III SEMETRO - III Seminário Internacional de Metrologia Elétrica

15-17 de setembro, 1998
Rio de Janeiro - RJ - Brasil

(Organizado pelo Inmetro)

Estatística Bayesiana e o Guia Internacional para Expressão da Incerteza de Medição

Gregory Amaral Kyriazis

INMETRO - Instituto Nacional de Metrologia, Normalização e Qualidade Industrial

Wolfgang Wöger

PTB - Physikalisch-Technische Bundesanstalt

1.0 RESUMO

O Guia Internacional para Expressão da Incerteza de Medição é discutido com base na estatística Bayesiana e no Princípio de Máxima Entropia de Informação.

2.0 PALAVRAS-CHAVE

Incerteza de Medição, Estatística Bayesiana, Princípio de Máxima Entropia de Informação.

3.0 INTRODUÇÃO

O teorema de Bayes é aplicado à inferência do valor esperado de uma distribuição de freqüência normal de valores observados. O Princípio de Máxima Entropia de Informação é usado para estabelecer uma densidade de probabilidade com base em informação não estatística disponível. Os *métodos estatísticos* (Tipo A) e os *outros métodos* (Tipo B) são ambos unificados e identificados com os métodos da estatística Bayesiana.

4.0 DUAS TEORIAS DA PROBABILIDADE

A escola de pensamento tradicional [1] considera a probabilidade de um evento como uma propriedade objetiva daquele evento, sempre passível, em princípio, de medição empírica por meio da observação das freqüências relativas em um experimento aleatório.

Por outro lado, a escola de pensamento Bayesiana [2] considera probabilidades como expressões da ignorância humana; a probabilidade de um evento é meramente uma expressão formal de nossa confiança de que o evento ocorreu ou ocorrerá, baseado em qualquer informação que esteja disponível.

No método de inferência Bayesiano, não há distinção entre erros aleatórios e sistemáticos ou entre valores aleatórios e não aleatórios. Esta unificação corresponde à realidade de que só raramente uma distinção clara pode ser feita entre os dois tipos de erros. O fato de um erro sistemático de valor fixo não ser uma grandeza aleatória não está em conflito com o enunciado de probabilidade sobre o erro sistemático, uma vez que este enunciado de probabilidade apenas descreve numericamente o que o metrologista aprendeu sobre o erro sistemático com base no experimento. O enunciado de probabilidade não contém qualquer enunciado sobre as freqüências relativas, embora medições repetidas geralmente contribuam para o enunciado de probabilidade. Os métodos Bayesianos não negam a existência das distribuições de freqüência. O que é negado é a identificação de freqüências relativas com probabilidades. Tanto informação na forma de freqüências relativas (ou mesmo a sua distribuição), quanto informação de outro tipo podem ser usadas para expressar probabilidades e avaliar o mensurando (o que não é possível na estatística tradicional).

A estatística Bayesiana não é muito bem conhecida pela comunidade metrológica. Mas é a estatística Bayesiana que fundamenta uma teoria da incerteza de medição [3] que considera de modo consistente a incerteza causada por efeitos sistemáticos.

5.0 INFORMAÇÃO e PROBABILIDADE

A idéia básica para inferir o valor de um mensurando é começar de qualquer informação relevante disponível sobre o mensurando. A informação pode consistir de observações obtidas de uma distribuição *de freqüência* que se supõe conhecida (informação Tipo A) ou de qualquer outro tipo como, por exemplo, apenas dois limites contendo o valor de interesse (informação Tipo B). A conhecida distribuição de probabilidade definida sobre todo o conjunto de valores que podem ser razoavelmente atribuídos ao mensurando descreve o *estado de conhecimento* do mensurando correspondente à informação dada. A descrição corresponde a uma dada informação incompleta sobre o mensurando. Uma informação completa seria representada por uma função impulso centrada no valor esperado do mensurando.

Como a informação é dada e é fixa, a distribuição de probabilidade do mensurando e seus parâmetros correspondentes são conhecidos e fixos. Isto está de acordo com a definição de incerteza do *Guia* [4]. Não há *incerteza de uma estimativa* ou *incerteza da incerteza* [5].

