

# Prognóstico de Câncer de Próstata: Probabilidade e Possibilidade

M.J.P. CASTANHO<sup>1</sup>, Departamento de Matemática, UNICENTRO, 85010-990 Guarapuava, PR, Brasil

L.C. BARROS<sup>2</sup>, Departamento de Matemática Aplicada, IMECC-UNICAMP, 13083-970 Campinas, SP, Brasil.

**Resumo.** Para decidir o tratamento do câncer de próstata é necessário prever seu estágio. A literatura urológica dispõe de vários nomogramas baseados na teoria de probabilidade para auxiliar o médico nas previsões. Nesse artigo, propomos um sistema baseado em regras *fuzzy* (SBRF) para realizar essa tarefa. Utilizamos a teoria dos conjuntos *fuzzy* para desenvolver o modelo, por sua capacidade em lidar com incertezas inerentes às condutas médicas. O sistema baseado em regras *fuzzy* fornece resultados em termos de possibilidade do paciente estar em determinado estágio patológico. Simulamos resultados para alguns pacientes e comparamos com as probabilidades fornecidas pelas Tabelas de Partin. Os resultados são consistentes quando se adota probabilidade ou possibilidade para estudar o estágio patológico da doença.

## 1. Introdução

Avaliar o estágio em que o câncer de próstata se encontra é necessário para poder decidir a terapia a ser aplicada. Se o tumor estiver confinado no órgão a prostatectomia radical ou radioterapia são os tratamentos mais efetivos. Porém, se desenvolveu metástase, tratamentos paliativos são indicados com o objetivo de reduzir o crescimento do tumor e aliviar os sintomas do paciente.

Para auxiliar o médico na determinação do estágio, vários nomogramas baseados na teoria da probabilidade foram desenvolvidos [9]. Em 1997, Partin *et al.* utilizaram as variáveis pré-cirúrgicas: nível de antígeno prostático específico (PSA), escore de Gleason e estado clínico, e por meio de uma regressão multinomial log-linear encontraram a probabilidade do indivíduo ter câncer confinado no órgão, envolvimento capsular, de vesículas seminais e/ou linfonodos pélvicos [7]. Essas tabelas foram atualizadas em 2001 [6]. Contudo, um estudo feito por Augustin *et al.* [1] comparando sua acurácia mostrou não haver diferença significativa entre as tabelas publicadas em 1997 e 2001.

---

<sup>1</sup>zeza@unicentro.br

<sup>2</sup>laeciocb@ime.unicamp.br

Tendo em vista que as variáveis utilizadas por Partin *et al.* são incertas, nesse artigo desenvolvemos um sistema baseado em regras utilizando a teoria dos conjuntos *fuzzy* [12]. Esse sistema é estruturado no conceito de conjunto *fuzzy*, ou seja, conjuntos com fronteiras indefinidas nos quais a transição de pertinência para não-pertinência é gradual ao invés de abrupta.

## 2. Sistema Baseado em Regras *Fuzzy*

Um sistema baseado em regras *fuzzy* (SBRF) contém quatro componentes principais: um codificador que representa as variáveis de entrada e saída em conjuntos *fuzzy*, uma base de conhecimento, um método de inferência, e um decodificador que transforma a saída em um valor numérico. A Figura 1 ilustra o esquema do SBRF aqui utilizado.

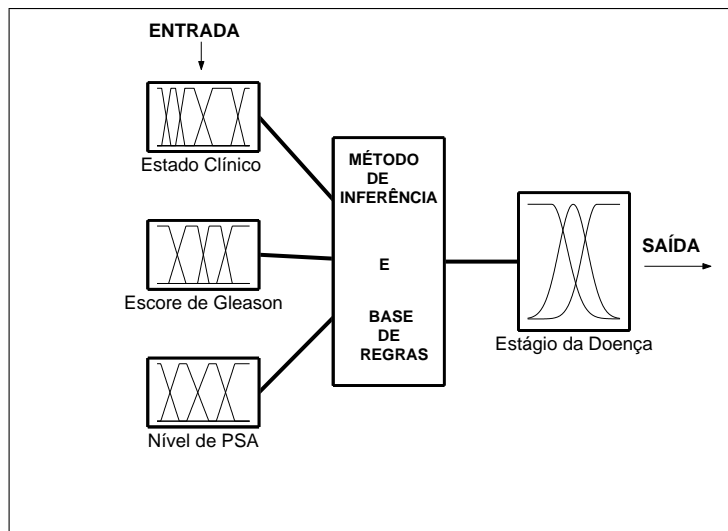


Figura 1: Estrutura do sistema baseado em regras *fuzzy* construído para prever o estágio patológico do câncer de próstata.

