

A REINCIDÊNCIA DE EVENTOS RAROS NAS ORGANIZAÇÕES- MODELAGEM PELA DISTRIBUIÇÃO DE POISSON

Prof. Newton Braga Rosa nbragarosa@gmail.com 051 9987 7904

“Eventos raros, quando ocorrem, tendem a se repetir”.

RESUMO:

Quando ocorre um evento raro numa organização, o senso comum poderia concluir que ele não deve acontecer novamente. Talvez este evento seja taxado como uma "fatalidade" e, portanto, não se justificaria o custo da implantação de estratégias e medidas preventivas. Este trabalho propõe um modelo para prever a reincidência de eventos raros nas corporações, fundamentado na distribuição de Poisson. Mais precisamente, na distribuição de probabilidades Exponencial dos tempos entre dois eventos raros consecutivos. O trabalho procura demonstrar que eventos raros aleatórios, decorrentes de variáveis incontrolláveis de dentro e de fora da organização, não aparecem uniformemente dispersos na linha do tempo. Ao contrário, eles formam grupos, separados entre si por longos períodos de "silêncio". A modelagem aqui proposta permite fazer afirmações probabilísticas sobre a reincidência de um determinado evento raro (um "outlier"), para subsidiar a tomada de decisões e a gestão de risco nas corporações. O modelo proposto demonstra estatisticamente que, ao contrário do que se possa presumir, eventos raros, quando ocorrem, tendem a se repetir.

ABSTRACT

When a rare event occurs in an organization, common sense would conclude that it should not happen again. Maybe this event would be understood as a "fatality" and therefore it would not justify the cost of deploying strategies and preventive measures. This paper proposes a model to predict the recurrence of rare events in corporations, based on Poisson distribution, more precisely, the exponential probability distribution of time between two consecutive rare events. According to this exponential distribution, random rare events due to uncontrollable variables from inside and outside the organization, do not appear uniformly dispersed in the timeline. At the contrary, they form groups, separated by long periods of "silence". The model proposed here allows one to make probability statements about the recurrence of a particular rare event (an "outlier"), supporting decision making and risk management in corporations. The

proposed model demonstrates statistically that, in opposition to what one might assume, rare events, when they occur, tend to repeat.

1- Introdução e contexto

A distribuição de probabilidades de Poisson é conhecida na literatura estatística por permitir a modelagem de eventos raros (ER). Portanto, apesar de aleatórios, eventos raros não são completamente imprevisíveis, por mais que isso contrarie o senso comum. A distribuição de Poisson é usada nas ciências da natureza (como engenharias e afins) para antever a chegada de eventos raros em servidores da Internet, em centrais telefônicas, estações de call centers, entre várias outras aplicações. Entretanto, pouco existe na área da gestão organizacional sobre técnicas e experiências empíricas de modelos não gaussianos, muito menos sobre a distribuição de Poisson. O desafio deste trabalho é estabelecer uma ponte entre a distribuição de Poisson, uma técnica estatística consagrada nas ciências naturais, e a gestão de risco e tomada de decisões nas organizações.

A modelagem aqui proposta demonstra que os eventos raros, inclusive os de alto impacto, não aparecem uniformemente dispersos na linha do tempo. Ao contrário, eles formam grupos, separados entre si por longos períodos de “silêncio”. Existem vários exemplos ilustrativos: acidentes aeronáuticos, apagões de energia, acidentes graves de trânsito, falhas críticas na execução de rotinas organizacionais, erros nos processos de gestão de dados quanto a confidencialidade, integridade ou disponibilidade, entre outros casos, alguns de consequências catastróficas.

Uma primeira motivação deste trabalho foi um artigo de Hansen (2004) que defende vantagens da estatística de Bayes, sobre métodos estatísticos convencionais, para fazer inferências preditivas envolvendo “outliers”. Hansen aborda as limitações da estatística convencional, a qual recomenda o expurgo de “outliers” das amostras em análise: *“The statistical methods used in applying the theory (about extraordinary performers or outliers) must account for individual firm differences, and not be based on means, which statistically neutralize firm differences”* (HANSEN, 2004, p.1280). *“If a researcher finds that an observation is an outlier, then the observation could justifiably be removed from the analysis”* (PERRY, 2005, p.308).

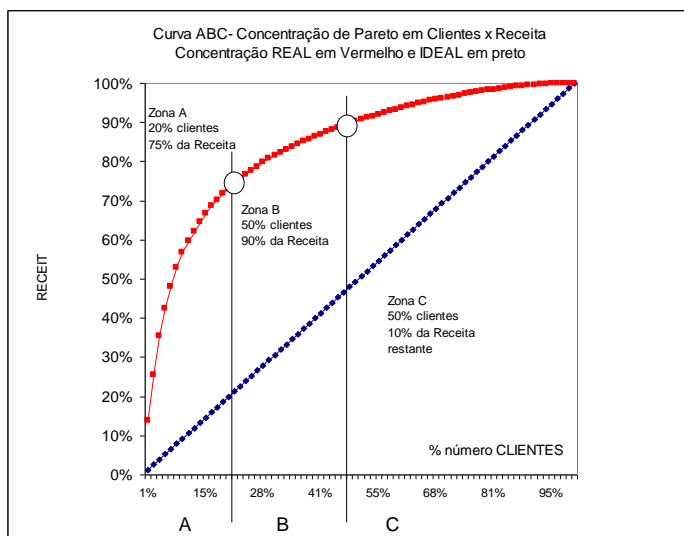
Nas comunidades biológicas, a dominância numérica de poucas espécies é uma das poucas regras gerais da ecologia. Vários modelos quantitativos foram propostos para

descrever este padrão de concentração (PRADO, 2010), (ASSUNÇÃO, 1998). No mundo corporativo, o teorema de Pareto e a lei de Benford modelam a concentração de fenômenos, permitindo fazer inferências probabilísticas nas organizações, apesar daquelas concentrações, eventualmente, contrariarem o senso comum. O presente trabalho trata da concentração de eventos raros na linha do tempo. A proposição deste trabalho é que a modelagem por Poisson pode subsidiar o processo de tomada de decisão e a gestão de risco nas organizações, da mesma forma que modelos de Pareto e Benford.

