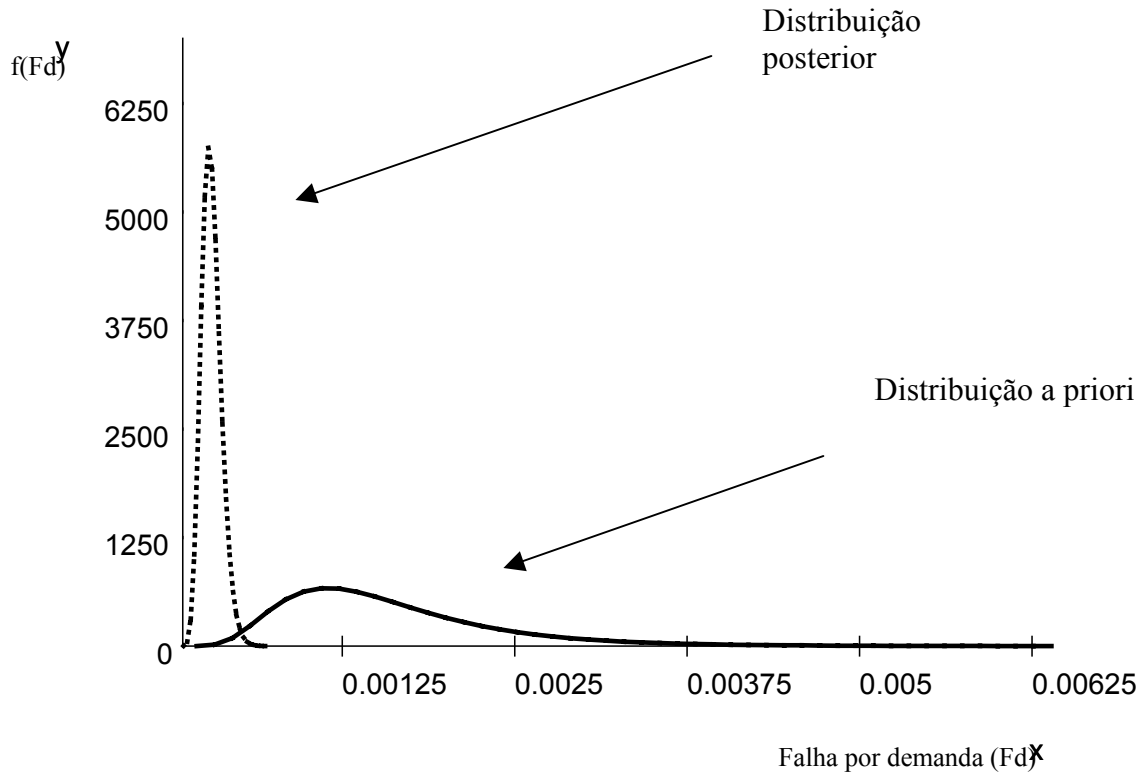


Figura 3: Gráficos das fdp´s antes da evidência da planta (priori) e depois (posteriori)

Podemos ainda observar que a curva da distribuição posteriori nos apresenta uma baixa dispersão em torno da média, o que demonstra que a nossa evidência (função de verossimilhança) é suficiente forte para a especialização da taxa de falha.

A freqüência anual do evento com o novo valor de λ_{v1} fica então:

$\lambda t \times \lambda_{v1}$ = 2,2 x 10E-05 x 4.4 = 9,68x10E-05, contra a evidência anterior de 7,14x10E-04.

Continuando o nosso exemplo de EAR, para ilustrar melhor a influência dessa nova freqüência de ocorrência do cenário de vazamento anteriormente descrito, mostramos a seguir as curvas obtidas a partir de uma simulação computacional efetuada com o *software* RISKAN®, da empresa SERENO Sistemas Ltda.

