Trabalho de Conclusão de Curso - Bacharelado em Matemática

Distribuição de Probabilidade do Prêmio de Risco de Seguros

Willian Peterson Lopes

Universidade Federal do ABC

Título: Distribuição de Probabilidade do Prêmio de Risco de Seguros

Autor: Willian Peterson Lopes
Orientador: Prof. Dr. Alexei Magalhães Veneziani

Trabalho de conclusão de curso apresentado como requisito parcial para obtenção do título de Bacharel em Matemática pela Universidade Federal do ABC.

Banca Examinadora:

Prof. Dr. Rafael de Mattos Grisi
Universidade Federal do ABC

Prof. Dr. Erika Alejandra Rada Mora
Universidade Federal do ABC

Santo André, 7 de agosto de 2018.

Aos meus pais, Neide e Luiz, e à minha esposa Juliana pelo amor incondicional.
Dedico.

"Uma mente que se abre a uma nova ideia jamais voltará ao seu tamanho original".
Albert Einstein.

Agradecimentos

Aos meus amados pais, Neide e Luiz, à minha melhor amiga e esposa Juliana e aos meus fiéis irmãos, Alan, Giulian e Joyce, pelas alegrias compartilhadas, por todo carinho, amor, incentivo, confiança e exemplo.

Ao meu orientador Alexei Magalhães Veneziani, pelas incansáveis noites em que trabalhamos juntos, por todos os ensinamentos, pela atenção e paciência.

Aos meus amigos da UFABC. Caio, pelas longas conversas filosóficas e pelo apoio mútuo durante boa parte do bacharelado. André, pela grande ajuda no final do curso e pelo grande exemplo de matemático que é. Wellington, pela enorme parceria dentro e fora da universidade e por todo o incentivo que me deu para finalizar este trabalho.

Aos meus amigos da Porto Seguro, por todos as horas compartilhadas todas as semanas, pelas grandes lições que me ensinaram e por toda a compreensão em relação aos meus dias sobrecarregados de trabalho.

Ao meu cachorro Panqueca, por sempre respeitar os momentos em que estive escrevendo este trabalho e me oferecer a mais fiel e sincera companhia.

A todos que, de alguma forma, fizeram parte desta árdua e compensatória trajetória que foi a graduação.

Resumo

Este projeto de pesquisa tem como objetivo encontrar uma distribuição de probabilidade que descreva os sinistros ocorridos em determinada carteira de seguros. Para isso, assumiu-se que a quantidade de sinistros ocorridos segue uma distribuição de Poisson, enquanto os valores destes sinistros seguem uma distribuição Gama, de tal forma que o montante de sinistros possa ser descrito por uma distribuição composta de Gama e Poisson. Adicionalmente, demonstrou-se que esta distribuição composta é equivalente à distribuição Tweedie que, por ser uma distribuição pertencente à família exponencial, pode ser utilizada na abordagem de modelos lineares generalizados para se obter a estimativa da variável aleatória do montante de sinistros ocorridos. Com essa estimativa e dada a exposição ao risco conhecida, pode ser calculado o prêmio de risco.

Palavras Chaves: Distribuição de Probabilidade, Modelo Linear Generalizado, Família Exponencial, Gama composta com Poisson, Distribuição Tweedie, Prêmio de Risco, Seguro

Abstract

This research project aims to find a probability distribution that describes the claims occurred in a particular insurance portfolio. For this, it was assumed that the amount of claims occurred follows a Poisson distribution, while the values of these claims follow a Gamma distribution, such that the amount of claims can be described by a composite distribution of Gamma and Poisson. In addition, it has been demonstrated that this composite distribution is equivalent to the Tweedie distribution, which, because it is a distribution belonging to the exponential family, can be used in the approach of generalized linear models to obtain the estimate of the random variable of the amount of claims occurred. With this estimate and given the known exposure to the risk, the risk premium can be calculated.

Keywords: Probability Distribution, Generalized Linear Model, Exponential Family, Composite Gamma with Poisson, Tweedie Distribution, Risk Premium,Insurance

Sumário

1 Introdução

Este projeto de pesquisa pretende encontrar uma distribuição de probabilidade adequada para ser utilizada na modelagem do montante de sinistros ocorridos de uma carteira de seguros, a fim de se estimar o prêmio de risco por unidade exposta. Para isso, articulou-se alguns conceitos básicos de probabilidade (ROSS, 2010; BILLINGSLEY, 1995) e de teoria de seguros (VILANOVA, 1969; FERREIRA, 2009; CORDEIRO FILHO, 2009). A partir da escolha da distribuição de Poisson para descrever a quantidade de sinistros e da distribuição Gama para descrever os valores assumidos (custo de cada sinistro), buscou-se construir uma variável aleatória que descrevesse o montante de sinistros ocorridos e mostrou-se que a distribuição de probabilidade desta variável aleatória é equivalente à distribuição de probabilidade Tweedie, conforme conceitos articulados por SMYTH (1996).

O seguro tem como um de seus princípios básicos o mutualismo. Características da mutualidade são observadas desde o início das civilizações. De acordo com NASCIMENTO, a história dos seguros no mundo remete a um período de mais de dois mil anos antes de Cristo, onde verificam-se registros de atividade similar ao seguro voltada para a proteção contra a perda de animais durante o transporte de mercadorias na Mesopotâmia. Além disso, o *foenus nauticus* – empréstimo marítimo a risco – era praticado entre gregos, fenícios e romanos. Estas formas rudimentares de seguro foram primordiais para garantir os valores relacionados às mercadorias que circulavam nesta época, mas as pessoas possuíam inúmeras dúvidas a respeito da credibilidade dos seguradores que assumiam os riscos. Todavia, o conceito de seguro se desenvolveu ao longo do tempo e em torno de 1347, foi escrito o primeiro contrato de seguros em Gênova. Este contrato possuía cláusulas que determinavam quando deveria ou não ocorrer o pagamento de uma indenização por parte do segurador.

No Brasil, a primeira companhia de seguros, denominada Boa-Fé, surgiu em 1808, o mesmo ano em que ocorreu a abertura dos portos brasileiros e, consequentemente, a intensificação das atividades marítimas, que constituíam o ramo de atuação desta companhia. Até o ano de 1850, a atividade seguradora no Brasil era controlada por Portugal, mas, nesse ano, foi promulgado o Código Comercial Brasileiro que regulamentava todos os aspectos do seguro marítimo no país. O Código Comercial Brasileiro juntamente com a instalação de sucursais de seguradoras estrangeiras no país, foram responsáveis pelo desenvolvimento do seguros nacionalmente. No ano de 1966, surge a Superintendência de Seguros Privados (SUSEP) como órgão fiscalizador oficial, de modo a estabelecer o Sistema Nacional de Seguros Privados.

Um dos pilares mais importantes de um contrato de seguro é o preço, tecnicamente conhecido como prêmio. Os prêmios individuais devem ser calculados de modo que, pelo mutualismo, possam cobrir os riscos de uma determinada carteira de seguros, além de ser suficiente para cobrir as despesas e gerar lucro para as seguradoras. Neste trabalho de pesquisa, o interesse se concentra em encontrar uma distribuição de probabilidade que seja adequada para descrever o montante de sinistros ocorridos em um carteira de seguros e, desta forma, que possa ser utilizada em um processo de modelagem para estimar estatisticamente o prêmio de risco em função de variáveis explicativas (características do bem segurado, por exemplo).

