



UNIVERSIDADE D  
COIMBRA

Diogo Filipe Santos Sousa

**MODELOS PARA SÉRIES TEMPORAIS DE  
CONTAGEM COM PERTURBAÇÃO EM ZERO**

**Dissertação no âmbito do Mestrado em Métodos Quantitativos em  
Finanças, orientada pela Professora Doutora Maria Esmeralda  
Elvas Gonçalves e apresentada ao Departamento de Matemática  
da Faculdade de Ciências e Tecnologia e à Faculdade de Economia.**

Abril de 2021



# Modelos para séries temporais de contagem com perturbação em zero

Diogo Filipe Santos Sousa



UNIVERSIDADE D  
COIMBRA

Mestrado em Métodos Quantitativos em Finanças

Master in Quantitative Methods in Finance

Dissertação de Mestrado | MSc Dissertation

April 2021



## **Agradecimentos**

Quero agradecer à Professora Doutora Maria Esmeralda Elvas Gonçalves. A sua energia inesgotável, firmeza, sabedoria e sugestões foram essenciais para a realização desta dissertação.

À minha família incrível, em especial aos meus Pais e Avós pelo encorajamento.

Ao Bruno, Inês, Rafael e Rui pelos valores e conhecimentos transmitidos.

Aos fantásticos colegas de curso que me acompanharam desde o início, Martim, Pedro e Sérgio.

Aos colegas de Mestrado, Patrícia por todas as ajudas, Luís, Simone e João.



## Resumo

As leis de probabilidade habitualmente utilizadas para descrever dados de contagem incluem em geral o valor zero como elemento do seu suporte. Mas, em diversas áreas de aplicações estatísticas, surgem frequentemente situações em que, seleccionando um determinado modelo para a característica de contagem em estudo, o número de zeros esperados difere significativamente do número de zeros realmente observados. Nesses casos, para uma melhor descrição da característica, torna-se necessário adaptar os modelos de contagem standard, inflacionando ou reduzindo a probabilidade inicialmente associada ao valor zero. Este trabalho envolve, assim, o problema da não compatibilidade entre o número esperado de ocorrências do valor zero e o correspondente número observado ao estudar uma variável de contagem. Abordamos o estudo de uma lei que generaliza a lei Geométrica [15] e que, através de um parâmetro adicional, permite alterar a probabilidade atribuída à observação zero. Analisadas características probabilísticas desta distribuição, é estabelecido o comportamento assintótico dos estimadores obtidos pelo método das proporções, dos momentos e da máxima verosimilhança, sendo o seu desempenho comparado por meio de simulações em amostras de dimensão média e elevada. A adequação desta lei a conjuntos de dados reais é avaliada, revelando-se mais eficaz do que com outras mais clássicas. O estudo prossegue com a introdução de um modelo para séries temporais de contagem com lei condicional ao passado pertencente à família destas leis. A estacionaridade à ordem um é estabelecida. A estacionaridade de segunda ordem é estudada para sub-classes das leis Geométricas e também para os modelos Geométricos inflacionados em zero, seguindo estudos recentes. A modelação da série relativa ao número de novos casos de infecção pelo Hantavírus por semana, num estado da Alemanha entre 2005 e 2018, conclui este estudo.

**Palavras chave:** Distribuição Geométrica; Séries Temporais de valores inteiros; Processos INGARCH.





# Conteúdo

<b>Lista de Figuras</b>	<b>ix</b>
<b>Lista de Tabelas</b>	<b>xi</b>
<b>1 Introdução</b>	<b>1</b>
<b>2 Lei Geométrica generalizada com perturbação em zero</b>	<b>3</b>
2.1 Modelo ZDGGD . . . . .	3
2.2 Função de distribuição e função característica . . . . .	4
2.3 Momentos . . . . .	5
2.4 Função quantil . . . . .	7
2.5 Relações com outros leis . . . . .	8
2.6 Estimação . . . . .	10
2.6.1 Definições e resultados . . . . .	11
2.6.2 Lei Geométrica . . . . .	13
2.6.3 Lei ZDGGD . . . . .	15
2.7 Simulação . . . . .	24
2.7.1 Estimação pontual . . . . .	24
2.7.2 Regiões de Confiança . . . . .	27
2.8 Aplicação a dados reais . . . . .	28
2.8.1 Modelos clássicos de contagem . . . . .	28
2.8.2 Nova Distribuição Discreta . . . . .	29
2.8.3 Lei de Poisson-Lindley . . . . .	30
2.8.4 Dados de Actuariado . . . . .	31
2.8.5 Outros Dados: greves e lesões . . . . .	33
<b>3 Modelos INARCH com lei condicional ZDGGD</b>	<b>35</b>
3.1 ZDGGD-INARCH . . . . .	35
3.2 Modelo Geométrico INARCH(p) . . . . .	37
3.3 Modelo Geométrico inflacionado em zero INARCH(p) . . . . .	41
3.4 Modelo G-INARCH(1): Máxima verosimilhança condicional . . . . .	45
3.5 Simulação . . . . .	46
3.6 Dados Reais . . . . .	47

<b>Bibliografia</b>	<b>51</b>
<b>Anexo A Cálculo de Momentos</b>	<b>53</b>
<b>Anexo B Códigos</b>	<b>55</b>