Concentração de Pareto

Vilfredo Pareto (1848-1923) foi um estatístico italiano que, ao fazer censos demográficos, constatou que a maior parte da riqueza das cidades estava concentrada em um pequeno número de indivíduos. Criou uma fórmula matemática para descrever a “concentração de grande parte dos resultados em poucos fatores”, também conhecida como princípio 80-20. Ou seja, 80% das consequências advêm de, apenas, 20% das causas.

Figura 1- Curva ABC da concentração de Pareto: um pequeno número de fatores (eixo X) representa grande parte dos resultados (eixo Y)



Fonte: Elaborado pelo autor

Posteriormente, o teorema de Pareto foi enunciado como um princípio universal para distinguir o "vital few from the useful many" (JURAN, 1951). Como exemplo, grande parte do faturamento está concentrado em um pequeno percentual dos clientes; a maior parte das despesas de uma empresa estão concentradas em um pequeno número

de itens; o valor do patrimônio de uma empresa está concentrado em um pequeno número de ativos, etc.

A curva ABC é uma ferramenta de aplicação prática, largamente difundida e de utilidade comprovada do princípio de Pareto. Ela classifica graficamente os itens de maior importância ou impacto, os quais são, normalmente, em menor número (CARVALHO, 2002). A identificação dos itens mais relevantes por Pareto permite que a empresa estabeleça prioridades baseadas em critérios técnicos. Na redução de custos, a distribuição de Pareto aponta os itens mais relevantes a serem trabalhados. A distribuição de Pareto pode servir como alerta para gestores de risco. Fraudes, por exemplo, em uma cadeia de lojas que estejam uniformemente distribuídas entre todas as filiais pode ser um indicativo para auditoria. Segundo Pareto, a maioria das fraudes deveria estar concentrada num pequeno número de filiais, se as causas da fraude forem “independentes”. Em uma empresa de software, Pareto prevê que um pequeno número de programadores provoca a maioria dos erros de programação. A distribuição uniforme de erros entre todos os programadores é uma anormalidade a ser investigada.

Concentração de Benford

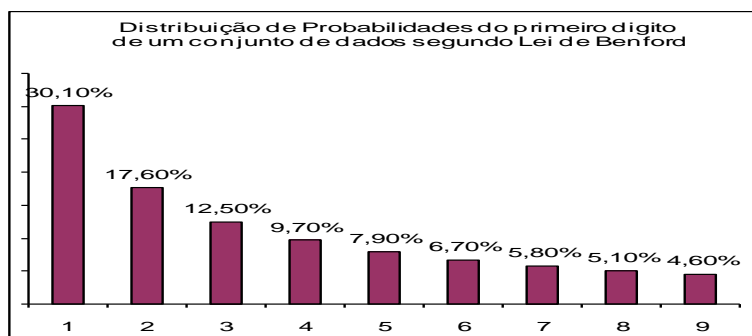
Outro exemplo de concentração no mundo organizacional é a Lei de Benford, também é conhecida por “Lei do Primeiro Dígito”. Ela permite identificar possíveis erros, potenciais fraudes e outras irregularidades em conjuntos de dados numéricos “naturais”. Trata-se de uma técnica analítica, inicialmente proposta por Newcomb, em 1881, e publicada sob um título: “*Note on the frequency of use of the different digits in natural numbers*”. Posteriormente, a técnica foi aperfeiçoada por um físico dos laboratórios da General Electric, Frank Benford (1938). Sua comprovação teórica só teria ocorrido recentemente (HILL, 1995). A lei de Benford tem sido crescentemente usada para “detecção científica de fraudes”.

A lei de Benford trata da frequência de ocorrência do primeiro dígito, do dígito mais à esquerda, dos números que compõem um conjunto de dados. O senso comum poderia inferir que existe uma distribuição uniforme entre os dados de um conjunto que são iniciados pelos algarismos de “1” a “9”. Cada um destes 9 dígitos deveria ser o primeiro dígito em cerca de 11% dos dados. Entretanto, Benford demonstrou que os dados iniciados pelo dígito “1” representam cerca de 30% dos dados totais, e não apenas 11% como se poderia supor; os iniciados pelo dígito “9”, são apenas 5%. Os demais

dígitos aparecem nos percentuais mostrados na figura 2, gerada a partir da fórmula de Benford:

$$P(d) = \log_{10}\left(1 + \frac{1}{d}\right), \text{ onde "d" é o dígito de 1 a 9 e } P(d) \text{ a probabilidade de ocorrência do}$$

Figura 2- Distribuição de probabilidades da Lei de Benford



Fonte: Elaborado pelo autor

Entretanto, se os dados forem manipulados por qualquer motivo, como em uma fraude, eles deixam de seguir a distribuição de probabilidades de Benford. A comparação da frequência observada com a frequência esperada pela Lei de Benford permite identificar conjuntos de dados anômalos em trabalhos de auditoria e de análise forense nas mais diversas situações como dados com valores de dívidas de clientes de um banco, saldo devedor dos clientes de um crediário, despesas na prestação de contas em projetos, custo do reparo de veículos segurados e resultado de eleições, como abordado por Battersby, em 2009, sobre uma eleição no Irã. Existem trabalhos sobre aplicação da lei de Benford na prestação de contas de entidades do terceiro setor aos órgãos públicos, na fiscalização do imposto sobre serviços (LAGIOIA, et al., 2011), entre outras aplicações.

Em resumo, a Lei de Pareto e a de Benford exemplificam que: a) concentrações não intuitivas ocorrem no mundo corporativo; b) que estas concentrações podem ser modeladas; c) que estes modelos explicam o fenômeno e permitem fazer previsões probabilísticas. As modelagens de Pareto e Benford ajudam a ilustrar o objetivo e proposição do presente trabalho, a modelagem da concentração, na linha do tempo, de eventos raros de alto impacto nas corporações.

2- Questão de Pesquisa e Proposição:

Este contexto e as referências levam a seguinte questão de pesquisa:

Como prever, probabilisticamente, quando ocorrerá a reincidência de um evento raro em uma organização?