Para criar o modelo *fuzzy*, inicialmente consideramos as variáveis que permitem a descrição do fenômeno em termos lingüísticos, isto é, ao invés de números reais, seus valores são números *fuzzy*. Então, modelamos as variáveis de entrada do sistema — *Estado Clínico*, *Escore de Gleason* e *Nível de PSA* — por conjuntos *fuzzy* em seus respectivos domínios.

O estado clínico é determinado pelo exame retal digital que depende da percepção e experiência do médico, complementado, ou não, por modalidades de imagem. Logo, *Estado Clínico* é uma variável subjetiva, para a qual foram considerados cinco estágios, traduzidos por conjuntos *fuzzy* com base na tabela TNM [7]. Os estágios *T1a*, *T1b* e *T1c* foram classificados apenas como *T1* (não-palpável), pois o

que os difere na tabela são o nível de PSA e o grau de diferenciação do tumor, que são, também, variáveis desse modelo.

Como o estado clínico está relacionado com o tamanho do tumor, o domínio dado pelo intervalo  $[0, 1]$  representa o percentual de comprometimento da próstata, ou seja, 0 representa uma próstata sem tumor enquanto 1 representa a próstata totalmente tomada pelo tumor. A partir do estágio  $T3$ , o tumor já se espalhou para além da próstata; portanto, os estágios seguintes não serão considerados no modelo. Assim, os termos lingüísticos empregados para a variável *Estado Clínico* são:  $T1$ , tumor não palpável;  $T2a$ , tumor palpável, circunscrito a menos da metade de um lobo;  $T2b$ , tumor palpável, afetando mais da metade de um lobo mas não os dois;  $T2c$ , tumor palpável, comprometendo os dois lobos e  $T3$ , tumor com extensão extracapsular.

O nível de PSA até 4 ng/ml é considerado normal, embora a linha de corte entre normal e alterado não seja bem definida. Os níveis entre 4 e 10 ng/ml são considerados levemente elevados, entre 10 e 20 ng/ml, moderadamente elevados e acima de 20 ng/ml, altamente elevados. Assim, o nível de PSA pode ser considerado uma variável lingüística, cujos valores são classificados segundo os termos *Normal*, *Levemente Elevado*, *Moderadamente Elevado* e *Altamente Elevado* de forma a propiciar uma transição gradual entre os níveis para evitar o corte abrupto como nas faixas das tabelas de probabilidade.

O escore de Gleason é atribuído por um patologista, após a análise de uma amostra do tumor, que é bastante heterogêneo; portanto, temos um valor preciso representando uma situação imprecisa. Para essa variável utilizamos os termos lingüísticos: *Bem Diferenciado*, *Moderadamente Diferenciado*, *Pouco Diferenciado* e *Indiferenciado*, de acordo com a classificação feita por Seabra [10]: lesões bem diferenciadas (que têm um comportamento menos agressivo) têm grau 2, 3 e 4; moderadamente diferenciadas, graus 5 e 6; pouco diferenciadas, grau 7 e as indiferenciadas (mais agressivas) são caracterizadas com graus 8, 9 e 10.

O método aqui adotado para reunir todas as informações e posteriormente produzir inferências foi o de Mamdani, no qual a variável de saída *Estágio da Doença* deve ser modelada por conjuntos *fuzzy*. Para essa variável atribuímos os seguintes termos lingüísticos: *Confinado*, indicando que todo o câncer está confinado na cápsula prostática; *Capsular*, se, mesmo com o câncer já fora da cápsula prostática, as vesículas seminais e linfonodos estiverem livres do tumor, e *Vesícula+Linfonodos* se o câncer já invadiu a parede muscular de uma ou ambas as vesículas seminais e/ou os linfonodos pélvicos. Como essa é uma variável qualitativa, escolhemos uma escala de 0 a 1 para indicar a extensão do tumor: quanto mais próximo de 0, menor o tamanho do tumor e quanto mais próximo de 1 maior o comprometimento das estruturas adjacentes à próstata. (Ver Figura 2).