# Lista de Figuras

2.1	Lei ZDGGD: Região de deflação em zero (esquerda); Região de assimetria negativa (centro); Região de subdispersão (direita). . . . .	6
2.2	Histogramas de realizações das leis: ZDGGD(0.2,-0.7) (esquerda); ZDGGD(0.8,0.5) (centro); ZDGGD(0.04,-0.925) (direita). . . . .	7
2.3	Eficiência relativa assintótica do estimador dos momentos em relação ao das proporções. . . . .	24
2.4	Histogramas de uma realização da lei $ND(-0.1, 0.25)$ (esquerda) e da lei $ND(0.1, 0.1)$ (direita). . . . .	30
2.5	Histogramas de uma realização da lei $PL(0.4)$ (esquerda) e da lei $PL(0.8)$ (direita) . .	31
2.6	Conjunto 2. . . . .	32
2.7	Conjunto 5. . . . .	32
3.1	Trajectória de um processo que verifica o modelo G-INARCH(1) com $a_0 = 1$ e $a_1 = 0.5$ . 41	41
3.2	Trajectória de um processo que verifica o modelo ZIG-INARCH(1) com $a_0 = 1$ , $a_1 = 0.5$ e $w = 0.4$ . . . . .	45
3.3	Histograma de $\sqrt{n}(\hat{a}_0 - a_0)$ (esquerda) e Histograma de $\sqrt{n}(\hat{a}_1 - a_1)$ (direita). . . .	47
3.4	Trajectória do número de novos casos de infecção por Hantavirus por semana. . . . .	47
3.5	Autocorrelações (esquerda) e autocorrelações parciais (direita) empíricas. . . . .	48
3.6	Autocorrelações (esquerda) e autocorrelações parciais (direita) dos resíduos de Pearson correspondentes ao modelo ZIG – INARCH(1). . . . .	49



# Lista de Tabelas

2.1	Estatísticas de amostras geradas. . . . .	7
2.2	Resultados dos estimadores do Método das Proporções. . . . .	25
2.3	Resultados dos estimadores do Método dos Momentos. . . . .	25
2.4	Resultados dos estimadores da Máxima Verosimilhança. . . . .	25
2.5	Comparação dos resultados empíricos com os teóricos pelo método das proporções. . . . .	26
2.6	Comparação dos resultados empíricos com os teóricos pelo método dos momentos. . . . .	26
2.7	Comparação dos resultados empíricos com os teóricos pelo método da máxima verosimilhança. . . . .	26
2.8	Estimativas das probabilidades de cobertura da região de confiança com confiança assintótica 0.95. . . . .	27
2.9	Resumo de algumas características das leis Poisson, negativa binomial, binomial, nova distribuição discreta, Poisson-Lindley e ZDGGD. . . . .	31
2.10	Estatísticas relativas aos pedidos de indemnização por apólice (conjunto 2). . . . .	32
2.11	Testes de adequação para o número de pedidos de indemnização por apólice. . . . .	32
2.12	Estatísticas relativas aos surtos de greve. . . . .	33
2.13	Testes de adequação para o número de surtos de greve. . . . .	33
2.14	Estatísticas relativas ao número de lesões. . . . .	34
2.15	Testes de adequação para o número de lesões. . . . .	34
3.1	Valores teóricos e empíricos de uma trajectória de um processo G-INARCH(1), $a_0 = 1$ e $a_1 = 0.5$ . . . . .	41
3.2	Valores teóricos e empíricos de uma trajectória de um processo ZIG-INARCH(1), $a_0 = 1$ , $a_1 = 0.5$ e $w = 0.4$ . . . . .	45
3.3	Resultados dos estimadores da máxima verosimilhança condicional para o modelo G-INARCH(1). . . . .	46
3.4	Histograma do número de novos casos de infecção por Hantavirus por semana e algumas estatísticas. . . . .	47
3.5	Estimativas dos parâmetros dos modelos e valores dos critérios -Log, AIC e BIC. . . . .	48
3.6	Média, variância e autocorrelação de primeira ordem para cada modelo e da amostra. Média ( $M_r$ ) e variância ( $V_r$ ) dos resíduos de Pearson para cada modelo. . . . .	49



# Capítulo 1

## Introdução

Vivemos num contexto histórico, aquando da elaboração desta dissertação, marcado pela pandemia provocada pelo COVID-19. Todos os dias somos confrontados com, e sentimos necessidade de verificar, o número de novos casos de infecção registados e o número de mortes ocorridas, ansiando pela sua redução sistemática. De facto, estes números são séries temporais de contagem relativas ao número de infecções e de mortes registados diariamente e é de forma natural que compreendemos a importância de termos modelos matemáticos capazes de descrever estas características para tentarmos estudar, e mesmo antecipar ou prever, a sua evolução. Mas, claramente, os exemplos de séries temporais não ficam por aqui. O número de transacções diárias num determinado mercado de acções ou o número de greves num sector social são exemplos que podemos acrescentar, entre muitos, de áreas como o Actuariado, Medicina, Economia ou Biologia.

As leis de probabilidade habitualmente utilizadas para descrever dados de contagem incluem em geral o valor zero como elemento do seu suporte. Mas, em diversas áreas de aplicações estatísticas, surgem frequentemente situações em que, seleccionando um determinado modelo para a característica de contagem em estudo, o número de zeros esperados difere significativamente do número de zeros realmente observados. Nesses casos, para uma melhor descrição da característica, torna-se necessário adaptar os modelos de contagem standard, inflacionando ou reduzindo a probabilidade inicialmente associada ao valor zero. A título de exemplo, podemos considerar o número de vendas de determinado produto numa loja online por dia. Notemos que o valor zero possui aqui uma particularidade que o distingue dos outros valores possíveis. De facto, podemos observar o valor zero porque o produto não está disponível ou porque está disponível mas não ocorreu alguma compra. Assim, por exemplo, se tivermos conhecimento prévio que um modelo específico descreve em geral bem o número de vendas, ao tentar ajustar o modelo a estes dados, o modelo poderá ser sensível aos zeros que estão a mais, tentando ajustar-se ao número de zeros observados e provocar um ajustamento mau no geral. Este trabalho envolve, assim, o problema da não compatibilidade entre o número esperado de ocorrências do valor zero e o correspondente número observado ao estudar uma variável de contagem.