Desta forma, este trabalho foi estruturado em quatro capítulos mais o apêndice. No capítulo 2, são apresentadas as definições básicas de seguros como, por exemplo, os tipos de prêmio descritos por FERREIRA (2009) e as características elaboradas por VILANOVA (1969) que tornam um risco segurável. Já no capítulo 3, encontram-se as definições básicas de probabilidade, em sua maioria baseadas em ROSS (2010), que serão essenciais para o correto entendimento dos demais capítulos do texto. No capítulo 4, discutem-se as distribuições de probabilidade adequadas para descrever as componentes do prêmio de risco, ou seja, atribui-se a distribuição de Poisson para descrever a quantidade de sinistros ocorridos, a distribuição Gama para os valores de sinistro e, enfim, é construída uma distribuição composta de Gama com Poisson, assim como feito por JORGENSEN e DE SOUZA (1994) e, posteriormente, articulando os conceitos de função geradora de cumulantes e, utilizando o corolário retirado de BILLINGSLEY (1995) e provado no apêndice, demonstra-se que esta distribuição é equivalente a uma distribuição Tweedie, corroborando com a conclusão de SMYTH e JORGENSEN (2002) de que o prêmio de risco segue tal distribuição. Finalmente, no capítulo 5, são apresentadas as considerações finais e indica-se ao leitor alguns possíveis softwares para a implementação do processo de modelagem a partir das distribuições encontradas.

2 Definições básicas de seguros

Este trabalho tem como objetivo encontrar uma distribuição de probabilidade para o montante de sinistros ocorridos a fim de se obter uma metodologia para estimativa do prêmio de risco por meio do uso de modelos lineares generalizados. Porém, antes é necessário que a definição de seguros e, por conseguinte, a de prêmio de risco sejam entendidas. Segundo CORDEIRO FILHO (2009), existem muitas definições de seguro, mas a que mais se enquadra à realidade do século XXI é a de HEMARD.

> "Seguro é uma operação pela qual uma das partes, o Segurado, obtém a promessa de outra parte, o Segurador, mediante o pagamento de uma remuneração, o Prêmio, em seu favor ou no de terceiro, no caso de se verificar o risco, uma prestação com que o Segurador, fazendo uso de um conjunto de Riscos, indenizar-lhe-á de acordo com as leis Estatísticas."
>
> (HEMARD apud VILANOVA, 1969, página 15)

Por esta definição, pode-se verificar que o risco é um fator primordial para qualquer tipo de seguro. Segundo VILANOVA (1969), risco "é o acontecimento aleatório em sua realização, na época de sua realização ou, ainda, no grau em que se realiza". No entanto, nem todos os riscos são objetos de seguro e, para que isso aconteça, o risco deve ser segurável. Para VILANOVA, o risco segurável possui as seguintes características:

1º) Afetar por igual a todos os componentes do grupo, podendo atingir a alguns, mas não a todos, simultaneamente.

2º) Existir homogeneidade dos componentes do grupo, que deve ser o mais numeroso possível.

3º) Sua realização deve ocasionar uma necessidade econômica.

4º) O benefício do seguro não deve constituir um lucro, mas, tão somente um ressarcimento de prejuízos sofridos.

5º) Possibilitar, estatisticamente, basear-se em experiência passada, para deduzir leis que permitam prever em casos futuros da mesma natureza, iguais situações, desde que persistam as mesmas condições e circunstâncias.

6º) Deve existir independência na realização dos acontecimentos e essa realização deve ocasionar necessidade econômica, jurídica e efetivamente ressarcível.

De maneira geral, quando o risco se realiza, ocorre o que é chamado de sinistro, ou seja, o sinistro é a materialização do risco. Quando o sinistro ocorre, é gerada uma indenização ao segurado a partir do contrato de seguro.

O valor a ser pago por uma pessoa que adquire um seguro é chamado de prêmio. Segundo FERREIRA (2005), existem três tipos de prêmios no processo de precificação de um seguro, sendo eles: prêmio de risco, prêmio puro e prêmio comercial.

O prêmio puro corresponde ao prêmio de risco acrescido de um carregamento de segurança estatístico (que visa suportar as oscilações de risco) e o prêmio comercial, por sua vez, corresponde ao prêmio puro carregado com despesas e margem de lucro das seguradoras.

A precificação começa com o cálculo do prêmio de risco que pode ser obtido através do quociente entre o valor esperado do montante de sinistros ocorridos e o número de unidades expostas ao risco.

$$\text{Prêmio de Risco} = \frac{\text{Valor Esperado do Montante de Sinistros Ocorridos}}{\text{Unidades expostas ao risco}} \tag{2.1}$$

O objetivo deste trabalho de pesquisa é apresentar uma distribuição de probabilidade adequada para descrever o montante de sinistros ocorridos. Desta forma, se torna possível calcular o valor esperado desta variável aleatória e dividí-lo pelas unidades expostas ao risco, a fim de se obter o prêmio de risco.

3 Definições básicas de probabilidade

A fim de que este trabalho de pesquisa seja auto contido, este capítulo tem como objetivo apresentar algumas definições básicas de probabilidade que serão utilizadas no decorrer desta dissertação.

Considere um experimento em que o espaço amostral é dado por S. Agora, assuma um número $P(E)$ para cada evento $E \subset S$ que satisfaça os seguintes axiomas:

AXIOMA 1

$$0 \leq P(E) \leq 1$$

AXIOMA 2

$$P(S) = 1$$

AXIOMA 3 Dada uma sequência de eventos exclusivos E_1, E_2... onde $E_i \cap E_j = \emptyset$ quando $i \neq j$, tem-se que

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

Desta forma, chamamos o número $P(E)$ de probabilidade do evento E.

Resumidamente, o Axioma 1 demonstra que a probabilidade de um determinado resultado E ocorrer em um experimento é dada por um número entre 0 e 1. Pelo Axioma 2, temos que a probabilidade de todos os possíveis resultados, ou seja, a probabilidade do espaço amostral S é 1. Por fim, o Axioma 3 diz que a probabilidade de pelo menos um evento de uma sequência de eventos exclusivos ocorrer é dada pela soma das probabilidades dos eventos pertencentes a esta sequência.

Ao realizar um experimento, frequentemente o interesse é voltado para uma função do resultado e não ao próprio resultado em si. Um exemplo disso é o experimento de jogar uma moeda, onde o interesse pode ser na quantidade de caras que irá sair e não necessariamente na sequência de caras e coroas que se terá. Essas grandezas de interesse ou, formalmente, funções reais definidas no espaço amostral são denominadas variáveis aleatórias (ROSS, 2010).

DEFINIÇÃO 1 Uma variável aleatória $X : S \to \mathbb{R}$ que assume no máximo um número contável de valores possíveis, ou seja, um conjunto infinito que possui, no máximo, a cardinalidade dos números naturais, é chamada de variável aleatória discreta e sua função de probabilidade $p(a)$ de X é dada por

$$p(a) = P_X\{X = a\} = P(X^{-1}(a))$$

onde $X^{-1}(a) \subset S$ é a imagem inversa da variável aleatória.

DEFINIÇÃO 2 X é uma variável aleatória contínua se existir uma função não negativa f_X, definida para todo real $x \in (-\infty, \infty)$ tal que, para todo conjunto R de números reais, temos

$$P_X\{X \in R\} = \int_R f_X(x)dx = P(X^{-1}(R))$$

onde f_X é chamada função densidade de probabilidade da variável aleatória X.