As regras *fuzzy* definem a conexão entre as variáveis *fuzzy* de entrada e saída. Foram elaboradas tendo em vista as informações contidas nos nomogramas construídos por Partin *et al.* [7].

A base de regras consiste de 116 regras do tipo:

1. “Se *Estado Clínico* é  $T1$  e *Escore de Gleason* é Moderadamente Diferenciado e *Nível de PSA* é Normal, então *Estágio da Doença* é Confinado.”

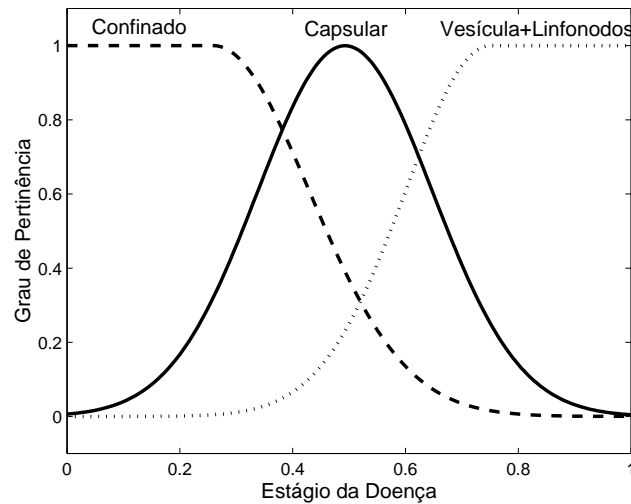


Figura 2: Funções de pertinência aos conjuntos *fuzzy* assumidos pela variável lingüística *Estágio da Doença*. A saída do sistema é um número no eixo das abscissas ao qual corresponde um valor de pertinência a cada um dos conjuntos: *Confinado*, *Capsular* e *Vesícula+Linfonodos*.

2. “Se *Estado Clínico* é T2b e *Escore de Gleason* é Bem Diferenciado e *Nível de PSA* é Levemente Elevado, então *Estágio da Doença* é Capsular.”
3. “Se *Estado Clínico* é T2c e *Escore de Gleason* é Indiferenciado e *Nível de PSA* é Moderadamente Elevado, então *Estágio da Doença* é Vesícula+Linfonodos.”

A base de regras e o conjunto das funções de pertinência que representam as variáveis de entrada e saída constituem a base de conhecimento.

A saída do método de inferência de Mamdani é um conjunto *fuzzy*. Um número real representativo desse conjunto é encontrado através do método do centro de gravidade.

A cada valor real, que representa a saída do sistema baseado em regras *fuzzy*, corresponde um grau de pertinência ao conjunto *fuzzy* que descreve determinado estágio da doença: *Confinado*, *Capsular* e *Vesícula+Linfonodos*.

A proposição “*Estágio da Doença* é *Confinado*”, por exemplo, permite ver esse grau como a possibilidade de que a doença esteja confinada no órgão. Dessa forma, essa proposição define uma distribuição de possibilidade no universo dos pacientes em um estágio da doença.

Para que possamos verificar se as informações dos nomogramas, dadas em termos probabilísticos, são consistentes com os resultados que obtivemos por meio do sistema baseado em regras *fuzzy*, dados em termos possibilísticos, uma transformação possibilidade-probabilidade é necessária.

### 3. Transformação Possibilidade-Probabilidade

Seja  $\Omega$  um espaço amostral de um experimento e  $\mathbf{P}(\Omega)$  o conjunto de todos os eventos de  $\Omega$ .

Uma medida de probabilidade  $P$  é uma função real que determina, para todo evento  $A$  em  $\Omega$ , uma probabilidade  $P(A)$  tal que:

1.  $P(A) \geq 0$ ;
2.  $P(\Omega) = 1$ ;
3.  $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$ , onde  $A_i$  é uma coleção de eventos disjuntos.