O trabalho defende a seguinte proposição:

É possível modelar a concentração temporal de eventos raros de uma organização pela distribuição de probabilidades de Poisson.

Mais precisamente, é possível modelar a concentração, na linha de tempo, de eventos raros de uma organização, que seguem a distribuição de Poisson, através da distribuição Exponencial dos tempos entre dois eventos consecutivos.

3- Objetivo

Subsidiar a tomada de decisões e a gestão de risco nas organizações através da modelagem de eventos raros de alto impacto, pela distribuição de probabilidades de Poisson. Este modelo tem por objetivo permitir afirmações probabilísticas sobre o momento da reincidência de eventos imponderáveis, de rara ocorrência (“outliers”) fora dos padrões Gaussianos (fora da distribuição Normal).

4- Alguns conceitos e referências sobre eventos raros nas organizações

A Teoria da Normalidade dos Acidentes (PERROW, 1999) explora o imponderável em sistemas complexos, onde ocorrem interações múltiplas, imprevisíveis e de alto impacto (GANDRA, 2004). O evento raro é um acontecimento improvável, mas não impossível, como o Cisne Negro de Taleb (2007). Posteriormente surgem explicações que fazem um “Cisne Negro” parecer menos aleatório e mais previsível, pois as pessoas seriam “*demasiado vulneráveis ao impulso de simplificar, narrar e classificar*” (TALEB, 2007). Eventos raros podem ser conceituados como interrupções inesperadas e não programadas, de baixa frequência, que geram discontinuidades nos processos organizacionais (LAMPEL et al., 2009).

A dificuldade do estudo de eventos raros (ER) reside, em parte, no fato das empresas não terem muitos motivos para aprender acerca do evento, já que o consideram como algo excepcional cuja possibilidade de recorrência é considerada remota. Existe a tendência de tratar o evento raro como uma experiência única e, portanto, sem utilidade para inferências e estimativas de probabilidades. Ao contrário, os ER podem ser um poderoso estímulo na melhoria contínua das atividades organizacionais revelando fragilidades e melhorando sua capacidade de resposta (CHRISTIANSON et al. 2009).

A avaliação do impacto do ER tem uma forte componente intersubjetiva, porque a opinião dos atores podem atribuir diferentes significados ao ER e às suas consequências, o que influi no processo de tomada de decisão. Além disso, como decorrência da dinâmica organizacional, a percepção do impacto pode variar ao longo do tempo. Se a organização superar a contento as consequências adversas do ER as pessoas terão uma percepção minimizada do risco e do próprio tamanho do impacto (VAREY; KANHEMAN, 1990), comprometendo a mobilização da organização para tomar medidas preventivas capazes de minimizar as consequências da reincidência, que de outra forma poderia deflagrar uma catástrofe organizacional.

Muitos eventos raros decorrem de fatores imponderáveis. E o imponderável é uma força importante na vida das empresas. *“Geralmente subestimamos os efeitos da aleatoriedade”*. *“A mente humana está preparada para identificar uma causa definida para cada acontecimento, podendo assim ter bastante dificuldade em aceitar a influência de fatores aleatórios ou não relacionados”* (MLODINOW, 2008).

Esse trabalho procura contribuir para o campo de estudos organizacionais no que se refere a mudanças drásticas e repentinas, não programadas e com impactos significativos, em períodos de tempo relativamente curto, explicados pela teoria do equilíbrio pontuado de Gersick (1995): a evolução incremental de uma organização é “pontuada” por rupturas. As mudanças ocorrem de forma incremental nos longos períodos de estabilidade, pontuadas por convulsões revolucionárias em alguns poucos períodos, curtos, de mudanças não lineares, de rupturas (GERSICK, 1995). A diferença entre mudanças incrementais de períodos em equilíbrio e mudanças revolucionárias é como mudar a goleira de um jogo de futebol. Mudança incremental é diminuir a largura da goleira; mudança revolucionária é retirar a goleira. A primeira mantém as regras e

estrutura do jogo intactas, enquanto que a segunda desmantela o jogo, inviabilizando as regras atuais. Não é possível mudar para o futebol sem goleira, segundo etapas incrementais. São necessárias novas regras. (GERSICK, 1991).

Este trabalho considera como evento raro as interrupções não programadas no processo evolucionário incremental de uma organização. Essas interrupções podem dar origem a grandes revoluções nas quais a empresa muda de estado, que pode ser para melhor ou para pior. O “estado 2” é sucedido pelo “estado 3”, mas ele não é necessariamente mais “avançado” do que o anterior. Estas mudanças de estado imprevisíveis, caóticas e sem uma relação linear com causas podem ser tratadas pela teoria do caos. Não há situações estáveis ou permanência que nos interessem (na teoria do caos), mas sim revoluções, crises e instabilidades (PRIGOGINE; STENGERS, 1984, 2007). Os autores criticam o paradigma da otimização, pois ele somente se aplica em situações de equilíbrio e de evolução incremental. Esses modelos ignoram a possibilidade de transformações radicais, que mudam a definição do problema e portanto o tipo de solução mais adequada. O foco do presente estudo são os ER associados a essas rupturas, as revoluções em curto espaço de tempo a que as organizações estão sujeitas, especialmente os empreendimentos inovadores.

O ER deve ser encarado não como uma fatalidade excepcional, mas sim como parte integrante do processo da evolução organizacional, onde ele não é único e, provavelmente, ocorrerá novamente. A natureza aleatória da crise favorece desculpas para não dedicar estudos a ela (ROUX-DUFORT, 2007). Neste trabalho, a causa do evento raro que disparou um processo revolucionário não é vista como um fato isolado, ela faz parte de uma complexa rede de relações causais que é irreprodutível e muitas vezes difícil de ser analisada, dada a complexidade dos sistemas organizacionais.