Observação: Como um dado valor da variável aleatória é determinado pelo resultado do experimento, que é um evento denotado por $X^{-1}(A)$, $A \subset \mathbb{R}$), podemos atribuir probabilidades a qualquer subconjunto de possíveis valores da variável aleatória.

DEFINIÇÃO 3 O valor esperado ou esperança de uma variável aleatória discreta X é dado, caso exista e for finita, por

$$E[X] = \sum_i x_i P_X\{X = x_i\}$$

DEFINIÇÃO 4 O valor esperado ou esperança de uma variável aleatória contínua X é dado, caso exista e for finita, por

$$E[X] = \int_{-\infty}^{\infty} x f_X(x)dx$$

Observação: O valor esperado de uma variável aleatória também é chamado de média.

DEFINIÇÃO 5 Se X é uma variável aleatória com média μ, então a variância de X, representada por $Var(X)$, é definida, caso exista e for finita, como

$$Var(X) = E[(X-\mu)^2]$$

DEFINIÇÃO 6 As variáveis aleatórias X e Y são independentes se, para quaisquer dois conjuntos de números reais A e B,

$$P_{XY}\{X \in A \quad e \quad Y \in B\} = P_X\{X \in A\}P_Y\{Y \in B\}$$

DEFINIÇÃO 7 Duas variáveis aleatórias X e Y são identicamente distribuídas se ambas tiverem a mesma distribuição de probabilidade.

DEFINIÇÃO 8 Dadas duas variáveis aleatórias X e Y, a função de distribuição de probabilidade conjunta de X e Y é dada por

$$F(a,b) = P_{XY}\{X \leq A \quad e \quad Y \leq B\}$$

onde $-\infty < a, b < \infty$.

DEFINIÇÃO 9 Sejam X e Y variáveis aleatórias conjuntamente discretas, a função de probabilidade condicional de X dado que $Y = y$ é definida, para todo y tal que $P\{Y = y\} > 0$, por

$$P_{X|Y}\{X = x|Y = y\} = \frac{P_{XY}\{X = x \quad e \quad Y = y\}}{P_Y\{Y = y\}}$$

DEFINIÇÃO 10 Sejam X e Y variáveis aleatórias conjuntamente discretas, a esperança condicional de X dado que $Y = y$, para todos os valores de y tais que $p_Y(y) > 0$, é definida como

$$E[X|Y = y] = \sum_x x P_{X|Y}\{X = x|Y = y\}$$

DEFINIÇÃO 11 Sejam X e Y variáveis aleatórias conjuntamente contínuas com função densidade de probabilidade conjunta f_{XY}, então a função densidade de probabilidade condicional de X dado que $Y = y$ é definida, para todos os valores de y tais que $f_Y(y) > 0$, por

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)}$$

DEFINIÇÃO 12 Sejam X e Y variáveis aleatórias conjuntamente contínuas, a esperança condicional de X dado que $Y = y$, é definida como

$$E[X|Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx$$

desde que $f_Y(y) > 0$.

4 Distribuições de probabilidade

Conforme explicitado na quinta característica citada por Vilanova para que os riscos sejam seguráveis, é necessário que estes possibilitem, de maneira estatística, prever comportamentos futuros em função de dados passados. Sendo assim, o prêmio de risco de um determinado risco coberto pela apólice de seguro pode ser estimado por meio de um modelo estatístico.

De acordo com McCullagh, um modelo estatístico é um conjunto de hipóteses sobre dados observados e dados semelhantes de diferentes fatores. Em termos matemáticos, os modelos estatísticos possuem duas componentes fundamentais e pode ser representado por um par (S, P), onde S é o espaço amostral (conjunto das observações possíveis) e P é uma família de distribuições de probabilidades em S.

Segundo Lindsey (1997), os modelos estatísticos envolvem variabilidade no resultado por conta de fatores aleatórios desconhecidos. A classe mais importante desses modelos contém os modelos lineares generalizados, os quais serão indicados para a construção da modelagem do prêmio de risco ao fim deste trabalho de pesquisa.

O primeiro passo do processo de modelagem estatística é encontrar uma distribuição de probabilidade que se ajuste bem aos dados do espaço amostral utilizado.

De acordo com Meyers (2009), modelar adequadamente dados de sinistros sempre foi problemático. Segundo ele, o problema é particularmente agudo para dados individuais onde a maioria dos sinistros são zero e, para casos com sinistros positivos, existe alta distorção. A maioria dos modelos tradicionais não lidam com uma mistura de dados discretos de zero e dados positivos contínuos. No entanto, o modelo que será proposto neste trabalho consegue lidar com essa mistura.

Como definido no primeiro capítulo, o prêmio de risco nada mais é do que o quociente entre o montante de sinistros ocorridos e as unidades expostas ao risco. As unidades expostas ao risco ou mais popularmente conhecido como exposição, é um valor conhecido durante o processo de modelagem, pois o modelo é construído com os sinistros gerados a partir destes expostos. Portanto, a única variável aleatória na equação (2.1) é o montante de sinistros ocorridos que, basicamente, é a soma da severidade (custo) de todos os sinistros ocorridos na carteira em determinado período de tempo.

Desta forma, o montante de sinistros ocorridos depende de duas componentes, sendo elas: a quantidade de sinistros e o quanto custa cada um deles. Sendo assim, é importante entender quais distribuições podem descrever estas componentes.

4.1. Distribuição da quantidade de sinistros

A quantidade de sinistros pode ser associada a uma contagem, isto é, à contagem de cada sinistro que ocorreu. Portanto, é necessário encontrar uma distribuição de probabilidade que seja utilizada para modelar dados de contagem. Alguns exemplos de modelos de contagem em seguros são:

- O número de mortes em relação à idade, sexo e profissão em estudos de mortalidade.

- Em seguro de saúde, o número de eventos gerados por segurados em função de variáveis explanatórias como idade, sexo e plano.

- Em seguro de automóvel, o número de sinistros gerados por veículos segurados em função da cor, ano de modelo, sinistros prévios e outros.

Suponha que possamos classificar a ocorrência de sinistro em determinada apólice de seguros como sinistro ocorrido ou sinistro não ocorrido. Tome $X = 1$ se o sinistro ocorreu e $X = 0$ se o sinistro não ocorreu. Logo, pode-se escrever a função de probabilidade de X como

$$\begin{cases} P[X=0] = 1-p \\ P[X=1] = p \end{cases} \tag{4.1}$$

onde p, tal que $0 \leq p \leq 1$, é a probabilidade do sinistro ter ocorrido.

DEFINIÇÃO 13 Uma variável aleatória discreta X é chamada de variável aleatória de Bernoulli se sua função de probabilidade for dada pelas equações 4.1 para algum $p \in (0,1)$. p é chamado de parâmetro da variável aleatória de Bernoulli.

Agora, suponha que se queira analisar a ocorrência de sinistros em n apólices completamente independentes, para cada qual denota-se a probabilidade do sinistro ter ocorrido como p e a probabilidade de não ter ocorrido sinistro como $1-p$. Desta forma, diz-se que B_n^p é uma variável aleatória binomial com parâmetros (n,p).

DEFINIÇÃO 14 Uma variável aleatória B_n^p é chamada de variável aleatória binomial de parâmetros (n,p) se e somente se

$$B_n^p = \sum_{i=1}^{n} X_i \tag{4.2}$$

onde cada $X_i \sim Bernoulli(p)$ é uma variável de Bernoulli independente e identicamente distribuída com parâmetro p.