A melhor estimativa de um mensurando X é determinada minimizando-se a expressão [6]

$$E(X - x')^2 = Var(X) + (x - x')^2 \qquad (1)$$

Aqui o valor esperado E(X) do estimador X é denotado por x. A raiz quadrada positiva de E(X - x')^2 é denominada de *incerteza padrão* $u(x')$ *associada à estimativa escolhida* x' *do valor do mensurando* X. Embora o mesmo símbolo seja usado tanto para o mensurando quanto para seu estimador, isto não significa que sejam idênticos: o estimador (variável aleatória) fornece uma estimativa para o mensurando e para a incerteza associada a esta estimativa. Como o mínimo ocorre para x' = E(X) = x, a incerteza padrão $u(x)$ é associada com a melhor estimativa, ou valor esperado, do estimador.

Portanto, uma vez estabelecida a distribuição de probabilidade, a raiz quadrada da variância desta distribuição descreve quantitativamente com que extensão o valor desconhecido do mensurando é conhecido e o valor esperado dá uma estimativa razoável do valor do mensurando. Tanto esta estimativa quanto sua incerteza associada dependem apenas da informação dada.

Se a única informação disponível são os valores observados obtidos de uma distribuição de *freqüência* que se supõe conhecida, a distribuição de probabilidade que descreve o estado de conhecimento do mensurando pode ser obtida de uma aplicação do teorema de Bayes (vide seção *6.0*). Se a única informação disponível for de um tipo tal que não possa ser expressa por métodos estatísticos, o Princípio de Máxima Entropia de Informação pode ser usado para estabelecer uma distribuição de probabilidade (vide seção *7.0*).

6.0 INFERÊNCIA BAYESIANA

Os fundamentos da inferência estatística Bayesiana [7] são apresentados nesta seção. De uma forma geral, uma pessoa tem um grau de confiança *anterior* acerca de várias hipóteses possíveis e então modifica este grau de confiança anterior à luz de dados relevantes que coleta com o fim de chegar a um grau de confiança *posterior*. Os métodos de inferência Bayesiana são basicamente uma descrição matemática do processo de aprendizagem. Matematicamente, isto é obtido pelo teorema de Bayes. Supondo-se n observações $\boldsymbol{x} = x_1, x_2, \ldots, x_n$ disponíveis, se o parâmetro θ denota os valores que podem ser razoavelmente atribuídos ao mensurando X, o teorema de Bayes pode ser escrito em termos de densidades de probabilidades como

$$p(\theta \mid \boldsymbol{x}) \propto p(\theta).l(\theta \mid \boldsymbol{x}) \qquad (2)$$

Aqui $p(\theta|\boldsymbol{x})$ é a densidade de probabilidade posterior (aos dados $\boldsymbol{x}$) de θ, $p(\theta)$ é a densidade de probabilidade anterior (aos dados $\boldsymbol{x}$) de θ e $l(\theta|\boldsymbol{x})$ é a verossimilhança de θ (para $\boldsymbol{x}$ dado). A verossimilhança é numericamente igual à densidade de probabilidade dos dados $\boldsymbol{x}$ (dado que θ é conhecido). Ela representa a informação adicional fornecida pelos dados.

O teorema de Bayes pode ser facilmente lembrado como: *Posterior* ∝ *Anterior* x *Verossimilhança*. Esta relação resume o modo pelo qual um pessoa deve modificar seu grau de confiança a fim de considerar os dados disponíveis. Do ponto de vista Bayesiano, as densidades anterior e posterior são descrições do estado de conhecimento de X e, portanto, não há restrição quanto a X gerar dados aleatórios ou não.

O resultado da inferência Bayesiana é univocamente determinado uma vez que a densidade anterior tenha sido escolhidas.

É útil ter uma distribuição anterior de referência (não informativa) para ajudar o discurso público em situações onde opiniões anteriores difiram ou não sejam muito fortes [7].