Empresas de seguro possuem estatísticas sobre a reincidência de eventos frequentes que seguem a distribuição Normal, de Gauss: “temos um acidente a cada 18 minutos em média”. Esta é uma informação gerencial essencial para capacitar a seguradora que deve tratar aquela quantidade de acidentes, independentemente de suas causas. Portanto, “não é necessário saber as causas de um evento frequente para modelar sua recorrência no tempo”. De forma análoga, na definição básica dos princípios de aplicação da modelagem por Poisson existe uma separação entre causa e efeito que pode ser enunciada de maneira análoga: não é necessário saber as causas de

um evento raro para modelar sua recorrência no tempo. Este trabalho se concentra na reincidência e não nas causas da reincidência do ER, não obstante a importância das causas para prevenção.

Estratégias de Ação

Segundo Miles e Snow (1978) existem quatro tipos básicos de estratégias na solução de problemas: prospectiva, analisadora, defensiva e reativa. A tabela abaixo mapeia estas estratégias segundo duas categorias de análise: frequência e impacto dos eventos na organização.

Quadro 1: Estratégias de Ação em função da Frequência *versus* Impacto de problemas.

ESTRATÉGIAS	Eventos FREQUENTES	Eventos RAROS
BAIXO IMPACTO	<p>Quadrante 1- Estratégias de MELHORIA INCREMENTAL, para redução das causas dos problemas</p> <p>Os eventos são de baixo impacto, mas não podem ser ignorados, porque são freqüentes. As estratégias devem focar nas mcausas dos problemas visando sua gradativa eliminação, segundo estratégias de melhoria contínua.</p>	<p>Quadrante 2- Estratégias de MANUTENÇÃO.</p> <p>Os eventos são de baixo impacto e raramente ocorrem. Portanto, tendem a ser irrelevantes. A estratégia é a manutenção do <i>status quo</i>.</p>
ALTO IMPACTO	<p>Quadrante 3 – Estratégias ADAPTATIVAS</p> <p>A empresa sujeita a eventos freqüentes de alto impacto, de alguma forma, já desenvolveu estratégias de sobrevivência (WINTER, 2005). As estratégias devem adaptar a empresa para enfrentar hostilidades do ambiente, como condição de sobrevivência e evolução.</p>	<p>Quadrante 4- Estratégias: Planos de CONTINGÊNCIA e GESTÃO DE RISCO</p> <p>A empresa deve se precaver através de estratégias de Cenários e Diagramas de Contingência para minimizar os efeitos dos eventos raros, imprevisíveis e de alto impacto. A probabilidade de reincidência do evento raro contribui na mobilização de “estados de prontidão”.</p>

Fonte: Organizado pelo autor, adaptado de Vanca (1998)

Este trabalho foca na modelagem de eventos do quadrante 4 (eventos raros de alto impacto). Estes eventos, por serem de alto impacto, podem ser determinantes para

competitividade e mesmo para a sobrevivência das empresas. A modelagem aqui proposta permite que as empresas façam inferências probabilísticas sobre a reincidência dos eventos raros para subsidiar a tomada de decisão e a gestão de risco. O grande objetivo é propor ferramentas estatísticas, pouco usuais na administração, para diminuir o grau de incerteza e avançar no uso de ferramentas para tomada de decisões técnicas e racionais.

5- A modelagem da concentração de Eventos Raros por Poisson:

A administração deve muito da sua evolução, como ciência aplicada, ao desenvolvimento de ferramentas estatísticas, devidamente adaptadas à natureza dos fatos da gestão organizacional. A comprovação empírica ganha força e consistência com um tratamento estatístico adequado, permitindo melhor entender e fazer previsões probabilísticas. Entretanto, a literatura acadêmica gravita muito em torno de fenômenos explicados pela distribuição Gaussiana. Pouco existe sobre o estudo estatístico de “*outliers*”, de situações longe da média, acima de dois desvios padrões.

Em termos estatísticos, evento raro é aquele que representa um pequeno percentual do total de eventos produzidos por uma ou mais fontes, num determinado período de tempo. No caso de acidentes aéreos, o evento frequente é a quantidade voos (“*n*”- número de voos no período). O evento raro é o voo que resulta num acidente. A observação deste fenômeno ao longo dos anos leva a uma média de, por exemplo, 10 acidentes graves por ano ($\lambda=10$). Outro exemplo de ER é o número de reclamações que chegam por dia a um Call Center (grande número de reclamações “*n*”), e destas, apenas algumas são graves. A observação destas reclamações pode levar a determinação de uma média (λ) de reclamações graves/dia.

5.1 Distribuição de Poisson

A distribuição de Poisson permite calcular a probabilidade da ocorrência de eventos aleatórios, imprevisíveis, oriundos de fontes independentes, múltiplas, internas ou externas à organização. A literatura sobre aplicação de Poisson na Administração de Empresas é deserta face a quantidade da literatura que trata da aplicação da distribuição de Gauss (Normal), por exemplo. A modelagem matemática por Poisson pode ser aplicada a vários fenômenos do mundo corporativo que satisfaçam as seguintes condições:

- as variáveis devem ser discretas, aleatórias e binomiais (só há dois resultados possíveis: “sucesso” ou “fracasso”, no sentido estatístico dos termos);
- as fontes geradoras dos eventos devem ser **independentes e incomunicáveis** entre si;
- (n): a quantidade de eventos produzidos deve ser grande;
- (p): a probabilidade de “sucesso” deve ser pequena e constante a cada experimento
- (n*p): a quantidade de eventos raros deve ser pequena. É possível o emprego da distribuição de Poisson quando $n \geq 100$ e $(n*p) < 10$ (FREUND; SIMON, 1995). Ou seja, segundo esses autores, o número de eventos por unidade de tempo para ser modelado por Poisson deve ser menor do que 10 ocorrências na unidade de tempo estabelecida (por hora, dia, semana, ano etc).

