

MÉTODOS ESTATÍSTICOS EM CIÊNCIA FORENSE

Palavras-Chave: EVIDÊNCIA, FATOR DE BAYES, FORENSE

Autores(as):

ERIC PAVARIM LIMA, IMECC – UNICAMP

Prof. Dr. ALEX RODRIGO DOS SANTOS SOUSA (orientador), IMECC - UNICAMP

INTRODUÇÃO:

O presente projeto tem como objetivo a revisão bibliográfica de aplicações de técnicas estatísticas em problemas forenses, com enfoque em diversos temas, de tal forma que, o foco principal é a compreensão dos pontos principais da rede de cada tema além de eventuais casos criminais reais para auxiliar a importância da quantificação correta [1] [11] e simulações estatísticas (com uso da linguagem R) de mesma natureza.

Os temas abordados foram, cronologicamente, concentração de compostos químicos em fragmentos de vidro (devido ao pioneirismo na estatística forense) [2], perfis de DNA juntamente com impressões digitais e, por último, espectroscopia.

METODOLOGIA:

No âmbito de fragmentos de vidro, o principal problema consiste na identificação de uma possível origem comum de diferentes fragmentos de vidros, comumente entre encontrados na roupa de algum suspeito e na cena do crime cujo vidro fora quebrado, conseqüentemente gera-se a conexão entre o crime em questão e o suspeito e a identificação do culpado. Para isso, percebe-se a suma importância do contexto bayesiano em aplicações forenses com o início em cálculos de fator de Bayes (ou fator de verossimilhança) [2] para quantificação e avaliação de evidências forenses. Vale comentar que o fator de Bayes é a ferramenta principal para comparação, estimação e medida de evidências (a favor ou contra um sujeito, por exemplo) pois trata-se de uma razão entre probabilidades da existência da evidência sob hipóteses complementares.

Seja H_d a hipótese do suspeito ser dono de dada evidência enquanto H_p o evento complementar. Além disso, E é o evento da evidência coletada e I todas informações prévias sobre o caso. A razão de verossimilhança, também nomeada no contexto como força da evidência, é definida como:

$$V = \frac{P(E|H_p, I)}{P(E|H_d, I)}$$

Para além disso, uma outra abordagem interessante é a “odds ratio” (ou razão de chances):

$$\theta = \frac{O_1}{O_2}$$

em outras palavras, a razão dos fatores de Bayes (O_1, O_2) para comparar, por exemplo, dois grupos contaminados de cédulas de cocaína, a razão indica que as chances a favor das cédulas de um primeiro grupo amostral estar contaminado é θ vezes maior que as chances para o segundo grupo.

Quando analisa-se o panorama de fragmentos de vidro, sua aplicação mais comum é validação dos chamados fragmentos (partículas não visíveis à olho nú) do material que fora encontrado em uma cena de crime, por exemplo, com o encontrado em suspeitos [7]. Para tal fim, a análise da concentração de compostos químicos torna-se necessária, visto que, há diferentes concentrações de certos elementos químicos em relação a dois painéis de vidro distintos, gerando, assim, uma forma de distinção entre vidros de origens diferentes.

Partindo para o estudo de DNA Profiling (Traçar perfis de DNA), o problema muda na quantificação da evidência, embora a motivação seja a mesma de encontrar uma correspondência entre o DNA encontrado em cena de crime e o de algum suspeito, daí pode-se fazer testes póstumos de DNA ou uso de banco de dados criminais já bem estabelecidos [3]. Embora haja uso forte da razão de verossimilhança e com ressalvas (como parentesco), quantificação também da evidência, há um cenário polêmico quanto à forma de coleta de DNA e a sua análise póstuma como evidência ao júri [3]. Analogamente, nota-se que a atribuição de valor às evidências de DNA e seus usos é semelhante à de impressões digitais [4], porém, a última opção tem mais relevância judicial.

Historicamente, houveram diversas formas de tentativas de quantificação de evidências de impressões digitais, tanto por funções de Kernel [5] quanto por mapeamentos geométricos de traços característicos [5]. Para além da quantificação, vale-se a discussão sobre a escolha do teste prévio à situação forense (aplicável tanto para DNA quanto para impressões digitais), isto é, se deseja-se testar indivíduos específicos (Specific Source Scenario) ou testar a evidência em comum (Common Source Scenario), pois é mostrado que a negligência da escolha do teste leva a conclusões catastróficas (exceto em casos de precisão instrumental exata e evidência intacta, como laboratorial) [9].