No Capítulo 2 iremos trabalhar sob a hipótese de independência das observações. Abordamos uma nova lei de probabilidade, proposta em [15], denominada distribuição Geométrica generalizada com distorção em zero (ZDGGD) a qual, como o nome indica, generaliza a distribuição Geométrica e permite aumentar ou reduzir a probabilidade inicialmente associada ao valor zero. Estudamos as principais propriedades desta lei, como a sua função distribuição, a função característica e os

momentos, e também relações existentes com outras leis de probabilidade. Depois, para que possamos responder ao problema de ajustamento desta lei a observações de uma característica de contagem torna-se necessário possuir métodos eficientes de estimação para os parâmetros. Assim, avaliamos e comparamos três métodos de estimação (proporções, momentos e máxima verosimilhança) deduzindo os estimadores correspondentes e estudando o seu comportamento assintótico segundo várias formas de convergência. Avaliaremos a eficiência destes estimadores em amostras de dimensão moderada ou elevada, desenvolvendo estudos de simulação de estimação pontual e intervalar.

No final do Capítulo 2 para além de ilustrarmos a flexibilidade que o modelo ZDGGD possui, comparamos o ajustamento a três conjuntos de dados obtido usando este modelo com o obtido considerando modelos como o de Poisson, Negativa Binomial, Binomial, Nova distribuição discreta e Poisson-Lindley. Concluimos que a lei ZDGGD se ajustou bem aos três conjuntos de dados e para além disso ajustou-se melhor que os modelos atrás mencionados.

No contexto das séries temporais a hipótese de independência das observações presente na Estatística clássica e no Capítulo 2 não é, em geral, verificada. Desta forma surge a necessidade de construir modelos que descrevam a evolução ao longo do tempo de séries de contagem. Os modelos lineares ARMA clássicos não são adequados para estas séries, pois o seu suporte é  $\mathbb{N}_0$  ou um seu subconjunto. Por outro lado, as séries de contagem apresentam frequentemente volatilidade condicional que depende do passado da série. É certamente com inspiração nos modelos GARCH que Ferland, Latour e Oraichi [3] propõem um modelo para séries temporais de valores inteiros não negativos onde a distribuição condicional ao passado é a lei Poisson, o qual é denominado modelo INGARCH. Assim o Capítulo 3 visa adaptar o modelo INARCH substituindo a lei condicional pela lei ZDGGD e efectuar um estudo probabilista do modelo ZDGGD-INARCH. Como perdemos a hipótese de independência precisamos de condições que façam aparecer estabilidade. Desta forma, iremos estabelecer condições de estacionaridade de primeira e segunda ordem. Ao abordar a condição de estacionaridade de segunda ordem deparamos-nos com a necessidade de particularizar o modelo. Assim, iremos estudar o modelo Geométrico INARCH, caso particular do modelo negativo binomial INARCH introduzido por Zhu em [21]. Ao efectuar esta particularização perdemos flexibilidade para a probabilidade de ocorrer zero. Logo, tornou-se relevante estudar o modelo Geométrico inflacionado em zero proposto por Zhu em [22]. Apresentamos trajetórias dos respectivos modelos e verificamos a existência de concordância entre os valores obtidos através do método de estimação Yule-Walwer e valores teóricos. Abordamos o método de estimação da máxima verosimilhança condicional para o modelo G-INARCH(1) ilustrando-o com simulações.

Finalmente, como exemplo de aplicação, é estudado o número de novos casos de infecção pelo Hantavírus por semana no estado federal Eslésvico-Holsácia da Alemanha entre 2005 e 2018. Face a características empíricas, em particular proporção de zeros e autocorrelações parciais, foram considerados ajustamentos por modelos INARCH Geométricos, de Poisson e as versões inflacionadas em zero correspondentes. Avaliado o ajustamento concluimos que o modelo que melhor reproduz a evolução desta série é o INARCH(1) Geométrico inflacionado em zero.

A concluir, registamos que teria sido interessante comparar a ZDGGD com outras propostas de generalizações da lei Geométrica como a introduzida em [7]. Completar o estudo probabilista do modelo ZDGGD-INGARCH, estabelecendo em particular condições de estacionaridade fraca e de existência de momentos, são igualmente desafios pendentes.



## Capítulo 2

# Lei Geométrica generalizada com perturbação em zero

Neste Capítulo apresentamos uma distribuição recentemente proposta em [15] que generaliza a lei Geométrica e que, à custa de um parâmetro adicional, permite perturbar a probabilidade em zero. Estudamos algumas das suas propriedades de natureza probabilista, como a função distribuição, a função característica e a existência de momentos, explorando também relações com outras leis de probabilidade seguindo a referencia anterior. Abordamos três métodos de estimação dos parâmetros (proporções, momentos e máxima verosimilhança) estabelecendo o correspondente comportamento assintótico. Este comportamento é ilustrado com estudos de simulação. A aplicação a três conjuntos de dados reais de áreas diversas (actuariado, sociais e saúde) mostra que esta família de leis permite obter melhor ajustamento quando comparada com outros modelos probabilistas discretos mais clássicos como a lei de Poisson ou binomial negativa, entre outros. Os códigos utilizados ao longo deste Capítulo foram feitos com recurso ao software Matlab.