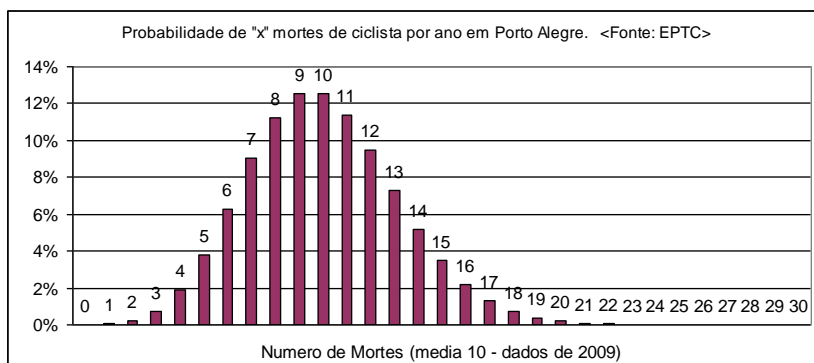
Fórmula da Função Densidade de Probabilidade de Poisson, onde:

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

- P(X) : probabilidade de ocorrer x “sucessos” no intervalo de tempo especificado
- x: número de “sucessos”
- e: a base do logaritmo natural (2,71828)
- λ : média = número de eventos / unidade de tempo- é o número médio de “sucessos” no intervalo de tempo especificado. Esta média resulta da observação de períodos anteriores.

Um indicativo para uso da modelagem por Poisson (e não a distribuição binomial, por exemplo) é quando se tem uma amostra com variância σ muito grande, próxima da média λ .

Figura 3 – Distribuição de probabilidades de Poisson para acidentes com ciclistas (EPTC, 2009)



Fonte: Elaborado pelo Autor a partir de dados da EPTC (2009)

A função densidade de probabilidade da figura 3 mostra a probabilidade de ocorrência de “x” mortes, variando de 0 a 30, para uma média $\lambda=10$ acidentes fatais/ano. Nove e 10 acidentes fatais por ano possuem as maiores probabilidades da curva (um pouco acima de 12%); a probabilidade de 6 fatalidades é a metade (por volta de 6%); dificilmente haverá 20 fatalidades no ano (a probabilidade é quase 0%), etc.

Quadro 2- Comparativo entre as distribuições de POISSON e EXPONENCIAL

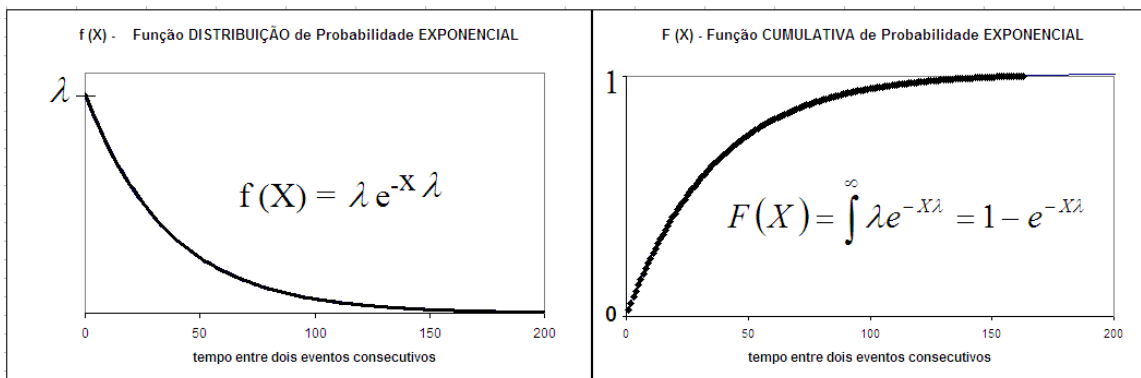
<p>POISSON- Variável Discreta</p> <p>Analisa número de eventos ocorridos num dado intervalo de tempo</p>	<p>EXPONENCIAL- Variável Contínua</p> <p>Analisa os tempos entre dois eventos consecutivos</p>
<p>Média λ. Exemplo: 10 eventos por mês</p>	<p>Média $1/\lambda$. Exemplo: 3 dias entre dois eventos consecutivos em média</p>
<p>Fórmula Matemática</p> $f(x) = \lambda^x \frac{e^{-\lambda}}{x!}$	<p>Formula Matemática</p> $f(X) = \lambda e^{-X\lambda}$
<p>Em ambas as distribuições, a Média λ é igual a Variância σ</p>	
<p>Questão típica: Qual a probabilidade de ocorrência de “x” eventos num certo intervalo de tempo?</p> <p>Ex: Se a média $\lambda=10$ eventos / mês, qual a probabilidade do evento raro ocorrer 6 vezes num mês qualquer?</p>	<p>Questão típica: Qual a probabilidade da reincidência de um evento raro “x” dias depois?</p> <p>Ex: Se o tempo médio entre duas ocorrências sucessivas é de 3 dias, qual é a probabilidade de reincidência 12 dias após o evento anterior?</p>

Fonte: Compilado pelo autor a partir de Siegel (1975)

O eixo X da figura 4 mostra o intervalo de tempo entre dois eventos consecutivos. A distribuição exponencial modela um **processo estocástico**, no sentido

genérico de Kac (1976) e Nelson (1985), ou seja, trata de probabilidades de uma variável aleatória ao longo do tempo. No caso da exponencial, o tempo considerado é o “intervalo de tempo” transcorrido entre a chegada de dois eventos raros consecutivos.

Figura 4- Função distribuição de Probabilidade Exponencial e a correspondente Função Cumulativa de Probabilidade



Os intervalos de tempo entre a chegada de dois eventos raros consecutivos de média λ (número de eventos / intervalo de tempo) formam uma distribuição EXPONENCIAL de média $1/\lambda$ (intervalo médio de tempo entre a chegada de dois eventos raros consecutivos). As duas distribuições Poisson e Exponencial estão relacionadas pela seguinte regra: o conjunto de tempos medidos entre duas ocorrências consecutivas de um fenômeno modelado por Poisson é segue uma distribuição Exponencial.