Uma medida de possibilidade  $\Pi$  em  $\Omega$  é uma função de  $\mathbf{P}(\Omega)$  em  $[0, 1]$ , dada por

$$\Pi(A) = \begin{cases} \sup_{x \in A} [\pi(x)] & \text{se } A \neq \phi \\ 0 & \text{se } A = \phi \end{cases}$$

onde  $\Pi : \Omega \rightarrow [0, 1]$  com  $\Pi(x) = 1$  para algum  $x \in \Omega$ . Dessa forma tem-se:

1.  $\Pi(\phi) = 0$ ;
2.  $\Pi(\Omega) = 1$ ;
3.  $\Pi(\bigcup_{i \in I} A_i) = \sup \Pi(A_i)$  para qualquer família  $\{A_i\}$  de subconjuntos de  $\Omega$ .

Quando informações sobre um fenômeno são dadas em termos probabilísticos e possibilísticos, as duas descrições devem ser, de alguma forma, consistentes [5], isto é, dada uma medida de probabilidade  $P$  e uma medida de possibilidade  $\Pi$ , em  $\Omega$ , as duas medidas devem satisfazer alguma condição de consistência, a qual revela concordância dos resultados.

Segundo Dubois e Prade [3], o grau de possibilidade de um evento é maior ou igual a seu grau de probabilidade, ou seja,

$$P(A) \leq \Pi(A), \quad \forall A \subset \Omega,$$

onde  $P$  é a medida de probabilidade. Quando  $\Omega$  é discreto e  $p(x)$  é a probabilidade de  $x \in \Omega$ ,

$$P(A) = \sum_{x \in A} p(x).$$

Por outro lado, como vimos antes,  $\Pi$  é a medida de possibilidade associada a  $\pi$ ,

$$\Pi(A) = \sup_{x \in A} [\pi(x)].$$

Obedecer o princípio de consistência acima é essencial para qualquer transformação de probabilidade/possibilidade. A seguir descreveremos uma das transformações mais comuns [5].

Sejam  $\Omega = \{x_1, x_2, \dots, x_n\}$ ,  $p_i = p(x_i)$  e  $\pi_i = \pi(x_i)$ , com  $i = 1, \dots, n$  e

$$\pi_1 = 1 \quad \text{e} \quad \sum_{i=1}^n p_i = 1.$$

Supomos que os elementos de  $\Omega$  estão ordenados de tal forma que as distribuições de possibilidade

$$\Pi = \langle \pi_1, \pi_2, \dots, \pi_n \rangle$$

e distribuições de probabilidade

$$P = \langle p_1, p_2, \dots, p_n \rangle$$

são sempre seqüências não crescentes. Então, as transformações mais simples,  $P \leftrightarrow \Pi$ , são expressas pelas equações:

$$\pi_i = \frac{p_i}{p_1}, \quad (3.1)$$

$$p_i = \frac{\pi_i}{\sum_{i=1}^n \pi_i}. \quad (3.2)$$

Além desse, diversos tipos de transformações têm sido sugeridos na literatura [2, 4, 5, 11].

## 4. Resultados

Para um paciente com dados pré-cirúrgicos conhecidos, é possível determinar a possibilidade dele estar em determinado estágio da doença. Fizemos simulações do modelo para alguns pacientes e transformamos a possibilidade encontrada em probabilidade utilizando a equação (3.2). Alguns resultados estão descritos na Tabela 1.

Seja, por exemplo, um paciente com estágio clínico T2a, isto é, com tumor palpável, circunscrito a menos da metade de um lobo; a biópsia revela um escore de Gleason igual a 7 e o nível de PSA é igual a 5,3 ng/ml. Segundo o SBRF, a possibilidade de que esse paciente tenha câncer confinado ao órgão é de 0,60, de que tenha envolvimento capsular é de 0,93 e de que tenha envolvimento de vesículas e/ou linfonodos é de 0,11. Transformando essas possibilidades em probabilidades, a partir da fórmula (3.2), tem-se 0,36, 0,57 e 0,07, respectivamente. Esse resultado é coerente com os nomogramas [7], nos quais as probabilidades são 0,33, 0,52 e 0,14, respectivamente. Em torno de 80% dos casos simulados o resultado encontrado estava no intervalo de confiança das tabelas de Partin.