### 2.1 Modelo ZDGGD

**Definição 2.1.** *Uma variável aleatória real (v.a.r.)  $X$  discreta de suporte  $S_X = \mathbb{N}_0 = \{0, 1, \dots\}$  segue uma lei Geométrica generalizada com perturbação em zero com parâmetros  $q \in ]0, 1[$  e  $\alpha \in [-1, +\infty[$  se*

$$P(X = k) = \begin{cases} 1 - q^{\alpha+1} & \text{se } k = 0 \\ (1 - q)q^{k+\alpha} & \text{se } k \in \mathbb{N}. \end{cases} \quad (2.1)$$

Escrevemos de modo abreviado  $X \sim \text{ZDGGD}(q, \alpha)$  <sup>(1)</sup> e observamos que se trata de facto de uma lei de probabilidade discreta pois

$$\sum_{k \in S_X} P(X = k) = 1 - q^{\alpha+1} + (1 - q)q^\alpha \sum_{k=1}^{+\infty} q^k = 1 - q^{\alpha+1} + \frac{(1 - q)q^{\alpha+1}}{1 - q} = 1.$$

---

<sup>(1)</sup> Da terminologia inglesa **Zero Distorted Generalized Geometric Distribution**.

Relativamente ao parâmetro  $\alpha$  assinalamos os aspectos seguintes:

1. Se  $\alpha = 0$ , então  $P(X = k) = (1 - q)q^k$ ,  $k \in \mathbb{N}_0$ , ou seja,  $X$  segue uma lei Geométrica de parâmetro  $q$ , o que denotamos por  $X \sim G(q)$ .
2. Se  $\alpha = -1$ , então  $P(X = 0) = 0$  e  $P(X = k) = (1 - q)q^{k-1}$ ,  $k \in \mathbb{N}$ .
3. Quando  $\alpha \rightarrow +\infty$  tem-se que  $P(X = 0)$  tende para 1, ou seja, a lei obtida é degenerada na origem.

Verifica-se, assim, que a lei *ZDGGD* generaliza a lei Geométrica. Notamos que o parâmetro  $\alpha$  permite perturbar a probabilidade no ponto 0. Podemos comparar o valor de  $P(X = 0)$  com o valor  $P(X = 1)$  dizendo-se que a lei está inflacionada em zero se  $P(X = 0) > P(X = 1)$  e deflacionada em zero se  $P(X = 0) < P(X = 1)$ , [15]. A análise desta condição para a lei *ZDGGD*( $q, \alpha$ ) mostra que

1. Se  $\alpha > -\frac{\log(q(2-q))}{\log(q)}$  a lei está inflacionada em zero.
2. Se  $-1 \leq \alpha < -\frac{\log(q(2-q))}{\log(q)}$  a lei está deflacionada em zero.

Observamos que podemos também concluir que a lei *ZDGGD* é unimodal com moda 0 ( respectivamente 1) se  $\alpha >$  ( respectivamente,  $<$ )  $-\frac{\log(q(2-q))}{\log(q)}$  já que, considerando a função  $f(x) = (1 - q)q^{x+\alpha}$  para  $x > 0$ , se tem  $f'(x) = (1 - q)q^{x+\alpha} \log(q) < 0$ ,  $\forall x > 0$ . Na Figura 2.1 podemos observar a região  $\left\{ (q, \alpha) \in ]0, 1[ \times ]-1, +\infty[ : -1 < \alpha < -\frac{\log(q(2-q))}{\log(q)} \right\}$  onde existe deflação em zero.

## 2.2 Função de distribuição e função característica

A função de distribuição da lei *ZDGGD* é dada por

$$\begin{aligned} \forall x \in \mathbb{R}, F_X(x) = P(X \leq x) &= \sum_{k \in S_X \cap ]-\infty, x]} P(X = k) \\ &= \begin{cases} 0, & \text{se } x < 0 \\ 1 - q^{\alpha+1}, & \text{se } 0 \leq x < 1 \\ 1 - q^{\alpha+1} + (1 - q)q^{1+\alpha} = 1 - q^{\alpha+2}, & \text{se } 1 \leq x < 2 \\ \dots & \dots \end{cases} \\ &= \begin{cases} 0, & \text{se } x < 0 \\ 1 - q^{\lfloor x \rfloor + \alpha + 1} & \text{se } x \geq 0, \end{cases} \end{aligned}$$

onde  $\lfloor x \rfloor$  representa a parte inteira de  $x$ . A sua função característica é, onde  $i$  é a unidade imaginária,

$$\begin{aligned} \forall t \in \mathbb{R}, \Phi_X(t) = E(e^{itX}) &= \sum_{k=0}^{+\infty} e^{itk} P(X = k) = 1 - q^{\alpha+1} + \sum_{k=1}^{+\infty} e^{itk} (1 - q)q^{k+\alpha} \\ &= 1 - q^{\alpha+1} + q^\alpha (1 - q) \sum_{k=1}^{+\infty} (e^{it} q)^k = 1 - q^{\alpha+1} + \frac{e^{it} (1 - q) q^{\alpha+1}}{1 - e^{it} q}. \end{aligned}$$

## 2.3 Momentos

Comecemos por observar que  $X$  admite momentos de todas as ordens,  $E(X^r)$ ,  $r \in \mathbb{N}$ . De facto,  $E(X^r)$  existe se a série  $\sum_{k=1}^{+\infty} k^r q^k$  for convergente pois

$$E(X^r) = \sum_{k=0}^{+\infty} k^r P(X = k) = \sum_{k=1}^{+\infty} k^r (1-q)q^{\alpha+k} = (1-q)q^\alpha \sum_{k=1}^{+\infty} k^r q^k. \quad (2.2)$$

Mas como

$$\lim_{n \rightarrow \infty} \frac{(n+1)^r q^{n+1}}{n^r q^n} = q < 1,$$

usando o critério d'Alembert conclui-se que a série é absolutamente convergente e consequentemente convergente.