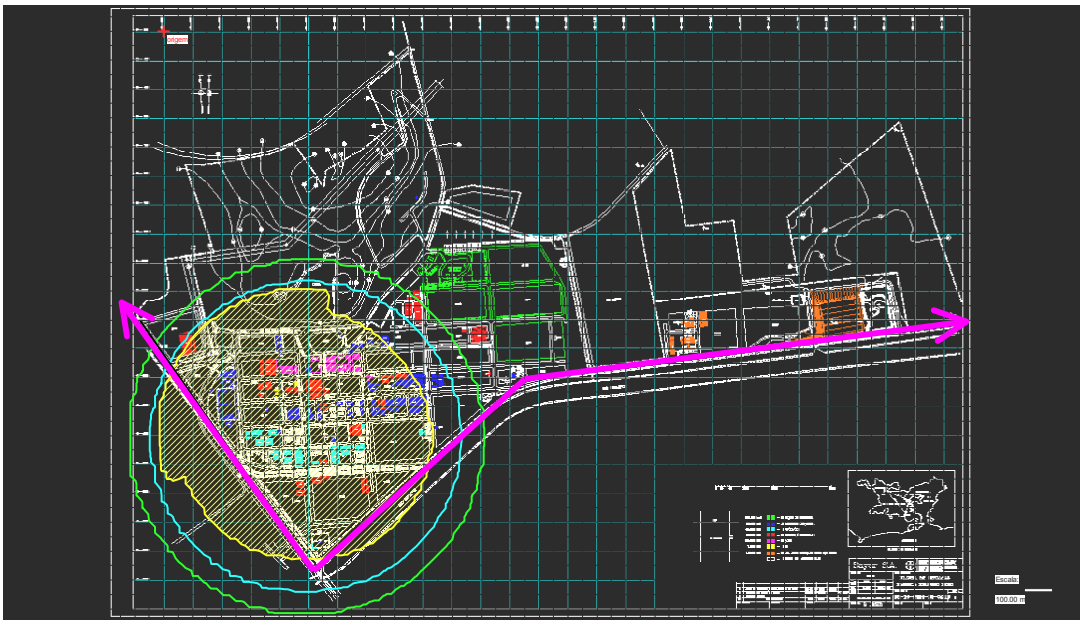
O cálculo do risco é efetuado pelo *software* , sendo que este é um produto entre o risco de fatalidades ligado à conseqüência do cenário em estudo e a freqüência anual de ocorrência do mesmo. O cálculo do risco de fatalidades da conseqüência pode ser feito por uma série de equações de *probit* [19]. Essas equações transformam uma variável física, exposição a uma nuvem tóxica de um produto, por exemplo, em uma probabilidade de fatalidade.

No Brasil os órgãos governamentais especializados em controle do meio ambiente tais como: FEEMA , CETESB, FEPAM, CRA e outros, possuem critérios de aceitabilidade de riscos. Um deles é o risco individual, ou seja, o risco que um indivíduo da comunidade vizinha a uma instalação industrial pode estar exposto, deve ser inferior a 10E-05 fatalidades/ano, para que a instalação seja aceita sem restrições.

Com as mesmas condições de conseqüência, os resultados do cálculo do risco, em forma de curvas de isorisco, com freqüências anuais antes (priori) e depois (posteriori) da

especialização dos dados pela metodologia bayesiana, foram traçadas em um mapa de uma instalação industrial, conforme as Figuras 4 e 5.

Podemos perceber que a curva de risco correspondente a 10E-05 (área hachuriada) com a freqüência a priori (Figura 4) fora da fronteira da instalação industrial com a comunidade ao redor da mesma, ou seja, inaceitável, e a curva com a freqüência posterior (Figura 5) totalmente dentro da instalação industrial, ou seja, aceitável.



