



**UnB**

Universidade de Brasília  
Instituto de Ciências Exatas  
Departamento de Estatística

Dissertação de Mestrado

## **Estudo da distribuição de Poisson generalizada**

Por:

Hemilhana Tolentina Pereira

Orientador: Prof. Dr. Gustavo L. Gilardoni

Brasília

2016

Hemilhana Tolentina Pereira

# **Estudo da distribuição de Poisson generalizada**

Dissertação apresentada ao Departamento de Estatística do Instituto de Ciências Exatas da Universidade de Brasília como requisito parcial à obtenção do título de Mestre em Estatística.

Brasília

2016



# Resumo

Esta dissertação se dedica ao estudo e à revisão bibliográfica da distribuição de Poisson generalizada. Essa distribuição é uma alternativa aos modelos discretos que apresentam sobredispersão. A flexibilização do valor da variância em relação ao da média é devido ao acréscimo de um parâmetro à distribuição de Poisson.

A distribuição de Poisson generalizada foi inicialmente introduzida como elemento da classe de distribuições Lagrangeanas. Neste trabalho é proposta uma parametrização em função da  $W$  de Lambert que permite relacioná-la à classe de distribuições de série de potência e, desse modo, valer-se das propriedades e das características dessa classe de distribuições. Além disso, serão apresentadas aplicações de diferentes áreas do conhecimento, métodos de estimação pontual e intervalar, testes de hipóteses e as funções do *software R* dedicadas a essa distribuição.

**Palavras-chaves:** classe de distribuições Lagrangeanas; classe de distribuições de séries de potência; distribuição de Poisson generalizada; expansão de Lagrange; função  $W$  de Lambert.

# Abstract

In this dissertation we present a review of the generalized Poisson distribution. This distribution is an alternative to allow over dispersion in discrete models. The flexibility of the value of the variance in relation to the average is due to the presence of an additional parameter with respect to the Poisson distribution.

The generalized Poisson distribution was first introduced as an element of the class of Lagrangian distributions. In this work it is proposed a new parametrization in terms of the Lambert  $W$  function that allows us to relate it to the class of power series distributions and, therefore, to use well known properties and characteristics of this class. In addition, we will present applications from different areas of knowledge, inference methods for both point and interval estimation and hypotheses testing. We also review some R packages which deal with this distribution.

**Key words:** Lagrangean class of distributions, power series class of distributions, generalized Poisson distribution, Lagrange expansion, Lambert  $W$  function.

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Distribuição de Poisson Generalizada</b>	<b>3</b>
2.1	Parametrizações	4
2.2	Representações gráficas	5
<b>3</b>	<b>Motivação ao estudo da DPG</b>	<b>9</b>
3.1	Aplicações	9
3.1.1	Áreas de estudo	9
3.1.2	Modelo de regressão Poisson generalizada	12
<b>4</b>	<b>Classes de distribuições de probabilidade</b>	<b>15</b>
4.1	Classe de distribuições Lagrangeanas	15
4.1.1	Expansão de Lagrange	15
4.1.2	Distribuições de Lagrangeanas	19
4.1.2.1	Função geradora de probabilidade	20
4.1.2.2	Distribuição Lagrangeana básica	20
4.1.2.3	Distribuição Lagrangeana delta	20
4.1.2.4	Distribuição Lagrangeana generalizada	21
4.2	$W$ de Lambert	21
4.3	Classe de Distribuições de Série de Potências	23
4.3.1	Relações de recorrência para os momentos	23
4.3.2	Função característica	26
4.4	DPG como um elemento da classe de distribuições de série de potência	27
4.5	Propriedades e características da DPG	29
4.5.1	Soma das probabilidades da DPG	29
4.5.2	Relações de recorrência para os momentos	30
4.5.3	Função geradora de probabilidade	32
4.5.4	Função geradora de momentos	33
4.5.5	Função característica	33
4.5.6	Coefficiente de assimetria	34
4.5.7	Coefficiente de curtose	34
4.5.8	Convolução	35
4.5.9	Unimodalidade	35
<b>5</b>	<b>Inferência estatística para a DPG</b>	<b>37</b>
5.1	Família exponencial	37

5.2	Suficiência . . . . .	38
5.3	Métodos de estimação pontual dos parâmetros da DPG . . . . .	39
5.3.1	Método dos momentos . . . . .	39
5.3.2	Método da Máxima Verossimilhança . . . . .	40
5.3.2.1	Propriedades dos estimadores de máxima verossimilhança . . . . .	43
5.3.2.2	Avaliação dos EMVs para a DPG . . . . .	44
5.3.2.3	EMV quando um dos parâmetros é conhecido . . . . .	46
5.3.3	Método de estimação baseado na média amostral e na frequência de zeros da amostra . . . . .	47
5.3.4	Método da discrepância empírica ponderada . . . . .	48
5.3.5	Estudo de Monte Carlo . . . . .	49
5.4	Testes de hipóteses . . . . .	52
5.4.1	Teste da razão de verossimilhanças . . . . .	52
5.4.2	Testes de qualidade de ajuste . . . . .	54
5.5	Estimação Intervalar . . . . .	55
<b>6</b>	<b>Ferramentas computacionais no <i>software R</i> para a DPG . . . . .</b>	<b>61</b>
6.1	GPseq . . . . .	61
6.2	HMMpa . . . . .	62
6.3	RMKdiscrete . . . . .	63
6.4	VGAM . . . . .	65
6.5	lamW . . . . .	65
6.6	LambertW . . . . .	65
<b>7</b>	<b>Conclusão . . . . .</b>	<b>67</b>
	<b>Apêndices . . . . .</b>	<b>69</b>
	<b>APÊNDICE A DPG Truncada . . . . .</b>	<b>71</b>
	A.1 Espaço paramétrico de $\lambda$ . . . . .	71
	A.2 DPG truncada . . . . .	72
	<b>APÊNDICE B Cálculo da média e variância . . . . .</b>	<b>73</b>
	B.1 Média e Variância . . . . .	73
	<b>APÊNDICE C Resultado das simulações de Monte Carlo . . . . .</b>	<b>75</b>
	C.1 Estudo de Monte Carlo - Viés e EQM dos parâmetros . . . . .	75
	<b>APÊNDICE D Algoritmo . . . . .</b>	<b>79</b>
	D.1 Função para o método dos momentos . . . . .	79
	D.2 Função para o método baseado na frequência de zeros e na média . . . . .	79
	D.3 Função para o método de máxima verossimilhança . . . . .	80

D.4 Função para o método de discrepância empírica ponderada . . . . .	81
<b>Referências</b> . . . . .	<b>83</b>



# Agradecimentos

Agradeço ao meu orientador, Prof. Dr. Gustavo L. Gilardoni, pela sugestão do tema, indicações bibliográficas, críticas, apoio, compreensão, paciência e disposição durante a realização deste trabalho.

À minha família pelo apoio e incentivo.

Aos membros da banca examinadora por aceitarem o convite.

Aos funcionários do Departamento de Estatística pela cordialidade e presteza.



# Lista de ilustrações

Figura 1 – Funções de probabilidade da DPG com diferentes valores para os parâmetros.	6
Figura 2 – Funções de probabilidade da DPG com diferentes valores para os parâmetros.	7
Figura 3 – Funções de probabilidade da DPG com a mesma média ( $\mu = 4$ ) e com diferentes valores para os parâmetros. . . . .	8
Figura 4 – Função $W$ de Lambert. . . . .	22
Figura 5 – Intervalos de confiança pelo método de eliminação de maximização da verossimilhança. . . . .	59

# Lista de tabelas

Tabela 1	– Distribuição de 402 tatuzinhos de jardim em 122 placas. . . . .	10
Tabela 2	– Distribuição de 102 aranhas em 240 placas. . . . .	10
Tabela 3	– Classe de lesões em células de coelhos segundo a dosimetria da droga <i>streptonigrin</i> . . . . .	11
Tabela 4	– Método de estimação com a menor divergência de Kullback-Leibler. Simulação com 20 unidades amostrais. . . . .	51
Tabela 5	– Método de estimação com a menor divergência de Kullback-Leibler. Simulação com 50 unidades amostrais. . . . .	51
Tabela 6	– Método de estimação com a menor divergência de Kullback-Leibler. Simulação com 70 unidades amostrais. . . . .	51
Tabela 7	– Método de estimação com a menor divergência de Kullback-Leibler. Simulação com 120 unidades amostrais. . . . .	52
Tabela 8	– Testes de qualidade de ajuste segundo o modelo Poisson e Poisson generalizado para o conjunto de dados de 102 aranhas em 240 placas. . . . .	55
Tabela 9	– Intervalo de confiança para os parâmetros da DPG para o número de cromossomos lesionados em células de coelhos pelo uso de <i>streptonigrin</i> com diferentes dosagens. . . . .	60

# 1 Introdução

A distribuição de probabilidade de uma variável aleatória pode exibir formas variadas. Encontram-se na literatura estatística muitos modelos teóricos de distribuições que buscam descrever o comportamento de um evento em relação a essas distribuições de frequências, de modo que para cada evento haverá uma probabilidade de ocorrência associada. Parte importante na realização de estudos ou trabalho de pesquisa é o de associar dados ou medições com certos atributos ou características a algum modelo teórico. Essa associação tem a vantagem de permitir uma série de inferências estatísticas.

Para dados discretos, um dos principais modelos probabilísticos é a distribuição de Poisson. Uma das características dessa distribuição é a equidispersão (média igual à variância). No entanto, diante da percepção de que a natureza é heterogênea e de que a equidispersão nem sempre é observada, foi preciso, portanto, valer-se de outras possibilidades. Para dados com sobredispersão, por exemplo, a distribuição de Poisson subestima a variância e, como alternativa, tem-se utilizado a distribuição binomial negativa. Entretanto, Consul e Jain (1970, 1973a), Consul e Jain (1973b) sugeriram o uso da Distribuição de Poisson Generalizada (DPG) como opção aos modelos discretos que não apresentam equidispersão. O acréscimo de um parâmetro (denominado nesta dissertação de  $\lambda$ ) à distribuição de Poisson possibilitou a flexibilização do valor da variância em relação ao da média. Destaca-se que, quando  $\lambda = 0$ , a DPG se reduz à distribuição de Poisson.

Esta dissertação se dedica ao estudo e à revisão bibliográfica da DPG desde sua concepção – apresentada por meio da Classe de Distribuições Lagrangeanas (CDL) por Consul e Shenton (1972) –, das suas propriedades e características – demonstradas mediante a classe de distribuições de séries de potência –, das suas principais aplicações e das inferências estatísticas. Além disso, de modo original, será verificada a relação da DPG com a função  $W$  de Lambert e como essa decorrência simplifica alguns resultados. Apesar da DPG apresentada por Consul e Jain ter as características de subdispersão e sobredispersão, nesta dissertação será tratada apenas a característica de sobredispersão, pois, conforme será apresentado, a DPG com subdispersão apresenta uma série de restrições e dificuldades.

O conteúdo deste trabalho está organizado em 6 capítulos e 4 apêndices. No Capítulo 1, será apresentada a DPG, as principais parametrizações e as representações gráficas para diferentes parâmetros da DPG. No Capítulo 2, serão apresentadas algumas aplicações que motivaram o estudo e o aprofundamento da DPG. No Capítulo 3, será introduzido o conceito da expansão de Lagrange e como esse resultado dá origem à classe de distribuições Lagran-

geanas e a  $W$  de Lambert. Além disso, será estabelecida uma parametrização da DPG em função da  $W$  de Lambert; essa nova parametrização pertence à classe de distribuições de série de potência. Essa classificação permite utilizar para a DPG as propriedades e características dessa classe. No Capítulo 4, verifica-se os principais métodos de estimação para os parâmetros da DPG, um estudo simulado que compara esses métodos de estimação, os testes estatísticos para os parâmetros, o teste qui-quadrado de bondade de ajuste e, por fim, os métodos de cálculos dos intervalos de confiança para os parâmetros da DPG. No capítulo 5, observa-se os pacotes estatísticos do *software R-project* que apresentam funções dedicadas à DPG e à  $W$  de Lambert. No Capítulo 6, é apresentada a conclusão desta dissertação. Por fim, nos Apêndices A, B, C e D, encontram-se as demonstrações de alguns resultados apresentados neste trabalho.

## 2 Distribuição de Poisson Generalizada

Um fenômeno aleatório que realiza uma contagem de qualquer tipo, presumivelmente, desde que se cumpram alguns pressupostos, pode ser expresso por meio de uma distribuição de Poisson. Um desses pressupostos é que a média seja igual à variância (equidispersão) e que haja independência dos eventos em intervalos de tempo sucessivos, ou seja, em um determinado período, a ocorrência ou não de um evento não influencia a ocorrência posterior.

A equidispersão nem sempre é observada em dados de contagem. Diante disso, muitos pesquisadores propuseram distribuições baseadas na de Poisson, dando origem a várias derivações dessa com sobredispersão (o valor da variância é maior do que o da média) ou subdispersão (o valor da variância é menor do que o da média).

Em geral, quando a variância é diferente da média, o mais usual é utilizar a Distribuição Binomial Negativa (DBN) ou, a Distribuição de Poisson Generalizada (DPG). A principal vantagem da DBN e da DPG é que elas possuem um parâmetro adicional, tornando-as mais flexíveis. No caso de DBN, o parâmetro introduz sobredispersão e, no caso da DPG, o parâmetro introduz subdispersão ou sobredispersão.

A DPG dispõe de dois parâmetros,  $\alpha$  e  $\lambda$ , e esse último indica o tipo de variabilidade dos dados, isto é, quando o valor de  $\lambda$  for positivo, nulo ou negativo a variância dessa distribuição será maior, igual ou menor do que a média, respectivamente.

A DPG é particularmente adequada para modelos em que as probabilidades dependem do número de ocorrências anteriores ou quando a ocorrência dos eventos não são homogêneas. Conforme Consul (1989), para o caso particular da teoria das filas, a DPG representa a distribuição de períodos ocupados de um único serviço quando as chegadas dos clientes têm distribuição de Poisson e quando o número de clientes que espera na fila antes do início do serviço também tem distribuição de Poisson.

Este capítulo trará a definição da DPG seguida por breve descrição de dois outros tipos de parametrizações comuns na DPG e, por fim, exemplos de representações gráficas da DPG para diferentes valores dos parâmetros  $\alpha$  e  $\lambda$ .

**Definição 1** *Uma variável aleatória  $X$  que assume valores inteiros não negativos segue uma DPG com parâmetros  $\alpha > 0$  e  $0 \leq \lambda < 1$  se sua função de probabilidade por*

$$P_x(\alpha, \lambda) = \begin{cases} \frac{\alpha(\alpha + x\lambda)^{x-1}e^{-(\alpha+x\lambda)}}{x!} & x = 0, 1, 2, \dots \\ 0 & \text{caso contrário.} \end{cases} \quad (2.1)$$

A média e a variância da DPG fornecem importante relação do parâmetro  $\lambda$  com a dispersão dos dados e são dadas por

$$E(X) = \frac{\alpha}{1 - \lambda} \quad \text{e} \quad (2.2)$$

$$Var(X) = \frac{\alpha}{(1 - \lambda)^3} = \frac{1}{(1 - \lambda)^2} E(X) = \phi E(X) \quad (2.3)$$

O termo  $\phi = \frac{1}{(1-\lambda)^2}$  é o fator de dispersão. Quando  $\lambda = 0$ , ocorre a equidispersão e a DPG será igual à distribuição de Poisson e se  $\lambda > 0$  ocorre a sobredispersão.

Há autores como Consul e Jain (1970, 1973a,b) que apresentam a DPG com o parâmetro  $\lambda$  variando entre -1 e 1. No entanto, quando  $\lambda$  é negativo, a DPG não pode ser utilizada plenamente, pois há casos em que a soma das probabilidades difere de um ( $\sum_{x=0}^{\infty} P_x(X = x) \neq 1$ ), veja Nelson (1975). Para esses casos, é sugerido que se utilize a DPG truncada. Detalhes e referências da DPG para dados com subdispersão ( $\lambda < 0$ ), tal como a DPG truncada, serão apresentados no Apêndice A desta dissertação.

## 2.1 Parametrizações

Alguns autores propuseram diferentes parametrizações para a DPG. A forma como a distribuição é apresentada está relacionada à interpretação dada pelo pesquisador aos parâmetros, bem como à facilidade dos cálculos.

Além da forma mais corrente, apresentada na equação (2.1), há outras duas principais parametrizações:

1. Parametrização para que  $\varphi$  seja igual à média.

Substituindo  $\alpha = \frac{\varphi}{1+\beta\varphi}$  e  $\lambda = \frac{\beta\varphi}{1+\beta\varphi}$  em (2.1), obtém-se

$$P_x(\beta, \lambda) = \frac{(1 + \beta x)^{x-1}}{x!} \left( \frac{\varphi e^{-\beta\varphi(1+\beta\varphi)^{-1}}}{1 + \beta\varphi} \right)^x e^{-\varphi(1+\beta\varphi)^{-1}} \quad x = 0, 1, \dots \quad (2.4)$$

Para  $\varphi > 0$  e  $\beta \geq 0$ .

2. Parametrização para que  $\lambda$  seja diretamente proporcional a  $\alpha$ . O parâmetro  $\alpha$  será a taxa média de ocorrência dos eventos.

Substituindo  $\lambda = \varphi\alpha$  em (2.1), obtém-se

$$P_x(\alpha, \varphi) = \frac{(1 + \varphi x)^{x-1} \alpha^x}{x!} e^{-\alpha - \varphi\alpha x} \quad x = 0, 1, \dots \quad (2.5)$$

para  $0 < \varphi < \frac{1}{\alpha}$  e  $\alpha > 0$ .

## 2.2 Representações gráficas

As Figuras 1 e 2 apresentam as distribuições de probabilidade da DPG. Observa-se que, a medida que  $\alpha$  aumenta, as distribuições dos conjuntos de dados tornam-se mais parecidas com um sino e, à medida que  $\lambda$  aumenta, as distribuições tendem a ficar achatadas e a probabilidade máxima do conjunto de dados será atingida em valores de  $x$  mais altos.

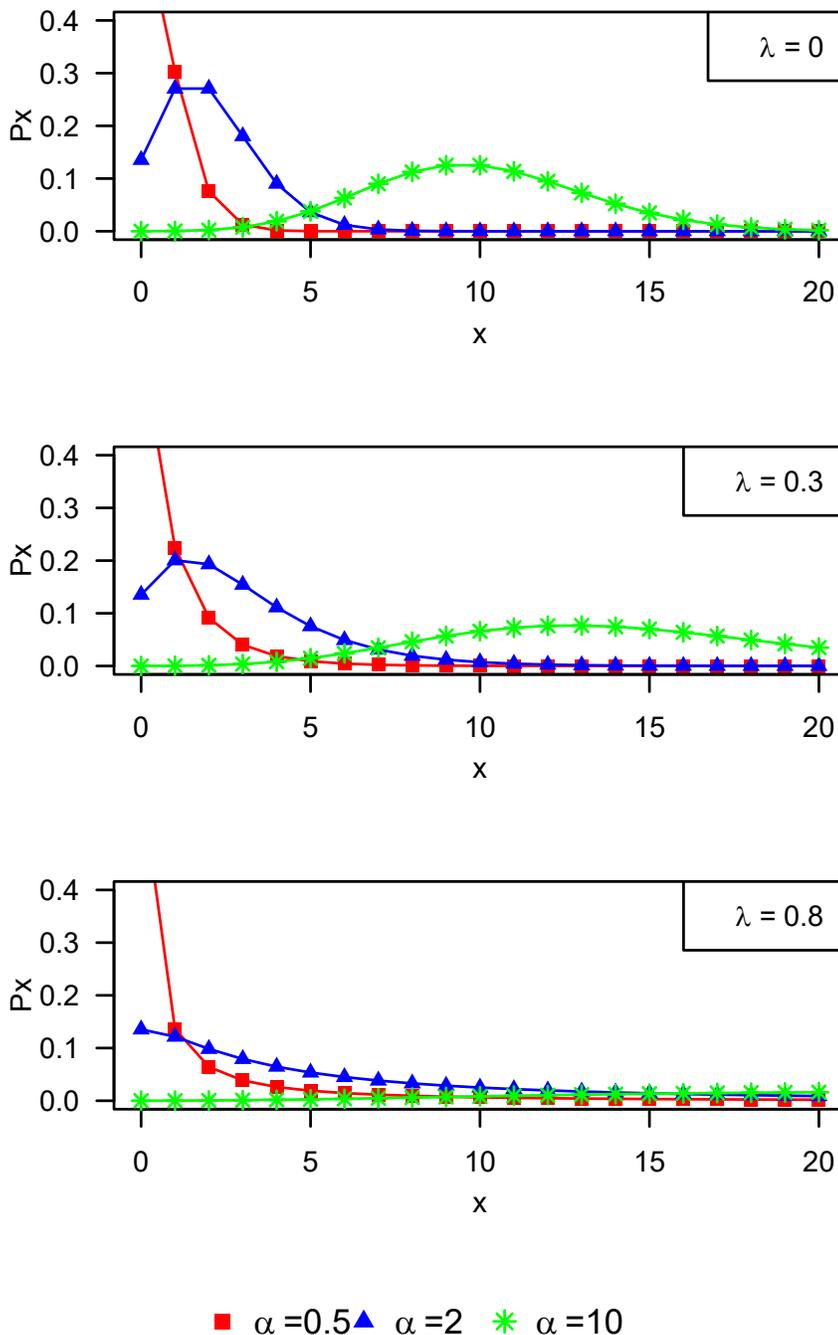


Figura 1 – Funções de probabilidade da DPG com diferentes valores para os parâmetros. Os gráficos se diferenciam pelos valores dos parâmetros  $\alpha$  e  $\lambda$ , a saber:  $\alpha = \{0.5, 2, 10\}$  e  $\lambda = \{0, 0.3, 0.8\}$ . No eixo das ordenadas, estão os valores da função de probabilidade da DPG,  $P_x(\alpha, \lambda)$ , e no eixo das abscissas, os valores de  $x$  variando de zero a vinte. Cada cor representa uma distribuição de probabilidade com diferentes valores de  $\alpha$ , conforme visto nas legendas.

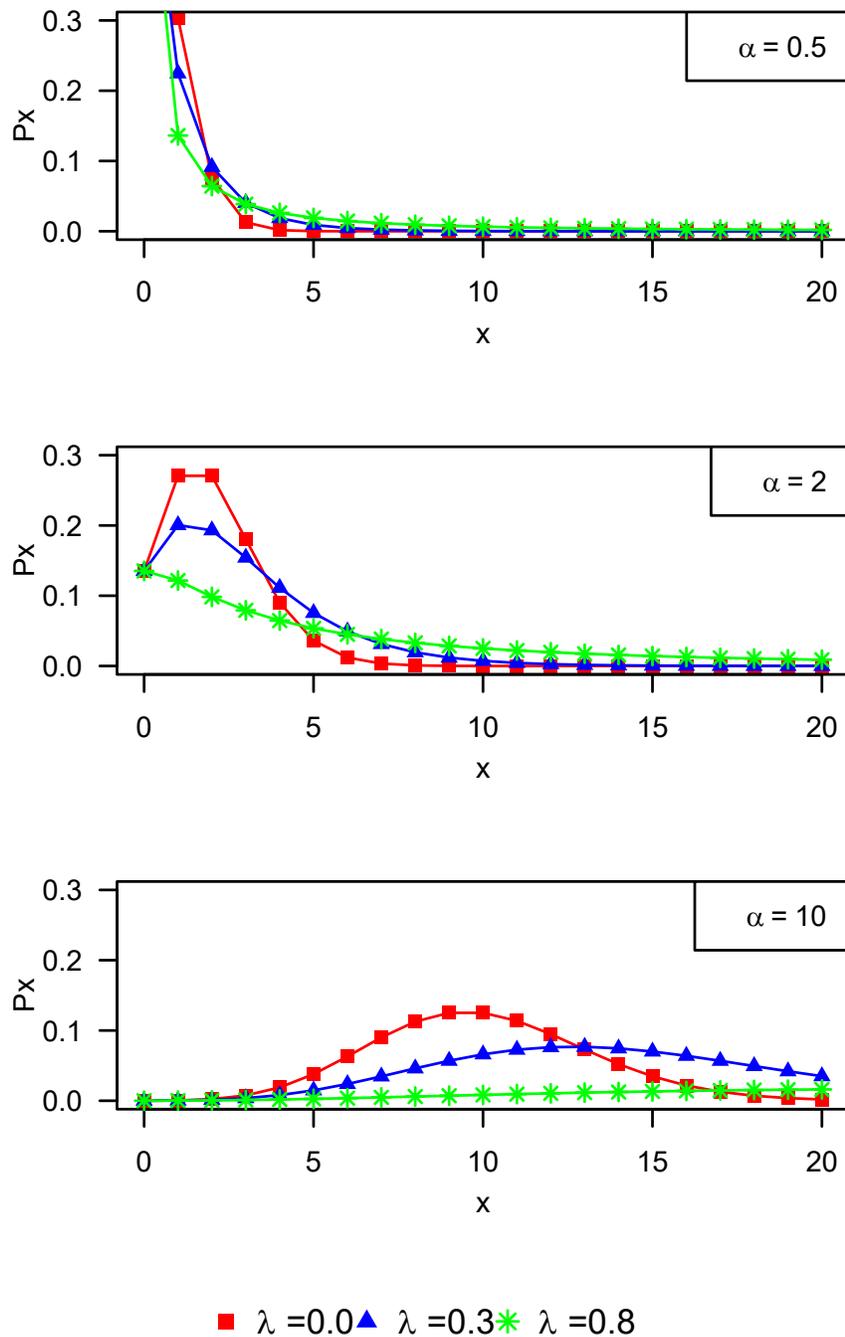


Figura 2 – Funções de probabilidade da DPG com diferentes valores para os parâmetros. Os gráficos se diferenciam pelos valores dos parâmetros  $\alpha$  e  $\lambda$ , a saber:  $\alpha = \{0.5, 2, 10\}$  e  $\lambda = \{0, 0.3, 0.8\}$ . No eixo das ordenadas, estão os valores da função de probabilidade da DPG,  $P_x(\alpha, \lambda)$ , e no eixo das abscissas, os valores de  $x$  variando de zero a vinte. Cada cor representa uma distribuição de probabilidade com diferentes valores de  $\lambda$ , conforme visto nas legendas.