- Função distribuição de probabilidade Exponencial:
- Média: $\bar{X} = 1/\lambda$
- Variância: $S^2 = 1/\lambda$ (a variância é igual à média)

5.2- Teste não paramétrico de Kolmogorov–Smirnov (KS)

Um teste paramétrico é um teste “exato”, que produz um valor característico da distribuição de probabilidades em análise como, por exemplo, a média e o desvio padrão. Entretanto, os testes paramétricos necessitam saber, a priori, vários dados sobre a distribuição que está sendo analisada (se ela é Normal, Logarítmica, Exponencial, etc.). Só então é possível saber qual teste paramétrico pode ser aplicado àquele tipo de

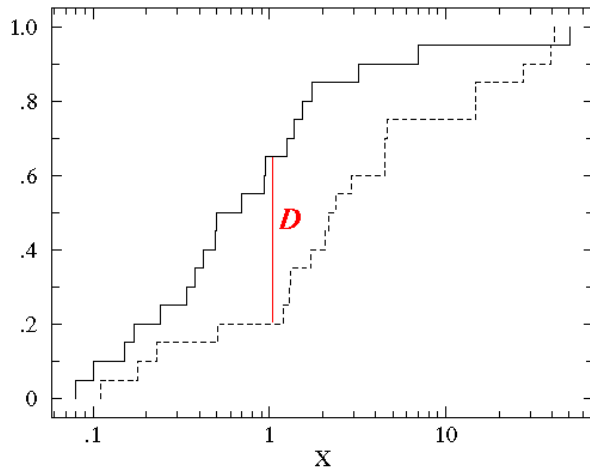
distribuição. Mesmo assim, há outras restrições. Por exemplo, alguns testes paramétricos falham quando os dados estão na cauda da curva Normal, longe da média.

O Coeficiente de Correlação R pode ser usado para determinar a relação existente entre “recursos” estratégicos e “desempenho econômico” da empresa, por exemplo. Entretanto, o coeficiente R só pode ser aplicado a fenômenos que seguem a distribuição Normal (distribuição de Gauss). O mesmo acontece com o coeficiente de determinação R^2 que mostra o quanto a variação de X (causa) explica a variação de Y (efeito).

Ao contrário, o teste não paramétrico independe do tipo da função distribuição de probabilidades, mas tem pelo menos uma limitação: só é possível realizar testes “comparativos”. Por exemplo, avaliar se a população sujeita a um tipo de medicamento (amostra teste) tem, estatisticamente, os mesmos resultados de outro grupo sujeito a outro tipo de medicamento (amostra de controle). Ou ainda, se uma determinada amostra segue ou não uma determinada distribuição teórica de probabilidades (gaussiana, exponencial, etc.). O teste não paramétrico, também denominado “*goodness of fit*” é um teste de aderência que mede a “distância” entre dados empíricos e uma referência. Esta referência pode ser uma amostra de controle ou um modelo teórico.

O teste não paramétrico Kolmogorov–Smirnov (KS) proposto permite avaliar se duas amostras de variáveis contínuas possuem ou não a mesma distribuição de probabilidades. Ele mede a “distância” entre uma distribuição de probabilidades empírica (amostra teste) e uma distribuição de probabilidades de referência (amostra de controle ou uma distribuição teórica). A **hipótese nula (H_0)** é que ambas as amostras possuam a mesma distribuição. Ou seja, os gráficos da distribuição empírica e da distribuição de controle estão muito “próximos”. Neste caso, se diz que o fenômeno empírico “segue” o modelo teórico (SIEGEL, 1975).

Figura 5- Teste Kolmogorov-Smirnov para comparação de duas distribuições (empírica x controle)



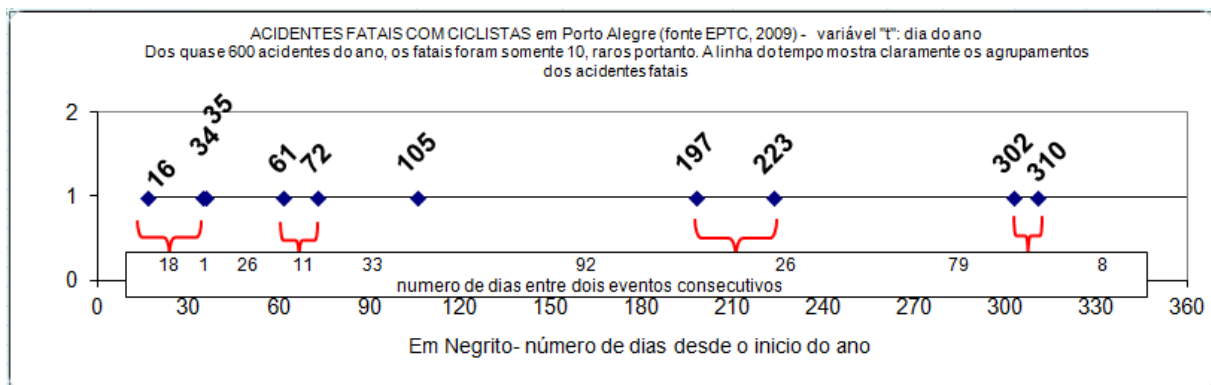
Fonte: (TRIOLA, 2005); www.itl.nist.gov/div898/handbook/eda/section3/eda35g.htm

O teste KS funciona bem mesmo para dados longe da média e para dados não Normais. Ao contrário, o teste “t” pode falhar se os dados não possuírem uma distribuição Normal. Uma “qualidade” importante do teste KS é que ele pode ser aplicado mesmo que o número de observações empíricas seja menor do que o número de dados do modelo teórico, como ocorre em muitos experimentos.

6- Uma aplicação da modelagem de eventos raros na Gestão Pública: mobilidade urbana

Este item tem por objetivo somente exemplificar a aplicação do modelo proposto. Testes empíricos mais apurados do modelo proposto estão além do escopo deste trabalho, neste momento. Assim, somente a título de exemplo, segue a modelagem de acidentes fatais com ciclistas em Porto Alegre. Os acidentes são raros ($n \cdot p$ é pequeno), são aleatórios, foram provocados por fontes independentes e, importante neste trabalho, estão fidedignamente documentados e disponíveis para consulta pública (EPTC, 2009).