Legenda:

Curvas de isorisco:

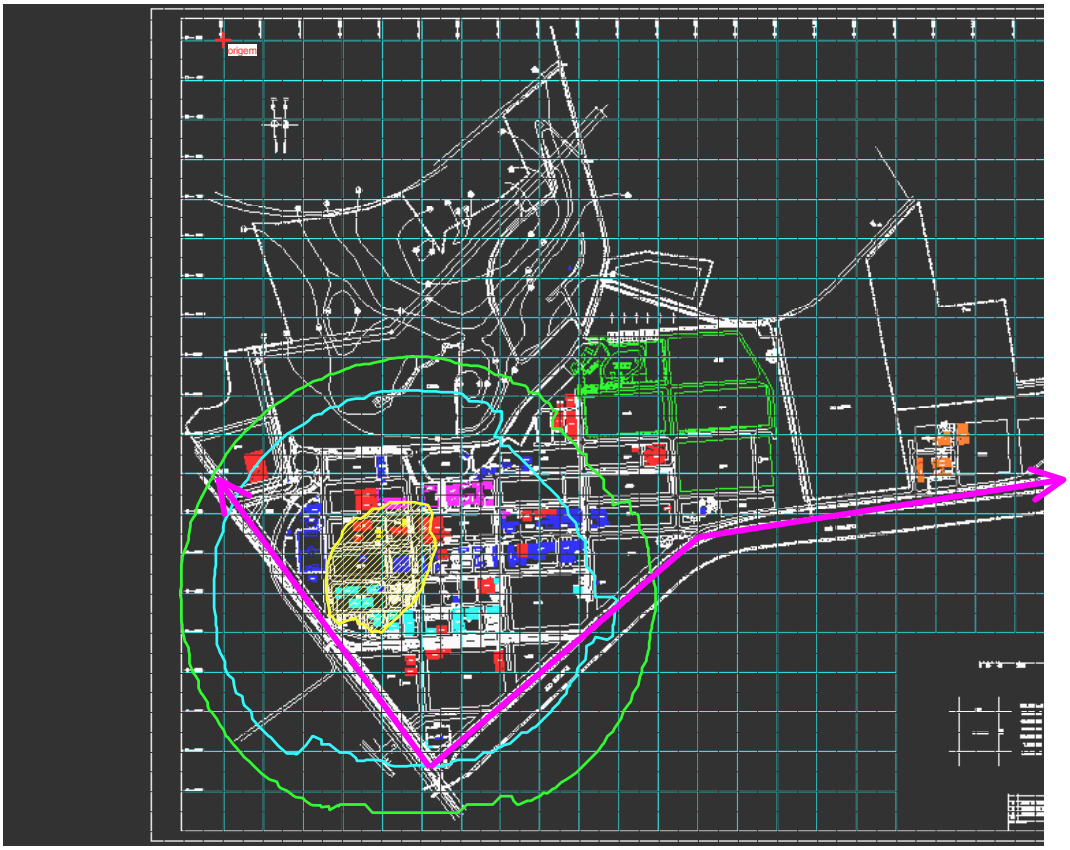
1,00 e-08/ano

1,00 e-06/ano

1,00 e-05/ano

Fronteira da instalação industrial com a comunidade ao redor.

Figura 4: Curvas do risco individual, antes (priori) da especialização da taxa de falha, pela inferência bayesiana.



Legenda:

Curvas de isorisco:

1,00 e-08/ano (risco individual)

1,00 e-06/ano (risco individual)

1,00 e-05/ano (risco individual)

Fronteira da instalação industrial com a comunidade ao redor.

Figura 5: Curvas de risco individual, após (posterior)especialização da taxa de falha, pela inferência bayesiana.

3.3- Aplicação a uma APS.

Para exemplificar o uso da metodologia bayesiana em plantas nucleares, mostraremos como a mesma é abordada num relatório de APS de uma usina nuclear similar a Angra 1.

As estimativas das freqüências de ocorrência dos eventos iniciadores da APS, foram obtidas a partir da incorporação aos dados genéricos de outras usinas nucleares similares, dados estes retirados do apêndice D do NUREG/CR 3862 [20] e também mostrados na referência [21]. A tabela apresentada no apêndice 2 reproduz então os dados genéricos considerados para a APS.

A metodologia usada é a mesma discutida no NUREG/CR-1110 [22] e nas referências [21] e [23] que a seguir apresentamos com alguns comentários no final.