A Figura 3 mostra três distribuições de probabilidade com a mesma média ( $\mu = 4$ ) e diferentes desvios-padrão, a saber  $\sigma = \{10, 4, 2.22\}$ . Na legenda da figura, DP representa desvio-padrão.

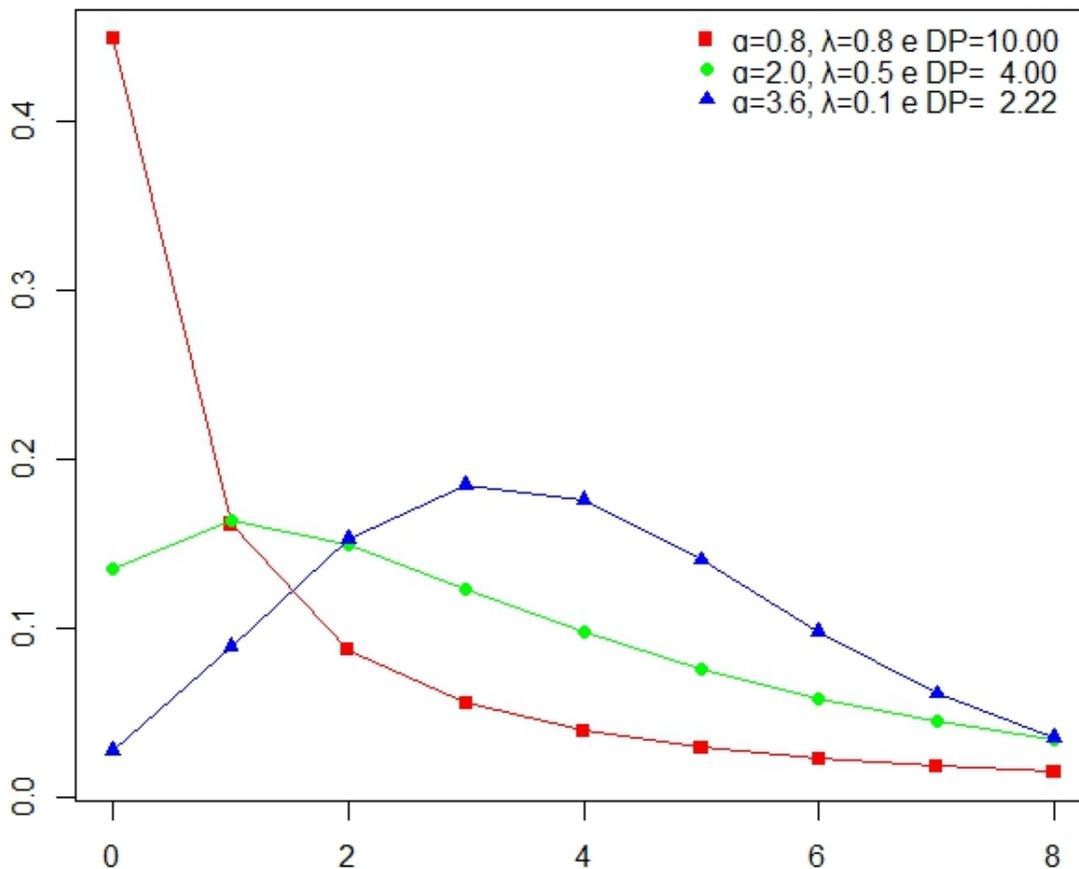


Figura 3 – Funções de probabilidade da DPG com a mesma média ( $\mu = 4$ ) e com diferentes valores para os parâmetros.

Cada cor representa uma distribuição de probabilidade com diferentes valores de  $\alpha$  e  $\lambda$ , conforme visto nas legendas. No eixo das ordenadas, estão os valores da função de probabilidade da DPG,  $P_x(\alpha, \lambda)$ , e no eixo das abscissas, os valores de  $x$  variando de zero a 8.

## 3 Motivação ao estudo da DPG

Consul e Jain (1970, 1973a) em um trabalho pioneiro, apresentaram pela primeira vez a DPG. Desde então, alguns autores, especialmente Famoye, Shoukri, Shenton e também Consul e Jain, se dedicaram ao estudo das propriedades e aplicações da DPG. Atualmente, é possível encontrar várias publicações utilizando a DPG em diferentes áreas de aplicação. No entanto, na língua portuguesa, há poucas referências, sobretudo a respeito das características e particularidades dessa distribuição. Diante disso, esta dissertação tem como propósito fazer uma revisão bibliográfica de conteúdos relevantes que evidenciem as singularidades da DPG e permitam sua compreensão. Ademais, de modo original, neste trabalho será demonstrada a DPG parametrizada em função da  $W$  de Lambert e, como consequência, sua classificação como distribuição de série de potência. Esses resultados simplificam e facilitam alguns cálculos. A seguir serão apresentadas algumas aplicações que motivaram o estudo dessa distribuição.

### 3.1 Aplicações

Um elemento diferenciador da DPG quando comparada a uma distribuição de Poisson é a inserção de um parâmetro, denominado nesta dissertação de  $\lambda$ . A natureza especial desse parâmetro está na possibilidade de se modelar dados discretos que apresentam demasiada dispersão relativa à média. Quando comparada com a distribuição binomial negativa, que também apresenta sobredispersão, a DPG geralmente exibe caudas mais pesadas. Além disso, ela é especialmente apropriada para modelos em que as probabilidades dependem do número de ocorrências anteriores ou quando a ocorrência dos eventos não são homogêneas. Em Consul (1989), a DPG é apresentada como derivada de alguns modelos de urna em que as distribuições de probabilidade dependem do número total de bolas nas urnas. Seguem, abaixo, alguns exemplos de várias áreas de estudo que foram relevantes para a divulgação da DPG.

#### 3.1.1 Áreas de estudo

***Difusão da informação*** – Em um processo de difusão de informação, a hipótese geral é que, a partir de uma fonte, uma ideia ou rumor é difundido para outros membros da população. Denomina-se geração ao conjunto de membros que conheceram o rumor em um período de tempo. O rumor passará de uma geração para outra até ser extinto. O número

de indivíduos que recebe informação em cada geração pode ser modelado por uma DPG. Interpretação análoga é dada para modelos de epidemia.

**Ecologia** – Na biologia, há exemplos de conjunto de dados que sugerem distribuição diferente da de Poisson. Por exemplo, a concentração ou a dispersão de ovos de pragas postos em folhas de plantas indicam comportamento que visa a obter vantagens na natureza. Esse comportamento pode acarretar variabilidade diferente do que se espera para uma distribuição de Poisson. Diante disso, Janardan *et al.* (1979) demonstraram que alguns padrões de comportamento podem ser explicados e descritos por uma DPG.

Consul (1989) apresentou a distribuição de 402 tatuzinhos de jardim deixados sob um conjunto de 122 placas. Conforme visto na Tabela 1, observa-se que em 28 placas não há nenhum inseto, há apenas 1 inseto em outras 28 placas e há 17 insetos em outras 2 placas. A Tabela 2 traz dados referentes a 102 aranhas colocadas sob 240 placas. Verifica-se que não há aracnídeos em 159 placas, há, no entanto, 64 placas com 1 aranha em cada e em outras 4 placas há 3 aranhas em cada. Inicialmente, havia suspeita de que os animais tivessem comportamentos opostos, os tatuzinhos de jardim tendessem ao agrupamento enquanto as aranhas à dispersão.

Tabela 1 – Distribuição de 402 tatuzinhos de jardim em 122 placas.

Quantidade por placa	Obs.	Quantidade por placa	Obs.
0	28	9	3
1	28	10	3
2	14	11	2
3	11	12	0
4	8	13	1
5	11	14	2
6	2	15	1
7	3	16	0
8	3	17	2
Total	N=122	$\bar{x} = 3.2951$	$s^2 = 15.0527$

Fonte: Consul (1989)

Tabela 2 – Distribuição de 102 aranhas em 240 placas.

Quantidade por placa	Obs.	Quantidade por placa	Obs.
0	159	2	13
1	64	3	4
Total	N=240	$\bar{x} = 0.425$	$s^2 = 0.4546$

Fonte: Consul (1989)

**Genética** – Para demonstrar a versatilidade da DPG, Consul (1989) apresentou alguns exemplos da DPG aplicada a dados de genética. Um desses exemplos é a avaliação das

lesões ocorridas nas células de coelhos submetidos a diferentes dosimetria da droga *streptonigrin*. Na Tabela 3, são apresentados os dados observados após as avaliações citogenéticas das células. Verifica-se, por exemplo, que em 500 células que receberam a dose de  $30\mu g/kg$ , 404 não apresentaram alterações nos cromossomos, 80 células apresentaram alteração em 1 cromossomo, 13 células em 2 cromossomos e 3 células em 3 cromossomos. A média e a variância do número de cromossomos danificados aumentam à medida que se eleva a dose da substância. Para qualquer dose, a variância é maior do que a média.

Tabela 3 – Classe de lesões em células de coelhos segundo a dosimetria da droga *streptonigrin*.

Dose ( $\mu g/kg$ )	Classe							Estatísticas
	0	1	2	3	4	5	6	
30	404	80	13	3				N=500 $\bar{x} = 0.230$ $s^2 = 0.27$
60	413	124	42	15	5	0	2	N=601 $\bar{x} = 0.474$ $s^2 = 0.74$
75	200	57	30	7	4	0	2	N=300 $\bar{x} = 0.553$ $s^2 = 0.95$
90	155	83	33	14	11	3	1	N=300 $\bar{x} = 0.853$ $s^2 = 1.37$

Fonte: Consul (1989)

**Processo de fila** – Segundo Consul e Shenton (1973), o número de clientes servidos em qualquer período ocupado de uma fila será uma variável aleatória  $X$  com distribuição de probabilidade Lagrangiana. Em um processo de fila no qual se aplica a DPG,  $X$  é uma variável aleatória que denota o número de clientes que esperam em uma fila antes do início do serviço.

Janardan e Schaeffer (1977) utilizaram a DPG para o estudo do número de aberrações cromossômicas nas células do linfócito humano. A aberração cromossômica é uma mudança no número ou na estrutura dos cromossomos, é produzida por agentes que podem elevar a frequência das mutações e é causada principalmente por radiações ou substâncias cancerígenas. Diariamente, os serem vivos são submetidos a agentes mutagênicos. Contudo, devido à existência de enzimas de reparação, há relativa estabilidade do material genético, pois as enzimas, em sua maioria, detectam e se reparam.

A determinação da dose de radiação absorvida é importante etapa na avaliação dos riscos associados à exposição de indivíduos à radiação. Para isso, tem-se utilizado a contagem de aberrações cromossômicas como bioindicador. Em células que foram submetidas à radiação, os cromossomos em boas condições formam fila no processo de produção, e os cromossomos danificados, uma fila no processo de reparação. Os autores consideraram que os cromossomos ficam dispostos em uma fila, D, onde passarão pelo processo de indução de aberração. Os cromossomos que passam pela fila podem ou não ser aberração cromossômica. No primeiro caso, o cromossomo danificado passa para outra fila, R, onde poderá ou não ser reparado; no segundo caso, o cromossomo passa diretamente para a fila D. Em ambos os casos eles retornam para a fila D. Em cada célula, as aberrações ocorrem de modo indepen-

dente e a uma determinada taxa uniforme. O período de tempo é dito ser constante devido à impossibilidade de um mesmo cromossomo sofrer mais de uma distorção em um período de tempo  $h$ .

Os cromossomos são danificados uma ou mais vezes no processo de produção de aberrações; zero ou mais passam pelo processo de restituição.

Seja  $X$  uma variável aleatória que representa o número líquido de aberrações que aguardam a restituição. A distribuição de probabilidade da variável aleatória  $X$ ,  $P(x)$ , é dada por uma DPG.

Janardan e Schaeffer (1977) demonstraram o bom ajuste da DPG em comparação às distribuições comumente utilizadas, a saber: Distribuição Binomial Negativa e Distribuição de Poisson.

**Modelos de risco** – Os modelos de dados do total dos sinistros é um dos mais importantes para a teoria de risco. Usualmente, quando a quantidade de indenizações apresenta média igual à variância, utiliza-se a distribuição de Poisson e, quando a variância é maior que a média, utiliza-se a distribuição binominal negativa. Consul (1993) comparou a DPG sugerida por Consul e Jain (1973b) com outras distribuições e concluiu que é plausível a utilização dessa distribuição.

### 3.1.2 Modelo de regressão Poisson generalizada

Consul e Famoye (1992) e Consul (1989) demonstraram como as covariáveis podem ser introduzidas em um modelo de regressão cuja variável resposta segue uma DPG. A relação entre a média da variável resposta  $Y$  e o preditor linear do modelo — representado pelas covariáveis  $\mathbf{X}$  e os parâmetros desconhecidos  $\beta$  — é apresentada por meio da função de ligação

$$\log \frac{\alpha_i}{1 - \lambda} = \sum_{r=1}^p x_{ir} \beta_r \quad (3.1)$$

na qual  $x_{ir}$  é a  $i$ -ésima observação da  $r$ -ésima covariável,  $p$  é o número de covariáveis no modelo e  $\beta_r$  é o  $r$ -ésimo parâmetro da regressão.

Em Islam *et al.* (2013) o número de crianças menores de cinco anos com desnutrição em Bangladesh foi ajustado segundo um modelo regressão de Poisson generalizado. Foram utilizadas medidas usualmente relacionadas à desnutrição (altura por idade, peso por altura e peso por idade de crianças menores de cinco anos) a dados socioeconômicos (educação e ocupação dos pais, tipo de fonte de água, local de residência, religião, acesso aos meios de comunicação, índice de riqueza, etc).

---

De acordo com os resultados encontrados em outros estudos daquele país, o modelo de regressão Poisson generalizada foi consistente, demonstrando ser alternativa eficiente como preditor do número de crianças subnutridas abaixo de cinco anos em famílias de Bangladesh.



## 4 Classes de distribuições de probabilidade

Pesquisadores observaram diferentes tipos de dados de contagens e tentaram encontrar padrões específicos em suas observações. Esses esforços levaram à descoberta de novas e mais complexas distribuições de probabilidade, gerando, portanto, um grande número de distribuições de probabilidade discretas. Diante disso, buscaram meios de classificá-las em classes e famílias de acordo com suas formas e características. Alguns autores se dedicaram a essa classificação. As publicações mais relevantes para o contexto desta dissertação foram apresentadas por Consul e Shenton (1972) e Noack (1950). Consul e Shenton descreveram uma classe baseada na expansão de Lagrange, chamada de Classe de Distribuições Lagrangeana (CDL). A partir dessa classe, os autores deduziram várias distribuições novas, com destaque para a DPG, que foi a mais divulgada, tendo merecido a publicação de um livro (CONSUL, 1989) cujo texto é integralmente dedicado ao seu estudo. Já Noack descreveu a Classe de Distribuições de Série de Potência (CDSP).

Devido aos cálculos matemáticos avançados, que não estão no propósito desta dissertação, será apresentada uma breve descrição da expansão de Lagrange e em seguida sua relação com a CDL. Um importante resultado mostrado neste capítulo, que será tratado como derivado da expansão de Lagrange, é a função  $W$  de Lambert, visto em Corless *et al.* (1996).

A  $W$  de Lambert é uma função definida como o inverso da função  $z = we^w$ . Essa função ocorre em muitas áreas e aplicações, como na matemática pura, em que a função  $W$  é útil na resolução de equações diferenciais. Neste trabalho, DPG será escrita em função da  $W$  de Lambert, o que permite relacioná-la com CDSP e obter de forma menos trabalhosa alguns resultados importantes para DPG.

### 4.1 Classe de distribuições Lagrangeanas

#### 4.1.1 Expansão de Lagrange

A expansão de Lagrange ou fórmula de Lagrange tem origem no problema de encontrar a função inversa de uma série de potência, ou seja, a partir de uma série  $\theta(w)$  obtém-se uma série de  $w$  em função de  $\theta$ ,  $w(\theta)$ . Essa expansão também pode ser aplicada para encontrar a função geradora de algumas sequências combinatórias, extrair os coeficientes de uma série de potência formal, calcular somas combinatórias e inverter identidades combinatórias.

**Teorema 4.1 Expansão de Lagrange** - Supondo que uma série de potência formal <sup>1</sup>  $\theta = \theta(w)$  seja definida pela relação  $w = \theta g(w)$ . Na qual  $g(\cdot)$  é uma função analítica <sup>2</sup> que também pode ser escrita na forma de uma série de potência e  $g(\cdot) \neq 0$ . Seja  $f(w)$  uma função analítica, então

i)

$$w = \sum_{k=1}^{\infty} \frac{\theta^k}{k!} \left[ \frac{\partial^{k-1}}{\partial w^{k-1}} (g(w))^k \right]_{w=0} \quad (4.1)$$

ii)

$$f(w) = f(0) + \sum_{k=1}^{\infty} \frac{\theta^k}{k!} \left[ \frac{\partial^{k-1}}{\partial w^{k-1}} (g(w))^x f'(w) \right]_{w=0} \quad (4.2)$$

iii)

$$\frac{f(w)}{1 - wg'(w)/g(w)} = f(0) + \sum_{k=1}^{\infty} \frac{\theta^k}{k!} \left[ \frac{\partial^k}{\partial w^k} (g(w))^x f(w) \right]_{w=0} \quad (4.3)$$

**Prova 1** Considere a equação

$$\theta = w/g(w).$$

A função  $w$  está definida implicitamente como função de  $\theta$ ,  $w(\theta)$ , e pode ser escrita na forma de uma série de Maclaurin. Logo,

$$w(\theta) = \theta g(w) \quad (4.4)$$

$$w(\theta) = \sum_{k \geq 0} \frac{\theta^k}{k!} \left[ \frac{\partial^k w(\theta)}{\partial \theta^k} \right]_{\theta=0}$$

$$w(\theta) = w(0) + \theta \left[ \frac{\partial w(\theta)}{\partial \theta} \right]_{\theta=0} + \frac{\theta^2}{2!} \left[ \frac{\partial^2 w(\theta)}{\partial \theta^2} \right]_{\theta=0} + \frac{\theta^3}{3!} \left[ \frac{\partial^3 w(\theta)}{\partial \theta^3} \right]_{\theta=0} + \dots \quad (4.5)$$

Substituindo cada termo de (4.5) por (4.4) e fazendo  $\theta = 0$ , obtém-se

- $w(0) = \theta g(w)|_{\theta=0} = 0$

- $\theta \left[ \frac{\partial}{\partial \theta} w(\theta) \right]_{\theta=0}$

$$\begin{aligned} &= \theta \left[ \frac{\partial}{\partial \theta} \theta g(w) \right]_{\theta=0} \\ &= \theta \left[ g(w) + \theta \frac{\partial}{\partial w} g(w) \frac{\partial w}{\partial \theta} \right]_{\theta=0} = \theta [g(w)]_{w=0} \end{aligned} \quad (4.6)$$

<sup>1</sup> Série de potência formal é uma generalização de um polinômio cujo número de termos pode ser infinito

<sup>2</sup> Função que pode ser escrita na forma de série de potência

$$\begin{aligned}
& \bullet \frac{\theta^2}{2!} \left[ \frac{\partial^2 w(\theta)}{\partial \theta^2} \right]_{\theta=0} \\
& \quad = \frac{\theta^2}{2!} \left[ \frac{\partial^2 \theta g(w)}{\partial \theta^2} \right]_{\theta=0} \\
& \quad = \frac{\theta^2}{2!} \left[ \frac{\partial g(w)}{\partial w} \frac{\partial w}{\partial \theta} + \frac{\partial g(w)}{\partial w} \frac{\partial w}{\partial \theta} + \theta \frac{\partial^2 g(w)}{\partial w^2} \left( \frac{\partial w}{\partial \theta} \right)^2 + \theta \frac{\partial g(w)}{\partial w} \left( \frac{\partial^2 w}{\partial \theta^2} \right) \right]_{\theta=0}
\end{aligned}$$

Diferenciando a equação (4.4) em relação a  $\theta$ , tem-se

$$\begin{aligned}
\frac{\partial w}{\partial \theta} &= g(w) + \theta \frac{\partial}{\partial w} g(w) \frac{\partial w}{\partial \theta} \\
\frac{\partial w}{\partial \theta} &= \frac{g(w)}{1 - \theta \frac{\partial g(w)}{\partial w}} \\
\frac{\partial w}{\partial \theta} \Big|_{\theta=0} &= g(w) \Big|_{w=0}
\end{aligned}$$

Portanto,

$$\frac{\theta^2}{2!} \left[ \frac{\partial^2 w(\theta)}{\partial \theta^2} \right]_{\theta=0} = \frac{\theta^2}{2!} \left[ 2g(w) \frac{\partial g(w)}{\partial w} \right]_{w=0} = \frac{\theta^2}{2!} \left[ \frac{\partial}{\partial w} g(w)^2 \right]_{w=0} \quad (4.7)$$

• Analogamente,

$$\begin{aligned}
& \frac{\theta^3}{3!} \left[ \frac{\partial^3 w(\theta)}{\partial \theta^3} \right]_{\theta=0} \\
& = \frac{\theta^3}{3!} \left[ \frac{\partial^3 \theta g(w)}{\partial \theta^3} \right]_{\theta=0} = \frac{\theta^3}{3!} \left[ \frac{\partial^2 g(w)^3}{\partial w^2} \right]_{w=0} \quad (4.8)
\end{aligned}$$

• Após sucessivas transformações, é possível obter um termo geral:

$$\begin{aligned}
& \frac{\theta^k}{k!} \left[ \frac{\partial^k w(\theta)}{\partial \theta^k} \right]_{\theta=0} \\
& = \frac{\theta^k}{k!} \left[ \frac{\partial^k \theta g(w)}{\partial \theta^k} \right]_{\theta=0} = \frac{\theta^k}{k!} \left[ \frac{\partial^{k-1}}{\partial w^{k-1}} g(w)^k \right]_{w=0} \quad (4.9)
\end{aligned}$$

Portanto,

$$w(\theta) = \sum_{k \geq 1} \frac{\theta^k}{k!} \left[ \frac{\partial^{k-1}}{\partial w^{k-1}} g(w)^k \right]_{w=0} \quad (4.10)$$

As demonstrações das equações (4.2) e (4.3) seguem os mesmos procedimentos descritos acima e podem ser vistos em Battin (1999).

**Exemplo 3.1:** Considere a função  $W$  de Lambert que é definida pela relação

$$\theta(w) = we^w.$$

Deseja-se obter  $w(\theta)$ . Suponha que  $w \geq \frac{-1}{e}$  e que  $\theta$  pertence ao conjunto dos números reais. Conforme a expansão de Lagrange, a partir da relação  $\theta = w/g(w)$ , obtém-se  $w$  escrita em função de  $\theta$  por meio da equação (4.1). Desse modo, para o exemplo dado,  $g(w) = e^{-w}$  e, portanto,  $w(\theta)$  será

$$\begin{aligned}
w(\theta) &= \sum_{k=1}^{\infty} \frac{\theta^k}{k!} \left[ \frac{\partial^{k-1}}{\partial w^{k-1}} (g(w))^k \right]_{w=0} \\
&= \sum_{k=1}^{\infty} \frac{\theta^k}{k!} \left[ \frac{\partial^{k-1}}{\partial w^{k-1}} e^{-wk} \right]_{w=0} \\
&= \theta - \frac{2\theta^2}{2!} + \frac{9\theta^3}{3!} - \frac{64\theta^4}{4!} + \dots + \frac{(-n)^{n-1}\theta^n}{n!} + \dots \\
&= \sum_{n=1}^{\infty} \frac{(-n)^{n-1}\theta^n}{n!}.
\end{aligned} \tag{4.11}$$

A série (4.11) converge para  $|\theta| < \frac{1}{e}$ .

Além disso, pode-se obter uma função  $f(w)$  escrita em termos de  $\theta$  por meio da equação (4.2), conforme apresentado abaixo,

$$\begin{aligned}
\theta &= we^w \\
\frac{w}{\theta} &= e^{-w} \\
\left(\frac{w}{\theta}\right)^b &= e^{-bw}, \quad b \in R.
\end{aligned} \tag{4.12}$$

Fazendo-se  $f(w) = e^{-bw}$  e utilizando-se a equação (4.1), obtém-se

$$\begin{aligned}
e^{-bw(\theta)} &= f(w) = f(0) + \sum_{k=1}^{\infty} \frac{\theta^k}{k!} \left[ \frac{\partial^{k-1}}{\partial w^{k-1}} (g(w))^k f'(w) \right]_{w=0} \\
&= 1 + \sum_{k=1}^{\infty} \frac{\theta^k}{k!} \left[ \frac{\partial^{k-1}}{\partial w^{k-1}} e^{-wk} (-be^{-bw}) \right]_{w=0} \\
&= 1 - b \sum_{k=1}^{\infty} \frac{\theta^k}{k!} \left[ \frac{\partial^{k-1}}{\partial w^{k-1}} e^{-w(k+b)} \right]_{w=0} \\
&= 1 - b\theta + \frac{b(2+b)\theta^2}{2!} - \frac{b(3+b)^2\theta^3}{3!} + \frac{b(4+b)^3\theta^4}{4!} + \dots + \frac{b(n+b)^{n-1}(-\theta)^n}{n!} + \dots \\
&= \sum_{n=0}^{\infty} \frac{b(n+b)^{n-1}(-\theta)^n}{n!}.
\end{aligned} \tag{4.13}$$

A série converge para  $|\theta| < \frac{1}{e}$ .