Utilizaremos agora a relação  $\Phi_X^{(r)}(t)|_{t=0} = r! E(X^r)$  para deduzir os valores dos momentos até à quarta ordem, estando os cálculos presentes no Anexo.

$$E(X) = \frac{q^{\alpha+1}}{1-q}, \quad E(X^2) = \frac{q^{\alpha+1}(1+q)}{(1-q)^2},$$

$$E(X^3) = \frac{q^{\alpha+1}(1+4q+q^2)}{(1-q)^3}, \quad E(X^4) = \frac{q^{\alpha+1}(1+11q+11q^2+q^3)}{(1-q)^4}.$$

Como foi notado em [15], os momentos podem ser obtidos também a partir da seguinte equação de recorrência onde  $\mu_r = E(X^r)$ ,

$$\begin{cases} \mu_1 = \frac{q^{\alpha+1}}{1-q} \\ \mu_{r+1} = q \frac{\partial}{\partial q} \mu_r - \frac{(1-q)\alpha - q}{(1-q)} \mu_r, \quad r \geq 1. \end{cases}$$

De facto, a partir de (2.2) tem-se

$$\begin{aligned} \frac{\partial}{\partial q} \mu_r &= \frac{\partial}{\partial q} (1-q)q^\alpha \sum_{k=1}^{+\infty} k^r q^k = -q^\alpha \sum_{k=1}^{+\infty} k^r q^k + (1-q)\alpha q^{\alpha-1} \sum_{k=1}^{+\infty} k^r q^k + (1-q)q^\alpha \sum_{k=1}^{+\infty} k^{r+1} q^{k-1} \\ &= \frac{-q^\alpha + (1-q)\alpha q^{\alpha-1}}{(1-q)q^\alpha} \mu_r + (1-q)q^\alpha \sum_{k=1}^{+\infty} k^{r+1} q^{k-1}. \end{aligned}$$

Multiplicando ambos os membros da igualdade anterior por  $q$ ,

$$q \frac{\partial}{\partial q} \mu_r = \frac{-q + (1-q)\alpha}{(1-q)} \mu_r + (1-q)q^\alpha \sum_{k=1}^{+\infty} k^{r+1} q^k = \frac{(1-q)\alpha - q}{(1-q)} \mu_r + \mu_{r+1},$$

obtemos a referida equação de recorrência.

Quanto aos momentos centrados,  $\mu'_r = E[(X - E(X))^r]$ ,  $r = 2, 3, 4$ , são

$$\mu'_2 = V(X) = \frac{q^{\alpha+1}(1+q-q^{\alpha+1})}{(1-q)^2}, \quad \mu'_3 = \frac{q^{\alpha+1}(1+4q+q^2) - 3q^{2(\alpha+1)}(1+q) + 2q^{3(\alpha+1)}}{(1-q)^3},$$

$$\mu_4' = \frac{q^{\alpha+1} (1 + 11q + 11q^2 + q^3) - 4q^{2(\alpha+1)} (1 + 4q + q^2) + 6q^{3(\alpha+1)} (1 + q) - 3q^{4(\alpha+1)}}{(1 - q)^4}.$$

O coeficiente de assimetria é dado por

$$s = \frac{\mu_3'}{(\mu_2')^{3/2}} = \frac{q^{\alpha+1} (1 + 4q + q^2) - 3q^{2(\alpha+1)} (1 + q) + 2q^{3(\alpha+1)}}{[q^{\alpha+1} (1 + q - q^{\alpha+1})]^{3/2}},$$

tendo-se

$$\begin{aligned} s < 0 &\Leftrightarrow q^{\alpha+1} (1 + 4q + q^2) - 3q^{2(\alpha+1)} (1 + q) + 2q^{3(\alpha+1)} < 0 \\ &\Leftrightarrow 1 + 4q + q^2 + 2q^{2(\alpha+1)} < 3q^{(\alpha+1)} (1 + q). \end{aligned}$$

A região  $\{(q, \alpha) \in ]0, 1[ \times ]-1, +\infty[ : 1 + 4q + q^2 + 2q^{2(\alpha+1)} < 3q^{(\alpha+1)} (1 + q)\}$ , onde a assimetria da lei é negativa, encontra-se na Figura 2.1 .<sup>(2)</sup>

O índice de dispersão é uma medida para avaliar a dispersão de uma lei de probabilidade dado por  $ID = V(X)/E(X)$ . Assim, se  $ID > 1$  a lei diz-se sobredispersa e se  $ID < 1$  diz-se subdispersa.

Se  $X \sim ZDGGD(q, \alpha)$  tem-se

$$ID = \frac{V(X)}{E(X)} = \frac{\frac{q^{\alpha+1}(1+q-q^{\alpha+1})}{(1-q)^2}}{\frac{q^{\alpha+1}}{1-q}} = \frac{1+q-q^{\alpha+1}}{1-q}.$$

Podemos então notar que

- se  $\alpha > \frac{\log(2)}{\log(q)}$  então  $ID > 1$ , logo a distribuição  $ZDGGD(q, \alpha)$  é sobredispersa,
- se  $\alpha < \frac{\log(2)}{\log(q)}$  então  $ID < 1$ , logo a distribuição  $ZDGGD(q, \alpha)$  é subdispersa.

Na Figura 2.1 encontra-se assinalada a região  $\{(q, \alpha) \in ]0, 1[ \times ]-1, +\infty[ : \alpha < \frac{\log(2)}{\log(q)}\}$  onde há subdispersão.

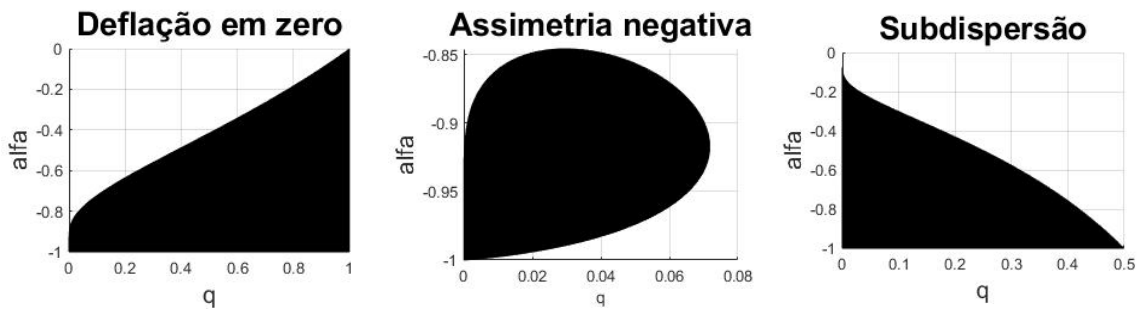


Fig. 2.1 Lei ZDGGD: Região de deflação em zero (esquerda); Região de assimetria negativa (centro); Região de subdispersão (direita).