O teorema de Bayes se apresenta, como já vimos, da seguinte forma

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

Podemos assumir a função verossimilhança pode ser dada pela distribuição de Poisson. neste caso teríamos:

$$P(B|A) = \frac{(p^t)e^{-pt}}{k!} \tag{5.3}$$

onde:

p = probabilidade do evento iniciador, assumido constante

K = número de ocorrências

t = período de tempo

A distribuição anterior, assumida com uma distribuição gama, pode ser descrita por sua média e variância, como segue:

$$m_a = \frac{k}{t}$$

- média da distribuição anterior
- variância da distribuição anterior $Va = \frac{k}{t^2}$

Um modo de expressar a variância da distribuição anterior é:

$$Va = m_a^2(e^{\sigma_{\ln}^2} - 1) \tag{5.4}$$

Onde $\sigma_{\ln}$ é o desvio padrão logarítmico, e é dado por :

$$\sigma_{\ln} = \frac{\ln(FE)}{1.645}$$

Se a média (m_a) e o fator de erro(FE) são fornecidos para a distribuição anterior, em vez de *k* e *t*, como ocorre freqüentemente, a informação da distribuição anterior aparece na forma de uma distribuição lognormal. Ela pode ser convertida em uma distribuição gama com parâmetros α e β pelo uso das seguintes expressões:

$$\alpha = \frac{(m_{a(\ln)})^2}{V_{a(\ln)}} \qquad \beta = \frac{V_{a(\ln)}}{(m_{a(\ln)})}$$

onde: $(m_{a(\ln)})$ = média da lognormal

$V_{a(\ln)}$ = variância da lognormal

Como a média da lognormal é igual à média da distribuição gama, combinando-se estes parâmetros com o número de ocorrência (*n*) e o período de tempo (*m*) dos novos dados, obtêm-se as seguintes equações para a média e a variância da distribuição posterior:

$$m_{A/B} = \frac{\alpha+n}{\beta+m}$$

$$V_{A/B} = \frac{\alpha+n}{(\beta+m)^2}$$

O fator de erro da distribuição posterior é dado por:

$$FE_{A/B} = e^{1.645\sigma_{A/B}}$$

onde,

$$\sigma_{A/B} = \left[\ln\left(\frac{V_{A/B}}{(m_{A/B})^2} + 1 \right) \right]^{\frac{1}{2}} \tag{5.5}$$

A Tabela do apêndice 3 apresenta o resultado da atualização bayesiana efetuada a partir destas considerações anteriores.

As aproximações anteriormente descritas das fdp's, ou seja, assumir uma fdp gama com sendo representativa de uma Poisson ou de uma lognormal nem sempre é possível, já que existem restrições de alguns parâmetros, tais como: fatores de forma, médias e variâncias. A não observância destas restrições pode levar a erros.

O uso da metodologia anteriormente apresentada no exemplo usado para o EAR de uma planta de processo químico, elimina este problema, uma vez que a comparação gráfica das fdp's a priori e a posteriori, permite avaliar diretamente este erro.

CAPÍTULO 4

Conclusões e Recomendações

4.0- Conclusões e recomendações.

Na aplicação da inferência bayesiana para o caso de um EAR podemos observar claramente a influência da especialização dos dados com uso dos dados de uma instalação industrial, mudando a curva de isorisco de aceitabilidade de risco individual em cerca de 250 a 300 m a partir do evento iniciador do risco. Isto significa que, pelo critério usado por orgãos ambientais (FEEMA,CETESB e outros) uma instalação anteriormente rejeitada, é perfeitamente aceitável sem restrições.

Entretanto, nem sempre os dados das instalações são disponíveis ou possíveis de serem usados. Estima-se que um número limitado de instalações tenha um tratamento adequado dos dados de falha de seus componentes. Neste caso, o resultado ideal vem do uso das evidências disponíveis na planta combinados com os dados genéricos disponíveis. Todavia, como já vimos, sempre que tivermos dados relevantes da instalação devemos usá-los ao invés de usarmos dados genéricos. Deve ser ressaltado também que zero falhas na planta, em um dado intervalo de tempo, não significa que não vai haver falha, e aí a inferência com dados genéricos ou de outras plantas irá ajudar consideravelmente o analista a quantificar o risco.