Dentre os 20% que diferiram, observamos que, pacientes com mesmo escore de Gleason, mesmo estado clínico e níveis de PSA próximos à fronteira das faixas da tabela (por exemplo iguais a 10,6 ng/ml e 9,3 ng/ml) obtiveram o mesmo resultado no SBRF. Segundo Partin *et al.* [7], eles têm probabilidades diferentes de estar em determinado estágio patológico, enquanto que um paciente com nível de PSA igual a 20 ng/ml tem a mesma probabilidade daquele com 10,6 ng/ml.

Augustin *et al* [1] utilizam a área sob a curva ROC [13] para testar a acurácia das tabelas de Partin, encontrando 0,727. Isso significa que esses nomogramas predizem corretamente a probabilidade de um estágio patológico 72% das vezes [8]. Em nosso caso, para uma amostra de 190 pacientes a área sob a curva é 0,76.

Tabela 1: Resultados obtidos pelo sistema baseado em regras *fuzzy*. Na terceira coluna estão as possibilidades obtidas pelo SBRF. A coluna *Probabilidade* é a transformação das possibilidades através da equação (3.2). Na última coluna estão as porcentagens constantes dos *Nomogramas* elaborados por Partin [7].

		SBRF			Partin <i>et al.</i>
Paciente		Estágio	Possibilidade	Probabilidade	Nomograma
Est. clínico	T1	Confinado	0,63	0,39	33(25-42)
Gleason	6	Capsular	0,91	0,55	49(38-59)
PSA	17	Ves+Linf	0,09	0,06	17(7-33)
Est. clínico	T2a	Confinado	0,97	0,66	68(63-72)
Gleason	5	Capsular	0,48	0,33	30(26-35)
PSA	2	Ves+Linf	0,01	0,01	2(1-4)
Est. clínico	T2b	Confinado	0,17	0,11	14(9-19)
Gleason	10	Capsular	0,84	0,55	46(39-53)
PSA	9	Ves+Linf	0,53	0,34	40(28-56)
Est. clínico	T2c	Confinado	0,14	0,09	15(11-19)
Gleason	7	Capsular	0,78	0,51	45(39-52)
PSA	15	Ves+Linf	0,60	0,40	40(28-54)
Est. clínico	T3a	Confinado	0,01	0,01	1(0-2)
Gleason	10	Capsular	0,19	0,16	17(11-26)
PSA	25	Ves+Linf	0,99	0,83	82(52-100)
Est. clínico	T2b	Confinado	0,61	0,38	29(20-40)
Gleason	2	Capsular	0,92	0,56	52(39-65)
PSA	23	Ves+Linf	0,10	0,06	17(4-40)
Est. clínico	T2c	Confinado	0,29	0,18	25(20-30)
Gleason	7	Capsular	0,97	0,61	48(42-54)
PSA	10	Ves+Linf	0,33	0,21	27(19-37)
Est. clínico	T2a	Confinado	0,60	0,36	33(29-38)
Gleason	7	Capsular	0,93	0,57	52(48-57)
PSA	5,3	Ves+Linf	0,11	0,07	14(11-19)
Est. clínico	T2a	Confinado	1,00	0,72	71(64-78)
Gleason	3	Capsular	0,37	0,27	26(19-33)
PSA	8,5	Ves+Linf	0,01	0,01	2(1-6)
Est. clínico	T2b	Confinado	0,18	0,12	14(9-19)
Gleason	8	Capsular	0,87	0,56	46(39-53)
PSA	9,7	Ves+Linf	0,50	0,32	40(28-56)

## 5. Conclusão

O sistema baseado em regras *fuzzy* aqui construído é uma alternativa para estimar o estágio patológico de pacientes com câncer de próstata. Pode ser útil ao médico para auxiliar em sua decisão com relação ao tratamento.