<sup>(2)</sup> A curtose é dada por  $\frac{\mu_4'}{(\mu_2')^2} = \frac{q^{\alpha+1} (1 + 11q + 11q^2 + q^3) - 4q^{2(\alpha+1)} (1 + 4q + q^2) + 6q^{3(\alpha+1)} (1 + q) - 3q^{4(\alpha+1)}}{q^{2(\alpha+1)} (1 + q - q^{\alpha+1})^2}$ . Notamos que a expressão correspondente em [15] é  $\frac{\mu_4'}{(\mu_2')^2}$  e não  $\frac{\mu_4'}{(\mu_2')^2}$ .

## 2.4 Função quantil

A função quantil de  $X \sim ZDGGD(q, \alpha)$ ,  $q_u$  com  $u \in ]0, 1[$ , é dada por

$$q_u = \inf \{x \in \mathbb{R} : F_X(x) \geq u\} \Leftrightarrow \inf \left\{ x \in \mathbb{R} : 1 - q^{[x] + \alpha + 1} \geq u \right\}$$

$$\Leftrightarrow \begin{cases} 0 & \text{se } u \leq 1 - q^{\alpha+1} \\ \inf \left\{ x \in \mathbb{R} : [x] \geq \frac{\log(1-u)}{\log(q)} - (\alpha + 1) \right\} & \text{se } u > 1 - q^{\alpha+1}. \end{cases}$$

A função quantil pode ser utilizada para simular variáveis aleatórias cuja lei de probabilidade é a  $ZDGGD(q, \alpha)$ , por exemplo com o método descrito em [14] para obter valores simulados de variáveis aleatórias discretas. Desta forma, para simular um valor  $x$  gera-se primeiro um número aleatório  $u$  de acordo com a lei  $U(]0, 1[)$  e depois considera-se  $x = q_u$ .

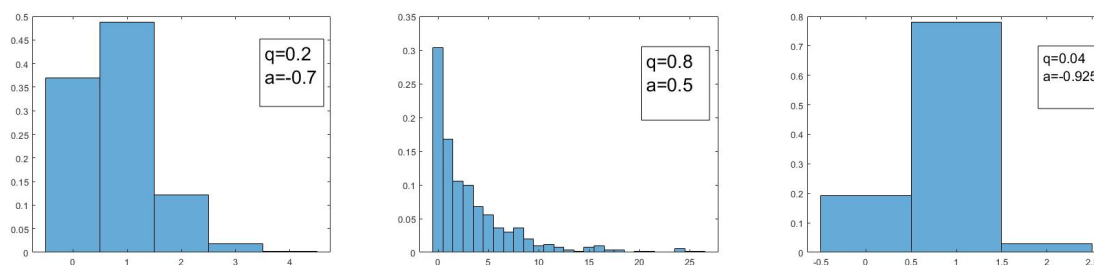


Fig. 2.2 Histogramas de realizações das leis:  $ZDGGD(0.2, -0.7)$  (esquerda);  $ZDGGD(0.8, 0.5)$  (centro);  $ZDGGD(0.04, -0.925)$  (direita).

Na Figura 2.2 apresentam-se os histogramas de algumas amostras geradas de tamanho 500 considerando  $(q, \alpha) \in \{(0.2, -0.7), (0.8, 0.5), (0.04, -0.925)\}$ <sup>(3)</sup>. É possível observar que para  $(q, \alpha) = (0.2, -0.7)$  existe deflação em zero e para  $(q, \alpha) = (0.8, 0.5)$  existe inflação. Na Tabela 2.1 figuram algumas estatísticas (índice de dispersão, assimetria e proporções de zero e de um) bem como os valores teóricos entre parêntesis. Pode-se observar de forma clara na Tabela 2.1 que para  $(q, \alpha) = (0.8, 0.5)$  existe sobredispersão e para os casos restantes subdispersão. Por fim, observamos também que, no caso  $(q, \alpha) = (0.04, -0.925)$  a assimetria é negativa enquanto que nos outros casos é positiva.

Tabela 2.1 Estatísticas de amostras geradas.

$q$	$\alpha$	ID	Assimetria	Proporção de 0	Proporção de 1
0.2	-0.7	0.6809(0.7287)	0.7372(1.0872)	0.3700(0.3830)	0.4880(0.4936)
0.8	0.5	5.8381(5.4223)	2.2888(2.0937)	0.3040(0.2845)	0.1680(0.1431)
0.04	-0.925	0.2314(0.2651)	-0.7611(-0.4846)	0.1920(0.2145)	0.7800(0.7541)

<sup>(3)</sup> Decidimos não manter na Figura 2.2 a mesma escala no eixo das ordenadas, porque algumas imagens perdiam informação.