Podemos afirmar que uma das grandes vantagens da inferência bayesiana é o fato de mesmo com um espaço amostral pequeno, disponível da planta, é possível aproveitar esta experiência e especializarmos os dados. O gráfico da fdp a priori e a posteriori , traçados na mesma escala na Figura 3, é uma maneira interessante de se verificar o comportamento da dispersão dos dados em torno do valor esperado da taxa de falha (variância),

principalmente na distribuição posterior afetada pela verossimilhança. Com esta comparação, podemos afirmar se os dados de nossa instalação (nossa verossimilhança) são suficientes ou não.

O uso da metodologia bayesiana requer o uso de métodos numéricos para a solução de integrais, porém, com o seu equacionamento a partir de fdp's conhecidas, podemos demonstrar que com o uso de programas de cálculo como MAPLE® para Windows®, usados em microcomputadores, é suficiente para obter os resultados, não necessitando de programação em linguagem computacional específica tipo FORTRAN® ou similar. Para o exemplo de EAR usado neste trabalho, tanto para o cálculo numérico quanto para a simulação computacional com o software RISKAN®, foi utilizado um computador portátil (notebook) IBM-Thinkpad (modelo 26453EP), com processador Intel- Pentium II® e sistema operacional Windows® 2000.

Algumas aproximações de fdp's em substituição a outras, para facilitar o uso da metodologia devem ser corretamente efetuadas, pois existem limitações de alguns parâmetros que podem permitir ou não aproximação. A não-observância destas condições pode levar a erros. Esta possibilidade de erro compromete a metodologia, posto que algumas referências [8] e[10], usadas neste trabalho, apresentam estas restrições com exemplos suficientes.

Recomenda-se a introdução nos manuais de referência para elaboração de EAR's usados pelos orgãos ambientais, critérios que permitam a especialização de dados de falha utilizando a inferência bayesiana, para as instalações em estudo. Quanto ao caso de APS de centrais nucleares, também é recomendável detalhar mais as restrições das aproximações efetuadas para as f.d.p.'s utilizadas de modo a se evitar erros que possam influenciar nas especializações efetuadas.

A principal recomendação fica para as indústrias de processo em geral, para terem um correto tratamento das suas taxas de falha, construindo bancos de dados consistentes e homogêneos que possuam um tratamento estatístico adequado com f.d.p's definidas, de modo a permitirem uma coerente especialização dos dados. O banco de dados deve ter também boas condições de ser auditado, com boa rastreabilidade. Sugerem-se para isso sistemas integrados ao gerenciamento da instalação, como por exemplo o sistema SAP® ou similar, particularmente a módulo PM (gerenciamento de manutenção) do mesmo. Com isto aproveitam-se rotinas já efetuadas, evitando-se criar mais trabalho para a mesma mão de obra existente, o que muitas vezes pode acabar inviabilizando a criação e a alimentação constante de informações no banco de dados. Uma simples abertura de ordem de manutenção ou teste periódico, ao ser encerrada, pode gerar a informação de que tanto precisamos para a especialização dos dados. Obviamente, devemos gastar esta energia com os pontos que tenham maior potencial de danos. Para se definir os equipamentos ou componentes aos quais vamos aplicar a metodologia. Podemos partir de análises de segurança qualitativas do tipo análise preliminar de perigos(APP), desvios operacionais (HAZOP) ou a escolha de eventos iniciadores indicados por orgãos reguladores de centrais nucleares com por exemplo o NUREG [12].

Quanto ao caso de uso em APS de plantas nucleares, ao invés de se usar aproximação de fdp's para se obter uma fdp posterior conjugada a priori, recomenda-se usar o cálculo direto, conforme o exemplo do EAR de uma indústria de processo, e depois comparar-se as curvas a priori e a posteriori.

Referências:

[1]- GIBELLI, S.M.O., MELO, P.F.F., OLIVEIRA, L.F.S., "*Análise de Segurança de Sistemas por árvore de falhas*", PEN-112, COPPE/UFRJ, Janeiro,1982.