Já a relação apresentada na equação (4.3) será

$$\begin{aligned} & \frac{f(w)}{1 - w[g'(w)/g(w)]} \\ &= \frac{e^{-bw}}{1 - w((-e^{-w})/e^{-w})} \\ &= \frac{e^{-bw}}{1 + w}. \end{aligned}$$

Portanto,

$$\begin{aligned} \frac{e^{-bw}}{1 + w} &= f(0) + \sum_{k=1}^{\infty} \frac{\theta^k}{k!} \left[ \frac{\partial^k}{\partial w^k} (g(w))^k f(w) \right]_{w=0} \\ &= 1 + \sum_{k=1}^{\infty} \frac{\theta^k}{k!} \left[ \frac{\partial^k}{\partial w^k} e^{-wk} (e^{-bw}) \right]_{w=0} \\ &= 1 + \sum_{k=1}^{\infty} \frac{\theta^k}{k!} \left[ \frac{\partial^k}{\partial w^k} e^{-w(k+b)} \right]_{w=0} \\ &= 1 - (b+1)\theta + \frac{(2+b)^2\theta^2}{2!} - \frac{(3+b)^3\theta^3}{3!} + \frac{(4+b)^4\theta^4}{4!} + \dots + \frac{(n+b)^n(-\theta)^n}{n!} + \dots \\ &= \sum_{n=0}^{\infty} \frac{(n+b)^n(-\theta)^n}{n!}. \end{aligned} \tag{4.14}$$

A série converge para  $|\theta| < \frac{1}{e}$ .

#### 4.1.2 Distribuições de Lagrangeanas

A CDL tem como origem a transformação de Lagrange e é dividida em três subclasses:

- Distribuição Lagrangeana básica;
- Distribuição Lagrangeana delta; e
- Distribuição Lagrangeana generalizada.

A seguir serão apresentadas a definição de uma fgp e a forma como ela se relaciona com a expansão de Lagrange para produzir as três subclasses de distribuição Lagrangeana. Conforme será visto à frente, no contexto da DPG, a subclasse mais importante é a distribuição Lagrangeana generalizada.

#### 4.1.2.1 Função geradora de probabilidade

Seja  $X$  uma variável aleatória discreta definida em um conjunto de valores inteiros não negativos e com função de probabilidade  $P_x$ , em que  $\sum_{x=0}^{\infty} P_x = 1$ , a sua fgp será dada por

$$h(t) = E(t^X) = \sum_{x=0}^{\infty} t^x P_x \quad (4.15)$$

na qual  $h(0) = P_0$  e  $h(1) = 1$ . A série converge se  $|t| \leq 1$ .

#### 4.1.2.2 Distribuição Lagrangeana básica

A partir da expansão de Lagrange dada na equação (4.1), a transformação de  $w = \theta g(w)$  gera uma fgp  $w = h(\theta)$ .

$$w = h(\theta) = \sum_{x=1}^{\infty} \frac{\theta^x}{x!} \left[ \frac{\partial^{x-1}}{\partial w^{x-1}} (g(w))^x \right]_{w=0} \quad (4.16)$$

na qual  $\frac{\partial^{x-1}}{\partial w^{x-1}} (g(w))^x|_{w=0} \geq 0$  para todos os valores de  $x$ .

Logo, a distribuição de probabilidade Lagrangeana básica será

$$P(X = x) = \frac{1}{x!} \left[ \frac{\partial^{x-1}}{\partial w^{x-1}} (g(w))^x \right]_{w=0}, \quad x \in N. \quad (4.17)$$

A função  $g(w)$  deve ser sucessivamente diferenciável,  $g(1) = 1$  e  $g(0) \neq 0$ . Não é necessário que  $g(w)$  seja uma fgp.

#### 4.1.2.3 Distribuição Lagrangeana delta

A partir da expansão de Lagrange dada na equação (4.1), a transformação de  $w = \theta g(w)$  gera uma fgp em que  $w^n = [h(\theta)]^n$ .

$$w^n = [h(\theta)]^n = \sum_{x=n}^{\infty} \frac{n\theta^x}{(x-n)!x!} \left[ \frac{\partial^{x-n}}{\partial z^{x-n}} (g(w))^x \right]_{w=0} \quad (4.18)$$

na qual  $\frac{\partial^i}{\partial w^i} (g(w))^{n+i}|_{w=0} \geq 0$  para  $i = 0, 1, 2, \dots$

Logo, a distribuição de probabilidade Lagrangeana delta será

$$P(X = x) = \frac{n}{(x-n)x!} \left[ \frac{\partial^{x-n}}{\partial w^{x-n}} (g(w))^x \right]_{w=0}, \quad x = n, n+1, n+2, \dots \quad (4.19)$$

A função  $g(w)$  é sucessivamente diferenciáveis,  $g(1) = 1$  e  $g(0) \neq 0$ . Não é necessário que  $g(w)$  seja uma fgp.

## 4.1.2.4 Distribuição Lagrangeana generalizada

A partir da expansão de Lagrange dada na equação (4.2), a transformação de  $w = \theta g(w)$  gera a fgp  $f(w)$  em que

$$f(w) = f(h(\theta))^n = \sum_{x=n}^{\infty} \frac{\theta^x}{x!} \left[ \frac{\partial^{x-1}}{\partial w^{x-1}} (g(w))^x f'(w) \right]_{w=0} \quad (4.20)$$

na qual  $\frac{\partial^{x-1}}{\partial w^{x-1}} (g(w))^x f'(w)|_{w=0} \geq 0$  para  $x \in N$ .

Logo, a distribuição de probabilidade Lagrangeana generalizada será

$$\begin{cases} P(X = 0) = f(0), & x = 0 \\ P(X = x) = \frac{1}{x!} \frac{\partial^{x-1}}{\partial z^{x-1}} (g(w))^x f'(w)|_{w=0} & x = 1, 2, 3, \dots \end{cases} \quad (4.21)$$

em que  $f(w)$  e  $g(w)$  sejam funções sucessivamente diferenciáveis em  $w$ ,  $g(1) = f(1) = 1$  e  $g(0) \neq 0$ .

A partir da escolha das funções  $f(w)$  e  $g(w)$  é possível obter grande quantidade de distribuições de probabilidade discretas, inclusive a DPG, conforme demonstrado a seguir.

Sejam  $f(w) = e^{\alpha(w-1)}$  e  $g(w) = e^{\lambda(w-1)}$ , em que  $\alpha > 0$  e  $0 < \lambda < 1$ , por meio da equação (4.21), fazendo as devidas substituições, obtém-se a DPG.

$$\begin{aligned} P(X = 0) &= f(0) = e^{-\alpha}, \quad \text{para } x = 0 \\ P(X = x) &= \frac{1}{x!} \frac{\partial^{x-1}}{\partial w^{x-1}} (g(w))^x f'(w)|_{w=0} \\ &= \frac{1}{x!} \frac{\partial^{x-1}}{\partial w^{x-1}} e^{x\lambda(w-1)} \alpha e^{\alpha(w-1)}|_{w=0} \\ &= \frac{\alpha e^{-\alpha-x\lambda}}{x!} \frac{d^{x-1}}{dw^{x-1}} e^{(\alpha+x\lambda)w}|_{w=0} \\ &= \frac{\alpha e^{-\alpha-x\lambda}}{x!} (\alpha + x\lambda)^{x-1}, \quad \text{para } x = 1, 2, 3, \dots \end{aligned} \quad (4.22)$$

4.2 *W* de Lambert

A função *W* de Lambert é definida como o inverso da função  $f(w) = we^w$ , ou seja, é a solução da equação

$$W(\theta)e^{W(\theta)} = \theta, \quad (4.23)$$

na qual  $W(\theta) = w$  e  $\theta \in R$ .

Conforme observado na Figura 4, as funções  $f(w)$  e  $W(\theta)$  não são injetivas. Por esse motivo, convencionou-se adotar o intervalo principal em que  $W(\theta) \in (-1, 1)$  por  $W_0(\theta)$  e o intervalo secundário em que  $W(\theta) \in (-\infty, -1)$  por  $W_{-1}(\theta)$ .

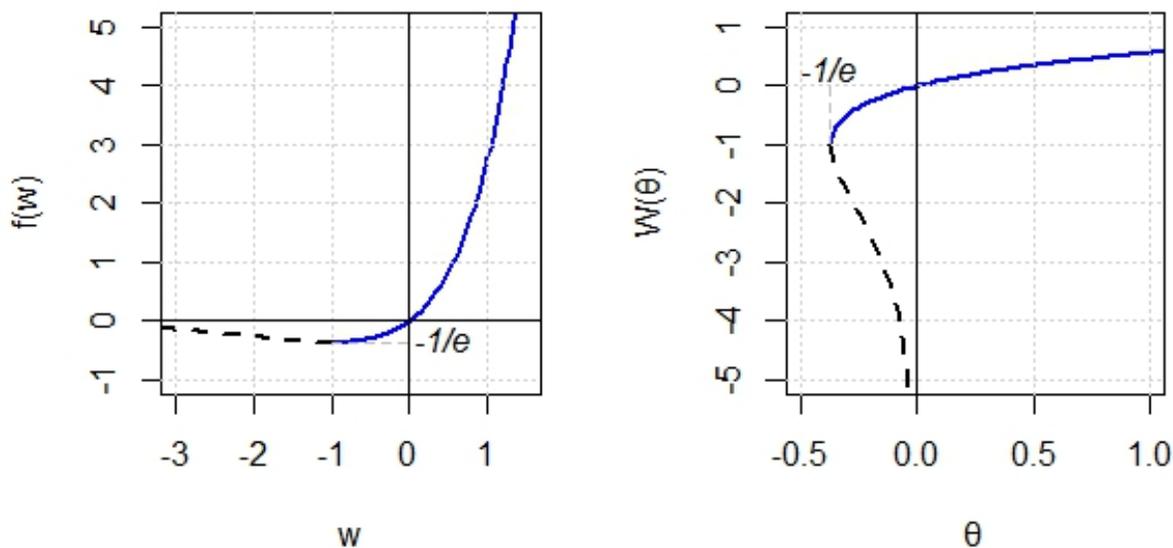


Figura 4 – Função  $W$  de Lambert.

A linha pontilhada é a representação de  $W_{-1}(\theta)$ , e a linha contínua de  $W_0(\theta)$ .

Alguns valores da função  $W(\theta)$  são fáceis de serem obtidos, tais como:  $W_0(-1/e) = -1$ ,  $W_0(0) = 0$ ,  $W_0(-\ln(2)/2) = -\ln(2)$  e  $W_0(e) = 1$ . Para os demais valores, utilizam-se métodos como de Newton Raphson, que buscam aproximações para obter as raízes de  $g(w) = we^w - \theta$ . Os resultados das equações (4.11), (4.13) e (4.14), demonstrados no exemplo 3.1, são formas alternativas para obter  $W_0(\theta)$ .

Muitos problemas que envolvem exponencial podem ser resolvidos com a função  $W$  de Lambert. A estratégia geral é mover os termos conhecidos para o lado direito da equação (4.23) e obter um valor para  $\theta$ . O lado esquerdo deve ter a forma da função  $we^w$ , logo, utilizando o procedimento adequado, é possível obter  $W_0(\theta)$ .

Importantes resultados e aplicações podem ser vistos em Corless *et al.* (1993). Neste trabalho, a função  $W$  de Lambert é utilizada para simplificar e resolver alguns cálculos referentes à DPG que, por vezes, podem ser dispendiosos.

### 4.3 Classe de Distribuições de Série de Potências

A classe de distribuições de séries de potência é uma forma generalizada de representar distribuições de probabilidades discretas. Além disso, permite obter a função geradora de momentos, relações de recorrência para os momentos e estimação dos parâmetros por meio de alguns resultados e propriedades. Inclui grande quantidade de distribuições, como binomial, binomial negativa e Poisson.

**Definição 2** *Uma variável aleatória com valores inteiros positivos segue uma distribuição da CDSP com parâmetro  $\theta$  se*

$$P(X = x) = \frac{a(x)\theta^x}{g(\theta)}, \quad x \in T \subset N \quad e \quad g(\theta) = \sum_{x \in T} a(x)\theta^x, \quad (4.24)$$

sendo  $g(\theta)$  uma função positiva, finita, diferenciável em  $\theta$ , podendo ser escrita como uma série de potência,  $a(x) \geq 0$  e  $\theta \in ]0, p]$ , em que  $p$  é o raio de convergência para a série  $g(\theta)$ .

Noack (1950) estabeleceu uma distribuição de série de potência em que a variável aleatória  $X$  assume valores inteiros não negativos ( $x \in N$ ). Já Patil (1962), estabeleceu a série de potência generalizada, em que a variável aleatória  $X$  assume valores inteiros não negativos em um conjunto  $T$ , sendo  $T$  subconjunto arbitrário não vazio de  $N$ . A restrição de  $x$  a um subconjunto dos números naturais é especialmente útil em situações em que o parâmetro  $\lambda$  da DPG é negativo e seja preciso utilizar a DPG truncada. Segundo Patil (1962), as propriedades da classe de série de potência podem ser deduzidas para a classe de série de potência generalizada.

#### 4.3.1 Relações de recorrência para os momentos

##### Média e Variância

Seja  $g(\theta) = \sum_{x \in T} a(x)\theta^x$  e seja  $X$  uma variável aleatória discreta que pertence à CDSP, então sua média e sua variância serão

$$\begin{aligned} \mu &= E(X) = \sum_{x \in T} xP_x \\ &= \sum_{x \in T} x \frac{a(x)\theta^x}{g(\theta)} \\ &= \frac{\theta}{g(\theta)} \sum_{x \in T} xa(x)\theta^{x-1} \\ &= \frac{\theta g'(\theta)}{g(\theta)} \end{aligned} \quad (4.25)$$

e

$$\sigma^2 = E(X)^2 - [E(X)]^2 = EX(X-1) + EX - [E(X)]^2$$

na qual

$$\begin{aligned}
 EX(X-1) &= \sum_{x \in T} x(x-1)P_x \\
 &= \sum_{x \in T} x(x-1) \frac{a(x)\theta^x}{g(\theta)} \\
 &= \frac{\theta^2}{g(\theta)} \sum_{x \in T} x(x-1)a(x)\theta^{x-2} \\
 &= \frac{\theta^2 g''(\theta)}{g(\theta)}.
 \end{aligned}$$

Logo,

$$\begin{aligned}
 \sigma^2 &= EX(X-1) + EX - [E(X)]^2 \\
 &= \frac{\theta^2 g''(\theta)}{g(\theta)} + \frac{\theta g'(\theta)}{g(\theta)} - \left[ \frac{\theta g'(\theta)}{g(\theta)} \right]^2.
 \end{aligned} \tag{4.26}$$

## Momentos

Seja  $X$  uma variável aleatória discreta e  $r$  um inteiro positivo, o  $r$ -ésimo momento de  $X$ , denotado por  $m_r$ , será dado por

$$m_r = E(X^r) = \sum_x x^r P_x(x).$$

O  $r$ -ésimo momento de uma CDSP é obtido de forma recursiva, conforme demonstração abaixo.

$$m_r = E(X^r) = \sum_x x^r \frac{a(x)\theta^x}{g(\theta)}$$

$$\frac{\partial m_r}{\partial \theta} = \sum_x x^r \frac{a(x)x\theta^{x-1}}{g(\theta)} - \sum_x x^r \frac{a(x)\theta^x g'(\theta)}{g(\theta)^2}$$

Multiplicando a equação anterior por  $\theta$

$$\begin{aligned}
 \theta \frac{\partial m_r}{\partial \theta} &= \theta \sum_x x^{r+1} \frac{a(x)\theta^{x-1}}{g(\theta)} - \theta \sum_x x^r \frac{a(x)\theta^x g'(\theta)}{g(\theta)^2} \\
 \theta \frac{\partial m_r}{\partial \theta} &= \sum_x x^{r+1} \frac{a(x)\theta^x}{g(\theta)} - \frac{\theta g'(\theta)}{g(\theta)} \sum_x x^r \frac{a(x)\theta^x}{g(\theta)} \\
 \theta \frac{\partial m_r}{\partial \theta} &= m_{r+1} - \frac{\theta g'(\theta)}{g(\theta)} m_r.
 \end{aligned}$$

Sendo  $m_0 = 1$  e  $m_1 = \frac{\theta g'(\theta)}{g(\theta)}$ , obtém-se

$$m_{r+1} = \theta \frac{\partial m_r}{\partial \theta} + m_1 m_r. \quad (4.27)$$

### Momentos centrais

Seja  $X$  uma variável aleatória discreta e  $r$  um inteiro positivo, o  $r$ -ésimo momento central de  $X$ , denotado por  $M_r$ , será dado por

$$M_r = E(X - m_1)^r = \sum_x (X - m_1)^r P_x(x). \quad (4.28)$$

O  $r$ -ésimo momento central de uma CDSP é obtido de forma recursiva, conforme demonstrado abaixo.

$$M_r = E(X - m_1)^r = \sum_x \frac{(x - m_1)^r a(x) \theta^x}{g(\theta)}.$$

Diferenciando  $M_r$  em relação  $\theta$ ,

$$\begin{aligned} \frac{\partial M_r}{\partial \theta} &= - \sum_x r(x - m_1)^{r-1} \frac{\partial m_1}{\partial \theta} \frac{a(x) \theta^x}{g(\theta)} \\ &\quad + \sum_x (x - m_1)^r a(x) x \theta^{x-1} g(\theta) - \sum_x \frac{(x - m_1)^r a(x) \theta^x g'(\theta)}{g(\theta)^2}. \end{aligned}$$

Multiplicando a equação anterior por  $\theta$ ,

$$\begin{aligned} \theta \frac{\partial M_r}{\partial \theta} &= -r \frac{\partial m_1}{\partial \theta} \theta \sum_x (x - m_1)^{r-1} \frac{a(x) \theta^x}{g(\theta)} \\ &\quad + \sum_x (x - m_1)^r \frac{a(x) x \theta^x}{g(\theta)} - \sum_x \frac{g'(\theta) \theta (x - m_1)^r a(x) \theta^x}{g(\theta)} \\ &= -r \frac{\partial m_1}{\partial \theta} \theta M_{r-1} + M_{r+1}. \end{aligned}$$

Portanto,

$$M_{r+1} = \theta \frac{\partial M_r}{\partial \theta} + r \frac{\partial m_1}{\partial \theta} M_{r-1},$$

sendo  $M_0 = 1$ ,  $M_1 = 0$  e  $M_2 = \theta \frac{\partial m_1}{\partial \theta}$ , a relação de recorrência para momento central será

$$M_{r+1} = \theta \frac{\partial M_r}{\partial \theta} + r M_2 M_{r-1}. \quad (4.29)$$

## Momentos fatoriais

Seja  $X$  uma variável aleatória discreta e  $r$  um inteiro positivo, o  $r$ -ésimo momento fatorial de  $X$ , denotado por  $M_{[r]}$ , será dado por

$$E(X^{[r]}) = M_{[r]} = E(X(X-1)\dots(X-r+1)) = \sum_x x(x-1)\dots(x-r+1)P(X=x).$$

O  $r$ -ésimo momento fatorial de uma CDSP é obtido de forma recursiva. Segue abaixo a demonstração desse resultado.

$$E(X^{[r]}) = \frac{1}{g(\theta)} \sum_x x^{[r]} a_x \theta^x.$$

Diferenciando a equação acima em relação a  $\theta$ , obtém-se

$$\frac{\partial M_{[r]}}{\partial \theta} = \frac{1}{g(\theta)} \sum_x x^{[r]} a_x x \theta^{x-1} - \frac{g'(\theta)}{g(\theta)^2} \sum_x x^{[r]} a_x \theta^x.$$

Fazendo  $x^{[r]}x = x^{[r+1]} + rx^{[r]}$ .

$$\frac{\partial M_{[r]}}{\partial \theta} = \frac{1}{g(\theta)} \sum_x (x^{[r+1]} + rx^{[r]}) a_x \theta^{x-1} - \frac{g'(\theta)}{g(\theta)^2} \sum_x x^{[r]} a_x \theta^x.$$

Simplificando a equação, obtém-se

$$\frac{\partial M_{[r]}}{\partial \theta} = \frac{1}{\theta} M_{[r+1]} + r \left[ \frac{1}{\theta} - \frac{g'(\theta)}{g(\theta)} \right] M_{[r]}.$$

Dessa forma, a relação de recorrência entre os momentos fatoriais é dada por

$$M_{[r+1]} = \theta \frac{\partial M_{[r]}}{\partial \theta} + M_{[r]} M_{[1]} - r M_{[r]}, \quad (4.30)$$

para  $r = 1, 2, 3, \dots$ .

### 4.3.2 Função característica

Seja  $X$  uma variável aleatória, sua função característica por meio da distribuição de série de potência será dada por

$$\begin{aligned} \psi_X(t) &= \sum_x e^{itx} P_x = \frac{1}{g(\theta)} \sum_x e^{itx} a_x \theta^x \\ &= \frac{1}{g(\theta)} \sum_x a_x (e^{it\theta})^x \\ &= \frac{g(e^{it\theta})}{g(\theta)}. \end{aligned} \quad (4.31)$$

## 4.4 DPG como um elemento da classe de distribuições de série de potência

Quando a DPG é escrita em função da  $W$  de Lambert, é possível classificá-la na CDSP e utilizar os resultados decorrentes dessa classe de distribuições.

Considere a parametrização vista na equação (2.1). Substituindo  $b = \alpha/\lambda$ ,  $\theta = \lambda e^{-\lambda}$  e  $W_0(-\theta) = -\lambda$ , a menos de uma constante de normalização ( $e^{-\alpha}$ ), obtém-se os termos da expansão em série de potência de  $e^{-bW_0(-\theta)}$  apresentada na equação (4.13).

De fato, temos de (2.1)

$$\begin{aligned} P_x(\alpha, \lambda) &= \frac{\alpha(\alpha + x\lambda)^{x-1} e^{-(\alpha+x\lambda)}}{x!} \\ &= e^{-\alpha} \frac{\alpha}{\lambda} \left(\frac{\alpha}{\lambda} + x\right)^{x-1} \frac{(\lambda e^{-\lambda})^x}{x!} \end{aligned} \quad (4.32)$$

e, com as substituições descritas acima, tem-se

$$P_x(b, \theta) = e^{bW_0(-\theta)} b(b+x)^{x-1} \frac{(\theta)^x}{x!}. \quad (4.33)$$

A série da equação (4.13) converge para  $|\theta| < \frac{1}{e}$ . Sendo  $-\theta = \lambda e^{-\lambda}$ , a série (4.32) converge para  $|\lambda e^{-\lambda}| < \frac{1}{e}$  e conseqüentemente será convergente para  $|\lambda| < 1$ . Este resultado condiz com a restrição de  $\lambda$  de uma DPG.

Diante desses resultados, a DPG reparametrizada em função da  $W$  de Lambert será

$$P_x(b, \theta) = \begin{cases} e^{bW_0(-\theta)} \frac{b(b+x)^{x-1} \theta^x}{x!} & x = 0, 1, 2, \dots \\ 0 & \text{caso contrário,} \end{cases} \quad (4.34)$$

para  $b > 0$  e  $0 < \theta < \frac{1}{e}$ .

A equação acima pertence à CDSP, sendo  $g(\theta) = e^{-bW_0(-\theta)}$  e  $a(x) = \frac{b(b+x)^{x-1}}{x!}$ .

Nas próximas seções estão demonstradas características e propriedades da DPG escrita em função da  $W$  de Lambert e em termos de  $\alpha$  e  $\lambda$  (parametrização apresentada em (2.1)). Para isso deve-se considerar os resultados abaixo.

- Diferenciação da função  $W$  de Lambert em relação a  $\theta$ .

Tem-se

$$-\theta = W_0(-\theta)e^{W_0(-\theta)}.$$

Derivando em relação a  $\theta$ , obtém-se

$$\begin{aligned} 1 &= \frac{\partial}{\partial \theta} \theta = W_0'(-\theta)e^{W_0(-\theta)} + W_0'(-\theta)W_0(-\theta)e^{W_0(-\theta)} \\ &= W_0'(-\theta)e^{W_0(-\theta)} [1 + W_0(-\theta)]. \end{aligned}$$

Substituindo  $e^{W_0(-\theta)}$  por  $\frac{-\theta}{W_0(-\theta)}$  e isolando  $W_0'(-\theta)$ , obtém-se

$$W_0'(-\theta) = -\frac{W_0(-\theta)}{\theta(1 + W_0(-\theta))}. \quad (4.35)$$

- Diferenciação da função  $g(\theta)$  em relação a  $\theta$ .

Tem-se

$$g(\theta) = e^{-bW_0(-\theta)}.$$

Logo,

$$g'(\theta) = be^{-bW_0(-\theta)}W'(-\theta).$$

Pela igualdade da equação (4.35), obtém-se

$$g'(\theta) = -e^{-bW_0(-\theta)} \left[ b \frac{W_0(-\theta)}{\theta(1 + W_0(-\theta))} \right]. \quad (4.36)$$

Analogamente,

$$g''(\theta) = bW_0^2(-\theta)e^{-bW_0(-\theta)} \frac{bW_0(-\theta) + W_0(-\theta) + b + 2}{[1 + W_0(-\theta)]^3\theta^2}. \quad (4.37)$$

- Diferenciação de  $\alpha = -bW_0(-\theta)$  em relação a  $\theta$ .

Tem-se

$$\alpha = -bW_0(-\theta).$$

Logo,

$$\frac{\partial \alpha}{\partial \theta} = bW_0'(-\theta).$$

Pela igualdade da equação (4.35), obtém-se

$$\frac{\partial \alpha}{\partial \theta} = -b \frac{W_0(-\theta)}{\theta(1 + W_0(\theta))}. \quad (4.38)$$

- Diferenciação de  $\lambda = -W_0(-\theta)$  em relação a  $\theta$ . Tem-se

$$\lambda = -W_0(-\theta).$$

Logo,

$$\frac{\partial \lambda}{\partial \theta} = W_0'(-\theta).$$

Pela igualdade da equação (4.35), obtém-se

$$\frac{\partial \lambda}{\partial \theta} = -\frac{W_0(-\theta)}{\theta(1 + W_0(\theta))}. \quad (4.39)$$

## 4.5 Propriedades e características da DPG

### 4.5.1 Soma das probabilidades da DPG

De acordo com Consul (1989), é difícil demonstrar que a soma de todas as probabilidades seja igual a 1. Diante disso, Lerner *et al.* (1997) fizeram essa demonstração utilizando funções analíticas, já Tuentner (2000) utilizou a equação de Euler apresentada na equação 5.12 de Gould (1978).

Outra forma de verificar a relação  $\sum_{x=0}^{\infty} P_x(\alpha, \lambda) = 1$  é por meio da função  $W$  de Lambert.