TEOREMA 1 A função densidade de probabilidade da variável aleatória B_n^p é

$$P[B_n^p = k] = \binom{n}{k} p^k (1-p)^{n-k} \qquad k = 0,1,\ldots,n \tag{4.3}$$

Demonstração: Note que a probabilidade de qualquer sequência de n resultados contendo k sucessos e $n-k$ fracassos é, pela independência que se supõe entre as tentativas, $p^k(1-p)^{n-k}$. Desta forma, obtém-se o resultado da equação 4.3, uma vez que há $\binom{n}{k}$ sequências diferentes de n resultados levando a k sucessos e $n-k$ fracassos. Para ilustrar este resultado, tome como exemplo $n=4$ e $k=2$, logo temos que existem $\binom{4}{2}=6$ maneiras diante das quais 4 tentativas podem gerar 2 sucessos, ou seja, os possíveis resultados são (s,s,f,f), (s,f,s,f), (s,f,f,s), (f,s,s,f), (f,s,f,s) e (f,f,s,s), onde s denota um sucesso e f um fracasso. Cada um desses resultados tem probabilidade $p^2(1-p)^2$ de ocorrer e, portanto, a probabilidade desejada de dois sucessos nas quatro tentativas é $\binom{4}{2}p^2(1-p)^2$. □

DEFINIÇÃO 15 Considere uma sequência de variáveis aleatórias Binomiais $B_n^{p_n}$ tal que $n \to \infty$ e $p_n \to 0$ com $np_n \to \lambda > 0$. A variável aleatória limite dessa sequência é chamada de variável aleatória de Poisson e denotada por N com parâmetro λ.

TEOREMA 2 A variável aleatória de Poisson tem função de probabilidade

$$P[N=k]=\frac{e^{-\lambda}\lambda^k}{k!}$$

Demonstração:

$$P[B_n^{p_n}=i]=\frac{n!}{(n-i)!i!}p_n^i(1-p_n)^{n-i}$$

$$=\frac{n!}{(n-i)!i!}\left(\frac{\lambda}{n}\right)^i\left(1-\frac{\lambda}{n}\right)^{n-i}$$

$$=\frac{n(n-1)...(n-i+1)}{n^i}\frac{\lambda^i}{i!}\frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^i}$$

Calculando o limite de $n\to\infty$ na expressão acima, tem-se que

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n=e^{-\lambda}$$

$$\lim_{n\to\infty}\frac{n(n-1)...(n-i+1)}{n^i}=1$$

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^i=1$$

Logo,

$$P[N=i]=e^{-\lambda}\frac{\lambda^i}{i!}$$

□

Observação: demonstra-se que

$$E[N]=\lambda \tag{4.4}$$

$$Var[N]=\lambda \tag{4.5}$$

De acordo com o segundo item da lista de VILANOVA (1969) que determina as características que tornam um risco segurável, o grupo deve ser o mais numeroso possível, ou seja, a aproximação $n \to \infty$ é justificada por esse princípio. Além disso, para que o prêmio não resulte em valores exorbitantes, a probabilidade p_n de que o sinistro ocorra deve ser suficientemente pequena. Essas hispóteses foram utilizadas no teorema anterior e justificam que a distribuição de Poisson é adequada para a modelagem de contagem de sinistros.

4.2. Distribuição dos valores de sinistro

Os valores dos sinistros ou sua severidade, como são conhecidos no mercado de seguros, correspondem à indenização que é paga ao segurado quando o risco se materializa, isto é, quando um sinistro ocorre. Desta forma, a severidade pode ser associada a uma variável contínua e, usualmente, de valores positivos.

De acordo com JONG e HELLER (2008), as opções para modelar esse tipo de variável são:

- Usar a transformação de normalidade e, então, empregar o modelo linear normal na variável transformada. Portanto, $g(\mu) \sim N(\mu, \delta^2)$ onde g é uma transformação e $\mu = x'\beta$.

- Modelar vias modelos lineares generalizados, usando uma distribuição que é concentrada em valores não-negativos. Exemplos de distribuições deste tipo são a Gama e a Gaussiana Inversa.

Porém, dado que os valores de severidade são positivos, a utilização do modelo linear para esta variável não é adequada. Sendo assim, para a modelagem da severidade, o emprego de modelos lineares generalizados é mais adequado porque podemos utilizar distribuições com suporte não-negativo. A distribuição que é utilizada mais comumente para este tipo de problema é a distribuição Gama.

DEFINIÇÃO 16 Uma variável aleatória X tem distribuição Gama com parâmetros $\alpha > 0$ e $\tau > 0$, denotando-se $X \sim Gama(\alpha, \tau)$ se sua função densidade de probabilidade for dada por

$$f_X(y;\alpha,\tau) = \frac{x^{\alpha-1}e^{\frac{-x}{\tau}}}{\tau^{\alpha}\Gamma(\alpha)} \tag{4.6}$$

onde

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1}e^{-x}dx$$

Observação: demonstra-se que

$$E[X] = \alpha\tau \tag{4.7}$$

$$Var[X] = \alpha\tau^2 \tag{4.8}$$

4.3. Distribuição do Montante de Sinistros Ocorridos

Uma vez encontradas as distribuições adequadas para estimar o número de sinistros ocorridos e também o valor dos sinistros, é possível construir uma distribuição para o montante dos valores de sinistros ocorridos, o que nada mais é do que a soma da severidade de cada sinistro que ocorreu na carteira.

DEFINIÇÃO 17 O montante de sinistros ocorridos Z é a soma de N variáveis aleatórias Gama independentes e identicamente distribuídas $X_i \sim Gama(\alpha,\tau)$, onde N é uma variável aleatória de Poisson $N \sim Poisson(\lambda)$, tal que

$$Z = X_1 + ... + X_N = \sum_{i=1}^{N} X_i \tag{4.9}$$

Adiante, mostrar-se-á que essa distribuição é equivalente, sob certas hipóteses, à distribuição Tweedie e, por esse motivo, a Tweedie foi denominada como composta de Poisson por FELLER (1968) e JORGENSEN (1987), como composta de Gama por JOHNSON e KOTZ (1971) e também é conhecida como Poisson-Gama, mistura Poisson de Gamas, soma Poisson de Gamas, entre outros.

DEFINIÇÃO 18 Seja X uma variável aleatória discreta. A função geradora de momentos de X é definida como

$$M_X(t) = E[e^{tX}] = \sum_x e^{tx} p(x) dx \tag{4.10}$$

DEFINIÇÃO 19 Seja X uma variável aleatória contínua. A função geradora de momentos de X é definida como

$$M_X(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f_X(x) dx \tag{4.11}$$

em que $f_X(x)$ é a função de probabilidade da variável X.

DEFINIÇÃO 20 A função geradora de cumulantes de uma variável aleatória X é definida por

$$C_X(t) = \log M_X(t) \tag{4.12}$$

onde $M_X(t)$ é a função geradora de momentos de X.