[2]- AIChE "*Guidelines for Chemical Process Quantitative Risk Analysis*", Center for Chemcial Process Safety of the American Institutr of Chemicals Enginerrs, . New York.,USA, 1989.

[3]- MIL-HDBK 217F, "*Reliability Prediction of eletronic Equipament*." U.S. Department of Defense Military Standardization Handbook ,1991.

[4]- SPJOTVOLL, E., "Estimation of failure rate from reliability data bases". *In society of Reliability Engineers Symposium* ,(1985: Trodheim)

[5]- OREDA-2002,. "*Offshore Reliability Data",.* prepared by SINTEF and market by DNV.

[6]- AMARAL NETTO, J.D., "*Metodologia Bayesiana para especialização e atualização de dados de falhas de centrais nucleares*"., M.Sc. (Tese), PEN-COPPE/UFRJ, 1981.

[7]- SANTOS, A. M., "*Uma abordagem Bayesiana para Estimar Taxas de Falha."* M.Sc. (Dissertação), Instituto de Matemática/UFRJ, 1996.

[8]- MARTZ, WALLER., "*Bayesian Reliability Analysis.* " 2, ed. Flórida : Krieger Publishing Company , 1991. 744p

[9]- AIChE, "*Guidelines for Process Equipament Reliability Data with Data Tables*", Center for Chemical Process Safety (CCPS) of the American Institute of Chemical Engineers, New York,USA, 1989

[10]- APOSTOLAKIS, G., et all, "Data Specialization for Plant Specific Risk Studies", *Nuclear Engineering and Design, 56,* North-Holland Publishing Company (1980), pág. 321-329.

[11]- USNRC (U.S. Nuclear Regulatory Comission), Reactor Safety Study – An assessment of Accident Risks in U.S. Commercial Nuclear Power Plants, WASH-1400, Apêndice III- Failure Data, NUREG-751014, 1975.

[12]- National Reliability Evaluation Program Procedure Guide, Suported by PAPAZOGLOU, I.A., Apendice C, NUREG/CR-2815, BNL-NUREG 51559, (1982).

[13]- MOSLEH, A., APOSTOLAKIS, G., "Models for the use of Expert Opinions", *Workshop on Low-Probability, High-Consequency Risk Analisys*, Arlington, Virginia, 1982.

[14]- LINSTONE, H.A. and TUROFF, M., "*The Delphi Method: Techniques and Apliccations* " Reading, 1975.

[15]- DeGROOT, M.H.,"*Probability and Statistics*"- 2nd Ed. Carnegie-Mellon University, Addison-Wesley Publishing Company, New York (1986).

[16]- NELSON,W.B. "*Applied Life Data Analisys*", John Willey & Sons, New York 1982.

[17]- HENLEY, E.J. e KUMAMOTO, H. "*Reliability Engineering and Risk Assessment*.", Prentice-Hall, Englewood Clifs. New Jersey, 1981

[18]- LEWIS, E.E.. "*Introduction to Reliability Engineering*". 2nd,ed, New York, John Wiley & Sons, Inc.. 1994, 435p. .

[19]- TNO- the Nettherlands Organization. 'Methods for Determining and Processing Probabilities- Red Book'. CPR 12E, 1997.

[20]- NUREG/CR 3862, U.S. Nuclear Regulatory Commission. Development of Transient Initiating Event Frequencies for use in Probabilistic Risk Assessments. Maio1985

[21]- Probabilistic Safety Assessment for Point Beach NPP, Units 1&2 Initianting Event Notebook. Revision 0. Wiscosin Public Service Co, May 1993.

[22]- NUREG/CR-1110. U.S. Nuclear Regulatory Commission ,Bayesian Analysis of Component Failure. November 1979.