Seja  $X$  uma DPG conforme equação (2.1). Escrevendo a DPG em função da  $W$  de Lambert, conforme apresentado na equação (4.34), obtém-se

$$\begin{aligned} \sum_{x=0}^{\infty} P_x(\theta, b) &= \sum_{x=0}^{\infty} e^{bW_0(-\theta)} \frac{b(b+x)^{x-1}\theta^x}{x!} \\ &= e^{bW_0(-\theta)} \sum_{x=0}^{\infty} \frac{b(b+x)^{x-1}\theta^x}{x!}, \end{aligned}$$

na qual  $\sum_{x=0}^{\infty} \frac{b(b+x)^{x-1}\theta^x}{x!}$  é uma forma correlata da equação (4.13)

Logo,

$$\sum_{x=0}^{\infty} P_x(\theta, b) = e^{bW_0(-\theta)} e^{-bW_0(-\theta)} = 1$$

Portanto, a DPG, como se espera de uma distribuição de probabilidade, tem a soma de todas as suas probabilidades igual a 1.

## 4.5.2 Relações de recorrência para os momentos

### Média e Variância

Seja  $X$  uma variável aleatória discreta que segue uma DPG, sua média e a sua variância serão denotadas, respectivamente, por  $\mu$  e  $\sigma^2$ . Para facilitar os cálculos, serão utilizadas as propriedades da CDSF, e a DPG será escrita em função de uma  $W$  de Lambert .

Utilizando as equações (4.36) e (4.37) nas equações (4.25) e (4.26), obtêm-se a média e a variância da DPG com a parametrização da  $W$  de Lambert

$$\mu = -\theta e^{W_0(-\theta)} e^{-W_0(-\theta)} \left[ b \frac{W_0(-\theta)}{\theta(1+W_0(-\theta))} \right] = -b \frac{W_0(-\theta)}{(1+W_0(-\theta))} \quad (4.40)$$

e

$$\begin{aligned} \sigma^2 &= bW_0^2(-\theta) \frac{bW_0(-\theta) + W_0(-\theta) + b + 2}{[1+W_0(-\theta)]^3\theta} - b \frac{W_0(-\theta)}{(1+W_0(-\theta))} - \left[ b \frac{W_0(-\theta)}{(1+W_0(-\theta))} \right]^2 \\ &= -b \frac{W_0(-\theta)}{[1+W_0(-\theta)]^3}. \end{aligned} \quad (4.41)$$

Substituindo nas equações acima  $b = \alpha/\lambda$  e  $W_0(-\theta) = -\lambda$ , obtêm-se a média e a variância da DPG com a parametrização descrita em (2.1),

$$\mu = \frac{\alpha}{1-\lambda} \quad (4.42)$$

e

$$\sigma^2 = \frac{\alpha}{(1-\lambda)^3}. \quad (4.43)$$

A forma mais usual para os cálculos da média e da variância é feita de modo recursivo e requer mais desenvolvimento do que o modo exposto acima. Essa forma está apresentada no Apêndice B desta dissertação.

### Momentos

A partir da equação (4.27) é possível obter a relação de recorrência dos momentos escrita em função da  $W$  de Lambert da seguinte forma

$$m_{r+1} = \theta \frac{\partial m_r}{\partial \theta} - \left[ b \frac{W_0(-\theta)}{(1+W_0(-\theta))} \right] m_r.$$

Para encontrar a relação de recorrência dos momentos escrita em função  $\alpha$  e  $\lambda$ , deve-se utilizar derivadas parciais e considerar as equações (4.38) e (4.39), conforme mostrado

abaixo.

$$m_{r+1} = \theta \left[ \frac{\partial m_r}{\partial \lambda} \frac{\partial \lambda}{\partial \theta} + \frac{\partial m_r}{\partial \alpha} \frac{\partial \alpha}{\partial \theta} \right] + m_1 m_r$$

$$m_{r+1} = (-\lambda e^{-\lambda}) \left[ \frac{\partial m_r}{\partial \lambda} \frac{\lambda}{(-\lambda e^{-\lambda})(1-\lambda)} + \frac{\partial m_r}{\partial \alpha} \frac{\alpha}{(-\lambda e^{-\lambda})(1-\lambda)} \right] + \frac{\alpha}{1-\lambda} m_r$$

Logo, a relação de recorrência será

$$(1-\lambda)m_{r+1} = \lambda \frac{\partial m_r}{\partial \lambda} + \alpha \frac{\partial m_r}{\partial \alpha} + \alpha m_r. \quad (4.44)$$

Os dois primeiros momentos da DPG serão

$$m_1 = EX = \frac{\alpha}{(1-\lambda)}, \quad (4.45)$$

e

$$m_2 = \frac{\alpha}{(1-\lambda)^3} + \frac{\alpha^2}{(1-\lambda)^2}. \quad (4.46)$$

## Momentos centrais

A relação de recorrência dos momentos para a CDSP é dada na equação (4.27). Utilizando os resultados obtidos em (4.40) e (4.41), o  $r$ -ésimo momento da DPG escrito em função da  $W$  de Lambert será

$$M_{r+1} = \theta \frac{\partial M_r}{\partial \theta} - rb \frac{W_0(-\theta)}{[1+W_0(-\theta)]^3} M_{r-1}. \quad (4.47)$$

O  $r$ -ésimo momento central para uma DPG em termos de  $\alpha$  e  $\lambda$  será

$$M_{r+1} = \theta \left[ \frac{\partial M_r}{\partial \lambda} \frac{\partial \lambda}{\partial \theta} + \frac{\partial M_r}{\partial \alpha} \frac{\partial \alpha}{\partial \theta} \right] - rb \frac{W_0(-\theta)}{[1+W_0(-\theta)]^3} M_{r-1}. \quad (4.48)$$

Substituindo  $b = \frac{\alpha}{\lambda}$  e  $W_0(-\theta) = \lambda$  em (4.48), obtém-se

$$M_{r+1} = \frac{\lambda}{1-\lambda} \frac{\partial M_r}{\partial \lambda} + \frac{\alpha}{1-\lambda} \frac{\partial M_r}{\partial \alpha} + \frac{r\alpha}{(1-\lambda)^3} M_{r-1}. \quad (4.49)$$

Logo,

$$M_2 = \frac{\alpha}{(1-\lambda)^3} + \left[ \frac{\alpha}{(1-\lambda)} \right]^2$$

$$M_3 = \frac{\alpha(1+2\lambda)}{(1-\lambda)^5}$$

$$M_4 = \frac{3\alpha^2}{(1-\lambda)^6} + \frac{\alpha(1+8\lambda+6\lambda^2)}{(1-\lambda)^7}$$

$$\sigma^2 = M_2 - M_1^2 = \frac{\alpha}{(1-\lambda)^3}. \quad (4.50)$$

## Momentos fatoriais

A relação de recorrência dos momentos para a CDSP é dada na equação (4.30). Utilizando os resultados obtidos em (4.40), o  $r$ -ésimo momento fatorial da DPG escrito em função da  $W$  de Lambert será

$$M_{[r+1]} = \theta \frac{\partial M_{[r]}}{\partial \theta} - b \frac{W_0(-\theta)}{1 + W_0(-\theta)} M_{[r]} - r M_{[r]}. \quad (4.51)$$

O  $r$ -ésimo momento fatorial para uma DPG em função de  $\alpha$  e  $\lambda$  será

$$M_{[r+1]} = \theta \left[ \frac{\partial M_{[r]}}{\partial \lambda} \frac{\partial \lambda}{\partial \theta} + \frac{\partial M_{[r]}}{\partial \alpha} \frac{\partial \alpha}{\partial \theta} \right] - b \frac{W_0(-\theta)}{1 + W_0(-\theta)} M_{[r]} - r M_{[r]}. \quad (4.52)$$

Substituindo as equações (4.38) (4.39) em (4.52) obtém-se

$$M_{[r+1]} = \frac{\lambda}{1 - \lambda} \frac{\partial M_{[r]}}{\partial \lambda} + \frac{\alpha}{1 - \lambda} \frac{\partial M_{[r]}}{\partial \alpha} + \frac{\alpha}{1 - \lambda} M_r - r M_{[r]}. \quad (4.53)$$

### 4.5.3 Função geradora de probabilidade

Seja  $X$  uma variável aleatória com distribuição de probabilidade de Poisson generalizada e  $t > 0$ , a fgp de  $X$  será dada por

$$\begin{aligned} E(t^X) &= G_x(t) = \sum_x t^x P(X = x) \\ &= \sum_{x=0}^{\infty} t^x \frac{\alpha(\alpha + \lambda x)^{x-1}}{x!} e^{-\alpha - \lambda x}. \end{aligned}$$

Utilizando a expansão de Lagrange vista na equação (4.2),

$$f(z) = f(0) + \sum_{x=1}^{\infty} \frac{t^x}{x!} \frac{\partial^{x-1}}{\partial z^{x-1}} [g(z)^x] f'(z) \Big|_{z=0} \quad (4.54)$$

em que, no caso da DPG,  $g(z) = e^{\lambda(z-1)}$ ,  $f(z) = e^{\alpha(z-1)}$  e  $t = \frac{z}{g(z)}$ . Portanto,  $f'(z) = \alpha e^{\alpha(z-1)}$ ,  $f(0) = e^{-\alpha}$  e

$$\begin{aligned} \frac{\partial^{x-1}}{\partial z^{x-1}} [g(z)^x] f'(z) \Big|_{z=0} &= \frac{\partial^{x-1}}{\partial z^{x-1}} e^{x\lambda(z-1)} \alpha e^{\alpha(z-1)} \Big|_{z=0} \\ &= \alpha e^{-x\lambda - \alpha} \frac{\partial^{x-1}}{\partial z^{x-1}} e^{z(x\lambda + \alpha)} \Big|_{z=0} \\ &= \alpha(\alpha + \lambda x)^{x-1} e^{-x\lambda - \alpha} \end{aligned}$$

que permite substituir na equação (4.54) e obter

$$e^{\alpha(z-1)} = e^{\alpha} + \sum_{x=1}^{\infty} \frac{t^x}{x!} \alpha(\alpha + \lambda x)^{x-1} e^{-x\lambda - \alpha}.$$

Fazendo  $z = tg(z)$ , ou seja,  $z = te^{\lambda(z-1)}$ , obtém-se

$$e^{\alpha(te^{\lambda(z-1)}-1)} = \sum_{x=0}^{\infty} t^x \frac{\alpha(\alpha + \lambda x)^{x-1} e^{-x\lambda - \alpha}}{x!}.$$

Portanto, a fgp para a DPG será

$$G_x(t) = e^{\alpha(te^{\lambda(z-1)}-1)}.$$

A fgp escrita em função da  $W$  de Lambert será

$$\begin{aligned} G_x(t) &= e^{bW_0(-\theta)} \sum_{x=0}^{\infty} b(b+x)^{x-1} \frac{(t\theta)^x}{x!} \\ &= \exp \{bW_0(-\theta) - b[W_0(-t\theta)]\}. \end{aligned} \quad (4.55)$$

#### 4.5.4 Função geradora de momentos

Seja  $X$  uma variável aleatória com distribuição de probabilidade Poisson generalizada e  $t > 0$ , sua função geradora de momentos (fgm) será dada por

$$E(e^{tX}) = M_x(t) = \sum_x e^{tx} P(X = x).$$

Uma forma simplificada de obter a fgm é por meio da função  $W$  de Lambert.

Utilizando as equações (4.34) e a (4.13), obtém-se

$$M_x(t) = e^{bW_0(-\theta)} \sum_{x=0}^{\infty} b(b+x)^{x-1} \frac{(e^t\theta)^x}{x!}. \quad (4.56)$$

Portanto,

$$M_x(t) = \exp \{bW_0(-\theta) - bW_0(-e^t\theta)\}. \quad (4.57)$$

#### 4.5.5 Função característica

Função característica para CDSP é dada na equação (4.31). Esse resultado para a DPG em termos da  $W$  de Lambert será

$$\psi_X(t) = e^{-b[W(-it\theta)+W(-\theta)]}.$$

### 4.5.6 Coeficiente de assimetria

O coeficiente de assimetria de momentos de uma variável aleatória  $X$  é dado por

$$\gamma_1 = \frac{M_3}{M_2^{3/2}},$$

na qual  $M_i$  é o  $i$ -ésimo momento central.

O coeficiente de assimetria  $\gamma_1$  mede a simetria de um conjunto de dados. Quando a medida é negativa, indica que o conjunto de dados apresenta uma cauda à esquerda maior do que a cauda à direita. Do mesmo modo, quando a medida é positiva, indica que o conjunto de dados apresenta cauda à direita maior do que a cauda da esquerda. O valor zero indica simetria, ou seja, as caudas têm o mesmo tamanho.

Seja  $X$  uma variável aleatória com distribuição Poisson generalizada, seu coeficiente de assimetria será

$$\gamma_1 = \frac{\alpha(1+2\lambda)}{(1-\lambda)^5} \left[ \frac{\alpha}{(1-\lambda)^3} \right]^{3/2} = \frac{(1+2\lambda)}{\sqrt{a(1-\lambda)}}.$$

A assimetria será negativa se  $\lambda < -1/2$  e tenderá a zero quando  $\alpha$  assumir valores grandes. Essas decorrências podem ser visualizadas nas Figuras (1) e (2).

### 4.5.7 Coeficiente de curtose

O coeficiente de curtose de momentos de uma variável aleatória  $X$  é dado por

$$\gamma_2 = \frac{M_4}{M_2^2}.$$

na qual  $M_i$  é o  $i$ -ésimo momento central.

O coeficiente de curtose é uma medida que busca identificar o peso das caudas em comparação à cauda de uma distribuição normal. Uma distribuição normal tem coeficiente de curtose igual a 3 e é chamada de mesocúrtica. Um conjunto de dados cuja distribuição tem coeficiente de curtose acima de 3 tem caudas pesadas quando comparada com a distribuição normal e é chamada de leptocúrtica. Do mesmo modo, uma distribuição cuja curtose é abaixo de 3 tem caudas menos pesadas e é chamada de platicúrtica.

Seja  $X$  uma variável aleatória com distribuição Poisson generalizada, seu coeficiente de curtose será

$$\gamma_2 = \frac{\frac{3\alpha^2}{(1-\lambda)^6} + \frac{\alpha(1+8\lambda+6\lambda^2)}{(1-\lambda)^7}}{\left[\frac{\alpha}{(1-\lambda)^3}\right]^2} = 3 + \frac{(1+8\lambda+6\lambda^2)}{\alpha(1-\lambda)}.$$

Para qualquer valor de  $\alpha$  e para  $0 < \lambda < 1$ , a DPG assumirá a forma leptocúrtica e assumirá a forma mesocúrtica ou platicúrtica para valores de  $\lambda$  negativos.

#### 4.5.8 Convolação

**Teorema 4.2 (Convolação).** *Sejam  $X$  e  $Y$  variáveis aleatórias independentes que seguem uma DPG, sendo,*

$$X \sim GPD(\alpha_1, \lambda)$$

e

$$Y \sim GPD(\alpha_2, \lambda)$$

A soma de  $X + Y$  também será uma DPG com parâmetros  $\alpha_1 + \alpha_2$  e  $\lambda$ . Isto é,

$$X + Y \sim GPD(\alpha_1 + \alpha_2, \lambda).$$

(CONSUL; FAMOYE, 1989).

#### 4.5.9 Unimodalidade

**Teorema 4.3 (Unimodalidade).** *A DPG será unimodal para todos os valores de  $\alpha$  e  $\lambda$ . A moda estará em  $x = 0$  se  $\alpha e^{-\lambda} < 1$  e simultaneamente nos pontos  $x = 0$  e  $x = 1$  quando  $\alpha e^{-\lambda} = 1$ . Para  $\alpha e^{-\lambda} > 1$  a moda estará em  $x = M_o$ , em que*

$$(\alpha - e^{-\lambda})(e^{-\lambda} - 2\lambda)^{-1} < M_o < a,$$

na qual  $a$  é o menor valor de  $M_o$  que satisfaz a condição:

$$\lambda^2 M_o^2 + M_o[2\lambda\alpha - (\alpha + 2\lambda)e^\lambda] + \alpha^2 > 0.$$

(CONSUL; FAMOYE, 1986).



## 5 Inferência estatística para a DPG

O conjunto de técnicas e procedimentos que permite dar ao pesquisador um grau de confiabilidade nas afirmações que faz para a população a partir dos resultados das amostras é nomeado de Inferência Estatística. De modo geral, várias metodologias de inferência já foram propostas, sendo a decorrente da teoria de verossimilhança uma das mais relevantes. Neste capítulo, serão abordadas algumas propriedades fundamentais dos estimadores, como a suficiência, eficiência, consistência e o viés; e serão apresentados alguns métodos de estimação pontual, intervalar e teste de hipóteses, todos com ênfase na teoria de verossimilhança.

Para as análises abaixo, considere uma amostra  $X_1, X_2, \dots, X_n$  iid de tamanho  $n$  de uma população que segue uma DPG com parâmetros  $\alpha$  e  $\lambda$ . Sejam  $0, 1, 2, \dots, k$  os valores observados dessa amostra com as frequências  $n_0, n_1, n_2, \dots, n_k$ , respectivamente. Desse modo, o maior valor observado na amostra será  $k$  e  $n = \sum_{x=0}^k n_x$ .

### 5.1 Família exponencial

Uma família de distribuições de probabilidade é dita da família exponencial, se puder ser expressa da forma

$$P_x(\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta})\exp\left[\sum_{i=1}^k w_i(\boldsymbol{\theta})t_i(x)\right] \quad (5.1)$$

com  $h(x) \geq 0$ ,  $t_1(x), \dots, t_k(x)$  são funções reais das observações de  $x$  (não dependem de  $\boldsymbol{\theta}$ ),  $c(\boldsymbol{\theta}) \geq 0$  e  $w_1(\boldsymbol{\theta}), \dots, w_k(\boldsymbol{\theta})$  são funções reais do parâmetro que possivelmente têm seu valor definido pelo vetor  $\boldsymbol{\theta}$  (não podem depender de  $x$ ) (CASELLA; BERGER, 2010).

Para a DPG, obtém-se

$$P_{x_1, \dots, x_n}(\alpha, \lambda) = \left[ \frac{1}{\prod_{i=1}^k (x_i!)} \right] \alpha e^{-n\alpha} \exp \left[ \sum_{i=1}^k (x_i - 1) \ln(\alpha + \lambda x_i) - \lambda \sum_{i=1}^k x_i \right]. \quad (5.2)$$

O termo  $\sum_{i=1}^k (x_i - 1) \ln(\alpha + \lambda x_i)$  da equação acima não permite que a DPG seja escrita da forma (5.1). Portanto, a DPG não pertence à família exponencial.

## 5.2 Suficiência

Determinar a existência de uma estatística suficiente para um parâmetro de interesse permite aplicar uma série de teoremas e definições estatísticas que darão subsídios a vários resultados importantes.

A estatística  $T(X)$  é dita suficiente se ela captura toda a informação sobre o parâmetro de interesse contida na amostra. Formalmente,  $T(X)$  é uma estatística suficiente se a distribuição condicional de  $X$  dado  $T(X)$  não depende do parâmetro.

**Teorema 5.1** (*Fatoração de Fisher-Neyman*) Para quaisquer valores dos parâmetros em seus espaços paramétricos e para  $P_{\mathbf{x}}(\alpha, \lambda)$  uma função de probabilidade conjunta de uma amostra  $\mathbf{X}$ , uma estatística  $T(\mathbf{X})$  é suficiente para um parâmetro de interesse se, e somente se, existem funções  $g(T(\mathbf{x})|\alpha, \lambda)$  e  $h(\mathbf{x})$  de forma que vale a igualdade:

$$P_{\mathbf{x}}(\alpha, \lambda) = g(T(\mathbf{x})|\alpha, \lambda)h(\mathbf{x}).$$

A função de probabilidade conjunta de  $X_1, \dots, X_n$  observações independentes e identicamente distribuídas de distribuições de Poisson generalizada será

$$P_{x_1, \dots, x_n}(\alpha, \lambda) = \left[ \frac{1}{\prod_{i=1}^n (x_i!)} \right] \alpha e^{-n\alpha} e^{-\lambda \sum_{i=1}^n x_i} \left[ \prod_{i=1}^n (\alpha + \lambda x_i)^{x_i-1} \right].$$

Supondo que a estatística suficiente seja  $T(\mathbf{X}) = \sum_{i=1}^n X_i$  e utilizando a propriedade de convolução dada no Capítulo 4, obtém-se

$$X_1 + \dots + X_n \sim DPG(n\alpha, \lambda).$$

Logo,

$$g(T(\mathbf{x})|\alpha, \lambda) = \frac{n\alpha(n\alpha + \lambda t)^{t-1}}{t!} e^{-n\alpha - \lambda t}.$$

na qual  $T(\mathbf{x}) = \sum_{i=1}^n x_i = t$

Se  $T(X)$  for uma estatística suficiente, então

$$h(\mathbf{x}) = \frac{P_{x_1, \dots, x_n}(\alpha, \lambda)}{g(T(\mathbf{x})|\alpha, \lambda)}$$

em que  $h(\mathbf{x})$  é função apenas de  $x$ .

$$h(\mathbf{x}) = \frac{n \prod_{i=1}^n (\alpha + \lambda x_i)^{x_i-1}}{(n\alpha + \lambda t)^{t-1}} \frac{t!}{\prod_{i=1}^n x_i!}.$$

Supondo que um dos parâmetros seja conhecido, não é possível escrever  $h(\mathbf{x})$  em função apenas de  $x$ . Desse modo, a estatística  $T(X)$  não é suficiente para qualquer um

dos parâmetros. Não há, também, uma outra estatística  $T(X)$  que seja suficiente para um dos parâmetros. Além disso, na DPG, em um espaço paramétrico de dimensão 2 não há estatísticas suficientes da mesma dimensão. Isto é consequência do fato da DPG não pertencer a família exponencial (LEHMANN; CASELLA, 1998, pág. 40)

## 5.3 Métodos de estimação pontual dos parâmetros da DPG

### 5.3.1 Método dos momentos

Os  $j$ -ésimos momentos amostrais e teóricos, representados por  $M_j$  e  $\mu_j$ , respectivamente, serão

$$M_j = \sum_{x=0}^n \frac{x^j}{n} \quad (5.3)$$

e

$$\mu_j = \sum_{x=0}^{\infty} x^j P_x(\alpha, \lambda) \quad (5.4)$$

Pela Lei dos Grandes Números, os momentos amostrais convergem quase certamente para os momentos teóricos à medida que o tamanho da amostra aumenta.

Sendo os dois primeiros momentos amostrais dados por

$$M_1 = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$$

e

$$M_2 = \frac{1}{n} \sum_{i=1}^n x_i^2$$

O método dos momentos (MMo) consiste em igualar os dois primeiros momentos amostrais aos dois primeiros teóricos. Resolvendo o sistema formado por essas equações, obtêm-se as estimativas dos parâmetros pelo método dos momentos

$$\hat{\alpha} = \frac{\bar{x}^{3/2}}{s} \quad (5.5)$$

e

$$\hat{\lambda} = 1 - \frac{\bar{x}^{1/2}}{s}. \quad (5.6)$$

Shoukri (1980) obteve os vieses e as variâncias das estimativas (5.5) e (5.6) fazendo uma expansão de Taylor de  $\hat{\alpha}$ ,  $\hat{\alpha}^2$ ,  $\hat{\lambda}$  e  $\hat{\lambda}^2$ , aplicando o método de Cramer (1946) e calculando suas expectâncias. Com uma aproximação de primeira ordem, Shoukri obteve

$$b(\hat{\alpha}) \approx \frac{1}{4n} \left[ 5\alpha + \frac{3\lambda(2+3\lambda)}{1-\lambda} \right],$$

$$b(\hat{\lambda}) \approx -\frac{1}{4n\alpha} [5\alpha(1-\lambda) + \lambda(10 + 9\lambda^2)],$$

$$\text{var}(\hat{\alpha}) \approx \frac{\alpha}{2n} \left( \alpha + \frac{2 - 2\lambda + 3\lambda^2}{1 - \lambda} \right),$$

$$\text{var}(\hat{\lambda}) \approx \frac{1 - \lambda}{2n\alpha} (\alpha - \alpha\lambda + 2\lambda + 3\alpha^2)$$

e

$$\text{cov}(\hat{\alpha}, \hat{\lambda}) \approx -\frac{1}{2n} [\alpha(1 - \lambda) + 3\lambda^2]$$

Os vieses, as variâncias e a covariância dos estimadores de momentos são inversamente proporcionais ao tamanho da amostra ( $n$ ), portanto, esses valores tendem a ser menores à medida que  $n$  aumenta. A covariância entre os estimadores dos parâmetros será sempre negativa, indicando que quanto maior o valor de  $\hat{\alpha}$  menor será o valor de  $\hat{\lambda}$ , ou seja, quando  $\hat{\alpha}$  é superestimado  $\hat{\lambda}$  tende a ser subestimado. Do mesmo modo, quando  $\hat{\lambda}$  é superestimado  $\hat{\alpha}$  tende a ser subestimado.

As estimativas pelo método dos momentos não possuem as propriedades ótimas do método de máxima verossimilhança. No entanto, são consistentes e as estimativas da DPG são obtidas de maneira analítica e, conforme será visto na próxima seção, podem ser utilizadas como valor inicial em algoritmos que calculam os EMVs.

**Exemplo 1:** Na aplicação de biologia e ecologia apresentada no capítulo 2, observam-se as distribuições de aranhas e tatuzinhos de jardim sobre placas. Considere  $\bar{x}_A = 0.425$  e  $s_A^2 = 0.4546$  respectivamente a média e a variância da distribuição das aranhas e  $\bar{x}_T = 3.2951$  e  $s_T^2 = 15.0527$  respectivamente a média e a variância da distribuição dos tatuzinhos de jardim. Logo, as estimativas dos parâmetros da DPG obtidas pelo método dos momentos serão

$$\hat{\alpha}_A = \frac{\bar{x}_A^{3/2}}{s_A} = 0.4109, \quad \hat{\lambda}_A = 1 - \frac{\bar{x}_A^{1/2}}{s_A} = 0.0331,$$

$$\hat{\alpha}_T = \frac{\bar{x}_T^{3/2}}{s_T} = 1.5417 \quad \text{e} \quad \hat{\lambda}_T = 1 - \frac{\bar{x}_T^{1/2}}{s_T} = 0.5321$$

Verifica-se que  $\hat{\alpha}_A < \hat{\alpha}_T$ , indicando que os tatuzinhos de jardim tendem a se agrupar mais do que as aranhas. Já os valores de  $\hat{\lambda}$ , indicam que os tatuzinhos formam grupos de tamanhos variados.