O cenário nomeado “Specific Source Scenario” visa testar se duas evidências diferentes vieram de alguma pessoa específica (conhecida), a hipótese nula, ou se vieram de alguma outra pessoa (desconhecida) que não seja a testada em questão. Temos, de forma breve, após cálculos que:

em que μ é a média geral da população, μ_d é a média do quantificação da impressão digital do indivíduo testado,

$$\begin{aligned} \begin{pmatrix} e_u \\ e_s \end{pmatrix} | H_{0SS} &\sim NMV\left(\begin{pmatrix} \mu_d \\ \mu_d \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_s^2 \end{pmatrix}\right) \\ \begin{pmatrix} e_u \\ e_s \end{pmatrix} | H_{1SS} &\sim NMV\left(\begin{pmatrix} \mu \\ \mu_d \end{pmatrix}, \begin{pmatrix} \sigma_u^2 + \sigma_d^2 & 0 \\ 0 & \sigma_s^2 \end{pmatrix}\right) \end{aligned}$$

e_u e e_s são as evidências, tal como σ_d , σ_s e σ_u as variâncias instrumentais, de desgaste, interferências das evidências coletadas e do dono das impressões.

No cenário “Common Source Scenario” o teste só tem interesse em saber se as evidências foram deixadas pela mesma pessoa (desconhecida) (H_0) ou por pessoas diferentes (desconhecidas) (H_1). Assim, seguindo cálculos e a natureza do problema, temos que:

$$\begin{aligned} \begin{pmatrix} e_u \\ e_s \end{pmatrix} | H_{1CS} &\sim NMV\left(\begin{pmatrix} \mu \\ \mu \end{pmatrix}, \begin{pmatrix} \sigma_u^2 + \sigma_d^2 & 0 \\ 0 & \sigma_s^2 + \sigma_d^2 \end{pmatrix}\right) \\ \begin{pmatrix} e_u \\ e_s \end{pmatrix} | H_{0CS} &\sim NMV\left(\begin{pmatrix} \mu \\ \mu \end{pmatrix}, \begin{pmatrix} \sigma_u^2 + \sigma_d^2 & \sigma_d^2 \\ \sigma_d^2 & \sigma_s^2 + \sigma_d^2 \end{pmatrix}\right) \end{aligned}$$

em que, para facilitar o entendimento manteve-se e_u e e_s , porém agora as duas significam evidências distintas de dono(s) desconhecido(s).

Após essa definição de distribuições, o poder da evidência torna-se, para ambos os casos (common source e specific source):

$$LR_{CS} = \frac{f(e_u, e_s | H_{0CS})}{f(e_u | H_{1CS})f(e_s | H_{1CS})}$$

$$LR_{SS} = \frac{f(e_u | H_{0SS})}{f(e_u | H_{1SS})}$$

Por fim, houvera o estudo de um problema de quimiometria, em que se deseja a partir das curvas de absorvância (quantificação de um certo feixe de luz refletido em algum objeto) dos constituintes de uma dada substância, estimar a curva de absorvância da substância total, assim podendo identificar alguma possível adulteração em amostras de substância subsequentes do mesmo tipo [6].

A absorvância de uma substância é definida por:

$$x = -\log_{10} \left(\frac{I}{I_0} \right)$$

em que I é a intensidade da luz transmitida, após testada na substância (“refletida”) e I_0 é a intensidade da luz incidente, antes da interação com a substância.

A lei de Beer-Lambert é definida da seguinte forma:

$$x_j = \sum_{l=1}^m y_l a_{lj} + \varepsilon_j$$

em que, brevemente, x_j é a absorvância da amostra no j -ésimo comprimento de onda ($j=1, \dots, k$), a_{lj} é a absorvância no l -ésimo constituinte puro no j -ésimo comprimento de onda, y_l é a concentração do l -ésimo constituinte e um erro aleatório ε_j para $l=1, \dots, m$ e $j=1, \dots, k$.

RESULTADOS E DISCUSSÃO:

A partir do uso do banco de dados de fragmentos de vidro, constituindo 18 variáveis de elementos químicos em 48 painéis de vidro de duas empresas distintas com um total de 166 observações (fragmentos) por painel [10], fora feita a análise descritiva extensa do mesmo, com foco em identificar possíveis padrões distinguíveis entre os vidros de diferentes origens (empresas) para, em um contexto forense, identificar quais concentrações de elementos químicos tornam-se mais relevantes para essa dissimilaridade. Observa-se na Tabela 1, que ambas as concentrações de dois painéis de vidro aleatórios por empresa coincidem, isto é, a composição percentual entre vidros é semelhante, o que altera é a quantidade absoluta de composto, como visto nas Figura 1 e 2, há compostos com concentrações drasticamente diferentes e há concentrações comuns, não servindo para distinção dos mesmos.