## 2.5 Relações com outras leis

Quando estamos a modelar dados de contagem pode ser relevante ter informação sobre o excesso acima de um limite ou sobre o mínimo de alguns valores. Por exemplo, ao estudar o número de pedidos de indemnização por apólice, podemos ter interesse em descrever também o excesso relativamente a um limiar ou, se possuímos uma carteira de apólices, descrever o mínimo dos números de pedidos de indemnização. Os teoremas seguintes contribuem para estas questões.<sup>(4)</sup>

**Teorema 2.1.** *Seja  $X$  uma v.a.r. tal que  $X \sim ZDGGD(q, \alpha)$  e  $t \in \mathbb{N}_0$  tal que  $t + \alpha > -1$ . Então a v.a.r.  $Y = (X - t)^+ \sim ZDGGD(q, \alpha + t)$ .*

*Prova.* Temos  $S_Y = \mathbb{N}_0$ . Seja  $k \in S_Y$  então

- se  $k = 0$ ,  $P(Y = 0) = P((X - t)^+ = 0) = P(X \leq t) = 1 - q^{(\alpha+t)+1}$
- se  $k > 0$ ,  $P(Y = k) = P((X - t)^+ = k) = P(X = t + k) = (1 - q)q^{(\alpha+t)+k}$ .

Logo  $Y \sim ZDGGD(q, \alpha + t)$ . ■

**Teorema 2.2.** *Sejam  $X_i$ ,  $i = 1, \dots, n$ ,  $n \in \mathbb{N}$ , v.a. reais independentes tais que  $X_i \sim ZDGGD(q_i, \alpha)$ . Então  $Y = \min(X_1, X_2, \dots, X_n) \sim ZDGGD(\prod_{i=1}^n q_i, \alpha)$ .*

*Prova.* Como  $S_{X_i} = \mathbb{N}_0$ ,  $i = 1, \dots, n$ , segue-se que  $S_Y = \mathbb{N}_0$ . Seja  $k \in \mathbb{N}_0$ . Vem  $P(Y = k) = P(Y \geq k) - P(Y \geq k + 1)$ , faltando assim determinar  $P(Y \geq k)$ ,  $k \in \mathbb{N}_0$ . Se  $k = 0$  então  $P(Y \geq k) = 1$ .

$$\begin{aligned} \text{Se } k \in \mathbb{N}, \quad P(Y \geq k) &= P(\min(X_1, X_2, \dots, X_n) \geq k) = P(X_1 \geq k, X_2 \geq k, \dots, X_n \geq k) \\ &= \prod_{i=1}^n P(X_i \geq k) = \prod_{i=1}^n (1 - 1 + q_i^{k+\alpha}) = \left( \prod_{i=1}^n q_i \right)^{k+\alpha}. \end{aligned}$$

$$\text{Consequentemente temos } P(Y = 0) = 1 - P(Y \geq 1) = 1 - \left( \prod_{i=1}^n q_i \right)^{\alpha+1}$$

$$\begin{aligned} \text{e } P(Y = k) &= P(Y \geq k) - P(Y \geq k + 1) = \left( \prod_{i=1}^n q_i \right)^{\alpha+k} - \left( \prod_{i=1}^n q_i \right)^{\alpha+k+1} \\ &= \left( 1 - \prod_{i=1}^n q_i \right) \left( \prod_{i=1}^n q_i \right)^{\alpha+k}. \end{aligned}$$

Logo por (2.1) tem-se  $Y \sim ZDGGD(\prod_{i=1}^n q_i, \alpha)$ . ■

Os teoremas anteriores permitem afirmar que se  $X \sim ZDGGD(q, \alpha)$  e  $t \in \mathbb{N}_0$  tal que  $t + \alpha > -1$  então as leis de  $(X - t)^+$  e de  $\min(X_1, X_2, \dots, X_n)$  com  $X_i \sim ZDGGD(q_i, \alpha)$ ,  $i = 1, \dots, n$ ,  $n \in \mathbb{N}$  independentes continuam a pertencer à família de leis  $ZDGGD$ . Apresentamos agora um resultado que permite relacionar a lei  $ZDGGD$  com a lei exponencial e a lei de Poisson generalizada através de um método para criar novas leis, chamado *mixing* de variáveis aleatórias, ([9], p.59-60).

<sup>(4)</sup> Relembramos a seguinte notação  $(X - t)^+ = \begin{cases} 0 & \text{se } X \leq t \\ X - t & \text{se } X > t. \end{cases}$

Recordemos que

- A v.a.r.  $X$  segue uma lei exponencial de parâmetro  $\theta > 0$ ,  $X \sim \mathfrak{E}(\theta)$ , se é absolutamente contínua de densidade  $f_X(x) = \theta e^{-\theta x} \mathbb{I}_{x>0}$ . <sup>(5)</sup>
- A v.a.r.  $Y$  segue uma lei de Poisson Generalizada,  $Y \sim GP(\lambda, \alpha)$ , [1], com  $\lambda > 0$  e  $0 \leq \alpha < 1$  se  $S_Y = \mathbb{N}_0$  e

$$P(Y = k) = \begin{cases} \frac{1}{\Gamma(1 + \alpha)} \int_{\lambda}^{+\infty} z^{\alpha} e^{-z} dz & \text{se } k = 0 \\ \frac{e^{-\lambda} \lambda^{k+\alpha}}{\Gamma(k + \alpha + 1)} & \text{se } k \in \mathbb{N} \end{cases}$$

onde  $\Gamma(t) = \int_0^{+\infty} x^{t-1} e^{-x} dx$ . Notemos que se  $\alpha = 0$ ,  $Y \sim P(\lambda)$ .