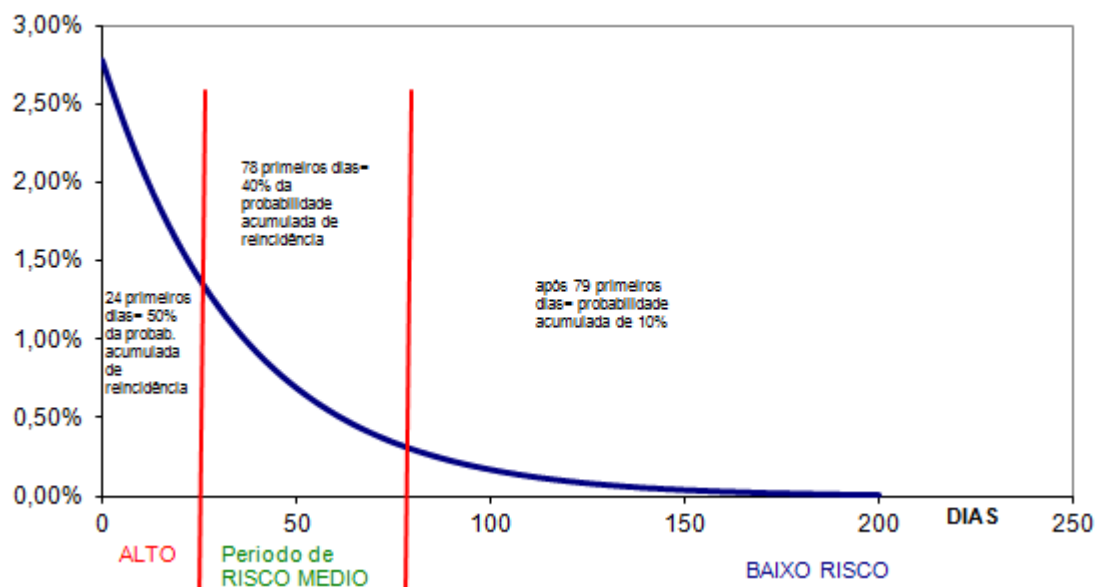
Figura 6 - Linha do tempo de acidentes fatais com ciclistas. Fonte Município de Porto Alegre



Fonte: Elaborado pelo autor a partir de dados da EPTC (2009)

A linha do tempo mostra que o primeiro acidente fatal ocorreu no 16º dia do ano. O seguinte ocorreu no 34º dia e o seguinte, apenas um dia após (no 35º dia do ano). A última fatalidade do ano ocorreu no 310º dia. A caixa do gráfico mostra o número de dias entre dois eventos consecutivos: {18, 1, 26, 11, 33, 92, 26, 79 e 8}. O gráfico mostra os agrupamentos dos eventos raros na linha do tempo, separados por longos períodos de “silêncio”. De 10 ocorrências, apenas uma delas (105º dia do ano) aconteceu “isoladamente”. Ela foi antecedida por um intervalo de 33 dias (maior do que os críticos 24 primeiros dias que acumulam 50% de toda a probabilidade da exponencial de reincidência- vide figura 7) e sucedido por um longo “silêncio” de 92 dias.

Figura 7- Modelagem da reincidência de um evento raro, com divisão da curva de risco em três áreas.



Fonte: Elaborado pelo autor, a partir de dados da EPTC (2009)

O gráfico síntese da figura 7 auxilia nas inferências preditivas sobre o fenômeno modelado. A figura 7 possui marcações de tempos associados à probabilidade de reincidência do evento raro do exemplo (EPTC, 2009). O gráfico foi propositalmente simplificado para ser uma ferramenta expedita, fácil de produzir, entender e aplicar. A curva está dividida em três partes, de acordo com o nível de risco associado a cada uma, através de dois cortes, como normalmente se faz na curva ABC associada à concentração de Pareto:

Área de Alto risco- O corte em x1 determina uma área, à sua esquerda (em vermelho), que representa 50% da área da função densidade de probabilidade acumulada. No exemplo modelado, isto ocorre no 24º dia. Ou seja, metade dos acidentes fatais do ano reincide, probabilisticamente, nos primeiros 24 dias depois do acidente imediatamente anterior.

Área de Médio risco- O corte em x2 (78 dias) delimita uma área acumulada de 90%, desde a origem (e de 40% desde x1). Ou seja, 90% dos acidentes fatais do ano reincidem, probabilisticamente, 78 dias após o acidente imediatamente anterior.

Área de Baixo risco- A Cauda Azul- A partir do 79º dia, a probabilidade acumulada total é de apenas 10%. A cauda azul é a zona “segura”, onde a probabilidade de reincidência é relativamente pequena e diminui exponencialmente a cada dia. A curva diz que a probabilidade de ocorrer um evento é muito pequena, mas não é zero, já que ela é assintótica ao eixo X. Ou seja, o acidente pode voltar a ocorrer mesmo depois de um longo período de silêncio. Entretanto, caso ocorra um novo acidente, a contagem dos tempos volta à origem do eixo X, onde a ordenada da probabilidade de reincidência é a maior de toda a curva exponencial (para $x=0$ a probabilidade de ocorrência é λ).

A figura 7 contribui para confirmar, graficamente, a hipótese básica deste trabalho: eventos raros quando ocorrem tendem a se repetir, no curto prazo, formando agrupamentos característicos na linha do tempo. Um gráfico como o da figura 7 pode modelar um fenômeno raro (“*outlier*”) de uma empresa, permitindo fazer inferências preditivas sobre o risco de reincidência deste fenômeno ao longo do tempo.

7- Comentários e Conclusões

Este trabalho procura trazer uma nova visão sobre os eventos aleatórios de rara ocorrência nas organizações. O princípio “20%-80%” de Pareto e a lei “*of anomalous numbers*” de Benford tratam de concentrações, não intuitivas, no mundo corporativo e são largamente usadas pelas organizações. A concentração de Pareto permite estabelecer critérios técnicos para planejamento estratégico. A Lei de Benford, usada para auditoria de empresas de seguro, por exemplo, permite responder se os dados formam um conjunto “natural” ou se houve alguma espécie de manipulação, como uma fraude. Nesta mesma linha, o presente trabalho propõe um modelo para a concentração, na linha do tempo, de eventos raros nas organizações. Permite responder qual a probabilidade de um evento raro reincidir “n” dias depois da última ocorrência.