TEOREMA 3 A função geradora de cumulantes de uma variável aleatória de Poisson é dada por

$$\lambda(e^t - 1) \tag{4.13}$$

Demonstração:

$$\begin{aligned} M_N(t) &= E[e^{tN}] \\ &= \sum_{n=0}^{\infty} \frac{e^{tn} e^{-\lambda} \lambda^n}{n!} \\ &= e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda e^t)^n}{n!} \end{aligned}$$

$$= e^{-\lambda} e^{\lambda e^t}$$
$$= e^{\lambda(e^t - 1)}$$

onde, na penúltima passagem, identificamos a série de Taylor da função exponencial. Logo,

$$C_N(t) = \ln M_N(t)$$
$$= \ln e^{\lambda(e^t-1)}$$
$$= \lambda(e^t - 1)$$

□

TEOREMA 4 A função geradora de cumulantes de uma variável aleatória Gama é dada por

$$\ln(1 - \tau t)^{-\alpha} \tag{4.14}$$

onde $t < \frac{1}{\tau}$.

Demonstração:

$$M_X(t) = E[e^{tX}] = \int_0^\infty e^{tx} \frac{x^{\alpha-1} e^{\frac{-x}{\tau}}}{\tau^\alpha \Gamma(\alpha)} dx$$
$$= \frac{1}{\tau^\alpha \Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{(t-\frac{1}{\tau})x} dx$$
$$= \frac{1}{\tau^\alpha \Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-\frac{(\tau-t)x}{\tau}} dx$$

Agora, faça a seguinte mudança de variável

$$u = \frac{(1-\tau t)x}{\tau} \Longrightarrow \frac{dx}{du} = \frac{\tau}{1-\tau t}$$

Logo, temos que

$$M_X(t) = \frac{1}{\tau^\alpha \Gamma(\alpha)} \int_0^\infty \left(\frac{\tau}{1-\tau t} u\right)^{\alpha-1} \frac{\tau}{1-\tau t} e^{-u} du$$
$$= \frac{1}{\Gamma(\alpha)} (1-\tau t)^{-\alpha} \int_0^\infty u^{\alpha-1} e^{-u} du$$
$$= (1-\tau t)^{-\alpha}$$

onde, foi utilizado a definição de função Gama na última passagem.

Portanto, temos que

$$C_X(t) = \ln(1-\tau t)^{-\alpha}$$

□

TEOREMA 5 Tome Z como na Definição 17, então a função geradora de cumulantes de Z é dada por

$$C_Z(t) = \ln M_Z(t) = \lambda[(1-\tau t)^{-\alpha} - 1]$$

Demonstração: Dado n inteiro positivo fixo,

$$E\left[e^{t\sum_i^N X_i}\middle| N = n\right]$$

$$= E\left[e^{t\sum_i^n X_i}\right]$$

$$= \prod_{i=1}^{n} E\left[e^{tX_i}\right]$$

$$= \left(E\left[e^{tX_i}\right]\right)^n$$

$$= [M_X(t)]^n$$

onde, na antepenúltima passagem, foi usado que as variáveis X_i são independentes, na penúltima, que elas são identicamente distribuídas e a última passagem decorre da própria definição de função geradora de momentos.

Assim,

$$M_X(t) = E\left[e^{tX_i}\right]$$

Portanto,

$$E\left[e^{tZ}\middle|N\right] = (M_X(t))^N$$

Logo,

$$M_Z(t) = E\left[(M_X(t))^N\right] = E\left[\left(E\left[e^{tX_i}\right]\right)^N\right] \tag{4.15}$$

Mas, por definição, temos que

$$C_X(t) = \ln M_X(t) \Longrightarrow e^{C_X(t)} = M_X(t) \tag{4.16}$$

Sendo assim, podemos substituir a primeira igualdade de 4.15 da seguinte forma

$$M_Z(t) = E[(e^{C_X(t)})^N] \tag{4.17}$$

Portanto, temos

$$C_Z(t) = \ln M_Z(t) = \ln E[e^{C_X(t)N}] = C_N(C_X(t)) \tag{4.18}$$

A partir da equação acima e utilizando os Teoremas 3 e 4, obtém-se que a função geradora de cumulantes de Z é

$$C_Z(t) = C_N(C_X(t))$$

$$= C_N(\ln(1-\tau t)^{-\alpha})$$

$$= \lambda[e^{\ln(1-\tau t)^{-\alpha}} - 1]$$

$$= \lambda[(1-\tau t)^{-\alpha} - 1] \tag{4.19}$$

□

De acordo com a Definição 17, o montante de sinistros ocorridos se trata de uma soma de distribuições Gama, onde o número de termos da soma é uma variável aleatória de Poisson.

JORGENSEN e DE SOUZA (1994) assumiram que o número de sinistros seguia a distribuição de Poisson, enquanto o valor de cada sinistro poderia ser descrito pela distribuição Gama, assim como elucidou-se anteriormente. Mais que isso, JORGENSEN e DE SOUZA modelaram diretamente o custo total esperado por unidade exposta ao risco, isto é, o prêmio de risco, utilizando esta distribuição de probabilidade.

Segundo SMYTH e JORGENSEN (2002), essas premissas implicam que o prêmio de risco segue uma distribuição Tweedie, que eles caracterizaram como uma distribuição composta de Poisson.

DEFINIÇÃO 21 Uma variável aleatória Y tem distribuição Tweedie se a sua média e a sua variância são dadas, respectivamente, por

$$E[Y] = \mu \tag{4.20}$$

para $\mu > 0$.

$$Var[Y] = \phi E[Y]^{\theta} = \phi\mu^{\theta} \tag{4.21}$$

para $\mu > 0$, $\phi > 0$, $\theta \in (-\infty, 0] \cup [1, \infty)$. μ, θ e ϕ são os parâmetros da variável aleatória de Tweedie.

Algumas distribuições conhecidas são casos especiais da Tweedie, por exemplo:

- Para $\theta = 0$, distribuição Normal (neste caso, não é necessário $\mu > 0$;
- Para $\theta = 1$ e $\phi = 1$, distribuição Poisson; (4.4) e (4.5)
- Para $\theta = 2$, distribuição Gama; (4.7) e (4.8)
- Para $\theta = 3$, distribuição Gaussiana Inversa.

No intervalo $\theta \in [1, 2]$, a Tweedie gradualmente perde sua massa em 0, enquanto passa de uma distribuição Poisson para uma distribuição Gama.

De acordo com DUNN e SMYTH (1997), para $\theta > 1$, a Tweedie tem a seguinte função densidade de probabilidade:

$$\frac{\partial \log f_Y}{\partial \mu} = \frac{y - \mu}{\phi\mu^{\theta}} \Longrightarrow \log f_Y(y; \mu, \theta, \phi) = \frac{1}{\phi}\left(y\frac{\mu^{1-\theta}}{1-\theta} - \frac{\mu^{2-\theta}}{2-\theta}\right) + c(y, \phi, \theta) \tag{4.22}$$

onde $c(y, \phi, \theta)$ não depende de μ.

DEFINIÇÃO 22 Diz-se que uma variável aleatória W tem distribuição pertencente à família exponencial se a função densidade de probabilidade (f.d.p.) puder ser escrita na forma

$$f(w; \nu, \phi) = exp\left\{\frac{y\nu - \kappa(\nu)}{a(\phi)} - c(y, \phi)\right\} \tag{4.23}$$

onde θ e ϕ são parâmetros escalares, $a(\cdot)$, $b(\cdot)$ e $c(\cdot,\cdot)$ são funções reais conhecidas.

TEOREMA 6 Seja Y uma variável aleatória de Tweedie com a função densidade de probabilidade dada por 4.22 e parâmetro θ fixo. Então Y pertence à família exponencial.