### 5.3.2 Método da Máxima Verossimilhança

A função de verossimilhança, denominada  $L(\alpha, \lambda; \mathbf{x})$ , será dada por

$$L(\alpha, \lambda | \mathbf{x}) = p_0(\alpha, \lambda)^{n_0} \times \dots \times p_k(\alpha, \lambda)^{n_k} = \prod_{x=0}^k p_x(\alpha, \lambda)^{n_x}$$

$$L(\alpha, \lambda | \mathbf{x}) = \prod_{x=0}^k \left[ \frac{\alpha(\alpha + \lambda x)^{x-1}}{x!} e^{-\alpha - \lambda x} \right]^{n_x}. \quad (5.7)$$

O método de máxima verossimilhança (MMV) assume que as estimativas  $\hat{\alpha}$  e  $\hat{\lambda}$  dos parâmetros  $\alpha$  e  $\lambda$ , respectivamente, são os valores que maximizam a função de verossimilhança,

$$\max\{L(\alpha, \lambda | \mathbf{x})\} = L(\hat{\alpha}, \hat{\lambda} | \mathbf{x}).$$

Para simplificar os cálculos, adota-se a função monótona crescente  $\ln(\cdot)$  na função de verossimilhança, uma vez que  $\max\{L(\alpha, \lambda | \mathbf{x})\} = \max\{\ln L(\alpha, \lambda | \mathbf{x})\}$ . Assim, o produtório é substituído pelo somatório, o que facilitará os cálculos subsequentes.

$$\begin{aligned} \ell(\alpha, \lambda | \mathbf{x}) &= \ln(L(\alpha, \lambda | \mathbf{x})) \\ &= n \ln(\alpha) - \sum_{x=0}^k n_x \ln(x!) \\ &\quad + \sum_{x=0}^k n_x (x-1) \ln(\alpha + \lambda x) - n\alpha - \lambda \sum_{x=0}^k n_x x. \end{aligned} \quad (5.8)$$

Os EMVs serão obtidos a partir da solução do sistema de equações formado pelos componentes da função escore igualados a zero

$$U(\alpha) = \frac{\partial \ell(\alpha, \lambda | \mathbf{x})}{\partial \alpha} = 0$$

e

$$U(\lambda) = \frac{\partial \ell(\alpha, \lambda | \mathbf{x})}{\partial \lambda} = 0.$$

Nessas equações, o parâmetro de interesse é isolado e escrito em função das informações obtidas da amostra.

$$U(\alpha) = \frac{n}{\alpha} + \sum_{x=0}^k n_x \frac{(x-1)}{\alpha + \lambda x} - n = 0 \quad (5.9)$$

e

$$U(\lambda) = \sum_{x=0}^k n_x \frac{(x-1)x}{\alpha + \lambda x} - \sum_{x=0}^k n_x x = 0 \quad (5.10)$$

Para simplificar as funções dos cálculos estimadores, as equações (5.9) e (5.10) são multiplicadas por  $\alpha$  e  $\lambda$ , respectivamente, e somadas.

$$\begin{aligned}
& n + \sum_{x=0}^k n_x \frac{(x-1)\alpha}{\alpha + \lambda x} - \alpha n + \sum_{x=0}^k n_x \frac{(x-1)\lambda x}{\alpha + \lambda x} - \lambda n \bar{x} \\
&= n - \alpha n - \lambda n \bar{x} + \sum_{x=0}^k n_x \frac{(x-1)(\alpha + \lambda x)}{\alpha + \lambda x} \\
&= n - \alpha n - \lambda n \bar{x} + \sum_{x=0}^k n_x (x-1) \\
&= -\alpha - \lambda \bar{x} + \bar{x} = 0.
\end{aligned} \tag{5.11}$$

Logo,

$$\hat{\alpha} = \bar{x}(1 - \lambda). \tag{5.12}$$

A estimativa de  $\hat{\lambda}$  é encontrada substituindo (5.12) na equação (5.10),

$$h(\hat{\lambda}) = \sum_{x=0}^k n_x \frac{(x-1)x}{\bar{x}(1-\hat{\lambda}) + \hat{\lambda}x} - n\bar{x} = 0. \tag{5.13}$$

Como pode ser verificado, a função (5.13) não é definida quando  $k = 0$  e não é igual a zero quando  $k = 1$ . Além disso, para esses mesmos valores de  $k$ , as derivadas segundas de  $\ell(\alpha, \lambda | \mathbf{x})$  não serão negativas – o sinal da derivada segunda é um teste que verifica se a raiz da função escore é ponto de máximo, mínimo ou de sela, ou seja, verifica se as estimativas dos parâmetros são de fato o ponto de máximo da função de verossimilhança. A partir desses resultados,

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial \alpha^2} &= -\frac{n}{\alpha^2} - \sum_{x=0}^k n_x \frac{(x-1)}{(\alpha + \lambda x)^2} = \frac{n - n_0}{\alpha^2} - \sum_{x=2}^k n_x \frac{(x-1)}{(\alpha + \lambda x)^2}, \\
\frac{\partial^2 \ell}{\partial \lambda^2} &= -\sum_{x=0}^k n_x \frac{(x-1)x^2}{(\alpha + \lambda x)^2} = -\sum_{x=2}^k n_x \frac{(x-1)x^2}{(\alpha + \lambda x)^2}
\end{aligned}$$

e

$$\frac{\partial^2 \ell}{\partial \lambda \partial \alpha} = -\sum_{x=0}^k n_x \frac{(x-1)x}{(\alpha + \lambda x)^2} = -\sum_{x=2}^k n_x \frac{(x-1)x}{(\alpha + \lambda x)^2}$$

conclui-se que os EMVs para a DPG existirão somente se  $k \geq 2$ .

Para a DPG, não é possível obter os EMVs de forma direta, pois o estimador de um parâmetro está em função das informações obtidas por meio da amostra e também do outro parâmetro. Em tais caso, utilizam-se métodos iterativos, como o de Newton-Raphson. Neste método, o valor de  $\hat{\lambda}$  será determinado pela relação de recorrência

$$\lambda^{(r+1)} = \lambda^{(r)} - \frac{h(\lambda^{(r)})}{h'(\lambda^{(r)})} \quad r = \{0, 1, \dots\}, \tag{5.14}$$

em que  $h'(\lambda^{(r)})$  é a derivada da função  $h(\lambda^{(r)})$  em (5.13).

Conforme mencionado anteriormente, os valores iniciais para as estimativas dos parâmetros ( $\alpha_0$  e  $\lambda_0$ ) podem ser obtidos por meio do método dos momentos, visto na seção anterior. O processo iterativo terminará quando, no  $m$ -ésimo passo, o valor de  $|\lambda^{(r+1)} - \lambda^{(r)}| < \epsilon$  (em que  $\epsilon$  é um valor pequeno e próximo de zero). O valor de  $\hat{\alpha}$  será obtido pela equação

$$\hat{\alpha} = \bar{x}(1 - \hat{\lambda}^{(m)}).$$

### 5.3.2.1 Propriedades dos estimadores de máxima verossimilhança

A função de verossimilhança possibilita o cálculo da função escore, da informação observada e da informação esperada de Fisher. Esses resultados são fundamentais para a construção de intervalos de confiança e testes de hipóteses e, por esse motivo, serão expostos a seguir.

- **Função escore**

A função escore é o vetor gradiente da função de log-verossimilhança.

$$\begin{aligned} U(\alpha, \lambda) &= \left[ \frac{\partial \ell(\alpha, \lambda | \mathbf{x})}{\partial \alpha}, \quad \frac{\partial \ell(\alpha, \lambda | \mathbf{x})}{\partial \lambda} \right]^T \\ &= \left[ \frac{n}{\alpha} + \sum_{x=0}^k n_x \frac{(x-1)}{\alpha + \lambda x} - n, \quad \sum_{x=0}^k n_x \frac{(x-1)x}{\alpha + \lambda x} - \sum_{x=0}^k n_x x \right]^T \end{aligned} \quad (5.15)$$

- **Informação observada**

A informação observada é definida como menos a derivada segunda da função de log-verossimilhança avaliada no ponto de máximo.

$$\begin{aligned} I_o(\alpha, \lambda) &= \begin{bmatrix} -\frac{\partial^2 \ell(\alpha, \lambda | \mathbf{x})}{\partial \alpha^2} & -\frac{\partial^2 \ell(\alpha, \lambda | \mathbf{x})}{\partial \alpha \partial \lambda} \\ -\frac{\partial^2 \ell(\alpha, \lambda | \mathbf{x})}{\partial \lambda \partial \alpha} & -\frac{\partial^2 \ell(\alpha, \lambda | \mathbf{x})}{\partial \lambda^2} \end{bmatrix}_{(\alpha, \lambda) = (\hat{\alpha}, \hat{\lambda})} \\ &= \begin{bmatrix} \frac{n}{\hat{\alpha}^2} + \sum_{x=0}^k n_x \frac{(x-1)}{(\hat{\alpha} + \hat{\lambda}x)^2} & \sum_{x=0}^k n_x \frac{x(x-1)}{(\hat{\alpha} + \hat{\lambda}x)^2} \\ \sum_{x=0}^k n_x \frac{x(x-1)}{(\hat{\alpha} + \hat{\lambda}x)^2} & \sum_{x=0}^k n_x \frac{x^2(x-1)}{(\hat{\alpha} + \hat{\lambda}x)^2} \end{bmatrix}. \end{aligned} \quad (5.16)$$

- **Informação de Fisher**

A informação de Fisher é definida como menos a esperança da derivada segunda da função log-verossimilhança

$$\begin{aligned}
 I(\alpha, \lambda) &= E[I_o(\alpha, \lambda)] = \begin{bmatrix} E \left[ -\frac{\partial^2 \ell(\alpha, \lambda | \mathbf{x})}{\partial \alpha^2} \right] & E \left[ -\frac{\partial^2 \ell(\alpha, \lambda | \mathbf{x})}{\partial \alpha \partial \lambda} \right] \\ E \left[ -\frac{\partial^2 \ell(\alpha, \lambda | \mathbf{x})}{\partial \lambda \partial \alpha} \right] & E \left[ -\frac{\partial^2 \ell(\alpha, \lambda | \mathbf{x})}{\partial \lambda^2} \right] \end{bmatrix} \\
 &= \begin{bmatrix} \frac{n[\alpha(1-\lambda) + 2\lambda]}{\alpha(\alpha + 2\lambda)} & \frac{n\alpha}{\alpha + 2\lambda} \\ \frac{n\alpha}{\alpha + 2\lambda} & \frac{n\alpha(\alpha + 2)}{(1-\lambda)(\alpha + 2\lambda)} \end{bmatrix}. \tag{5.17}
 \end{aligned}$$

Para detalhes do cálculo da informação de Fisher para a DPG veja Famoye e Consul (1990).

Cabe destacar que um importante resultado obtido por meio da informação de Fisher é  $var(U(\cdot)) = [I(\cdot)]^{-1}$ . Essa decorrência é especialmente útil quando se aproxima a DPG para uma distribuição normal, permitindo, portanto, o cálculo de intervalos de confiança para os parâmetros quando o tamanho da amostra é grande.

### 5.3.2.2 Avaliação dos EMVs para a DPG

Os EMVs possuem importantes propriedades, como invariância, consistência, eficiência assintótica, o fato de serem assintoticamente não viesados e de terem distribuições assintoticamente normais. Abaixo estão apresentados os resultados para os EMVs dos parâmetros da DPG.

- **Viés**

Os EMVs para a DPG são viesados e os cálculos para obtê-los são complexos. Consul e Shoukri (1984) demonstraram que os valores dos vieses são aproximados por

$$b(\hat{\alpha}) \approx -\frac{\alpha(5\alpha^2 + 28\alpha\lambda - 6\alpha\lambda^2 + 24\lambda^2)}{2n(1-\lambda)(\alpha + 2\lambda)^2(\alpha + 3\lambda)}$$

e

$$b(\hat{\lambda}) \approx \frac{5\alpha^3(1-\lambda) - 2\alpha^2\lambda(2\lambda^2 + 9\lambda - 13) + 4\alpha\lambda^2(11 - 6\lambda) + 24\lambda^2}{2n\alpha(1-\lambda)(\alpha + 2\lambda)^2(\alpha + 3\lambda)}$$

O viés do estimador de  $\alpha$  está em função de  $\lambda$ , assim como o viés do estimador de  $\lambda$  está em função de  $\alpha$ . Ambos os vieses são inversamente proporcionais a  $n$ , ou seja,

quanto maior o valor de  $n$  menor serão os vieses dos estimadores. Os EMVs da DPG são, portanto, assintoticamente não viesados.

- Erro quadrático médio (EQM)

Mede a diferença quadrática média entre o estimador e o parâmetro.

$$EQM_{\alpha,\lambda}(\hat{\alpha} - \alpha)^2 = var(\alpha) + b(\hat{\alpha})^2 \approx \frac{\alpha(2 + \alpha)}{2n} + \left[ \frac{\alpha(5\alpha^2 + 28\alpha\lambda - 6\alpha\lambda^2 + 24\lambda^2)}{2n(1 - \lambda)(\alpha + 2\lambda)^2(\alpha + 3\lambda)} \right]^2$$

$$\begin{aligned} EQM_{\alpha,\lambda}(\hat{\lambda} - \lambda)^2 &= var(\lambda) + b(\hat{\lambda})^2 \\ &\approx \frac{(\alpha + 2\lambda - \alpha\lambda)(1 - \lambda)}{2n\alpha} \\ &+ \left[ \frac{5\alpha^3(1 - \lambda) - 2\alpha^2\lambda(2\lambda^2 + 9\lambda - 13) + 4\alpha\lambda^2(11 - 6\lambda) + 24\lambda^2}{2n\alpha(1 - \lambda)(\alpha + 2\lambda)^2(\alpha + 3\lambda)} \right]^2. \end{aligned}$$

Os erros quadráticos médio são inversamente proporcional a  $n$ , ou seja, quanto maior o valor de  $n$  menor serão os EQMs dos parâmetros. A diferença quadrática média entre o estimador e o parâmetro tenderá a zero quando  $n \rightarrow \infty$ .

As variâncias dos estimadores podem ser obtidas por meio da inversa da informação de Fisher. Os cálculos detalhados podem ser vistos em Consul e Shoukri (1984).

- Consistência

Um estimador é consistente para estimar um parâmetro se, à medida que se amplia o tamanho da amostra, aumenta-se a precisão da estimativa. Assim, um estimador é dito fracamente consistente se

$$\lim_{n \rightarrow \infty} P(|\hat{\alpha} - \alpha| > \epsilon) = 0 \quad \forall \epsilon > 0$$

$$\lim_{n \rightarrow \infty} P(|\hat{\lambda} - \lambda| > \epsilon) = 0 \quad \forall \epsilon > 0$$

Uma forma de se verificar a consistência é por meio da desigualdade de Tchebychev, na qual

$$P(|\hat{\alpha} - \alpha| > \epsilon) < \frac{var(\hat{\alpha})}{\epsilon^2} = \frac{\alpha(2 + \alpha)}{2n\epsilon^2}.$$

Como

$$\lim_{n \rightarrow \infty} \frac{\alpha(2 + \alpha)}{2n\epsilon^2} = 0,$$

segue que

$$\lim_{n \rightarrow \infty} P(|\hat{\alpha} - \alpha| > \epsilon) = 0.$$

Analogamente para  $\lambda$ ,

$$P(|\hat{\lambda} - \lambda| > \epsilon) < \frac{\text{var}(\hat{\lambda})}{\epsilon^2} = \frac{(\alpha + 2\lambda - \alpha\lambda)(1 - \lambda)}{2n\alpha\epsilon^2}.$$

Como

$$\lim_{n \rightarrow \infty} \frac{(\alpha + 2\lambda - \alpha\lambda)(1 - \lambda)}{2n\alpha\epsilon^2} = 0,$$

segue que

$$\lim_{n \rightarrow \infty} P(|\hat{\lambda} - \lambda| > \epsilon) = 0.$$

Portanto, os EMVs, para a DPG são assintoticamente consistentes.

- Assintoticamente Normal

Seja  $[\hat{\alpha}, \hat{\lambda}]^t$  o vetor das estimativas de máxima verossimilhança dos parâmetros de uma DPG. Devido à propriedade de normalidade assintótica dos EMVs, pode-se aproximar a distribuição de  $[\hat{\alpha}, \hat{\lambda}]^t$  por uma distribuição normal bivariada quando  $n \rightarrow \infty$ , ou seja,

$$(\hat{\alpha}, \hat{\lambda}) \overset{\text{approx.}}{\sim} N_2 \left( (\alpha, \lambda), I(\alpha, \lambda)^{-1} \right)$$

em que  $I(\alpha, \lambda)^{-1}$  é a inversa de informação de Fisher.

### 5.3.2.3 EMV quando um dos parâmetros é conhecido

Eventualmente, um pesquisador pode ter a informação de um dos parâmetros. Nas subseções seguintes, são obtidos os EMVs quando um dos parâmetros é conhecido.

#### Quando $\lambda$ conhecido

Quando o valor de  $\lambda$  é conhecido, como feito na equação (5.12), obtém-se o valor de  $\hat{\alpha}$  em função de  $\lambda$  desconhecido,

$$\hat{\alpha} = \bar{x}(1 - \lambda). \quad (5.18)$$

#### Quando $\alpha$ é conhecido

Conforme feito acima, no caso de  $\alpha$  conhecido, a estimativa de  $\lambda$  será encontrada isolando o valor de  $\lambda$  na equação (5.11),

$$\hat{\lambda} = 1 - \frac{\alpha}{\bar{x}} \quad \text{para } 0 < \alpha \leq \bar{x}. \quad (5.19)$$

O MMV é considerado o método de estimação mais eficiente porque produz os estimadores de menor variância. Entretanto, para alguns casos, a maior eficiência do MMV é apenas assintótica, o que faz com que sua aplicação a amostras de pequeno tamanho produza estimadores de qualidade comparável ou inferior a outros métodos. Os EMVs são consistentes, suficientes e assintoticamente sem viés. Para amostras finitas, entretanto, os EMVs podem ser viesados, embora o viés possa ser corrigido. O MMV exige maior esforço computacional, por envolver soluções numéricas de sistemas de equações frequentemente não lineares e implícitas.

**Exemplo 1** (continuação) As estimativas da DPG para o método da máxima verossimilhança obtidas por meio da função `generalized_poisson_likelihood(.)` do pacote `GPSEQ` do R são:

$$\hat{\alpha}_A = 0.4114, \quad \hat{\lambda}_A = 0.0320,$$

$$\hat{\alpha}_T = 1.4446 \quad \text{e} \quad \hat{\lambda}_T = 0.5616.$$

### 5.3.3 Método de estimação baseado na média amostral e na frequência de zeros da amostra

Esse método de estimação é apropriado quando a frequência da classe zero é maior do que frequência das demais classes ou quando a distribuição dos dados tem a forma *L-shaped*.

As estimativas dos parâmetros serão:

$$\hat{\alpha} = -\frac{\ln(n_0)}{n} \tag{5.20}$$

e

$$\hat{\lambda} = 1 - \frac{\hat{\alpha}}{\bar{x}}, \tag{5.21}$$

na qual  $n_0$  é a quantidade de vezes que o valor zero é observado na amostra e  $\bar{x}$  é a média da amostra.

Em Consul (1989) é demonstrado que as variâncias dos estimadores são

$$\begin{aligned} \text{var}(\hat{\alpha}) &\approx \frac{e^\alpha - 1}{n} \\ \text{var}(\hat{\lambda}) &\approx \frac{1 - \lambda}{n\alpha^2} [(1 - \lambda)(e^\alpha - 1) + \alpha(2\lambda - 1)] \end{aligned}$$

$$\text{cov}(\hat{\alpha}, \hat{\lambda}) \approx -\frac{1 - \lambda}{n\alpha} (e^\alpha - 1 - \alpha)$$

O MMo e o método baseado na média amostral e na frequência de zero (FREQ0) são os dois únicos cujos estimadores da DPG podem ser obtidos de forma analítica. No entanto, esse desconsidera boa parte da informação da amostra para o cálculo de  $\hat{\alpha}$  e, para amostras de tamanho  $n$ , a variância desse estimador aumenta quando  $\alpha$  aumenta.

**Exemplo 1** (continuação) As estimativas da DPG obtidas pelo método baseado na média amostral e na frequência de zero serão

Para

$$n_{0A} = 159, \quad \bar{x}_A = 0.425,$$

$$n_{0T} = 28 \quad \text{e} \quad \bar{x}_T = 3.2951.$$

Têm-se

$$\hat{\alpha}_A = 0.4117, \quad \hat{\lambda}_A = 0.0312,$$

$$\hat{\alpha}_T = 1.4718 \quad \text{e} \quad \hat{\lambda}_T = 0.5533.$$

### 5.3.4 Método da discrepância empírica ponderada

No MMV, a função escore é ponderada pela frequência das classes ( $n_x$ ) da seguinte forma

$$\sum_{x=0}^k n_x \frac{\partial}{\partial \alpha} \log(p_x(\alpha, \lambda)) = \sum_{x=0}^k n_x \left[ \frac{x(\alpha + \lambda)}{\alpha(\alpha + \lambda x)} - 1 \right] = 0 \quad (5.22)$$

e

$$\sum_{x=0}^k n_x \frac{\partial}{\partial \lambda} \log(p_x(\alpha, \lambda)) = \sum_{x=0}^k n_x \left[ \frac{x(\alpha + \lambda)}{\alpha(\alpha + \lambda x)} - x \right] = 0. \quad (5.23)$$

Fazendo  $f_x = \frac{n_x}{n}$ , o MMV pode ser ponderado pela frequência relativa das classes  $x$ , ou seja,

$$\sum_{x=0}^k f_x \left[ \frac{x(\alpha + \lambda)}{\alpha(\alpha + \lambda x)} - 1 \right] = 0 \quad (5.24)$$

e

$$\sum_{x=0}^k f_x \left[ \frac{x(\alpha + \lambda)}{\alpha(\alpha + \lambda x)} - x \right] = 0. \quad (5.25)$$

Já no método de estimação das discrepâncias ponderadas a função escore é ponderada pela diferença entre a frequência relativa ( $f_x$ ) e a probabilidade ( $P_x$ ) em um ponto  $x$  conforme as equações

$$\sum_{x=0}^k (f_x - P_x) \left[ \frac{x(\alpha + \lambda)}{\alpha(\alpha + \lambda x)} - 1 \right] = 0 \quad (5.26)$$

e

$$\sum_{x=0}^k (f_x - P_x) \left[ \frac{x(\alpha + \lambda)}{\alpha(\alpha + \lambda x)} - x \right] = 0. \quad (5.27)$$

Para obter as estimativas pelo método discrepância ponderada empírica (MDEP), Felix e Lee (1992) consideraram a combinação desses dois métodos da seguinte forma

$$\sum_{x=0}^k f_x (f_x - P_x) \left[ \frac{x(\alpha + \lambda)}{\alpha(\alpha + \lambda x)} - 1 \right] = 0 \quad (5.28)$$

e

$$\sum_{x=0}^k f_x (f_x - P_x) \left[ \frac{x(\alpha + \lambda)}{\alpha(\alpha + \lambda x)} - x \right] = 0. \quad (5.29)$$

As equações (5.28) e (5.29) podem ser resolvidas de forma iterativa por meio do método de Newton-Raphson. Os valores iniciais de  $\alpha$  e  $\lambda$  podem ser obtidos por meio do MMo ou FREQ0. Para Consul e Famoye (2006), o viés dos estimadores obtidos pelo MDEP é menor ou igual ao viés obtido por MMV e MMo.

Lee e Famoye (1996) aplicaram vários métodos de estimação dos parâmetros da DPG para o caso específico do número de aberrações cromossômicas em células submetidas a diferentes doses de radiação. Eles compararam os métodos de momentos, MMV, método discrepância ponderada, método mínimo do qui-quadrado e MDEP. Nesse caso, os autores observaram que o MDEP tem o menor EQM e o menor erro médio absoluto para a maioria das doses de radiação.

**Exemplo 1** (Continuação) As estimativas da DPG pelo método da discrepância empírica ponderada foram obtidas por meio da função DEPDPG() desenvolvida no R. O algoritmo dessa função encontra-se no Apêndice D deste trabalho.

$$\begin{aligned} \hat{\alpha}_A &= 0.4118, & \hat{\lambda}_A &= 0.02808, \\ \hat{\alpha}_T &= 1.4500 & \text{e} & \hat{\lambda}_T = 0.5622. \end{aligned}$$

### 5.3.5 Estudo de Monte Carlo

Simulamos 4800 replicações da DPG para cada triáde formada pelo tamanho da amostra  $n = \{20, 50, 70, 120\}$  e pelos parâmetros  $\alpha = \{0.8, 1, 2, 5, 7, 15, 20, 50\}$  e  $\lambda = \{0.1, 0.2, 0.4, 0.6, 0.8, 0.95\}$ . Ao todo, foram 921.600 amostras geradas da DPG pela função rLGP do pacote RMKdiscrete do R. Essas amostras foram submetidas aos 4 métodos de estimação de parâmetros citados neste capítulo. As estimativas dos parâmetros pelo MMV foram obtidas

por meio da função `cal_chisq_statistic` do pacote GPSEQ do R e as funções para as estimativas pelo MMo, FREQ0 e MDEP foram desenvolvidas no *software* R durante esta pesquisa e encontram-se no Apêndice D deste trabalho.