Considere-se que nada se sabe sobre o valor esperado θ (parâmetro de localização) de uma distribuição normal e se deseje descrever esta ignorância na forma de uma distribuição de probabilidade $p(\theta)$. Somando θ a um fator positivo constante λ uma nova localização θ´ = θ+λ é obtida. Se $p(\theta)$ é a densidade de θ, então θ´ deve ter a densidade $g(\theta') = p(\theta).|d\theta/d\theta'| = p(\theta'-\lambda)$ (mudança de variáveis) ou $p(\theta) = g(\theta')$. . Isto é uma conseqüência da teoria das probabilidades. Como o estado de conhecimento de θ´ é exatamente o mesmo que o de θ, p e g devem ser a mesma função: $p = g$. Então, p deve satisfazer a equação funcional $p(\theta) = p(\theta+\lambda)$, ou seja, a distribuição anterior é invariante por translação. A solução geral desta equação funcional é a distribuição anterior de referência $p(\theta)$ = const.

Por outro lado, considere-se que nada se saiba sobre a variância φ (parâmetro de escala) de uma distribuição normal e se deseje descrever esta ignorância na forma de uma distribuição de probabilidade $p(\varphi)$. Multiplicando φ por um fator positivo constante λ uma nova escala $\varphi' = \lambda\varphi$ é obtida. Se $p(\varphi)$ é a densidade de φ, então φ' deve ter a densidade $g(\varphi') = p(\varphi).|d\varphi/d\varphi'| = p(\varphi'/\lambda)/\lambda$ (mudança de variáveis) ou $p(\varphi) = \lambda.g(\varphi')$. Isto é uma conseqüência da teoria das probabilidades. Como o estado de conhecimento de φ' é exatamente o mesmo que o de φ, p e g devem ser a mesma função: $p = g$. Então, p deve satisfazer a equação funcional $p(\varphi) = \lambda.\ p(\lambda\varphi)$, ou seja, a distribuição anterior é invariante por escala. A solução geral desta equação funcional é a distribuição anterior de referência $p(\varphi) = \text{const.}\ \varphi^{-1}$.

6.1 Informação Tipo A

Em metrologia, os valores observados são vistos como sendo amostrados de uma população que geralmente se supõe representada por uma distribuição de freqüência normal. *Sem efeitos sistemáticos*, o valor do mensurando é igual ao valor esperado da distribuição de freqüência. Pode-se usar o teorema de Bayes para inferir este parâmetro desconhecido da distribuição de freqüência, considerando a informação estatística observada e a informação de que o parâmetro de interesse pertence a uma distribuição de freqüência normal. O resultado dá o estado de conhecimento sobre o parâmetro, onde este estado de conhecimento é representado quantitativamente por uma distribuição de probabilidade posterior (no caso, a distribuição t-Student).

Considere-se o problema em que se colhe uma amostra de uma distribuição de freqüência normal e se deseja inferir dos valores observados $\boldsymbol{x} = x_1, \ldots, x_n$, obtidos independentemente, o valor esperado $\hat{\mu}$, que é o parâmetro de localização da distribuição. Não há nenhuma informação *anterior* sobre $\hat{\mu}$ e sobre a variância $\hat{\varphi}$, que é o parâmetro de escala da distribuição de freqüência. Os valores possíveis de $\hat{\mu}$ e $\hat{\varphi}$ serão denotados por θ e φ, respectivamente. A verossimilhança é

$$\begin{aligned} l(\theta,\varphi \mid \boldsymbol{x}) &\propto p(\boldsymbol{x} \mid \theta,\varphi) \\ &\propto \varphi^{-n/2} \exp\left[-\sum_{i=1}^{n}(x_i-\theta)^2/2\varphi\right] \\ &= \varphi^{-n/2} \exp\left[-\left\{S + n(\bar{x}-\theta)^2\right\}/2\varphi\right] \end{aligned} \qquad (3)$$