Demonstração: Usando as seguintes substituições

$$\nu = \frac{\mu^{1-\theta}}{1-\theta} \tag{4.24}$$

$$\kappa(\nu) = \frac{[(1-\theta)\nu]^{\frac{(2-\theta)}{(1-\theta)}}}{(2-\theta)} = \frac{\mu^{2-\theta}}{2-\theta} \tag{4.25}$$

temos que

$$\log f_Y(y;\mu,\phi) = \frac{1}{\phi}\left(y\frac{\mu^{1-\theta}}{1-\theta} - \frac{\mu^{2-\theta}}{2-\theta}\right) + c(y,\phi)$$

$$= \frac{1}{\phi}(y\nu - \kappa(\nu)) + c(y,\phi) \tag{4.26}$$

Portanto,

$$f_Y(y;\mu,\phi) = exp\left\{\frac{1}{\phi}(y\nu - \kappa(\nu)) + c(y,\phi)\right\} \tag{4.27}$$

com $a(\phi) = \phi$.

□

TEOREMA 7 Seja Y uma variável aleatória de Tweedie com a função densidade de probabilidade dada por (4.22). Então a função geradora de cumulantes de Y é dada por

$$C_Y(t) = \frac{1}{\phi}\frac{\mu^{2-\theta}}{2-\theta}\left[(1+t\phi(1-\theta)\mu^{\theta-1})^{\frac{2-\theta}{1-\theta}} - 1\right]$$

Demonstração: Pelo teorema anterior, sabemos que Y é da família exponencial. A função geradora dos momentos da família exponencial é dada por

$$M_Y(t) = E[e^{tY}]$$

$$= \int e^{ty} e^{\frac{1}{\phi}(y\nu - \kappa(\nu)) + c(y,\phi)} dy$$

$$= \int e^{\frac{1}{\phi}[y(\nu + t\phi) - \kappa(\nu)] + c(y,\phi)} dy$$

Somando e subtraindo $\kappa(\nu + t\phi)$ no argumento da exponencial, temos

$$e^{\frac{1}{\phi}[\kappa(\nu+t\phi)-\kappa(\nu)]} \int e^{\frac{1}{\phi}[y(\nu+t\phi)-\kappa(\nu+t\phi)]+c(y,\phi)} dy$$

Como t é suficientemente pequeno de forma que a função geratriz dos momentos está bem definida, a integral acima é igual a 1 devido à normalização da família exponencial (DASGUPTA, 2011), e, portanto, temos que

$$M_Y(t) = e^{\frac{1}{\phi}[\kappa(\nu+t\phi)-\kappa(\nu)]}$$

Agora, utilizando as correspondências de (4.24) e (4.25), temos que a função geradora de cumulantes é

$$\begin{aligned}
C_Y(t) = \log M_Y(t) &= \frac{1}{\phi}[\kappa(\nu + t\phi) - \kappa(\nu)] \\
&= \frac{1}{\phi}\left[\frac{[(1-\theta)(\nu + t\phi)]^{\frac{2-\theta}{1-\theta}}}{2-\theta} - \frac{\mu^{2-\theta}}{2-\theta}\right] \\
&= \frac{1}{\phi}\left[\frac{\left[(1-\theta)\left(\frac{\mu^{1-\theta}}{1-\theta} + t\phi\right)\right]^{\frac{2-\theta}{1-\theta}}}{2-\theta} - \frac{\mu^{2-\theta}}{2-\theta}\right] \\
&= \frac{1}{\phi}\left[\frac{[\mu^{1-\theta} + t\phi(1-\theta)]^{\frac{2-\theta}{1-\theta}} - \mu^{2-\theta}}{2-\theta}\right] \\
&= \frac{1}{\phi}\frac{\mu^{2-\theta}}{2-\theta}\left[(1 + t\phi(1-\theta)\mu^{\theta-1})^{\frac{2-\theta}{1-\theta}} - 1\right] \qquad (4.28)
\end{aligned}$$

□

TEOREMA 8 Considere a variável aleatória Y de Tweedie conforme (4.22), para $1 < \theta < 2$ fixo, e Z a variável aleatória composta de Gama com Poisson conforme Definição 17. Então Z e Y têm a mesma distribuição de probabilidade.

Demonstração: ($\Longrightarrow$) Da equação (4.19), temos que a função geradora de cumulantes de Z é dada por

$$C_Z(t) = \lambda[(1-\tau t)^{-\alpha} - 1]$$

Agora, utilize as seguintes correspondências

$$\lambda = \frac{1}{\phi}\frac{\mu^{2-\theta}}{2-\theta}$$

$$\alpha = \frac{2-\theta}{\theta-1}$$

$$\tau = \phi(\theta-1)\mu^{\theta-1}$$

Portanto, temos que

$$\begin{aligned} C_Z(t) &= \lambda[(1-\tau t)^{-\alpha}-1] \\ &= \frac{1}{\phi}\frac{\mu^{2-\theta}}{2-\theta}\left[(1-\phi(\theta-1)\mu^{\theta-1}t)^{-\left(\frac{2-\theta}{\theta-1}\right)}-1\right] \\ &= \frac{1}{\phi}\frac{\mu^{2-\theta}}{2-\theta}\left[(1+t\phi(1-\theta)\mu^{\theta-1})^{\frac{2-\theta}{1-\theta}}-1\right] = C_Y(t) \end{aligned}$$

Logo, pelo Corolário 1 enunciado e provado no apêndice, temos que Z e Y possuem a mesma função geradora de cumulantes e, consequentemente, a mesma distribuição.

($\Longleftarrow$) Da equação (4.28), temos que a função geradora de cumulantes de Y é dada por

$$\begin{aligned} C_Y(t) &= \frac{1}{\phi}\frac{\mu^{2-\theta}}{2-\theta}\left[(1+t\phi(1-\theta)\mu^{\theta-1})^{\frac{2-\theta}{1-\theta}}-1\right] \\ &= \frac{1}{\phi}\frac{\mu^{2-\theta}}{2-\theta}\left[(1-\phi(\theta-1)\mu^{\theta-1}t)^{-\left(\frac{2-\theta}{\theta-1}\right)}-1\right] \end{aligned}$$

Agora, utilize as seguintes correspondências

$$\mu = \lambda\alpha\tau$$

$$\theta = \frac{\alpha+2}{\alpha+1}$$

$$\phi = \frac{\lambda^{1-\theta}(\alpha\tau)^{2-\theta}}{2-\theta}$$

Temos que

$$2-\theta = 2-\left(\frac{\alpha+2}{a+1}\right) = \frac{\alpha}{\alpha+1}$$

$$\theta-1 = \frac{\alpha+2}{a+1} - 1 = \frac{1}{\alpha+1}$$

$$\phi = \frac{\lambda^{-\frac{1}{\alpha+1}}(\alpha\tau)^{\frac{\alpha}{\alpha+1}}}{\alpha}(\alpha+1)$$

$$\frac{1}{\phi}\frac{\mu^{2-\theta}}{2-\theta} = \frac{\alpha}{\lambda^{-\frac{1}{\alpha+1}}(\alpha\tau)^{\frac{\alpha}{\alpha+1}}(\alpha+1)}\frac{(\lambda\alpha\tau)^{\frac{\alpha}{\alpha+1}}(\alpha+1)}{\alpha} = \frac{\lambda^{\frac{\alpha}{\alpha+1}}}{\lambda^{-\frac{1}{\alpha+1}}} = \lambda \tag{4.29}$$

$$\phi(\theta-1)\mu^{\theta-1} = \frac{\lambda^{-\frac{1}{\alpha+1}}(\alpha\tau)^{\frac{\alpha}{\alpha+1}}(\alpha+1)}{\alpha}\left(\frac{1}{\alpha+1}\right)(\lambda\alpha\tau)^{\frac{1}{\alpha+1}} = \tau \tag{4.30}$$

$$\frac{2-\theta}{\theta-1} = \frac{\alpha}{\alpha+1}\frac{\alpha+1}{1} = \alpha \tag{4.31}$$