[23]- INPO/FGR-363.Isntitute of Nuclear Power Operations, B.LOGAN. A, layman's Guide to Probabilistic Risk Assessments·

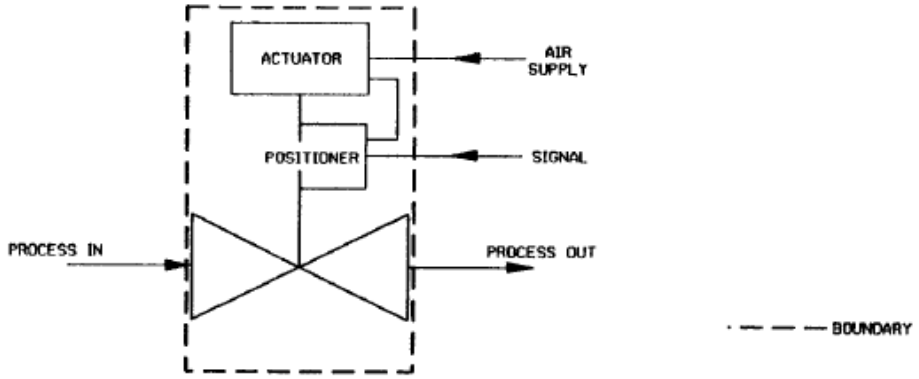
DATA ON SELECTED PROCESS SYSTEMS AND EQUIPMENT

Taxonomy No. 3.5.3.3	Equipment Description VALVES-OPERATED-PNEUMATIC
Operating Mode	Process Severity UNKNOWN

Population	Samples	Aggregated time in service (10^6 hrs)		No. of Demands
		Calendar time	Operating time	

Failure mode	Failures (per 10^6 hrs)			Failures (per 10^3 demands)		
	Lower	Mean	Upper	Lower	Mean	Upper
CATASTROPHIC a. External Leakage b. Internal Leakage >1% c. Spurious Operation d. No Change of Position on Demand	0.274	3.59	12.3	0.306	2.2	6.62
DEGRADED a. Delayed Actuation						
INCIPIENT a. Wall Thinning b. Embrittlement c. Cracked or Flawed d. Internal Leakage						

Equipment Boundary



Data Reference No. (Table 5.1): 8, 8.1, 8.2, 8.3, 8.4, 8.7, 8.10, 8.12, 8.14, 8.15

Apêndice 1: Taxa de falha de válvula automática, acionada remotamente por atuador pressurizado com ar, a partir de sinal de instrumento de controle do processo,AIChE [9].

Apêndice 2: dados genéricos considerados para APS de usinas similares a Angra 1 [21].

FREQÜÊNCIA DE EVENTOS INICIADORES EM OUTROS APSs

Evento Iniciador	Salem 1	Seabrook 1 & 2	Ginna (h)	Kewaunee	Oconee 3 (g)	Zion 1	Sequoyah 1	Surry 1	Crystal River 3 (g)	Point Beach 1&2	Millstone 3	Wash-1400(a)
Pequeno LOCA	1.0E-3	1.8E-2	2.0E-2 (c)	5.1E-3	3.0E-3	3.5E-2	1.0E-3	1.0E-3	3.0E-3	3.0E-3	9.1E-3	1.0E-3
Médio LOCA	1.0E-3	4.7E-4	1.0E-3	2.4E-3	-	9.4E-4	1.0E-3	1.0E-3	-	1.0E-3	6.1E-4	3.0E-4
Grande LOCA	5.0E-4	2.0E-4	1.0E-4	5.0E-4	9.3E-4	9.4E-4	5.0E-4	5.0E-4	5.0E-4	2.5E-4	3.9E-4	1.0E-4
LOCA Excessivo	3.0E-7	2.7E-7	-	3.0E-7	1.1E-6	-	-	-	-	7.0E-7	3.0E-7	1.0E-7
Ruptura de Tubo do Gerador de Vapor	1.0E-2	2.8E-4	1.4E-2 (c)	6.4E-3	8.6E-3	1.0E-2	1.0E-2	1.0E-2	8.6E-3	6.6E-3	3.9E-2	-
LOCA de Interface	1.0E-7	7.8E-6(c)	-	2.9E-7	1.4E-7	1.1E-7	6.5E-7	1.6E-6	-	7.2E-8	2.2E-7	4.0E-6
Ruptura de Linha de Vapor Fora da Contenção	5.0E-4	6.0E-3	2.0E-4	2.5E-3	2.6E-3 (f)	9.4E-4	-	-	-	8.0E-4	3.8E-2	-
Ruptura de Linha de Vapor Dentro da Contenção	1.5E-3	4.7E-4	5.5E-4(c)	-	5.4E-4 (f)	9.4E-4	-	-	2.1E-2	7.0E-4	3.9E-4	-
Perda do SRC	5.4E-2	2.9E-3(e)	1.8E-2	2.0E-3	-	9.4E-4	-	-	2.2E-3	1.6E-3	-	-
Perda do SAS	3.6E-6	3.6E-3(e)	1.8E-2	5.6E-5	4.0E-3	9.4E-4	-	-	5.2E-3	1.7E-3	1.8E-2 (e)	-
Perda de uma Barra DC	1.5E-2	3.2E-3	3.9E-4	2.4E-3	2.0E-2	-	5.0E-3	5.0E-3	-	9.3E-4	3.9E-3	-
Perda de Potencia Elétrica Externa	6.0E-2	4.9E-2(c)	1.1E-1(d)	4.4E-2	1.7E-1 (c)	7.8E-2	9.1E-2	7.7E-2	3.5E-2	6.0E-2	1.1E-1	2.0E-1