Um evento raro não é um evento único. A reincidência de um evento raro tende a ser negligenciada, mesmo que tenha provocado prejuízos importantes, porque o senso comum leva, erroneamente, a tratá-lo como único, e que ele não deve ocorrer novamente. A curva da probabilidade da reincidência de um evento raro é uma exponencial decadente (*decay function*). Ela mostra como a probabilidade de reincidência decai com o passar do tempo. A modelagem aqui proposta prevê, probabilisticamente, quando um evento raro vai ocorrer novamente. Isto permite

mobilizar recursos destinados a minimizar consequências adversas, subsidiando o processo de tomada de decisão e a gestão de risco.

Em resumo, este trabalho propõe um modelo probabilístico sobre a reincidência de eventos raros nas organizações, ancorado em bases teóricas consagradas da estatística, a distribuição de Poisson e a distribuição Exponencial.

A inquietante reincidência de eventos raros pode ser modelada de forma simples, desmistificada e universal, conferindo a este trabalho um amplo campo de aplicação no mundo corporativo.

BIBLIOGRAFIA

ASSUNÇÃO, R. M.; Barreto, S. M.; Guerra, H. L.; Sakurai, E.; **Mapas de taxas epidemiológicas: uma abordagem Bayesiana**. Caderno de Saúde Pública, v. 14, n. 4, p.713-723, 1998.

BENFORD, F. ; **The law of anomalous numbers**. Proceedings of the American Philosophical Society 78 (4): 551–572, 1938.

[http://links.jstor.org/sici?sici=0003-](http://links.jstor.org/sici?sici=0003-049X(19380331)78%3A4%3C551%3ATLOAN%3E2.0.CO%3B2-G)

[049X\(19380331\)78%3A4%3C551%3ATLOAN%3E2.0.CO%3B2-G.](http://links.jstor.org/sici?sici=0003-049X(19380331)78%3A4%3C551%3ATLOAN%3E2.0.CO%3B2-G)

<http://www.fooledbyrandomness.com/>

CARVALHO, J. M.C.; **Logística**. 3ª ed. Lisboa: Edições Silabo, 2002. ISBN 978-972-618-279-5, 2002

CONNIE J.; GERSICK G. **Revolutionary change theories: a multilevel exploration of the punctuated equilibrium paradigm**. University of California, Los Angeles. 1995.

CHRISTIANSON, M. K; FARKAS, M.T.; SUTCLIFFE, K. M.; WEICK, K. E. Learning Through Rare Events: Significant Interruptions at the Baltimore & Ohio Railroad Museum. Organization Science, vol. 20, n. 05, p. 846-860, Sept-oct, 2009.

DESARBO, W. S.; BENEDETTO, C. A. di; SONG, M.; SINHA, I. Strategic Management Journal, v. 26, n. 1, p. 47-74, 2005

EPTC, Empresa Pública de Transportes e Circulação. Município de Porto Alegre. 2009.

FREUND, J. E.; SIMONS, G.; **Statistics: a first course**. Prentice-Hall, 1995.

GANDRA. VII SEMEAD- Seminários em administração- FEA-USP; **Acidentes do trabalho: Evoluindo do modelo de causalidade centrada no indivíduo para o modelo de cultura organizacional**; 10 e 11 agosto 2004 p.8

HANSEN, M. , PERRY, L. T. , REESE, C. A Bayesian Operationalization of the

Resource-based view. *Strategic Management Journal*. *Strat. Mgmt. J.*, 25: 1279–1295. Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/smj.432, 2004.

HILL, T.; **"The first digit phenomenon"** (PDF). *American Scientist* 86: 358, 1995
<http://www.tphill.net/publications/BENFORD%20PAPERS/TheFirstDigitPhenomenonAmericanScientist1996.pdf>.

JURAN. **Quality Control Handbook**, New York, New York: McGraw-Hill, 1951, OCLC 1220529

LAGIOIA, U. et al. **Revista de Contabilidade de Finanças**. USP, São Paulo, v. 22, n. 56, p. 203-224, maio/jun./jul./ago. 2011

LAMPEL, J; SHAMSIE, J.; SHAPIRA, Z. Experiencing the Improbable: Rare Events and Organizational Learning. *Organization Science*, vol. 20, n. 05, p. 835-845, Sept.-oct., 2009.

MLODINOW, L.; **O Andar do Bêbado - Como o Acaso Determina Nossas Vidas**. JORGE ZAHAR, 2008

MORETTIN,; **Estatística Básica - Probabilidade**. 7ª. Ed., Makron Books, São Paulo, SP. 1999.

PERROW, Charles. *Normal Accidents: Living with high-risk technologies*. New York: Basic Books Inc., Publishers. 1999.

PERRY. **Diversification and Focus: A Bayesian Application of the Resource-Based View**. *Schmalenbach Business Review*, vol57, pp 304-319, 2005.

PRADO, L. **Distribuições de Abundâncias de Espécies: Avanços analíticos para entender um padrão básico em ecologia**. *Revista Ciência e Ambiente*. Depto de Ecologia, Instituto de Biociencias, USP- prado@ib.usp.br , 2010.

PRIGOGINE, Ilya. **O fim das certezas: tempo, caos e as leis da natureza**. São Paulo: UNESP, 1996.

PRIGOGINE, I. R., STENGERS, I., FARIA, M., TRINCHEIRA, M. J. M., & MENDES, J. P.. **A nova aliança: metamorfose da ciência**, 1984.

ROUX-DUFORT, C. Is Crisis Management (only) a Management of Exceptions? *Journal of Contingencies and Crisis Management*, Vol. 15, n. 2, jun., 2007.

SIEGEL, S. **Estatística não-paramétrica para as ciências do comportamento**. São Paulo: Makron Books: McGraw-Hill, 1975.

TALEB. X. **The Black Swan: The Impact of the Highly Improbable**: Amazon (US), Barnes and Noble (US). Book Reviews. 2007, 2nd Ed. 2010

TRIOLA, M. **Introdução à estatística**. 9ª edição - 2005 682 pgs. ISBN 85-2161431-4 , 2005

VANCA, P.M. **A Importância do Gerenciamento de Riscos de Negócios**. Revista BQI – Brazilian Quality Index. 4ª ed. S^o Paulo: Quinta Essência, 1998.

VAREY, C.; KAHNEMAN, D. **Experiences Extended across Time: Evaluation of Moments and Episodes**. Journal of Econometrics, L (1990), 69-89