onde se define

$$S = \sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad (4)$$

Como não há nenhuma informação *anterior* sobre θ e sobre a variância φ, faz-se uso de uma distribuição anterior de referência. É comum usar

$$p(x \mid \theta,\varphi) \propto 1/\varphi \qquad (5)$$

que é o produto da distribuição anterior de referência $p(\theta) \propto 1$ para θ pela distribuição anterior de referência $p(\varphi) \propto 1/\varphi$ para φ. A justificativa para a independência entre os parâmetros θ e φ é que se muito pouco é conhecido sobre o valor esperado e sobre a variância, parece improvável que a aquisição de informação sobre um afete o julgamento a respeito do outro.

Considerando a distribuição anterior de referência estabelecida acima, o teorema de Bayes fornece a distribuição posterior

$$p(\theta,\varphi \mid x) \propto \varphi^{-n/2-1} \exp\left[-\left\{S + n(\bar{x}-\theta)^2\right\}/2\varphi\right] \qquad (6)$$

É conveniente estabelecer

$$\nu = n - 1 \qquad (7)$$

na potência de φ, mas não na exponencial, de modo que

$$p(\theta,\varphi \mid x) \propto \varphi^{-(\nu+1)/2-1} \exp\left[-\left\{S + n(\bar{x}-\theta)^2\right\}/2\varphi\right] \qquad (8)$$

Agora, o que interessa neste problema é o valor esperado θ. A exigida distribuição (posterior) *marginal* de θ é [7]

$$p(\theta \mid \boldsymbol{x}) \propto \left\{S + n(\bar{x}-\theta)^2\right\}^{-(\nu+1)/2} \qquad (9)$$

Entretanto, esta não a maneira mais conveniente de expressar o resultado. A distribuição de θ depende de $\bar{x}$ e S, bem como de ν, e consequentemente é mais complicada para tabelar. A fim de reduzir os parâmetros envolvidos é usual definir

$$t = \frac{\theta - \bar{x}}{s/\sqrt{n}} \qquad (10)$$

onde

$$s^2 = S/(n-1) = S/\nu \qquad (11)$$

A densidade posterior para t é dada pela regra usual de mudança de variável

$$p(t \mid x) = p(\theta \mid x)\left|\mathrm{d}\theta/\mathrm{d}t\right| \qquad (12)$$

Como $|\mathrm{d}\theta/\mathrm{d}t|$ é uma constante, a densidade posterior de t é dada por

$$\begin{aligned} p(t \mid x) &\propto \left\{\nu s^2 + (st)^2\right\}^{-(\nu+1)/2} \\ &\propto \left\{1 + t^2/\nu\right\}^{-(\nu+1)/2} \end{aligned} \qquad (13)$$

Esta é a densidade de uma variável aleatória com distribuição t-Student de ν graus de liberdade (na estatística Bayesiana, grau de liberdade é somente um termo técnico para um parâmetro específico de algumas distribuições de probabilidade como, por exemplo, a distribuição t-Student ou Qui-quadrado), de modo que é possível escrever $t \sim t_\nu$. Como a distribuição de t depende de um único parâmetro ν, é sensato expressar o resultado em termos desta distribuição em vez da distribuição do próprio θ.

A expressão para a densidade de probabilidade posterior do estimador θ do valor esperado $\hat{\mu}$ especifica o estado de conhecimento do valor $\hat{\mu}$ após os dados ***x*** terem sido obtidos. Na prática, $\hat{\mu}$ é igual ao valor daquela grandeza de entrada X que aparece no modelo de avaliação (vide seção *8.0*) como uma grandeza para a qual se tem uma informação Tipo A (valores observados obtidos individualmente) e uma informação de que os dados foram amostrados de uma população normal.