Composto	Empresa A	Empresa B
Na23	0.5259	0.5133
Ca42	0.3286	0.3476
Mg25	0.1194	0.1229
...
Hf178	0.000006	0.000005
Pb208	0.000006	0.000004

Tabela 1: Concentrações de compostos químicos entre dois painéis de vidro da empresa A e da empresa B

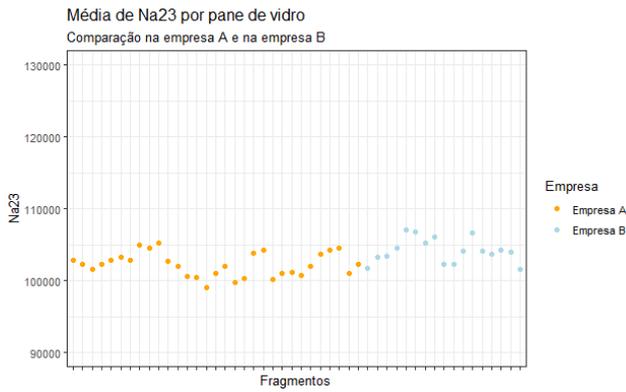


Figura 1: Gráfico de dispersão da concentração média de Na23 por fragmento de vidro em cada empresa

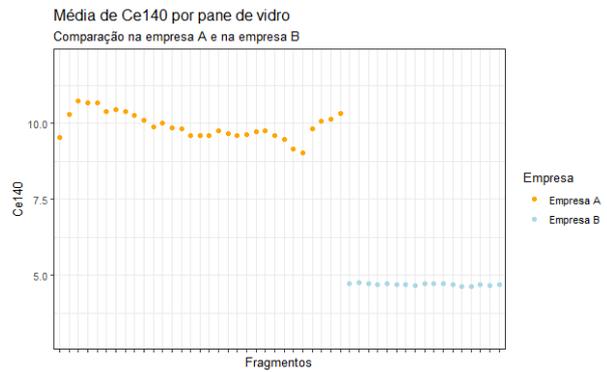


Figura 2: Gráfico de dispersão da concentração média de Ce140 por fragmento de vidro em cada empresa

Para finalizar a parte de simulações, fora replicada a simulação [8] (com uso do pacote 'mvtnorm') para distinguir a escolha do panorama investigativo, isto é, se selecionamos a hipótese de testar o indivíduo ser o dono ou não da impressão digital ou testar se a evidência faz parte ou não do mesmo dono, mesmo que desconhecido nessa segunda opção. Nota-se que ao realizar as simulações variando σ_s (tido como variância instrumental ou de coleta da evidência) e μ_d (indica a média da evidência em comparação com μ que é a média populacional geral, portanto, a proximidade das duas estatísticas indicam quão comum é a evidência ou não se escolhida aleatoriamente), percebemos nas figuras 3,4,5 e 6 que, sob hipóteses diferentes, não ocorre a convergência de razões de verossimilhança, apenas quando σ_s é negligenciável.

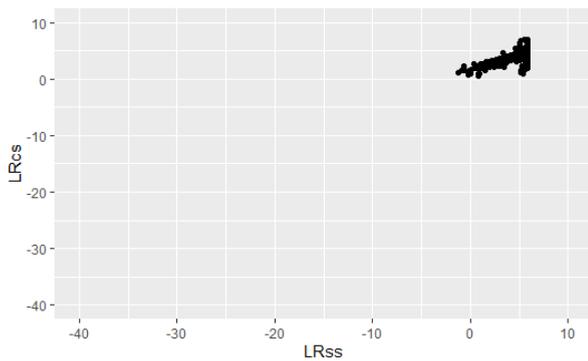


Figura 3: Gráfico de dispersão de razões de verossimilhança sob a hipótese nula de dois cenários forenses diferentes com $\mu_d = 0$ e $\sigma_s = 1$