Para avaliar os métodos de estimação, foram feitas comparações utilizando o método da divergência de Kullback-Leibler (DKL). Esse método verifica a diferença entre a função de probabilidade com os parâmetros reais e a função de probabilidade com parâmetros estimados por meio da relação

$$DKL_{jk} = \sum_{x=0}^{\infty} (P_x(\alpha, \lambda) [\ln(P_x(\alpha, \lambda)) - \ln(P_x(\hat{\alpha}_{jk}, \hat{\lambda}_{jk}))]),$$

na qual  $j$  representa a replicação ( $j = \{1, \dots, 4800\}$ ) e  $k$  o método de estimação ( $k = \{1, 2, 3, 4\}$ ). Quanto menor o valor de DKL, menor é a discrepância entre a função de probabilidade com os parâmetros reais e a função de probabilidade com parâmetros estimados. Isto posto, considerou-se como o melhor estimador aquele com a menor média de  $DKL_{jk}$  nas replicações, em que

$$MDKL_k(\hat{\alpha}, \hat{\lambda}) = \frac{1}{4800} \sum_{j=1}^{4800} DKL_{jk}.$$

Além disso, foi verificado o método de estimação com menor variabilidade nos  $DKL_{jk}$ ,

$$VAR(DKL_k) = \frac{1}{4800} \sum_{j=1}^{4800} [DKL_{jk} - MDKL_k(\hat{\alpha}, \hat{\lambda})]^2.$$

O conjunto de tabelas apresentadas a seguir trazem os resultados obtidos com os dados simulados. Observa-se nas tabelas que, de acordo com os parâmetros  $\alpha$  e  $\lambda$  e o tamanho amostral, os métodos de estimação que alcançaram os menores médias dos  $DKL_{jk}$ . Cada método de estimação é representado por uma cor. Observa-se, por exemplo, que para amostras de tamanho 20,  $\alpha = 50$  e  $\lambda = 0.6$ , o método dos momentos obteve a menor média das divergências de Kullback-Leibler. Nota-se que à medida que o tamanho da amostra aumenta, o MMV tende a ser o de menor média para modelos com  $\alpha$  baixos e  $\lambda$  altos e também modelos com  $\alpha$  altos e  $\lambda$  baixos. O MDEP foi o de menor média especialmente quando  $\lambda = 0.4$ .

No Apêndice C deste trabalho é apresentada a avaliação dos métodos de estimação pelo viés e pelo EQM para cada um dos parâmetros. Verifica-se que, em algumas situações, o método mais indicado para o parâmetro  $\alpha$  não é o mais indicado para o parâmetro  $\lambda$ . Por exemplo, para  $\alpha = 0.8$  o método FREQ0 é o menos viesado para o parâmetro  $\alpha$ , já para o parâmetro  $\lambda$  o MDEP é o menos viesado. Os EQMs dos parâmetros apresentaram resultados semelhantes. O MMo tem o menor EQM para boa parte das situações quando o tamanho da amostra é pequeno, à medida que o tamanho da amostra aumenta o método de máxima verossimilhança torna-se o mais frequente, ou seja, o que tem o menor EQM.

Tabela 4 – Método de estimação com a menor divergência de Kullback-Leibler. Simulação com 20 unidades amostrais.

$\alpha$	$\lambda$					
	0,1	0,2	0,4	0,6	0,8	0,95
0,8	MMo	MMo	MDEP	MDEP	MDEP	MDEP
1	MMo	MMo	MDEP	MDEP	MDEP	MMV
2	MMo	MMV	MDEP	FREQ0	MMV	MMV
5	MMV	MMV	MDEP	MMV	MMV	MMV
7	MMV	MMV	MDEP	MMo	MMV	MDEP
10	MMV	MMV	MDEP	MMo	MMV	MDEP
15	MMV	MMV	MMV	MMo	MMV	MDEP
20	MMV	MMV	MMV	MMo	MMo	MDEP
50	MMV	MMV	MDEP	MMo	MMo	MMV

Tabela 5 – Método de estimação com a menor divergência de Kullback-Leibler. Simulação com 50 unidades amostrais.

$\alpha$	$\lambda$					
	0,1	0,2	0,4	0,6	0,8	0,95
0,8	MMo	MMo	MDEP	MDEP	MMV	MMV
1	MDEP	MMo	MDEP	MDEP	MMV	MMV
2	MMV	MDEP	MDEP	MMV	MMV	MMV
5	MMV	MMV	MDEP	MMV	MMV	MMV
7	MMV	MMV	MDEP	MMV	MMV	MMV
10	MMV	MMV	MDEP	MMo	MMV	MMV
15	MMV	MMV	MDEP	MMo	MMV	MMV
20	MMV	MMV	MDEP	MMo	MMV	MMV
50	MMV	MMV	MDEP	MMo	MMo	MDEP

Tabela 6 – Método de estimação com a menor divergência de Kullback-Leibler. Simulação com 70 unidades amostrais.

$\alpha$	$\lambda$					
	0,1	0,2	0,4	0,6	0,8	0,95
0,8	MDEP	MDEP	MDEP	MMV	MMV	MMV
1	MDEP	MDEP	MDEP	MMV	MMV	MMV
2	MMV	MDEP	MDEP	MMV	MMV	MMV
5	MMV	MMV	MDEP	MMV	MMV	MMV
7	MMV	MMV	MDEP	MMV	MMV	MMV
10	MMV	MMV	MDEP	MMV	MMV	MMV
15	MMV	MMV	MDEP	MMo	MMV	MMV
20	MMV	MMV	MDEP	MMo	MMV	MMV
50	MMV	MMV	MDEP	MMo	MMo	MDEP

Tabela 7 – Método de estimação com a menor divergência de Kullback-Leibler. Simulação com 120 unidades amostrais.

$\alpha$	$\lambda$					
	0,1	0,2	0,4	0,6	0,8	0,95
0,8	MDEP	MDEP	MDEP	MMV	MMV	MMV
1	MDEP	MDEP	MDEP	MMV	MMV	MMV
2	MMV	MDEP	MDEP	MMV	MMV	MMV
5	MMV	MMV	MDEP	MMV	MMV	MMV
7	MMV	MMV	MDEP	MMV	MMV	MMV
10	MMV	MMV	MDEP	MMV	MMV	MMV
15	MMV	MMV	MMV	MMo	MMV	MMV
20	MMV	MMV	MMV	MMo	MMV	MMV
50	MMV	MMV	MDEP	MMo	MMo	MDEP

## 5.4 Testes de hipóteses

### 5.4.1 Teste da razão de verossimilhanças

Seja uma amostra com função de verossimilhança definida em (5.7). Para testar as hipóteses

$$\begin{cases} H_0 : \lambda = \lambda_0 \\ H_a : \lambda \neq \lambda_0 \end{cases} \quad (5.30)$$

a estatística do teste da razão de verossimilhança será

$$L_r(\lambda) = \frac{L(\lambda_0, \tilde{\alpha}(\lambda_0))}{L(\hat{\lambda}, \hat{\alpha})}, \quad (5.31)$$

na qual  $\tilde{\alpha}(\lambda_0) = \bar{x}(1 - \lambda)$  é a raiz da equação (5.10) para  $\lambda = \lambda_0$ .

Portanto,

$$\begin{aligned} L_r(\lambda) &= \frac{l(\lambda_0, \tilde{\alpha}(\lambda_0))}{l(\hat{\lambda}, \hat{\alpha})} \\ &= \left[ \frac{\tilde{\alpha}(\lambda_0)}{\hat{\alpha}} \right]^n \exp\{n[\hat{\alpha} - \tilde{\alpha}(\lambda_0) + (\hat{\lambda} - \lambda_0)\bar{x}]\} \prod_{i=1}^n \left[ \frac{(\tilde{\alpha}(\lambda_0) + \lambda_0 X_i)}{\hat{\alpha} + \hat{\lambda} X_i} \right]^{x_i - 1} \end{aligned} \quad (5.32)$$

Casella e Berger (2010, pág. 438) dizem que  $-2\ln(L_r)$  se aproxima assintoticamente

de uma distribuição Qui-Quadrado com  $d$  graus de liberdade,

$$D(\lambda) = -2\{\ln[L_r(\lambda)]\} \sim \chi_{\gamma, d}^2, \quad (5.33)$$

em que  $d$  é a dimensão do espaço paramétrico menos a dimensão da hipótese nula, que aqui é  $2-1=1$ .

Então, para um teste de hipótese em que se deseja testar um parâmetro, a função  $D(\lambda)$  se aproxima de uma distribuição qui-quadrado com 1 grau de liberdade,

$$D_r(\lambda) = -2\{\ln[L(\lambda_0, \tilde{\alpha}(\lambda_0))] - \ln[L(\hat{\alpha}, \hat{\lambda})]\} \stackrel{approx.}{\sim} \chi_{\gamma, 1}^2. \quad (5.34)$$

Logo,

$$\begin{aligned} D_r(\lambda) = & -2n \left\{ \ln[\tilde{\alpha}(\lambda_0)] - \ln(\hat{\alpha}) - (\tilde{\alpha}(\lambda_0) - \hat{\alpha}) - \bar{x}[\lambda_0 - \hat{\lambda}] \right\} \\ & - 2 \sum_{i=1}^n (x_i - 1) \ln \left[ \frac{\tilde{\alpha}(\lambda_0) + \lambda_0 x_i}{\hat{\alpha} + \hat{\lambda} x_i} \right]. \end{aligned} \quad (5.35)$$

A hipótese nula será rejeitada ao nível de significância  $\gamma$  se  $D_r(\lambda) > \chi_{\gamma, 1}^2$ .

Os cálculos para testar as hipóteses do parâmetro  $\alpha$  são semelhantes.

Sejam as hipóteses

$$\begin{cases} H_0 : \alpha = \alpha_0 \\ H_a : \alpha \neq \alpha_0. \end{cases} \quad (5.36)$$

A estatística do teste é dada por

$$\begin{aligned} D_r(\alpha) = & -2n \left\{ \ln(\alpha_0) - \ln(\hat{\alpha}) - (\alpha_0 - \hat{\alpha}) - \bar{x}[\tilde{\lambda}(\alpha_0) - \hat{\lambda}] \right\} \\ & - 2 \sum_{i=1}^n (x_i - 1) \ln \left[ \frac{\alpha_0 + \tilde{\lambda}(\alpha_0) x_i}{\hat{\alpha} + \hat{\lambda} x_i} \right] \end{aligned} \quad (5.37)$$

na qual  $\tilde{\lambda}(\alpha_0) = 1 - \frac{\alpha_0}{\bar{x}}$  é o valor de  $\lambda$  que maximiza a equação (5.10) quando  $\alpha = \alpha_0$ .

A hipótese nula será rejeitada ao nível de significância  $\gamma$  se  $D_r(\alpha) > \chi_{\gamma, 1}^2$ .

**Exemplo 1** (continuação). Conforme visto, o valor da estimativa de  $\lambda_T$  assume valores próximos de zero, levantando a suspeita de que a distribuição mais adequada para esse conjunto de dados seja a distribuição de Poisson. Logo, sob as hipóteses

$$\begin{cases} H_0 : \lambda_A = 0 \\ H_a : \lambda_A \neq 0 \end{cases} \quad (5.38)$$

a estatística do teste da razão de verossimilhança será

$$\begin{aligned} D_r(\lambda_A) &= -2[\ln(L_r(\lambda_A))] \\ &= -2\ln \left[ \frac{L(\lambda_{A0}, \tilde{\alpha}_A(\lambda_{A0}))}{L(\hat{\lambda}_A, \hat{\alpha}_A)} \right] = -2\{\ln [L(0, 0.425)] - \ln [L(0.0320, 0.4114)]\} \\ &= 0.48036. \end{aligned} \tag{5.39}$$

Obtendo  $\chi_{0.05,1}^2 = 3.84$ ,

$$D_r(\lambda) < \chi_{0.05,1}^2$$

Portanto, para um nível de significância de 0.05, aceita-se  $H_0$ . O teste indica que a distribuição de Poisson é mais adequada para descrever a distribuição de aranhas em placas.

### 5.4.2 Testes de qualidade de ajuste

Para dados discretos, o teste mais usual para verificar a adequação ou discrepância dos dados a uma distribuição é o teste qui-quadrado de aderência. Além de testar o ajuste dos dados a DPG, é possível averiguar se os dados se ajustam à distribuição Poisson e à binomial negativa.

Seja uma amostra aleatória  $X_1, X_2, \dots, X_n$  iid de tamanho  $n$ .

As hipóteses do teste serão

$$\begin{cases} H_0: \text{A variável aleatória } X \text{ segue uma distribuição } P_x \\ H_1: \text{A variável aleatória } X \text{ não segue uma distribuição } P_x. \end{cases} \tag{5.40}$$

A estatística utilizada para testar a qualidade do ajuste é dada por

$$Q = \sum_{x=0}^k \frac{(n_x - nP_x)^2}{nP_x}.$$

Seja  $P_x$  a probabilidade associada a classe  $x$ , tal que  $x = \{0, \dots, k\}$ , e seja  $k_c$  o total de classes de  $x$ ,  $k_c$  será igual a  $k$  se não existir a classe  $x = 0$ . A estatística do teste  $Q$  se aproxima a uma distribuição qui-quadrado com  $k_c - t - 1$  graus de liberdade, em que  $t$  é o número de parâmetros estimados no modelo.

$$Q \sim \chi_{\gamma, k_c - t - 1}^2$$

Com um nível de significância de  $\gamma$  e com  $k_c - t - 1$  graus de liberdade, obtém-se o valor da distribuição qui-quadrado padrão que define a região de rejeição.

Se  $Q > \chi_{\gamma, k_c - t - 1}^2$ , rejeita-se a hipótese nula.

Se  $Q < \chi_{\gamma, k_c - t - 1}^2$ , aceita-se a hipótese nula.

Valores de  $Q$  altos indicam distância maior entre os valores esperado e observado e, portanto, um ajuste mais pobre dos dados à distribuição.

**Exemplo 1** (continuação). A Tabela 8 apresenta os testes de qualidade de ajuste que verificam se a distribuição das aranhas em placas se adéqua ao modelo de Poisson generalizado e ao modelo Poisson. As hipóteses serão:

$$\left\{ \begin{array}{l} H_0 : \lambda_A = \lambda_{A0}, \quad \text{as aranhas se organizam sobre placas segundo} \\ \quad \text{determinada distribuição (DPG ou Poisson)} \\ H_a : \lambda_A \neq \lambda_{A0}, \quad \text{as aranhas não se organizam sobre placas segundo} \\ \quad \text{determinada distribuição (DPG ou Poisson)} \end{array} \right. \quad (5.41)$$

Tabela 8 – Testes de qualidade de ajuste segundo o modelo Poisson e Poisson generalizado para o conjunto de dados de 102 aranhas em 240 placas.

Quantidade por placa	Observado	Esperado	
		DPG	Dist. de Poisson
0	159	159.053	156.905
1	64	63.374	66.685
2	13	14.590	14.170
3	4	2.551	2.007
$\geq 4$	0	0.433	0.233
Total	240	240	240
EMV		$\alpha_A = 0.41, \lambda_{A0} = 0.03$	$\alpha_A = 0.425, \lambda_{A0} = 0$
$Q$		0.5257	1.615
g.l.		1	2
$\chi_{0.05, gl}^2$		3.84	5.99
Decisão		Accepta $H_0$	Accepta $H_0$

Com nível de significância de 0.05, os testes indicam que o modelo de organização das aranhas se ajusta à distribuição de Poisson e à DPG. Como o modelo de Poisson é o mais simples e o valor de  $\lambda$  na DPG é próximo de zero, indica-se assumir que o modelo de organização das aranhas segue a distribuição de Poisson.

## 5.5 Estimação Intervalar

Em Famoye e Consul (1990), são discutidos o intervalo de confiança (IC) para a DPG. Os autores consideraram o tamanho da amostra e o conhecimento de um dos parâmetros para observar o IC sob diferentes cenários.

Considere uma amostra com média  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ , desvio-padrão  $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})$  e função de verossimilhança  $L(\alpha, \lambda; \mathbf{x}) = \prod_{i=1}^n P_{x_i}(\alpha, \lambda)$ . Os resultados serão obtidos para o nível de confiança  $(1 - \gamma)100\%$ .

### IC quando um dos parâmetros é conhecido e a amostra é grande

Em Consul e Shenton (1973), verifica-se que, quando  $\alpha$  assume valores acima de 10, a DPG se aproxima de uma distribuição normal.

Supondo que não há conhecimento da média e do desvio-padrão da população e que  $\lambda$  é conhecido e assume o valor  $\lambda_0$ , o IC para  $\alpha$  será dado por

$$\begin{aligned}
 1 - \gamma &= P_x \left( -z_{\gamma/2} < \frac{\sqrt{n}(\bar{X} - \mu)}{s} < z_{\gamma/2} \right) \\
 &= P_x \left( \bar{X} - \frac{s z_{\gamma/2}}{\sqrt{n}} < \mu < \bar{X} + \frac{s z_{\gamma/2}}{\sqrt{n}} \right) \\
 &= P_x \left( \bar{X} - \frac{s z_{\gamma/2}}{\sqrt{n}} < \frac{\alpha}{1 - \lambda_0} < \bar{X} + \frac{s z_{\gamma/2}}{\sqrt{n}} \right) \\
 &= P_x \left( (1 - \lambda_0) \left[ \bar{X} - \frac{s z_{\gamma/2}}{\sqrt{n}} \right] < \alpha < (1 - \lambda_0) \left[ \bar{X} + \frac{s z_{\gamma/2}}{\sqrt{n}} \right] \right), \quad (5.42)
 \end{aligned}$$

na qual  $z_{\gamma/2}$  é o ponto em que a probabilidade acumulada de distribuição normal padrão é igual a  $\gamma/2$ , isto é,  $\int_{-\infty}^z \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2}) dx = \frac{\gamma}{2}$ . Logo, o IC de  $\alpha$  com nível de significância de  $\gamma$  será

$$IC \left[ (1 - \lambda_0) \left( \bar{x} - \frac{s z_{\gamma/2}}{\sqrt{n}} \right); (1 - \lambda_0) \left( \bar{x} + \frac{s z_{\gamma/2}}{\sqrt{n}} \right) \right]. \quad (5.43)$$

Supondo agora que a variância da população (vista na equação (3)) é conhecida e que o valor de  $\lambda$  também é conhecido e é igual a  $\lambda_0$ . Desse modo, o IC para  $\alpha$  será dado por

$$\begin{aligned}
1 - \gamma &= P_x \left( -z_{\gamma/2} < \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} < z_{\gamma/2} \right) \\
&= P_x \left( -z_{\gamma/2} < \frac{\sqrt{n}(\bar{X}(1 - \lambda_0) - \alpha)(1 - \lambda_0)^{\frac{1}{2}}}{\sqrt{\alpha}} < z_{\gamma/2} \right) \\
&= P_x \left( -\frac{z_{\gamma/2}}{\sqrt{n}(1 - \lambda_0)^{\frac{1}{2}}} < \frac{(\bar{X}(1 - \lambda_0) - \alpha)}{\sqrt{\alpha}} < \frac{z_{\gamma/2}}{\sqrt{n}(1 - \lambda_0)^{\frac{1}{2}}} \right) \\
&= P_x \left( \bar{X}(1 - \lambda_0) + \frac{z_{\gamma/2}^2}{2n(1 - \lambda_0)} \right. \\
&\quad \left. - \frac{z_{\gamma/2}}{2\sqrt{n}(1 - \lambda_0)^{\frac{1}{2}}} \sqrt{4\bar{X}(1 - \lambda_0) + \frac{z_{\gamma/2}^2}{n(1 - \lambda_0)}} < \alpha < \right. \\
&\quad \left. \bar{X}(1 - \lambda_0) + \frac{z_{\gamma/2}^2}{2n(1 - \lambda_0)} \right. \\
&\quad \left. + \frac{z_{\gamma/2}}{2\sqrt{n}(1 - \lambda_0)^{\frac{1}{2}}} \sqrt{4\bar{X}(1 - \lambda_0) + \frac{z_{\gamma/2}^2}{n(1 - \lambda_0)}} \right). \tag{5.44}
\end{aligned}$$

IC quando os dois parâmetros são desconhecidos e o tamanho da amostra é grande

Para se obter o IC de um dos parâmetros, deve-se eliminar da inferência o parâmetro que não é de interesse. Para a parametrização descrita em (2.1), Famoye e Consul (1990) sugeriram o método de eliminação de maximização da verossimilhança.

Seja a função de verossimilhança dada por

$$L(\alpha, \lambda) = \alpha^n e^{n\alpha + \lambda \sum_{i=1}^n x_i} \prod_{i=1}^n \left[ \frac{(\alpha + \lambda x_i)^{x_i - 1}}{x_i!} \right].$$

Sejam os estimadores de máxima verossimilhança  $\hat{\alpha}$  e  $\hat{\lambda}$ , então a razão de verossimilhança será

$$\begin{aligned}
L_r(\alpha, \lambda) &= \frac{L(\alpha, \lambda)}{L(\hat{\alpha}, \hat{\lambda})} \\
&= \left[ \frac{\alpha}{\hat{\alpha}} \right]^n \exp\{n[\hat{\alpha} - \alpha + (\hat{\lambda} - \lambda)\bar{x}]\} \prod_{i=1}^n \left[ \frac{(\alpha + \lambda x_i)}{\hat{\alpha} + \hat{\lambda} x_i} \right]^{x_i - 1}. \tag{5.45}
\end{aligned}$$

Para o IC de  $\alpha$ ,  $\lambda$  será substituído por  $\tilde{\lambda}(\alpha)$ , em que esse é, para um dado valor de  $\alpha$ ,

o valor que maximiza a equação (5.10),

$$\begin{aligned} L_r(\alpha) &= \frac{L(\alpha, \tilde{\lambda}(\alpha))}{L(\hat{\alpha}, \hat{\lambda})} \\ &= \left[ \frac{\alpha}{\hat{\alpha}} \right]^n \exp\{n[\hat{\alpha} - \alpha + (\hat{\lambda} - \tilde{\lambda}(\alpha))\bar{x}]\} \prod_{i=1}^n \left[ \frac{(\alpha + \tilde{\lambda}(\alpha)x_i)}{\hat{\alpha} + \hat{\lambda}x_i} \right]^{x_i-1}. \end{aligned} \quad (5.46)$$

Dessa forma, aplicando a função  $\ln(\cdot)$  em (5.46) e utilizando a aproximação mostrada na função (5.33), obtém-se

$$D_r(\alpha) = -2\{\ln[L(\alpha, \tilde{\lambda}(\alpha))] - \ln[L(\hat{\alpha}, \hat{\lambda})]\} \sim \chi_{\gamma,1}^2 \quad (5.47)$$

Logo,

$$\begin{aligned} D_r(\alpha) &= -2n \left\{ \ln \left( \frac{\alpha}{\hat{\alpha}} \right) - (\alpha - \hat{\alpha}) - \bar{x}[\tilde{\lambda}(\alpha) - \hat{\lambda}] \right\} \\ &\quad - 2 \sum_{i=1}^n (x_i - 1) \ln \left[ \frac{\alpha + \tilde{\lambda}(\alpha)x_i}{\hat{\alpha} + \hat{\lambda}x_i} \right]. \end{aligned} \quad (5.48)$$

Com nível de significância de  $\gamma$ , o limite superior ( $\alpha_S$ ) e o inferior ( $\alpha_I$ ) do IC de  $\alpha$  são obtidos pela interseção da função  $D_r(\alpha)$  com a linha em que  $D_r(\alpha) = \chi_{\gamma,1}^2$ , conforme visto na Figura 5.

O IC para o parâmetro  $\lambda$  é encontrado de forma análoga,

$$\begin{aligned} L_r(\lambda) &= \frac{L(\tilde{\alpha}(\lambda), \lambda)}{L(\hat{\alpha}, \hat{\lambda})} \\ &= \left[ \frac{\tilde{\alpha}(\lambda)}{\hat{\alpha}} \right]^n \exp\{n[\hat{\alpha} - \tilde{\alpha}(\lambda) + (\hat{\lambda} - \lambda)\bar{x}]\} \prod_{i=1}^n \left[ \frac{(\tilde{\alpha}(\lambda) + \lambda x_i)}{\hat{\alpha} + \hat{\lambda}x_i} \right]^{x_i-1}. \end{aligned} \quad (5.49)$$

$$\begin{aligned} D_r(\lambda) &= -2n \left\{ \ln \left( \frac{\tilde{\alpha}(\lambda)}{\hat{\alpha}} \right) - (\tilde{\alpha}(\lambda) - \hat{\alpha}) - \bar{x}[\lambda - \hat{\lambda}] \right\} \\ &\quad - 2 \sum_{i=1}^n (x_i - 1) \ln \left[ \frac{\tilde{\alpha}(\lambda) + \lambda x_i}{\hat{\alpha} + \hat{\lambda}x_i} \right]. \end{aligned} \quad (5.50)$$

Observa-se na Figura 5 que os intervalos de confiança, representados pela linha vermelha pontilhada, são definidos pela interseção da função  $D_r(\cdot)$  e a reta  $D_r(\cdot) = \chi_{\gamma,1}^2 = 3.84$ .

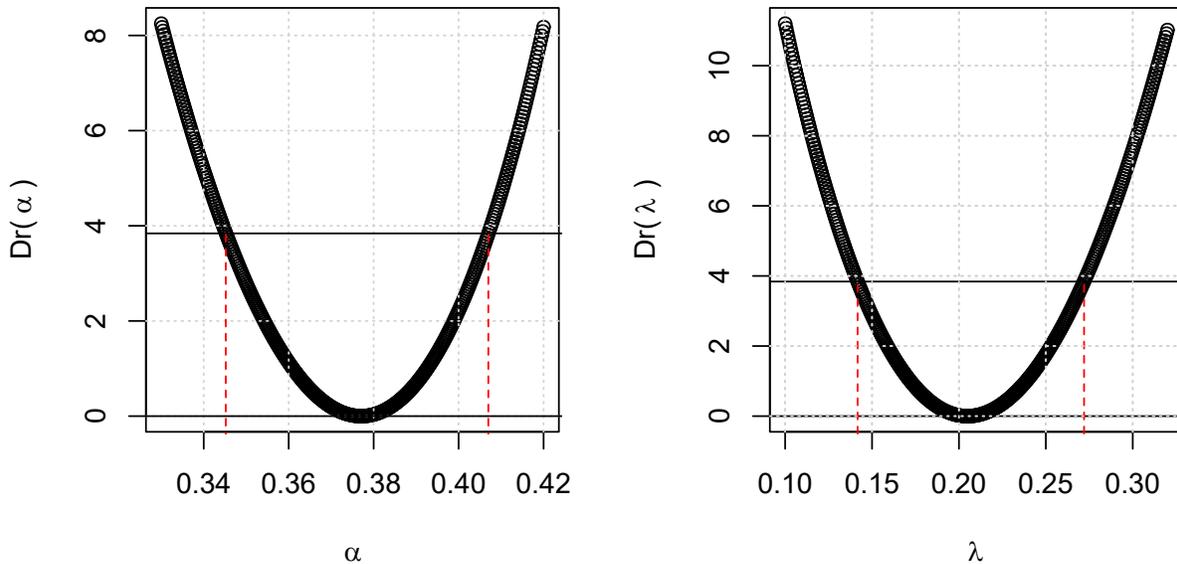


Figura 5 – Intervalos de confiança pelo método de eliminação de maximização da verossimilhança.