Apendice 3: Exemplo do resultado da atualização Bayesiana dos eventos iniciadores de uma APS de uma usina nuclear a partir de dados genéricos de usinas nucleares similares a Angra1 [21].

DADOS PARA OS EVENTOS INICIADORES

DESCRIÇÃO	DADOS GENÉRICOS				DADOS DA PLANTA		ATUALIZAÇÃO POR BAYES			DADOS SELECIONADOS	
	#	MÉDIA	VARIÂNCIA	FATOR DE ERRO	#	MÉDIA	MÉDIA	VARIÂNCIA	FATOR DE ERRO	MÉDIA	FONTE
Loss of Circulating Water/Turbine Cooling	14	3.75E-02	5.98E-03	8.3	2	1.90E-01	1.33E-01	7.92E-03	2.72	1.33E-01	B
Loss of Component Cooling Water	5	3.41E-02	2.35E-02	17.7	0	0.0	4.15E-03	3.47E-04	17.7	4.15E-03	B
Loss of Service Water System	1	2.84E-03	3.71E-04	25.2	0	0.0	1.20E-03	6.60E-05	25.2	1.20E-03	B
Turbine Trip, Valve Closure, EHC Probl.	379	1.02E-00	4.08E-01	2.6	7	6.67E-01	7.33E-01	5.67E-02	1.68	7.33E-01	B
Generator Trip Generator Caused Faults	138	4.59E-01	2.20E-01	4.0	2	1.90E-01	2.35E-01	1.87E-02	2.43	2.35E-01	B
Loss of All Off-Site Power (LOSP)	43	1.02E-01	2.32E-02	5.9	4	3.81E-01	2.98E-01	1.99E-02	2.10	2.98E-01	B
Pressurizer Spray Failure	9	2.02E-02	3.27E-03	11.5	0	0.0	7.45E-03	4.48E-04	11.5	7.45E-03	B
Loss of Power to Necessary Plant Systems	35	1.29E-01	4.02E-02	6.2	10	9.52E-01	7.60E-01	5.55E-02	1.65	7.60E-01	B
Spurious Trips - Cause Unknown	18	3.94E-02	9.59E-03	10.1	0	0.0	1.10E-02	7.57E-04	10.1	1.10E-02	B
Auto Trip - No Transient Condit.	448	1.32E-00	9.27E-01	2.9	0	0.0	1.61E-01	1.34E-02	2.90	1.61E-01	B
Manual Trip - No Transient Condition	161	5.14E-01	7.33E-01	6.7	2	1.90E-01	2.11E-01	1.88E-02	2.66	2.11E-01	B