Portanto, utilizando as correspondências (4.29), (4.30) e (4.31), temos que

$$\begin{aligned} C_Y(t) &= \frac{1}{\phi}\frac{\mu^{2-\theta}}{2-\theta}\left[(1-\phi(\theta-1)\mu^{\theta-1}t)^{-\left(\frac{2-\theta}{\theta-1}\right)} - 1\right] \\ &= \lambda[(1-\tau t)^{-\alpha} - 1] = C_Z(t) \end{aligned}$$

Logo, pelo Corolário 1 enunciado e provado no apêndice, temos que Y e Z possuem a mesma função geradora de cumulantes e, consequentemente, a mesma distribuição.

□

Observação: O fato de que o parâmetro α da distribuição Gama deve ser positivo, implica que a representação de Y como uma composta de Poisson e Gama é válida somente para $1 < \theta < 2$.

5 Considerações Finais

Uma vez encontrada a distribuição de probabilidade adequada, é possível modelar estatisticamente o prêmio de risco. Uma das características mais importantes da distribuição Tweedie é que ela pertence à família exponencial, como foi demonstrado no capítulo anterior. Esta característica é essencial para um processo de modelagem estatística que utiliza uma técnica conhecida como modelos lineares generalizados. Esta classe de modelo é justamente a classe que deixa-se indicada neste trabalho de pesquisa em caso de implementação.

Os modelos lineares generalizados possuem uma formulação que generaliza o erro do modelo linear clássico e seu desenvolvimento foi feito por Nelder e Wedderburn visando a unificação de modelos como a regressão logísitca, a regressão linear clássica e a de Poisson. O artigo original (NELDER; WEDDERBURN, 1972) que descreve este modelo data de 1972. Todavia, os trabalhos utilizando a aplicação destes modelos foram realizados somente após a década de 1980, em virtude do desenvolvimento computacional. Portanto, dado que atualmente tem-se uma variedade de softwares computacionais que permitam a aplicação deste tipo de abordagem, indica-se esta técnica para a estimativa do montante de sinistros ocorridos.

Segundo LINDSEY (1997), os modelos lineares generalizados constituem a classe mais importante de modelos estatísticos. De acordo com POUSINHO (2013), eles possuem como objetivo estudar a relação entre variáveis, ou seja, o efeito que uma ou mais variáveis explicativas provocam sobre a variável resposta.

Conforme foi apresentado por MCGULLAGH e NELDER (1989), um modelo linear generalizado possui três componentes, sendo elas:

- **Componente aleatória:** as variáveis aleatórias Y_i $(i = 1, 2, ..., 3)$ são independentes e possuem distribuição de probabilidade pertencente à família exponencial, conforme a definição 22.

- **Componente sistemática:** Suponha um conjunto de p coeficientes desconhecidos β e um conjunto de variáveis explicativas $X_{n\times p} = [x_1^T, x_2^T, ..., x_n^T]^T$, tal que

$$\eta = X\beta$$

onde X é a matriz das variáveis explicativas e β representa o vetor de parâmetros, de dimensão p.

- **Função de ligação:** a ligação entre as componentes aleatórias e as componentes sistemáticas é feita através da função de ligação g que, por sua vez, é monótona e diferenciável em todo seu domínio:

$$\eta = g(\mu) = g(E[Y])$$

Dado que, pelo Teorema 6, a Tweedie pertence à família exponencial, basta escolher uma função de ligação adequada e, para um dado conjunto de variáveis explicativas, o processo de modelagem utilizando a técnica de modelos lineares generalizados se torna viável. Desta forma, é possível estimar o prêmio de risco dadas as características do risco coberto.

Este trabalho de pesquisa não tem como objetivo explorar os algoritmos para a estimação dos parâmetros a partir do modelo, mas sim abordar, de maneira geral, aspectos relevantes para a estimativa do prêmio de risco, ou seja, as distribuições e técnicas adequadas para tal. Sendo assim, recomenda-se ao leitor que, em caso de implementação do modelo, utilize um software computacional adequado como, por exemplo, o SAS ("Statistical Analysis System"), o Emblem desenvolvido pela Towers Watson ou até mesmo pacotes específicos do R. Todos estes softwares são capazes de utilizar a distribuição Tweedie dentro da abordagem de modelos lineares generalizados, o que permitirá ao usuário, a partir de seu conjunto de dados, estimar de maneira assertiva a expectativa total de sinistros de uma carteira de seguros em determinado período de tempo, o que resultará, dada a exposição conhecida ao risco, no prêmio de risco por unidade exposta.

A Apêndice

PROPOSIÇÃO 1 A desigualdade de Markov
Se X é uma variável aleatória que apresenta valores não negativos então, para qualquer $a > 0$,

$$P\{X \geq a\} \leq \frac{E[X]}{a}$$

Demonstração: Para $a > 0$, suponha

$$I = \begin{cases} 1 & se \quad X \geq a \\ 0 & se \quad X < a \end{cases}$$

Além disso, dado que $X \geq 0$, temos que

$$I \leq \frac{X}{a}$$

Agora, tomando o cálculo do valor esperado em ambos os lados da desigualdade, temos

$$E[I] = P\{X \geq a\} \leq \frac{E[X]}{a}$$

□

PROPOSIÇÃO 2 A desigualdade de Chebyshev
Seja X uma variável aleatória com média finita μ e variância σ^2, então para $k > 0$, temos

$$P\{|X - \mu| \geq k\} \leq \frac{\sigma^2}{k^2}$$

Demonstração: Considere a variável aleatória não negativa $(X-\mu)^2$. Aplicando a desigualdade de Markov para esta variável e tomando $a = k^2$, temos

$$P\{(X-\mu)^2 \geq k^2\} \leq \frac{E[(X-\mu)^2]}{k^2}$$

Mas, pela definição de variância, temos

$$P\{|X-\mu| \geq k\} \leq \frac{E[(X-\mu)^2]}{k^2} = \frac{\sigma^2}{k^2}$$

pois $k > 0$. □

TEOREMA 9 Sejam μ e ν medidas de probabilidade em $[0,\infty)$. Se

$$\int_0^\infty e^{-sx}\mu(dx) = \int_0^\infty e^{-sx}\nu(dx),$$

onde $s_0 \geq s \geq 0$, então $\mu = \nu$.

Demonstração: Para cada λ, seja Y_λ uma variável aleatória com a distribuição de Poisson e parâmetro λ. Uma vez que Y_λ tem média e variância λ, a desigualdade de Chebyshev é dada por

$$P\left[\left|\frac{Y_\lambda - \lambda}{\lambda}\right| \geq \epsilon\right] \leq \frac{\lambda}{\lambda^2\epsilon^2} \to 0, \qquad \lambda \to \infty$$

Seja G_λ a função de distribuição de $\frac{Y_\lambda}{\lambda}$, então temos

$$G_\lambda(t) = \sum_{k=0}^{[\lambda t]} e^{-\lambda}\frac{\lambda^k}{k!}$$

Este resultado pode ser reescrito como

$$\lim_{\lambda\to\infty} G_\lambda(t) = \begin{cases} 1 & \text{se t>1,} \\ 0 & \text{se t<1.} \end{cases} \tag{A.1}$$

Agora considere uma densidade de probabilidade μ concentrada no intervalo $[0,\infty)$. Seja F a função de probabilidade correspondente. Defina

$$M(s) = \int_0^\infty e^{-sx}\mu(dx), \tag{A.2}$$

para $s \geq 0$.