Como Et = 0, a melhor estimativa do valor esperado $\hat{\mu}$ é

$$\mathrm{E}\theta = \overline{x} \tag{14}$$

Como

$$Var(t) = \frac{n-1}{n-3} = \frac{Var(\theta)}{s^2 / n} \qquad (n \geq 4) \tag{15}$$

tem-se então que a incerteza padrão associada à melhor estimativa do valor esperado $\hat{\mu}$ é

$$u(\overline{x}) = \sqrt{(n-1)/(n-3)}.s/\sqrt{n} > s/\sqrt{n} \quad (n \geq 4) \tag{16}$$

Quando n é grande, pode-se escrever

$$u(\overline{x}) = s/\sqrt{n} \tag{17}$$

O *Guia* estabelece que o número de observações deve ser suficientemente grande (vide Notas da seção 4.2.3 do *Guia* [4]) para que (17) seja válida. A estatística Bayesiana fornece uma expressão (16) para a incerteza padrão quando o número de observações é pequeno ($n \geq 4$). Isto está estritamente de acordo com a seção **4.3.2** do *Guia* [4].

7.0 PRINCÍPIO DE MÁXIMA ENTROPIA DE INFORMAÇÃO

Dada alguma informação incompleta sobre um mensurando, duas pessoas diferentes poderiam estabelecer diferentes distribuições de probabilidade ou estados de conhecimento. Isto só pode significar que pelo menos uma das pessoas fez uso (mediante uma representação inconsciente) de mais informação do que realmente foi fornecida.

Este problema de especificação de probabilidades, em casos onde pouca ou nenhuma informação está disponível, é tão antigo quanto a teoria das probabilidades. O "Princípio da Razão Insuficiente" de Laplace foi uma tentativa de suprir um critério de escolha, no qual se diz que probabilidades iguais devem ser atribuídas a dois eventos se não há qualquer razão para pensar de outra maneira. Entretanto, exceto nos casos onde há um elemento de simetria evidente que claramente resulta em eventos "igualmente possíveis", esta suposição pode parecer tão arbitrária quanto qualquer outra que possa ser feita. Desde o tempo de Laplace, esta maneira de formular problemas foi abandonada, devido à falta de qualquer princípio construtivo que desse uma razão para preferir uma distribuição de probabilidade em vez de outra nos casos em que ambas concordassem igualmente bem com a informação disponível.

Em 1949, Claude Shannon [8] propôs uma medida importante da "quantidade de incerteza" ou *entropia de informação* que está contida em mensagens enviadas ao longo de uma linha de transmissão. Esta medida foi subseqüentemente aplicada em campos distantes da engenharia de comunicações. Jaynes [9], mais tarde, propôs que a maximização da entropia de informação, sujeita à quaisquer restrições impostas pela aplicação específica, é a melhor maneira de estabelecer a distribuição de probabilidade de uma maneira menos tendenciosa (*Princípio de Máxima Entropia de Informação).* A idéia tem sido usada como um princípio geral de inferência estatística em campos como a física, engenharia, economia e em outros campos onde métodos estatísticos são exigidos [10].

O Princípio de Máxima Entropia de Informação determina univocamente uma distribuição de probabilidade que é livre de preconceitos, já que é maximamente não comprometida com a informação que está faltando. O formalismo não será discutido aqui.

Dois casos são de grande interesse em metrologia:

7.1 Densidade de probabilidade retangular

Suponha-se que se tem a informação de que um mensurando X está contido entre dois limites a e b ($a < b$) e que esta seja a única informação disponível. A densidade $p(X)$, que dá uma máxima entropia de informação quando sujeita à restrição de que $a \leq X \leq b$ e à condição de normalização, é

$$p(X) = \frac{1}{b-a} \qquad (a \leq X \leq b) \tag{18}$$

Este é apenas um enunciado de indiferença. É a única maneira honesta de descrever o estado do conhecimento correspondente à informação dada. Portanto, de acordo com o Princípio de Máxima Entropia de Informação, o estado de conhecimento correspondente é dado por uma densidade de probabilidade retangular.