No eixo das ordenadas, estão os valores da função  $D_r(\cdot)$  e no eixo das abscissas os valores dos parâmetros. As linhas pontilhadas na cor vermelha são os valores da função  $D_r(\cdot)$  que interceptam a reta  $D_r(\cdot) = \chi_{\gamma,1}^2 = 3.84$  (Nível de significância de 0.05) e definem, portanto, os intervalos de confiança dos parâmetros.

**Exemplo 2:** Na aplicação apresentada na Tabela 3 do Capítulo 2, verifica-se a quantidade de cromossomos lesionados em células de coelhos que receberam diferentes doses da droga *streptonigrin*.

Comparando os quatro conjuntos de dados com diferentes dosagens, observa-se que, quanto maior a dosagem, maiores são a média e a variância, ou seja, quanto maior a dosagem maior o número de cromossomos lesionados em uma célula. Os EMVs do parâmetro  $\alpha$  são maiores para os conjuntos de dados com as maiores dosagens. Nas dosagens acima de  $60\mu g/kg$ , os EMVs de  $\lambda$  são próximos e os intervalos de confiança apresentam interseção. Esses resultados sugerem que, em dosimetria acima de  $60\mu g/kg$ , acréscimos na dosagem da droga interfere diretamente no aumento do parâmetro  $\alpha$ , no entanto, há pouca diferença nos valores de  $\lambda$ .

Tabela 9 – Intervalo de confiança para os parâmetros da DPG para o número de cromossomos lesionados em células de coelhos pelo uso de *streptonigrin* com diferentes dosagens.

Classes	Dose ( $\mu g/kg$ )			
	30	60	75	90
0	404	413	200	155
1	80	124	57	83
2	13	42	30	33
3	3	15	7	14
4	0	5	4	11
5	0	0	0	3
6	0	2	2	1
Total	500	601	300	300
$\bar{x}$	0.23	0.47	0.55	0.85
$s^2$	0.27	0.74	0.95	1.37
EMV de $\alpha$	0.2136	0.377	0.4136	0.6627
$IC_{0.05,\alpha}$	(0.194,0.229)	(0.345,0.407)	(0.362,0.462)	(0.592,0.731)
EMV de $\lambda$	0.0715	0.205	0.2525	0.2234
$IC_{0.05,\lambda}$	(0.005,0.154)	(0.142,0.272)	(0.165,0.345)	(0.144,0.306)

## 6 Ferramentas computacionais no *software R* para a DPG

A linguagem R é uma ferramenta útil em ambientes de análises numéricas. Trata-se de um ambiente na qual as técnicas estatísticas são implementadas. Tem uma linguagem de programação simples e eficaz que inclui condicionais, *loops*, funções recursivas definidas pelo usuário e recursos de entrada e saída. A vantagem do *software R* é a possibilidade da criação de funções, a expansão de funcionalidades por meio de pacotes, a utilização de operadores para cálculos em matrizes, a facilidade para a manipulação de dados e a grande quantidade de colaboradores, o que permite muitas contribuições na construção e na revisão de suas funcionalidades. O sistema R está disponível como *Software Livre* sob os termos da *Free Software Foundation's General Public License* em forma de código fonte. Ele compila e roda nas plataformas UNIX, Windows e Macintosh.

A linguagem R é hoje umas das mais populares no campo da estatística e, por esse motivo, este capítulo será dedicado aos pacotes e às funcionalidades que o R oferece à DPG.

### 6.1 GPseq

O GPseq foi construído por Srivastava e Chen (2011) e disponibilizado pela primeira vez em 2011. O desenvolvimento desse pacote ocorreu como forma alternativa ao modelo de Poisson para a detecção de genes diferencialmente expressos em diferentes condições de *splicing* de éxons (O RNA é constituído por uma longa cadeia formada pela união de éxons intercalados por introns. O processo de *splicing* retira os introns dessa cadeia e liga os éxons, formando o RNA mensageiro funcional ou maduro). Éxons são elementos codificantes, ou seja, que carregam o código que dá origem à proteína. A exclusão ou o acréscimo de um único nucleotídeo em um éxon pode levar à alteração da fase de leitura e à produção de uma proteína completamente diferente da original.

O pacote conta com 14 funções, a maioria delas relacionadas ao contexto descrito acima. Contudo, destacam-se duas funções para a DPG que podem ser utilizadas por diferentes tipos de dados. A parametrização da DPG desse pacote é a mesma descrita em (2.1).

- `calc_chisq_statistic(x, λ, α)` – Dado um conjunto de observações,  $x$ , e os parâmetros  $\alpha$  e  $\lambda$  da DPG, a função estima o parâmetro do modelo de Poisson pelo método de máxima

verossimilhança e calcula o teste qui-quadrado de bondade de ajuste comparando os ajustes à DPG e à distribuição de Poisson.

- `generalized_poisson_likelihood(y)` – A partir de um conjunto de observações,  $y$ , a função calcula as estimativas dos parâmetros  $\alpha$  e  $\lambda$  da DPG pelo método de máxima verossimilhança auxiliado pelo método iterativo de Newton Raphson.

## 6.2 HMMpa

O pacote HMMpa foi construído por Witowski e Foraita (2014) e disponibilizado pela primeira vez em 2013.

Esse pacote foi desenvolvido para classificar o grau de atividade física realizada por uma pessoa. O grau de atividade é separado por classes que são divididas por pontos de cortes. A escolha desses pontos de corte depende de diferentes componentes, como a idade do sujeito e o tipo de acelerômetro utilizado (aparelho com sensor de movimento que mede a quantidade de atividade física realizada pelo seu usuário). Os cortes podem ser calculados pelo método de ponto de corte e pelo método baseado em cadeias de Markov ocultas.

O pacote conta com várias funções relacionadas ao contexto descrito. No entanto, existem 3 funções que podem ser aplicadas a qualquer conjunto de dados que siga a DPG. A parametrização da DPG desse pacote é a mesma descrita em (2.1).

- `dgenpois(x,α,λ)` - Calcula a função de probabilidade  $f(x)$  em  $x$  a partir dos argumentos  $\alpha$  e  $\lambda$ .
- `pgenpois(x,α,λ)` - Calcula a função de probabilidade acumulada  $F(x)$  em  $x$  a partir dos argumentos  $\alpha$  e  $\lambda$ .
- `rgenpois(n,α,λ)` - Gera uma amostra aleatória de tamanho  $n$  com distribuição de probabilidade Poisson generalizada a partir dos argumentos  $\alpha$  e  $\lambda$ . Utiliza o método aceitação/rejeição para a maior parte dos conjuntos de dados gerados. Uma explicação deste método é vista em Frey e Neto (2005, pág. 87). As exceções são os dados cujos parâmetros estão nos intervalos ( $\alpha \geq 10$ ;  $\lambda < 0$ ) e ( $\alpha \geq 30$ ;  $0 < \lambda < 0.2$ ) que são obtidos por:

$$x_i = \max(0, \text{floor}(\mu + (\sigma^2 * \text{rnorm}(1, 0, 1) + 0, 5))) \quad i = 1, \dots, n$$

em que `rnorm(1,0,1)` é uma função do R que gera, segundo os argumentos fornecidos, um número aleatório de uma distribuição normal padrão, e `floor(.)` é uma função que, a partir do valor fornecido como argumento, retorna o menor valor inteiro menor ou igual ao argumento.

Um limitante da função `pgenpois` desse pacote é que ela não permite o emprego de vetores no argumento de entrada  $x$ .

## 6.3 RMKdiscrete

O pacote `RMKdiscrete` foi construído por Kirkpatrick (2014) e disponibilizado pela primeira vez em 2014.

Esse pacote foi desenvolvido para a implementação de modelos univariados e bivariados das distribuições de probabilidade de Poisson generalizada (nomeada por Kirkpatrick de distribuição Lagrangiana de Poisson) e binomial negativa. O pacote conta com 16 funções, 8 delas referentes à DPG univariada e outras 3 à DPG bivariada.

Considera a parametrização descrita em (2.1). No entanto, permite que  $\lambda$  assumam valores negativos, ou seja,  $\max(-1, -\alpha/m) < \lambda < 1$ . Quando  $\lambda$  é negativo, há um limite superior,  $m$ , da variável aleatória  $X$ , em que  $m = -\alpha/\lambda$  é arredondado para o próximo menor inteiro. Caso  $\alpha = 0$ , define-se a distribuição como tendo massa unitária, no caso  $X = 0$ .

- `dLGP(x,  $\alpha$ ,  $\lambda$ , nc=NULL, log=FALSE)` - Calcula a função de probabilidade  $f(x)$  em  $x$  a partir dos argumentos  $\alpha$  e  $\lambda$ .
- `pLGP(x,  $\alpha$ ,  $\lambda$ , nc=NULL, lower.tail=TRUE, log.p=FALSE, add.carefully=FALSE)` - Calcula a função de probabilidade acumulada  $F(x)$  em  $x$  a partir dos argumentos  $\alpha$  e  $\lambda$ .
- `qLGP(p,  $\alpha$ ,  $\lambda$ , nc=NULL, lower.tail=TRUE, log.p=FALSE, add.carefully=FALSE)` - Calcula o quantil que corresponde à probabilidade  $p$ .
- `rLGP(n,  $\alpha$ ,  $\lambda$ )` - Gera uma amostra aleatória de tamanho  $n$  com distribuição de probabilidade Poisson generalizada com parâmetros  $\alpha$  e  $\lambda$ .
- `sLGP( $\alpha$ ,  $\lambda$ , nc=NULL, do.numerically=FALSE, add.carefully=FALSE)` - Calcula a média, mediana, moda, variância, desvio-padrão, terceiro momento central, quarto momento central, coeficiente de assimetria de Pearson e de curtose.
- `LGP.findmax( $\alpha$ ,  $\lambda$ )` - Calcula o limite superior da variável aleatória  $X$  quando  $\lambda$  assume valores negativos.
- `LGP.get.nc( $\alpha$ ,  $\lambda$ , nctol=1e-14, add.carefully=FALSE)` - Calcula o inverso da constante de normalização e fornece a soma de todas as probabilidades. Quando  $\lambda$  é positivo, a

função retorna o valor 1. Essa função é especialmente útil quando  $\lambda$  é negativo e a soma de todas as probabilidades não é igual a 1.

- `LGPMVP( $\mu, \sigma^2, \alpha, \lambda$ )` - Com a informação de 2 dos 4 argumentos, fornece os valores dos outros 2 argumentos. A função utiliza o método dos momentos para encontrar um dos parâmetros e com esse resultado encontra o outro argumento utilizando o método de máxima verossimilhança.

A DPG bivariada é construída a partir de três variáveis aleatórias independentes,  $X_0$ ,  $X_1$  e  $X_2$ .

$$X_0 \sim DPG(\alpha_0, \lambda_0)$$

$$X_1 \sim DPG(\alpha_1, \lambda_1)$$

$$X_2 \sim DPG(\alpha_2, \lambda_2)$$

As variáveis de interesse  $Y_1$  e  $Y_2$  são escritas em função de  $X_i$ ,  $i = 1, 2, 3$ :

$$Y_1 = X_0 + X_1$$

$$Y_2 = X_0 + X_2$$

Assim, a dependência entre  $Y_1$  e  $Y_2$  está no termo comum  $X_0$ . A função de probabilidade conjunta de  $Y_1$  e  $Y_2$  é derivada da função de probabilidade conjunta das três variáveis independentes. Dessa forma, as funções relacionadas à DPG bivariada são:

- `dbiLGP( $y, \alpha, \lambda, nc=NULL, log=FALSE, add.carefully=FALSE$ )` - Calcula a função de probabilidade da DPG bivariada em  $y$  a partir dos argumentos  $\alpha$  e  $\lambda$ .
- `biLGP.logMV( $\alpha, \lambda, nc=NULL, const.add=1, tol=1e-14, add.carefully=FALSE$ )` - Calcula numericamente as médias, as variâncias e as covariâncias de uma distribuição DPG bivariada.
- `rbiLGP( $n, \alpha, \lambda$ )` - Gera amostras aleatórias de tamanho  $n$  com distribuição de probabilidade Poisson generalizada bivariada.

Na qual  $y = [Y_1 \ Y_2]$ ,  $\alpha = [\alpha_0 \ \alpha_1 \ \alpha_2]$  e  $\lambda = [\lambda_0 \ \lambda_1 \ \lambda_2]$ .

## 6.4 VGAM

O pacote VGAM foi construído por Yee (2016) e teve sua última versão publicada em maio de 2016.

Suas principais funcionalidades são para os ajustes de modelos lineares generalizados e aditivos (VGLMs e VGAMs), bem como modelo rank reduzido VGLMs (RR-VGLMs) e quadrático RR-VGLMs (QRR-VGLMs). Atualmente, VGAM está disponível em CRAN no site do R e funciona sob R 3.1.0 ou superior.

Funções relacionadas à DPG:

- `dgenpois(x, λ, α, log = FALSE)` - Calcula a função de probabilidade em  $x$  a partir dos argumentos  $\alpha$  e  $\lambda$ .
- `genpoisson(llambda = "rhobit", ltheta = "loge", ilambda = NULL, itheta = NULL, use.approx = TRUE, imethod = 1, ishrinkage = 0.95, zero = "lambda")` - É utilizado como argumento das funções `vglm` e `vgam` quando se quer ajustar um modelo linear generalizado ou aditivo segundo a família de DPG. As funções `vglm` e `vgam` também pertencem ao pacote do R VGAM.

A função que calcula  $f(x)$  nos pacotes HMMpa e VGAM tem o mesmo nome, no entanto, a ordem de entrada dos argumentos dos parâmetros é diferente. O pacote `RMKdiscrete` é o único que ajusta os valores de  $f(x)$  e  $F(x)$  com a constante de normalização quando  $\lambda$  é negativo e a soma de todas as probabilidades é diferente de 1.

## 6.5 lamW

O pacote `lamW` foi construído por Adler (2016) e disponibilizado pela primeira vez no CRAN do *software R* em março de 2016. O pacote conta com apenas duas funções, uma para o cálculo da função  $W$  de Lambert no intervalo principal  $[\frac{1}{e}, \infty)$  e outra para o intervalo secundário  $[-\frac{1}{e}, 0)$ .

- `lambertW0(x)` - retorna o valor da função  $W$  de Lambert em  $x$  do intervalo principal.
- `lambertWm1(x)` - retorna o valor da função  $W$  de Lambert em  $x$  do intervalo secundário.

## 6.6 LambertW

O pacote `LambertW` foi construído por Goerg (2016) e disponibilizado pela primeira vez no CRAN do *software R* em março de 2016. O pacote é a continuidade dos estudos re-

alizados por Goerg (2011) a respeito da variável aleatória  $W$  de Lambert. Entre as muitas funções do pacote, ressaltam-se as que simula amostras aleatórias (`rLambertW`), calcula  $f(x)$  e  $F(x)$  (`dLambertW` e `pLambertW`), estima os parâmetros pelo método de máxima verossimilhança (`IGMM` e `MLE_LambertW`) e calcula o quantil correspondente à probabilidade  $p$  (`qLambertW`). Infelizmente essas funções são exclusivas para distribuições de probabilidade contínuas. Desse modo, no âmbito da DPG, destaca-se a função que calcula o valor da função  $W$  de Lambert:

- `W(x, branch = 0)` - retorna o valor da função  $W$  de Lambert em  $x$ . O argumento `branch=0` indica a função no intervalo principal e `branch=-1` no intervalo secundário.

## 7 Conclusão

A DPG foi proposta por Consul e Jain em 1970 e 1973, desde então, publicações de diversas áreas têm demonstrado sua versatilidade e sua eficiência como alternativa a distribuição de Poisson para dados com sobredispersão. Neste trabalho estudamos a expansão de Lagrange como precursora para o desenvolvimento da CDL, que inclui a DPG como uma das distribuições mais relevantes dessa classe, e como importante resultado para o cálculo da função  $W$  de Lambert. Propomos uma parametrização da DPG em função da  $W$  de Lambert. Essa parametrização permitiu alocar a DPG na CDSP e utilizar as propriedades e características dessa classe de distribuição. Devido a complexidade do cálculo do somatório para a DPG, classificá-la na CDSP permitiu obter vários resultados de modo facilitado. Além disso, realizamos um estudo de monte Carlo comparando os métodos de estimação dos momentos, da máxima verossimilhança, da frequência de zeros e da discrepância empírica ponderada. Os resultados dessa simulação indicam que, de acordo com o tamanho amostral, existem regiões formadas pelos espaços paramétricos de  $\alpha$  e  $\lambda$  em que há predominância de um método de estimação sobre os demais. Esse resultado sugere estudos mais aprofundados que possam ser conclusivos em determinar o método de estimação mais apropriado para cada região.



# Apêndices



# APÊNDICE A – DPG Truncada

## A.1 Espaço paramétrico de $\lambda$

Os primeiros artigos de Consul e Jain relativo à DPG foram publicados em 1970 e 1973. Nos anos subsequentes, esses autores fizeram novas publicações especialmente para alterar a amplitude do intervalo do parâmetro  $\lambda$ . Em seu primeiro artigo, Consul e Jain definiram que  $\lambda$  poderia assumir valores negativos desde que  $\alpha + m\lambda > 0$ . Quando  $\lambda$  é negativo,  $m$  é o maior valor que  $x$  pode assumir para que expressão  $\alpha + m\lambda > 0$  seja válida.

A restrição do domínio da variável aleatória  $X$  quando  $\lambda$  assume valores negativos não foi suficiente para evitar que a soma de todas as probabilidades seja diferente de 1, conforme mostrado nos exemplos abaixo:

- Para  $\alpha = 0.55$  e  $\lambda = -0.5$ , os valores possíveis para a variável aleatória  $X$  serão  $x = \{0,1\}$ , então

$$\sum_{x=0}^1 P_x(\alpha, \lambda) = 0.5769498 + 0.5231762 = 1.100126$$

- Para  $\alpha = 1.6$  e  $\lambda = -0.75$ , os valores possíveis para a variável aleatória  $X$  serão  $x = \{0,1,2\}$ , então

$$\sum_{x=0}^2 P_x(\alpha, \lambda) = 0.2018965 + 0.6838639 + 0.07238699 = 0.9581474$$

Diante disso, Consul e Shoukri (1984) mudaram o intervalo de  $\lambda$  para  $0 \leq \lambda < 1$ , retirando, portanto, a característica de subdispersão e deixando órfãos vários conjuntos de dados que haviam se ajustado bem à DPG.

Para corrigir o problema da soma das probabilidades ser diferente de 1, foi proposta uma correção feita habitualmente em modelos truncados.

$$g_x(\alpha, \lambda) = \frac{P_x(\alpha, \lambda)}{\sum_{x=0}^m P_x(\alpha, \lambda)} = P_x(\alpha, \lambda) F_m(\alpha, \lambda)^{-1},$$

Em outro artigo, Consul e Shoukri (1985) fizeram um estudo simulado para verificar os efeitos do fator de correção  $F_m(\alpha, \lambda)^{-1}$  nas probabilidades, na média e na variância de uma DPG com  $\lambda$  negativo. Os autores concluíram que se o número de acontecimentos distintos com probabilidades não nulas fosse pelo menos cinco ou se o valor estimado de  $\alpha$  é menor que

0.7 ou maior que 4.5, é desnecessário empregar o termo  $F(\alpha, \lambda)^{-1}$  na DPG. Desse modo, foi sugerido que o fator de correção não fosse utilizado quando  $m \geq 4$ . Nos demais casos, deve-se aplicar o modelo corrigido, apesar da dificuldade de se determinar estimativas razoáveis para  $\alpha$  e  $\lambda$ .

Por esse motivo, muitos artigos apresentam a DPG da equação (2.1) com o domínio de  $\lambda$  sendo  $\max(-1, -\alpha/m) < \lambda < 1$  acrescido da restrição de que  $m \geq 4$  quando  $\lambda$  negativo.

## A.2 DPG truncada

A DPG não pode ser utilizada plenamente quando o segundo parâmetro,  $\lambda$ , é negativo, pois, em alguns casos, a soma das probabilidades em todo o intervalo da variável aleatória  $X$  não será igual a 1. Para evitar esses problemas, foi proposta, como alternativa, a Distribuição de Poisson Generalizada Truncada (DPGT).

$$P(X = x) = g_x(\alpha, \lambda) = \frac{P_x(\alpha, \lambda)}{K_m}$$

em que  $x = 0, 1, 2, \dots, m$  e

$$P_x(\alpha, \lambda) = \frac{\alpha(\alpha + x\lambda)^{x-1}}{x!} e^{-\alpha - x\lambda}$$

$$K_m = \sum_{x=0}^m P_x(\alpha, \lambda).$$

Portanto,

$$P_x(\alpha, \lambda) = \begin{cases} \frac{\alpha(\alpha + x\lambda)^{x-1}}{x!} e^{-\alpha - x\lambda} & \text{para } x = 0, 1, 2, \dots, m \\ \sum_{x=0}^m \frac{\alpha(\alpha + x\lambda)^{x-1}}{x!} e^{-\alpha - x\lambda} & \\ 0 & \text{caso contrário} \end{cases} \quad (\text{A.1})$$

na qual  $x = 0, 1, 2, \dots, m$ ,  $\alpha + m\lambda > 0$  e  $\alpha > 0$

As estimativas dos parâmetros para a DPGT envolvem técnicas numéricas e iterativas mais complexas do que as utilizadas na DPG.

# APÊNDICE B – Cálculo da média e variância

## B.1 Média e Variância

A média e a variância da DPG podem ser obtidas de várias formas. No Capítulos 4, foram obtidas pelo por meio dos momentos da CDSP escrito em função da  $W$  de Lambert. De modo a demonstrar o esforço que é necessário para o cálculo da maneira mais usual, neste apêndice a média e a variância serão calculadas da forma recursiva.

$$\begin{aligned}
 E(X) &= m_1(\alpha, \lambda) = \sum_{x=0}^{\infty} xP(x) = \sum_{x=1}^{\infty} \frac{x\alpha(\alpha + \lambda x)^{x-1}}{x!} e^{-\alpha-\lambda x} \\
 &= \sum_{x=1}^{\infty} \frac{\alpha(\alpha + \lambda x)^{x-1}}{(x-1)!} e^{-\alpha-\lambda x} \\
 &= \sum_{x=0}^{\infty} \frac{\alpha(\alpha + \lambda + \lambda x)^x}{x!} e^{-\alpha-\lambda-\lambda x} = \sum_{x=0}^{\infty} \frac{\alpha(\alpha + \lambda + \lambda x)(\alpha + \lambda + \lambda x)^{x-1}}{x!} e^{-\alpha-\lambda-\lambda x} \\
 &= \alpha \sum_{x=0}^{\infty} \frac{(\alpha + \lambda)(\alpha + \lambda + \lambda x)^{x-1}}{x!} e^{-\alpha-\lambda-\lambda x} + \frac{\alpha\lambda}{\alpha + \lambda} \sum_{x=0}^{\infty} \frac{x(\alpha + \lambda)(\alpha + \lambda + \lambda x)^{x-1}}{x!} e^{-\alpha-\lambda-\lambda x}
 \end{aligned} \tag{B.1}$$

$$\begin{aligned}
 &= \alpha + \frac{\alpha\lambda}{\alpha + \lambda} m_1(\alpha + \lambda, \lambda) \\
 &= \alpha + \frac{\alpha\lambda}{\alpha + \lambda} \left[ \alpha + \lambda + \frac{(\alpha + \lambda)\lambda}{\alpha + 2\lambda} m_1(\alpha + 2\lambda, \lambda) \right] \\
 &= \alpha + \alpha\lambda + \frac{\alpha\lambda^2}{\alpha + 2\lambda} m_1(\alpha + 2\lambda, \lambda) \\
 &= \alpha + \alpha\lambda + \frac{\alpha\lambda^2}{\alpha + 2\lambda} \left[ (\alpha + 2\lambda) + (\alpha + 2\lambda)\lambda + \frac{(\alpha + 2\lambda)\lambda^2}{\alpha + 2\lambda} m_1(\alpha + 3\lambda, \lambda) \right] \\
 &= \alpha + \alpha\lambda + \alpha\lambda^2 + \alpha\lambda^3 + \frac{\alpha\lambda^4}{\alpha + 3\lambda} m_1(\alpha + 3\lambda, \lambda)
 \end{aligned}$$

Após sucessivas transformações, obtém-se

$$\begin{aligned}
 &= \alpha(1 + \lambda + \lambda^2 + \lambda^3 + \dots) \\
 &= \alpha \sum_{k=0}^{\infty} \lambda^k
 \end{aligned}$$

Para  $|\lambda| < 1$ , obtém-se  $\sum_{k=0}^{\infty} \lambda^k = \frac{1}{1 - \lambda}$

Portanto,

$$E(X) = \frac{\alpha}{1 - \lambda}$$

A variância é obtida por procedimentos similares ao da média.

$$\begin{aligned}
E(X^2) &= m_2(\alpha, \lambda) = \sum_{x=0}^{\infty} x^2 P(x) = \sum_{x=1}^{\infty} \frac{x^2 \alpha (\alpha + \lambda x)^{x-1}}{x!} e^{-\alpha - \lambda x} \\
&= \sum_{x=0}^{\infty} \frac{(x+1) \alpha (\alpha + \lambda + \lambda x)^x}{x!} e^{-\alpha - \lambda - \lambda x} \\
&= \sum_{x=0}^{\infty} \frac{x \alpha (\alpha + \lambda + \lambda x)^x}{x!} e^{-\alpha - \lambda - \lambda x} + \sum_{x=0}^{\infty} \frac{\alpha (\alpha + \lambda + \lambda x)^x}{x!} e^{-\alpha - \lambda - \lambda x} \\
&= \sum_{x=0}^{\infty} \frac{x \alpha (\alpha + \lambda + \lambda x) (\alpha + \lambda + \lambda x)^{x-1}}{x!} e^{-\alpha - \lambda - \lambda x} + \frac{\alpha}{1 - \lambda} \\
&= \frac{\alpha}{1 - \lambda} + \alpha \sum_{x=0}^{\infty} \frac{x (\alpha + \lambda) (\alpha + \lambda + \lambda x)^{x-1}}{x!} e^{-\alpha - \lambda - \lambda x} \\
&\quad + \frac{\alpha \lambda}{(\alpha + \lambda)} \sum_{x=0}^{\infty} \frac{x^2 (\alpha + \lambda) (\alpha + \lambda + \lambda x)^{x-1}}{x!} e^{-\alpha - \lambda - \lambda x} \\
&= \frac{\alpha}{1 - \lambda} + \alpha \frac{\alpha + \lambda}{1 - \lambda} + \frac{\alpha \lambda}{\alpha + \lambda} m_2(\alpha + \lambda, \lambda) \\
&= \frac{\alpha}{1 - \lambda} + \alpha \frac{\alpha + \lambda}{1 - \lambda} + \frac{\alpha \lambda}{1 - \lambda} + \frac{\alpha \lambda (\alpha + 2\lambda)}{1 - \lambda} + \frac{\alpha \lambda^2}{\alpha + 2\lambda} m_2(\alpha + 2\lambda, \lambda) \\
&= \frac{\alpha}{1 - \lambda} + \alpha \frac{\alpha + \lambda}{1 - \lambda} + \frac{\alpha \lambda}{1 - \lambda} + \frac{\alpha \lambda (\alpha + 2\lambda)}{1 - \lambda} + \frac{\alpha \lambda^2}{1 - \lambda} \\
&\quad + \frac{\alpha \lambda^2 (\alpha + 3\lambda)}{1 - \lambda} + \frac{\alpha \lambda^3}{1 - \lambda} + \frac{\alpha \lambda^3 (\alpha + 4\lambda)}{1 - \lambda} + \frac{\alpha \lambda^4}{\alpha + 3\lambda} m_2(\alpha + 3\lambda, \lambda)
\end{aligned}$$

Após sucessivas transformações, obtém-se

$$\begin{aligned}
&= \frac{\alpha}{1 - \lambda} [1 + (\alpha + \lambda) + \lambda + \lambda(\alpha + 2\lambda) + \lambda^2 + \lambda^2(\alpha + 3\lambda) + \lambda^3 + \lambda^3(\alpha + 4\lambda) + \dots] \\
&= \frac{\alpha}{1 - \lambda} [(\alpha + 1)(\lambda + \lambda^2 + \lambda^3 + \dots) + (\lambda + 2\lambda^2 + 3\lambda^3 + 4\lambda^4 \dots)] \\
&= \frac{\alpha}{-\lambda} \left[ (\alpha + 1) \sum_{k=1}^{\infty} \lambda^k + \sum_{k=1}^{\infty} k \lambda^k \right] \\
&= \frac{\alpha^2}{(1 - \lambda)^2} + \frac{\alpha \lambda}{(1 - \lambda)^3} + \frac{\alpha}{(1 - \lambda)^2} \\
&= \frac{\alpha^2}{(1 - \lambda)^2} + \frac{\alpha}{(1 - \lambda)^3}
\end{aligned}$$

A variância da função de distribuição Poisson generalizada será dada por:

$$\sigma^2 = E(X^2) - [E(X)]^2 = \frac{\alpha}{(1 - \lambda)^3}$$

# APÊNDICE C – Resultado das simulações de Monte Carlo

## C.1 Estudo de Monte Carlo - Viés e EQM dos parâmetros



Método de estimação com o menor EQM para o parâmetro  $\alpha$ .  
Simulação com 20 unidades amostrais.