De acordo com a definição, essa é a função geradora de momentos, mas o argumento foi refletido através da origem. Além disso, isto é uma transformada de Laplace unilateral, definida para todo s não negativo.

Para s positivo, de A.2, temos

$$M^{(k)}(s) = (-1)^k \int_0^\infty y^k e^{-sy} \mu(dy) \tag{A.3}$$

Logo, para x e s positivos, temos

$$\sum_{k=0}^{[sx]} \frac{(-1)^k}{k!} s^k M^{(k)}(s) = \int_0^\infty \sum_{k=0}^{[sx]} e^{-sy} \frac{(sy)^k}{k!} \mu(dy) = \int_0^\infty G_{sy}\left(\frac{x}{y}\right) \mu(dy) \tag{A.4}$$

Fixe $x > 0$. Se $0 \leq y < x$, então $G_{sy} \to 1$ na medida em que $s \to \infty$ de acordo com A.1. Já no caso $y > x$, o limite é 0. Se $\mu\{x\} = 0$, o integrando da direita em A.4 converge na medida em que $s \to \infty$ para $I_{[0,x]}(y)$, exceto para um conjunto de $\mu - medida$ 0. Pelo Teorema da Convergência Limitada, temos

$$\lim_{s\to\infty} \sum_{k=0}^{[sx]} \frac{(-1)^k}{k!} s^k M^{(k)}(s) = \mu[0,x] = F(x) \tag{A.5}$$

Portanto, $M(s)$ determina o valor de F em x se $x > 0$ e $\mu\{x\} = 0$, que cobre todos os valores contáveis de x em $[0,\infty)$. Desde que F seja contínua à direita, F e μ podem ser determinadas por A.5 pela função $M(s)$. De fato, μ é determinado, de acordo com A.5, pelos valores de $M(s)$, para s dado um s_0 arbitrário. $\square$

COROLÁRIO 1 Sejam f_1 e f_2 funções reais em $[0,\infty)$ tal que $f(x) = F'(x)$, onde F é função de distribuição em $\mathbb{R}$. Se

$$\int_0^\infty e^{-sx} f_1(x) dx = \int_0^\infty e^{-sx} f_2(x) dx,$$

onde $s_0 \geq s \geq 0$, então $f_1 = f_2$ fora de um conjunto de medida zero de Lebesgue.

Referências Bibliográficas

[1] SMYTH, G. K.; *Regression Analysis of Quantity Data with Exact Zeroes*, Department of Mathematics, University of Queensland, Brisbane, Australia, 1996.

[2] LINDSEY, J. K.; *Applying Generalized Linear Models*, Springer Verlag, New York, 1977.

[3] de JONG, Pietz; HELLER, Gillian Z.; *Generalized Linear Models for Insurance Data*, Cambridge University Press, 2008.

[4] MEYERS, Glenn; *Pure Premium Regression with the Tweedie Model*, The Actuarial Review, 2009.

[5] JORGENSEN, B.; de SOUZA, M. C. Paes; *Fitting Tweedie's Compound Poisson Model to Insurance Claims Data*, Scandinavian Actuarial Journal, 1994.

[6] DUNN, P. K.; SMYTH, G. K.; *Series Evaluation of Tweedie Exponential DispersionModel Densities*, Statistics and Computing, 2005.

[7] ROSS, Sheldon; *Probabilidade: um curso moderno com aplicações*, Bookman, 2010.

[8] BILLINGSLEY, Patrick; *Probabiliy and Measure*, Wiley series in probability and mathematical statistics. Probability and mathematical statistics, 1995.

[9] VIEIRA, Antonio Fernando de Castro; *Análise da média e dispersão em experimentos fatoriais não replicados para otimização de processos industriais*, Tese (doutorado) – Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Industrial, 2004.

[10] VILANOVA, Wilson; *Matemática atuarial: destinado aos cursos de ciências econômicas, contábeis e atuariais*, Livraria Pioneira, Editora da Universidade, 1969.

[11] PINTO, M. G. F.; *Modelo Linear Generalizado,* Projeto de Iniciação Científica de Mateus Gonzalez de Freitas Pinto sob a orientação do Prof. Dr. Alexei Magalhães Veneziani, Universidade Federal do ABC, 2018.

[12] FILHO, Antônio Cordeiro; *Cálculo atuarial aplicado: teoria e aplicações: exercícios resolvidos e propostos,* Atlas, 2009.

[13] FERREIRA, Paulo Pereira; *Modelos de Precificação e Ruína para Seguros de Curto Prazo,* Funenseg, 2005.

[14] QUIJANO, Oscar; Garrido, José; *Generalised linear models for aggregate claims; to Tweedie or not?,* Concordia University, Montreal, 2014.

[15] BOLAND, Philip J.; *Statistical and Probabilistic Methods in Actuarial Science,* University College Dublin, 2007.

[16] McCULLAGH, P.; NELDER, J. A.; *Generalized Linear Models,* Chapman and Hall, 1989.

[17] NELDER, J. A.; WEDDERBURN, R. W. M.; *Generalized Linear Models,* Journal of the Royal Statistical Society, 1972.

[18] POUSINHO, A. P.; *Modelos Lineares Generalizados Tweedie Aplicados ao Cálculo de Provisões para Sinistros,* Tese (mestrado) - Universidade Técnica de Lisboa, Instituto Superior de Economia e Gestão, 2013.

[19] FELLER, W.; *An Introduction to Probability Theory and Its Applications,* Wiley, New York, 1968.

[20] JORGENSEN, B.; *Exponential Dispersion Models,* J. R. Statist. Soc., 1987.

[21] JOHNSON, N. L.; KOTZ, S.; *Univariate continuous distributions: distributions in statistics Vol 1 and 2,* Wiley, New York, 1971.

[22] JORGENSEN, B.; *The theory of exponential dispersion models and analysis of deviance,* Monografias de Matemátika Nº51, IMPA, Rio de Janeiro, 1992.

[23] DASGUPTA, A.; *Probability for Statistics and Machine Learning: Fundamentals and Advanced Topics,* Springer, 2011.

[24] NASCIMENTO, L. C.; *Breves Apontamentos sobre a História dos Seguros em Portugal e no Mundo,* Credimédia Seguros, Portugal, 2015.

[25] CNSeg; *História do Seguro No Mundo*, Disponível em: <http://cnseg.org.br/cnseg/mercado/historia-do-seguro/no-mundo.html> Acesso em: 11/07/2018.

[26] Tudo Sobre Seguros; *História do Seguro*, Disponível em: <http://www.tudosobreseguros.org.br/tss-um-pouco-de-historia/> Acesso em: 11/07/2018.

[27] Tudo Sobre Seguros; *Cálculo de Prêmio de Seguros*, Disponível em: <http://www.tudosobreseguros.org.br/tss-calculo-de-premios-de-seguros/> Acesso em: 03/04/2018.