A vantagem de nossa metodologia com relação às tabelas de probabilidade é considerar uma transição gradual nas faixas da tabela e entre os estágios. É mais abrangente no sentido de que o médico pode avaliar o estado clínico numa escala contínua entre 0 e 1 ao invés de classificá-lo numa das 7 categorias da Tabela TNM. Como a cada valor do intervalo  $[0,1]$  corresponde uma pertinência aos conjuntos  $T1$ ,  $T2a$ ,  $T2b$ ,  $T2c$  e  $T3$ , ele pode considerar essa pertinência como o seu grau de confiança em classificar o paciente naquele estágio. Dessa forma, a teoria dos conjuntos *fuzzy* permite incluir a subjetividade existente nas variáveis lingüísticas utilizadas.

Um sistema baseado em regras *fuzzy* pode ser construído com o auxílio de especialistas, ou baseado em dados estatísticos, ou ambos. Aqui, utilizamos informações da literatura e verificamos que as simulações forneceram resultados coerentes em termos probabilísticos (tabelas de Partin) e possibilísticos (SBRF).

**Abstract.** Prostate cancer is one of the most common carcinoma in men in the world. Predicting the stage of the cancer is necessary to decide the appropriate therapy. There are several nomograms in the urological literature based on theory of probabilities to help physicians in their predictions. In this paper we propose a based-rule fuzzy system (BRFS) to predict pathological stage of prostate cancer. The system was developed based on fuzzy sets theory because it is suitable to deal with uncertain information. The results of the BRFS are given in possibilities of the patient being in some pathological stage. We simulated results for some patients and compared them with the probabilities given by Partin tables. We found consistent results when choose probabilities or possibilities to estimate the stage of the disease.

## Referências

- [1] H. Augustin, T. Eggert, S. Wenske, P.I. Karakiewicz, J. Palisaar, F. Daghofer, H. Huland, M. Graefen, Comparison of Accuracy between the Partin Tables of 1997 and 2001 to predict final pathological stage in clinically localized prostate cancer, *J. Urol.*, **171** (2004), 177-81.
- [2] M.R. Civanlar, H.J. Trussell, Constructing membership functions using statistical data, *Fuzzy Sets Syst.*, **18** (1986), 1-13.
- [3] D. Dubois, H. Prade, "Fuzzy Sets and Systems - Theory and Applications", Academic Press, Inc., San Diego, CA, USA, 1980.
- [4] D. Dubois, H. Prade, S. Sandri, On Possibility/Probability Transformations, <http://citeseer.nj.nec.com/cache/papers/cs>, (1993).
- [5] G.K. Klir, B. Yuan, "Fuzzy Sets and Fuzzy Logic - Theory and Applications", Prentice Hall Inc., New Jersey, USA, 1995.



- [6] A.W. Partin, L.A. Mangold, D.M. Lamm, P.C. Walsh, J.I. Epstein, J.D. Pearson, Contemporary update of prostate cancer staging nomograms (Partin Tables) for the New Millennium, *Urology*, **58**, No. 6 (2001), 843-8.
- [7] A.W. Partin, M.W. Kattan, E.N.P. Subong, P.C. Walsh, K.J. Wojno, J.E. Oesterling, P.T. Scardino, J.D. Pearson, Combination of Prostate-Specific Antigen, Clinical Stage, and Gleason Score to Predict Pathological Stage of Localized Prostate Cancer, *JAMA*, **277**, No. 18 (1997), 1445-51.
- [8] T.J. Polascik, J.E. Oesterling, A.W. Partin, Prostate Specific Antigen: a decade of discovery - what we have learned and where we are going, *J. Urol.*, **162** (1999), 293-306.
- [9] P.L. Ross, P.T. Scardino, M.W. Kattan, A catalog of prostate cancer nomograms, *J. Urol.*, **165** (2001), 1562-8.
- [10] D.D.G. Seabra, "Câncer de Próstata - Programa Educacional", Parte integrante da Tese de Doutorado, Universidade Federal de São Paulo, UNIFESP-EPM, São Paulo, SP, 2000.
- [11] T. Sudkamp, On probability-possibility transformations, *Fuzzy Sets Syst.*, **51** (1992), 73-81.
- [12] L.A. Zadeh, Fuzzy sets, *Inform. Control*, **8** (1965), 338-53.
- [13] M.H. Zweig, G. Campbell, Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine, *Clin. Chem.*, **39**, No. 4 (1993), 561-77.