$\alpha$	$\lambda$					
	0,1	0,2	0,4	0,6	0,8	0,95
0,8	MMo	MMo	MMV	MMV	MMV	MMV
1	MMo	MMo	MMo	MMV	MMV	MMV
2	MMo	MMo	MMo	MMV	MMV	MMV
5	MMo	MMo	MMo	MMV	MMV	MMV
7	MMo	MMo	MMo	MMV	MMV	MMV
10	MMo	MMo	MMo	MMo	MMV	MMV
15	MMo	MMo	MMo	MMo	MMV	MMV
20	MMo	MMo	MMo	MMo	MMV	MMV
50	MMo	MMo	MMo	MMo	MMo	MMV

Método de estimação com o menor EQM para o parâmetro  $\lambda$ .  
Simulação com 20 unidades amostrais.

$\alpha$	$\lambda$					
	0,1	0,2	0,4	0,6	0,8	0,95
0,8	MMo	MMo	MMo	MMV	MDEP	FREQ0
1	MMo	MMo	MMo	MMV	MDEP	MMV
2	MMo	MMo	MMo	MMV	MMV	MMV
5	MMo	MMo	MMo	MMV	MMV	MMV
7	MMo	MMo	MMo	MMV	MMV	MMV
10	MMo	MMo	MMo	MMo	MMV	MMV
15	MMo	MMo	MMo	MMo	MMV	MMV
20	MMo	MMo	MMo	MMo	MMV	MMV
50	MMo	MMo	MMo	MMo	MMo	MMV

Método de estimação com o menor EQM para o parâmetro  $\alpha$ .  
Simulação com 50 unidades amostrais.

$\alpha$	$\lambda$					
	0,1	0,2	0,4	0,6	0,8	0,95
0,8	MMo	MMo	MMV	MMV	MMV	MMV
1	MMo	MMo	MMV	MMV	MMV	MMV
2	MMo	MMo	MMV	MMV	MMV	MMV
5	MMo	MMo	MMV	MMV	MMV	MMV
7	MMo	MMo	MMo	MMV	MMV	MMV
10	MMo	MMo	MMo	MMV	MMV	MMV
15	MMo	MMo	MMo	MMV	MMV	MMV
20	MMo	MMo	MMo	MMV	MMV	MMV
50	MMo	MMo	MMo	MMo	MMV	MMV

Método de estimação com o menor EQM para o parâmetro  $\lambda$ .  
Simulação com 50 unidades amostrais.

$\alpha$	$\lambda$					
	0,1	0,2	0,4	0,6	0,8	0,95
0,8	MMo	MMo	MMV	MMV	MMV	FREQ0
1	MMo	MMo	MMV	MMV	MMV	MMV
2	MMo	MMo	MMV	MMV	MMV	MMV
5	MMo	MMo	MMV	MMV	MMV	MMV
7	MMo	MMo	MMo	MMV	MMV	MMV
10	MMo	MMo	MMo	MMV	MMV	MMV
15	MMo	MMo	MMo	MMV	MMV	MMV
20	MMo	MMo	MMo	MMV	MMV	MMV
50	MMo	MMo	MMo	MMo	MMV	MMV

Método de estimação com o menor EQM para o parâmetro  $\alpha$ .  
Simulação com 70 unidades amostrais.

$\alpha$	$\lambda$					
	0,1	0,2	0,4	0,6	0,8	0,95
0,8	MMo	MMo	MMV	MMV	MMV	MMV
1	MMo	MMo	MMV	MMV	MMV	MMV
2	MMo	MMo	MMV	MMV	MMV	MMV
5	MMo	MMo	MMV	MMV	MMV	MMV
7	MMo	MMo	MMV	MMV	MMV	MMV
10	MMo	MMo	MMV	MMV	MMV	MMV
15	MMo	MMo	MMo	MMV	MMV	MMV
20	MMo	MMo	MMo	MMV	MMV	MMV
50	MMo	MMo	MMo	MMo	MMV	MMV

Método de estimação com o menor EQM para o parâmetro  $\lambda$ .  
Simulação com 70 unidades amostrais.

$\alpha$	$\lambda$					
	0,1	0,2	0,4	0,6	0,8	0,95
0,8	MMo	MMo	MMV	MMV	FREQ0	FREQ0
1	MMo	MMo	MMV	MMV	MMV	MMV
2	MMo	MMo	MMV	MMV	MMV	MMV
5	MMo	MMo	MMV	MMV	MMV	MMV
7	MMo	MMo	MMV	MMV	MMV	MMV
10	MMo	MMo	MMV	MMV	MMV	MMV
15	MMo	MMo	MMo	MMV	MMV	MMV
20	MMo	MMo	MMo	MMV	MMV	MMV
50	MMo	MMo	MMo	MMo	MMV	MMV

Método de estimação com o menor EQM para o parâmetro  $\alpha$ .  
Simulação com 120 unidades amostrais.

$\alpha$	$\lambda$					
	0,1	0,2	0,4	0,6	0,8	0,95
0,8	MMo	MMV	MMV	MMV	MMV	MMV
1	MMo	MMV	MMV	MMV	MMV	MMV
2	MMo	MMV	MMV	MMV	MMV	MMV
5	MMo	MMo	MMV	MMV	MMV	MMV
7	MMo	MMo	MMV	MMV	MMV	MMV
10	MMo	MMo	MMV	MMV	MMV	MMV
15	MMo	MMo	MMV	MMV	MMV	MMV
20	MMo	MMo	MMV	MMV	MMV	MMV
50	MMo	MMo	MMo	MMV	MMV	MMV

Método de estimação com o menor EQM para o parâmetro  $\lambda$ .  
Simulação com 120 unidades amostrais.

$\alpha$	$\lambda$					
	0,1	0,2	0,4	0,6	0,8	0,95
0,8	MMo	MMV	MMV	MMV	MMV	MMV
1	MMo	MMV	MMV	MMV	MMV	FREQ0
2	MMo	MMV	MMV	MMV	MMV	MMV
5	MMo	MMo	MMV	MMV	MMV	MMV
7	MMo	MMo	MMV	MMV	MMV	MMV
10	MMo	MMo	MMV	MMV	MMV	MMV
15	MMo	MMo	MMV	MMV	MMV	MMV
20	MMo	MMo	MMV	MMV	MMV	MMV
50	MMo	MMo	MMo	MMV	MMV	MMV



## APÊNDICE D – Algoritmo

Funções desenvolvidas no *software* R para a estimação dos parâmetros da DPG pelo método baseado na frequência de zeros e na média, dos momentos, da discrepância empírica ponderada e da máxima verossimilhança.

### D.1 Função para o método dos momentos

```
MomDPG=function(y){
  Media<-mean(y)
  Variancia<-var(y)
  if(Variancia==0){
    return(list(mark = 0, theta = NaN, lambda = NaN))
  }
  else {
    alpha<-Media*(sqrt(Media/Variancia))
    lambda<-1-(Media/Variancia)^(0.5)
    return(list(mark = 1, theta = alpha, lambda = lambda))
  }
}
```

### D.2 Função para o método baseado na frequência de zeros e na média

```
FREQ0DPG=function(y){
  a=table(y)
  N.o=0
  N.o=a[names(a)==0]
  if (sum(names(a)==0)==0){
    return(list(mark = 0, theta = NaN, lambda = NaN))
  }
  else {
    theta=-log(N.o/length(y))
    lambda=1-theta/mean(y)
    return(list(mark = 1, theta = theta, lambda = lambda))
  }
}
```

### D.3 Função para o método de máxima verossimilhança

```

EMVDPG=function(x){
  J <- matrix(0,2,2)
  B<- matrix(0,2,1)
  Y<- matrix(0,2,1)
  e<-1
  t<-0
  Y[1,1]<-mean(x)*(sqrt(mean(x)/var(x)))
  Y[2,1]<-1-(mean(x)/var(x))^(0.5)
  if(var(x)==0){
    return(list(mark = 0, theta = NaN, lambda = NaN,t=0))
  }
  else{
    Ym<-Y
    n<-length(x)
    xm<-c(0:max(x))
    obs<-rep(0,max(x)+1)
    for (i in 1:n) {
      obs[x[i] + 1] = obs[x[i] + 1] + 1
    }
    while(e>0.0000001 & t<1000){
      t<-t+1
      Px= try(exp((log(Y[1,1])+(xm-1)*log(Y[1,1]+Y[2,1]*xm)-Y[1,1]-Y[2,1]*xm)-
      lfactorial(xm)),silent = TRUE)
      if(sum(is.nan(Px))==0){
        J[1,1]<-(n*(Y[1,1]*(1-Y[2,1])+2*Y[2,1])/(Y[1,1]*(Y[1,1]+2*Y[2,1])))
        J[1,2]<-(n*Y[1,1])/(Y[1,1]+Y[2,1]*2)
        J[2,2]<-(n*Y[1,1]*(Y[1,1]+2))/(Y[2,1]*(Y[1,1]+2*Y[2,1]))
        J[2,1]<-J[1,2]
        B[1,1]<-(n/Y[1,1]+sum((xm-1)*obs/(Y[1,1]+Y[2,1]*xm))-n)
        B[2,1]<-(sum((xm-1)*xm*obs/(Y[1,1]+Y[2,1]*xm))-n*mean(x))
        Yn<-Y+solve(J)%*%B
        e<-((Y[1,1]-Yn[1,1])^2)+((Y[2,1]-Yn[2,1])^2)
        Y<-Yn
      }

      else{
        return(list(mark = 0, theta = NaN, lambda = NaN))
      }
    }
    return(list(mark=1, theta= Y[1,1], lambda = Y[2,1], t = t,erro = e))}

```

## D.4 Função para o método de discrepância empírica ponderada

```

DEPDPG=function(x){
  J <- matrix(0,2,2)
  B<- matrix(0,2,1)
  Y<- matrix(0,2,1)
  e<-1
  t<-0
  Y[1,1]<-mean(x)*(sqrt(mean(x)/var(x)))
  Y[2,1]<-1-(mean(x)/var(x))^(0.5)
  if(var(x)==0){
    mark = 0
  }
  else{
    Ym<-Y
    n<-length(x)
    xm<-c(0:max(x))
    obs<-rep(0,max(x)+1)
    for (i in 1:n) {
      obs[x[i] + 1] = obs[x[i] + 1] + 1
    }
    fx=obs/n
    t=0
    while(e>0.00000001 && t<10000){
      t<-t+1
      if(Y[1,1]>0 && sum((Y[1,1]+Y[2,1]*xm)<0)==0){
        Px= try(exp((log(Y[1,1])+(xm-1)*log(Y[1,1]+Y[2,1]*xm)-Y[1,1]-
        Y[2,1]*xm)-lfactorial(xm)),silent = TRUE)
        if(sum(is.nan(Px))==0){
          mark=1
          d=(Y[1,1]+Y[2,1]*xm)
          J[1,1]<-sum(((fx*((-1/Y[1,1]^2)-(xm-1)/d^2))-
          (Px*((1/Y[1,1]+(xm-1)/d)*(1/Y[1,1]+((xm-1)/d)-1)+(-1/(Y[1,1]^2)-
          (xm-1)/d^2))))))
          J[1,2]<-sum(((fx*(-xm*(xm-1)/d^2))-
          (Px*((1/Y[1,1]+(xm-1)/d)*((xm*(xm-1)/d)-xm))-((xm-1)*xm/(d^2))))))
          J[2,2]<-sum(((fx*(-((xm-1)*(xm^2))/(d^2)))-
          (Px*((((xm*(xm-1)/d)-xm)^2)-((xm-1)*(xm^2)/(d^2))))))
          J[2,1]<-J[1,2]
          B[1,1]<-sum(fx*(fx-Px)*(1/Y[1,1]+((xm-1)/d)-1))
          B[2,1]<-sum(fx*(fx-Px)*(((xm-1)*xm/d)-xm))
          Yn<-Y-solve(J)%*%B
          e<-((Y[1,1]-Yn[1,1])^2)+((Y[2,1]-Yn[2,1])^2)
          Y<-Yn
        }
      }
    }
  }
}

```

```
else{
mark = 0
break
}
}
else{
mark = 0
break
}
}}
if(mark==0){
theta = NaN
lambda = NaN
}
if(mark==1){
theta = Y[1,1]
lambda = Y[2,1]
}
return(list(mark=mark, theta=theta, lambda = lambda, t = t,erro = e))}
```

## Referências

- ADLER, A. *lamW: Lambert-W Function*. [S.l.], 2016. R package version 1.1.1. Disponível em: <<https://CRAN.R-project.org/package=lamW>>. Citado na página 65.
- BATTIN, R. *An Introduction to the Mathematics and Methods of Astrodynamics*. American Institute of Aeronautics & Astronautics, 1999. ISBN 9781600860263. Disponível em: <<https://books.google.com.br/books?id=OjH7aVhiGdcC>>. Citado na página 17.
- CASELLA, G.; BERGER, R. L. *Inferência estatística-tradução da 2ª edição norte-americana*. [S.l.]: São Paulo: Cengage Learning, 2010. Citado 2 vezes nas páginas 37 e 52.
- CONSUL, P. *Generalized Poisson Distributions: Applications and Properties*. 1rd. ed. New York: Marcel Dekker, 1989. Citado 8 vezes nas páginas 3, 9, 10, 11, 12, 15, 29 e 47.
- CONSUL, P. A model for distributions of injuries in auto-accidents. *Insurance: Mathematics and Economics*, North-Holland, v. 13, n. 2, p. 147, 1993. Citado na página 12.
- CONSUL, P.; FAMOYE, F. On the unimodality of generalized poisson distribution. *ASTIN Bulletin*, v. 40, n. 3, p. 117–122, 1986. Citado na página 35.
- CONSUL, P.; FAMOYE, F. Minimum variance unbiased estimation for the lagrange power series distributions. *Statistics*, v. 20, n. 3, p. 407–415, 1989. Disponível em: <<http://dx.doi.org/10.1080/02331888908802188>>. Citado na página 35.
- CONSUL, P.; FAMOYE, F. Generalized poisson regression model. *Communications in Statistics - Theory and Methods*, v. 21, n. 1, p. 89–109, 1992. Disponível em: <<http://dx.doi.org/10.1080/03610929208830766>>. Citado na página 12.
- CONSUL, P.; SHENTON, L. Some interesting properties of lagrangian distributions. *Communications in Statistics*, v. 2, n. 3, p. 263–272, 1973. Disponível em: <<http://dx.doi.org/10.1080/03610927308827073>>. Citado 2 vezes nas páginas 11 e 56.
- CONSUL, P.; SHOUKRI, M. Maximum likelihood estimation for the generalized poisson distribution. *Communications in statistics. Theory and methods*, Taylor & Francis, v. 13, n. 12, p. 1533–1547, 1984. Citado 2 vezes nas páginas 44 e 45.
- CONSUL, P. C.; FAMOYE, F. *Lagrangian probability distributions*. [S.l.]: Springer, 2006. Citado na página 49.
- CONSUL, P. C.; JAIN, G. C. A Generalization of the Poisson Distribution. *Technometrics*, [Taylor & Francis, Ltd., American Statistical Association, American Society for Quality], v. 15, n. 4, p. 791–799, 1970, 1973a. Citado 2 vezes nas páginas 1 e 9.
- CONSUL, P. C.; JAIN, G. C. On some interesting properties of the generalized poisson distribution. *Biometrische Zeitschrift*, WILEY-VCH Verlag, v. 15, n. 7, p. 495–500, 1973b. ISSN 1521-4037. Disponível em: <<http://dx.doi.org/10.1002/bimj.19730150707>>. Citado 2 vezes nas páginas 1 e 12.

- CONSUL, P. C.; SHENTON, L. R. Use of Lagrange Expansion for Generating Discrete Generalized Probability Distributions. *SIAM Journal on Applied Mathematics*, Society for Industrial and Applied Mathematics, v. 23, n. 2, p. 239–248, 1972. Citado 2 vezes nas páginas 1 e 15.
- CORLESS, R. M.; GONNET, G. H.; HARE, D. E. G.; JEFFREY, D. J. Lambert's  $w$  function in maple. *Maple Technical Newsletter*, v. 9, 1993. Citado na página 22.
- CORLESS, R. M.; GONNET, G. H.; HARE, D. E. G.; JEFFREY, D. J.; KNUTH, D. E. On the lambert  $w$  function. *Advances in Computational Mathematics*, v. 5, n. 1, p. 329–359, 1996. ISSN 1572-9044. Disponível em: <<http://dx.doi.org/10.1007/BF02124750>>. Citado na página 15.
- CRAMER, H. Mathematical methods of statistics. *Princeton University Press, Princeton*, 1946. Citado na página 39.
- FAMOYE, F.; CONSUL, P. Interval estimation and hypothesis testing for the generalized poisson distribution. *American Journal of Mathematical and Management Sciences*, Taylor & Francis, v. 10, n. 1-2, p. 127–158, 1990. Citado 3 vezes nas páginas 44, 55 e 57.
- FELIX, F.; LEE, C. M.-S. Estimation of generalized poisson distribution. *Communications in Statistics-Simulation and Computation*, Taylor & Francis, v. 21, n. 1, p. 173–188, 1992. Citado na página 49.
- FREY, A.; NETO, F. C. *Elementos de Estatística Computacional usando plataformas de software Livre*. [S.l.], 2005. 25o. Coloquio Brasileiro de Matematica, IMPA. Disponível em: <[http://www.impa.br/opencms/pt/biblioteca/pm/PM\\\_21.pdf](http://www.impa.br/opencms/pt/biblioteca/pm/PM\_21.pdf)>. Citado na página 62.
- GOERG, G. M. Lambert  $w$  random variables a new family of generalized skewed distributions with applications to risk estimation. *The Annals of Applied Statistics*, JSTOR, p. 2197–2230, 2011. Citado na página 66.
- GOERG, G. M. *LambertW: An R package for Lambert W x F Random Variables*. [S.l.], 2016. R package version 0.6.4. Disponível em: <<http://CRAN.R-project.org/package=LambertW>>. Citado na página 65.
- GOULD, H. W. Euler's formula for  $n$ th differences of powers. *The American Mathematical Monthly*, Mathematical Association of America, v. 85, n. 6, p. 450–467, 1978. ISSN 00029890, 19300972. Disponível em: <<http://www.jstor.org/stable/2320064>>. Citado na página 29.
- ISLAM, M. M.; ALAM, M.; TARIQUZAMAN, M.; KABIR, M. A.; PERVIN, R.; BEGUM, M.; KHAN, M. M. H. Predictors of the number of under-five malnourished children in bangladesh: application of the generalized poisson regression model. *BMC Public Health*, v. 13, n. 1, p. 11, 2013. ISSN 1471-2458. Disponível em: <<http://dx.doi.org/10.1186/1471-2458-13-11>>. Citado na página 12.
- JANARDAN, K. G.; KERSTER, H. W.; SCHAEFFER, D. J. Biological applications of the lagrangian poisson distribution. *BioScience*, [American Institute of Biological Sciences,

Oxford University Press], v. 29, n. 10, p. 599–602, 1979. ISSN 00063568, 15253244. Disponível em: <<http://www.jstor.org/stable/1307766>>. Citado na página 10.

JANARDAN, K. G.; SCHAEFFER, D. J. Models for the analysis of chromosomal aberrations in human leukocytes. *Biometrical Journal*, WILEY-VCH Verlag, v. 19, n. 8, p. 599–612, 1977. ISSN 1521-4036. Disponível em: <<http://dx.doi.org/10.1002/bimj.4710190804>>. Citado 2 vezes nas páginas 11 e 12.

KIRKPATRICK, R. M. *RMKdiscrete: Sundry Discrete Probability Distributions*. [S.l.], 2014. R package version 0.1. Disponível em: <<https://CRAN.R-project.org/package=RMKdiscrete>>. Citado na página 63.

LEE, C. M.-S.; FAMOYE, F. A comparison of generalized poisson distribution for modeling chromosome aberrations. *Biometrical journal*, Wiley Online Library, v. 38, n. 3, p. 299–313, 1996. Citado na página 49.

LEHMANN, E.; CASELLA, G. *Theory of Point Estimation*. [S.l.]: Springer Verlag, 1998. ISBN 0387985026. Citado na página 39.

LERNER, B.; LONE, A.; RAO, M. On the generalized poisson distribution. *Probability and Mathematical Statistics*, p. 17:377–385, 1997. Citado na página 29.

NELSON, D. L. Some remarks on generalizations of the negative binomial and poisson distributions. *Technometrics*, [Taylor & Francis, Ltd., American Statistical Association, American Society for Quality], v. 17, n. 1, p. 135–136, 1975. Citado na página 4.

NOACK, A. A class of random variables with discrete distributions. *Ann. Math. Statist.*, The Institute of Mathematical Statistics, v. 21, n. 1, p. 127–132, 03 1950. Disponível em: <<http://dx.doi.org/10.1214/aoms/1177729894>>. Citado 2 vezes nas páginas 15 e 23.

PATIL, G. P. Certain properties of the generalized power series distribution ii. *Annals of the Institute of Statistical Mathematics*, v. 14, n. 1, p. 179–182, 1962. ISSN 1572-9052. Disponível em: <<http://dx.doi.org/10.1007/BF02868639>>. Citado na página 23.

SHOUKRI, M. *Estimation problems for some generalized discrete distributions*. Tese (Doutorado) — Ph. D. thesis, 1980. Citado na página 39.

SRIVASTAVA, S.; CHEN, L. *GPseq: gpseq: Using the generalized Poisson distribution to model sequence read counts from high throughput sequencing experiments*. [S.l.], 2011. R package version 0.5. Disponível em: <<https://CRAN.R-project.org/package=GPseq>>. Citado na página 61.

TUENTER, H. J. H. On the generalized poisson distribution. *Statistica Neerlandica*, Blackwell Publishers Ltd, v. 54, n. 3, p. 374–376, 2000. ISSN 1467-9574. Disponível em: <<http://dx.doi.org/10.1111/1467-9574.00147>>. Citado na página 29.

WITOWSKI, V.; FORAITA, R. *HMMpa: Analysing accelerometer data using hidden Markov models*. [S.l.], 2014. R package version 1.0. Disponível em: <<https://CRAN.R-project.org/package=HMMpa>>. Citado na página 62.

YEE, T. W. *VGAM: Vector Generalized Linear and Additive Models*. [S.l.], 2016. R package version 1.0-2. Disponível em: <<http://CRAN.R-project.org/package=VGAM>>. Citado na página 65.