7.2 Densidade de probabilidade Gaussiana

Suponha-se agora que se tenha a informação da melhor estimativa (valor esperado) e da incerteza associada (variância) de um mensurando X. A densidade $p(X)$, que dá uma máxima entropia de informação quando sujeita à restrição de que o valor esperado e a variância de X sejam, respectivamente, fixados em x e $u^2(x)$ e à condição de normalização, é

$$p(X) = \frac{1}{\sqrt{2\pi}\ u(x)} \exp\left[-(X-x)^2 / 2u^2(x)\right] \qquad (19)$$

Portanto, de acordo com o Princípio de Máxima Entropia de Informação, o estado de conhecimento correspondente é dado por uma densidade de probabilidade normal ou Gaussiana.

Generalizando para n dimensões, no caso em que se conhece somente as melhores estimativas (valores esperados) $\boldsymbol{x}$ e a *matriz de incertezas* (covariâncias) $\boldsymbol{U}$ das grandezas $\boldsymbol{X} = X_1, \ldots, X_n$,

$$p(\boldsymbol{X}) = \frac{\exp\left[-(\boldsymbol{X}-\boldsymbol{x})^T \boldsymbol{U}^{-1}(\boldsymbol{X}-\boldsymbol{x})/2\right]}{\sqrt{(2\pi)^n \det \boldsymbol{U}}} \qquad (20)$$

onde $\boldsymbol{X} = X_1, \ldots, X_n$ é o vetor coluna das variáveis aleatórias, $\boldsymbol{x} = x_1, \ldots, x_n$ é o vetor coluna dos valores esperados de $\boldsymbol{X}$, e $(\boldsymbol{X} - \boldsymbol{x})^T$ é a transposta do vetor coluna $(\boldsymbol{X} - \boldsymbol{x})$.

Portanto, de acordo com o Princípio de Máxima Entropia de Informação, o estado de conhecimento correspondente é dado por uma densidade de probabilidade Gaussiana n-dimensional com covariâncias $u^2(x_i, x_j)$.

8.0 PROPAGAÇÃO DE INCERTEZAS

Em uma tarefa de avaliação, é necessário estabelecer o *modelo* de avaliação. Este consiste de uma representação matemática do modo pelo qual uma *grandeza de saída Y* deve ser inferida com base na informação disponível a respeito das *grandezas de entrada* $X_1,...,X_n$. As grandezas de entrada são representadas no que se segue por um vetor coluna $\boldsymbol{X}$. O modelo é geralmente da forma

$$Y = G(\boldsymbol{X}) \qquad (21)$$

A densidade combinada dos estimadores (de entrada e de saída) é estabelecida usando-se o Teorema de Bayes e considerando-se o modelo em conjunto com os dados. As fontes de informação são muito diferentes e independentes. A distribuição anterior é então normalmente introduzida pelo produto $p(\boldsymbol{X})\delta(Y\text{-}G(\boldsymbol{X}))$, onde os dados são considerados pela *densidade de dados anterior* $p(\boldsymbol{X})$ dada por uma Gaussiana generalizada (20) e onde o modelo é considerado pela *densidade de modelo anterior* $\delta(Y\text{-}G(\boldsymbol{X}))$ [3]. A função impulso unitário $\delta(t)$ retorna um valor zero se o modelo não é satisfeito. Se nenhuma restrição adicional é exigida, a verossimilhança é constante e a densidade combinada dos estimadores (de entrada e de saída) pode ser calculada de

$$p(Y,\boldsymbol{X}) \propto p(\boldsymbol{X}).\delta(Y - G(\boldsymbol{X})) \qquad (22)$$

A integração de $p(Y,\boldsymbol{X})$ sobre todos os $\boldsymbol{X}$ fornece a densidade $p(Y)$ do estimador de saída, isto é,

$$p(Y) \propto \int p(\boldsymbol{X}).\delta(Y - G(\boldsymbol{X}))\mathrm{d}\boldsymbol{X} \qquad (23)$$

A distribuição do estimador de um mensurando não é necessariamente uma Gaussiana. Se a distribuição de $Y = G(\boldsymbol{X})$ é calculada a partir da distribuição de $\boldsymbol{X}$ dada, ela em geral não será Gaussiana, mesmo que a distribuição de $\boldsymbol{X}$ seja. A informação completa sobre o que é conhecido sobre Y consiste da sua distribuição, que idealmente deveria ser repassada ao usuário. Entretanto, exceto para casos simples, na prática somente o valor esperado $y = \mathrm{E}Y$ e a variância $Var(Y) = u^2(y)$ são estabelecidos e nada mais. Mas para o usuário, esta informação parcial, de fato, corresponde a um estado de conhecimento sobre Y que, de acordo com o Princípio de Máxima Entropia de Informação, é descrito por uma Gaussiana.