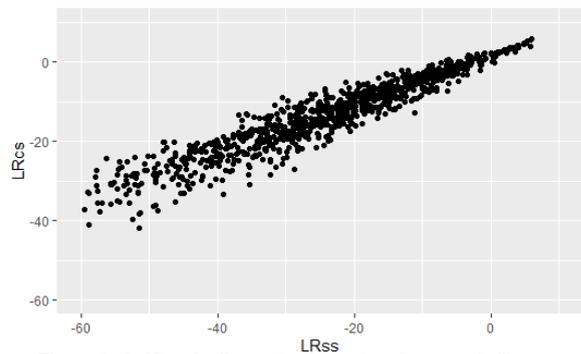


Figura 4: Gráfico de dispersão de razões de verossimilhança sob a hipótese alternativa de dois cenários forenses diferentes com $\mu_d = 0$ e $\sigma_s = 1$

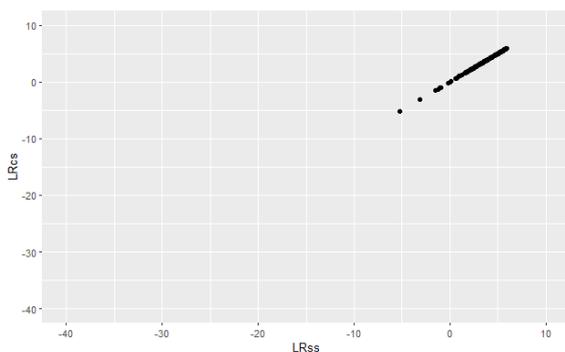


Figura 5: Gráfico de dispersão de razões de verossimilhança sob a hipótese nula de dois cenários forenses diferentes com $\mu_d = 0$ e $\sigma_s = 0,000001$

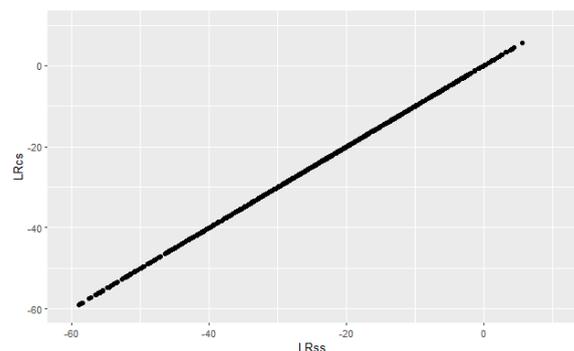


Figura 6: Gráfico de dispersão de razões de verossimilhança sob a hipótese alternativa de dois cenários forenses diferentes com $\mu_d = 0$ e $\sigma_s = 0,000001$

CONCLUSÕES:

Em suma, a aplicação estatística em confronto com as metodologias mais antigas está gradualmente tomando espaço devido a sua eficácia e custo. Em diversos contextos, entretanto, cabe ao estatístico, cientista em cargo tomar os devidos cuidados na quantificação de evidências das mais diferentes naturezas (desde fragmentos de vidro à impressões digitais), pois a escolha incorreta do panorama em questão acarreta em consequências gravíssimas, como conclusões errôneas [12].

BIBLIOGRAFIA

- [1] ROYALL, R., **On the Probability of Observing Misleading Statistical Evidence**. Alexandria, American Statistical Association, 2000
- [2] LINDLEY, D.V. , **A Problem in Forensic Science**. Oxford, Oxford University Press on behalf of Biometrika Trust, 1977
- [3] BALDING, D.J. e DONNELLY, P., **Evaluating DNA Profile Evidence When the Suspect Is Identified Through a Database Search**. Journal of Forensic Sciences, 1996
- [4] Rise v. State of Oregon
- [5] NEUMANN, C. et al, **Quantifying the weight of evidence from a forensic fingerprint comparison: a new paradigm**. Oxford, Journal of the Royal Statistical Society - Series A, 2012
- [6] SARAIVA, MA, **Análise não-paramétrica de dados funcionais : uma aplicação a quimiometria**. Campinas, Universidade Estadual de Campinas, 2009
- [7] EVETT, I.W e BUCKLETON, J., **The interpretation of glass evidence. A practical approach**. Journal of the Forensic Science Society, 1990
- [8] BANKS, D.L. et al, **Handbook of Forensic Statistics**. Chapman & Hall/CRC, 2001
- [9] AITKEN, C.G.G. e TARONI, F. **Statistics and the Evaluation of Evidence for Forensic Scientists**. Wiley, 2004
- [10] CSAFE - Center for Statistics and Applications in Forensic Evidence
- [11] NOBLES, R. e SCHIFF, D., **Misleading statistics within criminal trials: The Sally Clark case**. Significance magazine, 2005
- [12] KERN, J., **Fingerprinting: A Study in Cognitive Bias and its Effects on Latent Fingerprint Analysis**. Undergraduate Honors College Theses, Long Island University, 2016