Uma vez que o modelo $Y = G(\boldsymbol{X})$ foi estabelecido, o problema da propagação das incertezas em princípio está resolvido, bastando apenas calcular a melhor estimativa y e a incerteza padrão $u(y)$ da grandeza de saída Y. Usando-se (23) e trocando-se a ordem das integrações pode-se facilmente deduzir

$$y = \mathrm{E}Y = \int G(\boldsymbol{X}).p(\boldsymbol{X}).\mathrm{d}\boldsymbol{X} \qquad (24)$$

$$u(y) = \sqrt{Var(Y)} = \sqrt{\int (G(\boldsymbol{X}) - y)^2.p(\boldsymbol{X}).\mathrm{d}\boldsymbol{X}} \qquad (25)$$

O cálculo das integrais deve geralmente ser feito em um computador, mas isto não é necessário nos casos em que se pode linearizar o modelo.

Considere-se que um modelo é linearizado nas vizinhanças do vetor coluna *conhecido* $\boldsymbol{x}$. Então se $G(\boldsymbol{X})$ é uma função razoavelmente bem comportada e o vetor coluna $\boldsymbol{X}$ não está muito longe do vetor coluna $\boldsymbol{x}$, o teorema de Taylor implica em que

$$Y - G(\boldsymbol{x}) \cong \left[\frac{\partial G}{\partial \boldsymbol{x}}\right]^T.(\boldsymbol{X} - \boldsymbol{x}) \qquad (26)$$

onde $[\partial G/\partial \boldsymbol{x}]^T$ é a transposta do vetor coluna que contém os *coeficientes de sensibilidade* [4]. Estes coeficientes descrevem como o estimador de saída varia com alterações nos valores dos estimadores de entrada. Podem ser avaliados da função do modelo ou por métodos numéricos. Note-se que em vez de linearizar em torno de estimativas conhecidas, na teoria dos erros tradicional [11], a linearização é feita em torno de valores esperados desconhecidos de distribuições de freqüência, e somente com a suposição de ausência de efeitos sistemáticos pode-se obter os resultados abaixo.

Substituindo (26) nas expressões (24) e (25) obtém-se:

$$y = \mathrm{E}Y = G(\boldsymbol{x}) \tag{27}$$

$$u(y) = \sqrt{\sum_{i=1}^{n}\left(\frac{\partial G}{\partial x_i}\right)^2 .u^2(x_i) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\frac{\partial G}{\partial x_i}\frac{\partial G}{\partial x_j}.u(x_i, x_j)} \tag{28}$$

Este resultado é a *lei de propagação de incertezas* apresentada no *Guia* [4]. Ela é naturalmente obtida da aplicação rígida do formalismo Bayesiano.

9.0 CONCLUSÃO

A teoria da incerteza de medição resumida neste trabalho é inteiramente baseada na estatística Bayesiana e no Princípio de Máxima Entropia de Informação. Os métodos de avaliação de incerteza Tipos A e B não são mais distinguidos. A distinção feita na recomendação do BIPM [12] é apenas uma questão de conveniência. Os *métodos estatísticos* e os *outros métodos* são ambos unificados e identificados com os métodos da estatística Bayesiana.

O teorema de Bayes é usado para inferir o valor esperado desconhecido da distribuição de freqüência, considerando a informação estatística observada e a informação de que o valor esperado pertence a uma distribuição de freqüência normal. O resultado da inferência dá o estado de conhecimento sobre o valor esperado, onde este estado de conhecimento é representado quantitativamente por uma distribuição t-Student.

O Princípio de Máxima Entropia de Informação é usado para estabelecer uma densidade de probabilidade no caso em que a única informação disponível for de um tipo tal que não possa ser expressa por métodos estatísticos. Se em uma determinada etapa da avaliação dispõe-se das melhores estimativas e incertezas padrão associadas das grandezas de entrada relevantes, este estado de conhecimento é representado por uma Gaussiana generalizada.

O teorema de Bayes é então usado para inferir a densidade de probabilidade (e seus parâmetros) do estimador de saída com base na informação anterior disponível do modelo e dos dados. Mostra-se que a lei de propagação de incertezas recomendada pelo *Guia* [4] é válida para o caso específico em que se pode linearizar o modelo (a maioria dos casos encontrados na prática).

9. AGRADECIMENTOS

O primeiro autor agradece a todos os que, direta ou indiretamente, o ajudaram na elaboração deste trabalho. O Prof. Dr. Maurício Nogueira Frota, Diretor de Metrologia Científica e Industrial do INMETRO, e Luiz Carlos Gomes dos Santos, Chefe da Divisão de Metrologia Elétrica do INMETRO, o incentivaram a participar de cursos sobre o assunto promovidos pelo Programa RH Metrologia em 1997. O primeiro autor é especialmente grato ao segundo autor, Wolfgang Wöger, do Physikalisch-Technische Bundesanstalt (PTB), por seus esforços em transmitir seus conhecimentos adquiridos como participante ativo da elaboração do *Guia*. O autor se beneficiou também das discussões com o Prof. Ignacio Lira, da Pontificia Universidad Católica de Chile, durante seu período sabático nos laboratórios do INMETRO no corrente ano. O autor também agradece aos funcionários das bibliotecas da Universidade Federal do Rio de Janeiro (UFRJ), cujo acervo e pronto atendimento foram fundamentais à execução deste trabalho.

10. REFERÊNCIAS BIBLIOGRÁFICAS

1. FELLER, W., *An Introduction to Probability Theory and Its Applications*, John Wiley & Sons, Inc., New York, third edition, 1950.
2. JEFFREYS, H., *Theory of Probability*, Oxford University Press, London, third edition, 1961.
3. WEISE, K. & WÖGER, W., A Bayesian Theory of Measurement Uncertainty, *Meas. Sci. Technol.*, 4, pág. 1 - 11, 1993.
4. *Guide to the Expression of Uncertainty in Measurement*, first edition, 1993, corrected and reprinted 1995, International Organization for Standardization, Geneva, Switzerland. Disponível no idioma português, 1997, INMETRO.
5. WÖGER, W., Bayesian Statistics and the International Guidelines for Expressing Uncertainty in Measurement, *Advanced School of Metrology: Evaluation of Uncertainty in Measurement*, Angra dos Reis, RJ, Brazil, 1997.
6. OLESKO, K., The Measuring of Precision: The Exact Sensibility in Early Nineteenth-century Germany. *The Values of Precision*, Ed. N. Wise, New Jersey, Princeton University Press, 1994.
7. LEE, P.M., *Bayesian Statistics: An Introduction*, Edward Arnold, London, 1989.
8. SHANNON, C., A Mathematical Theory of Communication, *Bell Syst. Tech. J.* **27**, pág. 379-423 e 623-655, 1948.
9. JAYNES, E.T., Information Theory and Statistical Mechanics, *Phys. Rev.* **106**, pág. 620-630, 1957.
10. SMITH, C.R. & GRANDY, W.T., (eds.), *Maximum Entropy and Bayesian Methods in Inverse Problems*, Dordrecht, Reidel, 1985.
11. KU, H.H., Notes on the Use of Propagation of Error Formulas, *J. Res. Nat. Bur. Stand. - C. Eng. And Instrum.* Vol. **70C**, No.4, October-December, 1966.
12. *Recommandation INC-1*, BIPM, 1980. Disponível nos idiomas francês, inglês e português como parte da referência [4].