**Teorema 2.3.** *Seja  $\Lambda$  uma v.a.r. tal que  $\Lambda \sim \mathfrak{E}(\frac{p}{q})$  e  $q = 1 - p$ ,  $0 < p < 1$ . Considere-se agora a v.a.r.  $X$  tal que condicionada por  $\Lambda = \lambda$  segue a lei  $GP(\lambda, \alpha)$ , ou seja,  $X|\Lambda = \lambda \sim GP(\lambda, \alpha)$ . Então  $X \sim ZDGGD(q, \alpha)$ .*

*Prova.* Temos

$$\begin{aligned} P(X = 0) &= \int_{-\infty}^{+\infty} P(X = 0|\Lambda = \lambda) f_{\Lambda}(\lambda) d\lambda = \int_0^{+\infty} \frac{1}{\Gamma(1 + \alpha)} \left( \int_{\lambda}^{+\infty} z^{\alpha} e^{-z} dz \right) \frac{p}{q} e^{-\frac{p}{q}\lambda} d\lambda \\ &= \int_0^{+\infty} \frac{z^{\alpha} e^{-z}}{\Gamma(1 + \alpha)} \left( \int_0^z \frac{p}{q} e^{-\frac{p}{q}\lambda} d\lambda \right) dz = \int_0^{+\infty} \frac{z^{\alpha} e^{-z}}{\Gamma(1 + \alpha)} \left( 1 - e^{-\frac{p}{q}z} \right) dz \\ &= \int_0^{+\infty} \frac{z^{\alpha} e^{-z}}{\Gamma(1 + \alpha)} dz - \frac{q^{\alpha+1}}{\Gamma(1 + \alpha)} \frac{1}{q^{\alpha+1}} \int_0^{+\infty} z^{\alpha} e^{-(1+\frac{p}{q})z} dz = 1 - q^{\alpha+1} > 0 \end{aligned}$$

e para  $k \in \mathbb{N}$

$$\begin{aligned} P(X = k) &= \int_{-\infty}^{+\infty} P(X = k|\Lambda = \lambda) f_{\Lambda}(\lambda) d\lambda = \int_0^{+\infty} \frac{e^{-\lambda} \lambda^{k+\alpha}}{\Gamma(k + \alpha + 1)} \frac{p}{q} e^{-\frac{p}{q}\lambda} d\lambda \\ &= \frac{1}{\Gamma(k + \alpha + 1)} \frac{p}{q} \int_0^{+\infty} e^{-(1+\frac{p}{q})\lambda} \lambda^{k+\alpha} d\lambda \\ &= \frac{q^{k+\alpha+1}}{\Gamma(k + \alpha + 1)} \frac{1}{q} \frac{q}{q^{k+\alpha+1}} \int_0^{+\infty} e^{-(1+\frac{p}{q})\lambda} \lambda^{k+\alpha} d\lambda = \frac{p}{q} q^{k+\alpha+1} = (1 - q) q^{k+\alpha} > 0. \end{aligned}$$

Logo, como  $\sum_{k \in \mathbb{N}_0} P(X = k) = 1$ , temos  $S_X = \mathbb{N}_0$  e concluímos que  $X \sim ZDGGD(q, \alpha)$ . ■

O teorema seguinte permite identificar a lei de  $[X]$  quando a lei de  $X$  é uma determinada mistura de leis.

**Teorema 2.4.** *Seja  $X$  uma v.a.r. com lei de probabilidade  $P_X$  a qual é uma mistura de leis  $P_W$  e  $P_Z$ , onde  $W$  é uma v.a.r. degenerada na origem e  $Z \sim \mathfrak{E}(\theta)$  com proporções respectivamente  $1 - e^{-\theta\alpha}$  e  $e^{-\theta\alpha}$ ,  $\theta > 0$ . Então a v.a.r.  $Y = [X] \sim ZDGGD(e^{-\theta}, \alpha)$ .*

<sup>(5)</sup>  $\mathbb{I}_{x>0} = \mathbb{I}_{\{0, \infty\}}(x) = \begin{cases} 1 & \text{se } x \in ]0, \infty[ \\ 0 & \text{se } x \notin ]0, \infty[. \end{cases}$

*Prova.* Começamos por notar que  $S_Y = \mathbb{N}_0$ . Seja  $k \in S_Y$ , então

$$\begin{aligned} \text{Se } k = 0, \quad P(Y = 0) &= P([X] = 0) = P_X([0, 1]) = (1 - e^{-\theta\alpha})P_W([0, 1]) + e^{-\theta\alpha}P_Z([0, 1]) \\ &= (1 - e^{-\theta\alpha}) + e^{-\theta\alpha} \int_0^1 \theta e^{-\theta x} dx \\ &= 1 - e^{-\theta\alpha} + e^{-\theta\alpha} (1 - e^{-\theta}) = 1 - e^{-\theta(1+\alpha)} = 1 - (e^{-\theta})^{1+\alpha}. \end{aligned}$$

$$\begin{aligned} \text{Se } k > 0, \quad P(Y = k) &= P([X] = k) = P_X([k, k+1]) = (1 - e^{-\theta\alpha})P_W([k, k+1]) + e^{-\theta\alpha}P_Z([k, k+1]) \\ &= e^{-\theta\alpha} \int_k^{k+1} \theta e^{-\theta x} dx = e^{-\theta\alpha} (e^{-\theta k} - e^{-\theta(k+1)}) \\ &= e^{-\theta(\alpha+k)} (1 - e^{-\theta}) = (e^{-\theta})^{\alpha+k} (1 - e^{-\theta}). \end{aligned}$$

Logo  $Y \sim \text{ZDGGD}(e^{-\theta}, \alpha)$ . ■

O Teorema 2.4 pode ser útil numa situação do tipo que a seguir se descreve. Suponhamos que estamos interessados em estudar o valor pedido para uma indemnização  $X$  numa determinada carteira de apólices. O pedido de indemnização pode depender da ocorrência de um evento. Assumindo que a probabilidade de não ocorrer o evento é dada por  $1 - e^{-\theta\alpha}$  e que se não ocorrer o valor pedido é 0, e se ocorrer o valor pedido é descrito por uma variável aleatória  $Z \sim \mathfrak{E}(\theta)$  então o Teorema 4 mostra que  $[X] \sim \text{ZDGGD}(q, \alpha)$ .

De seguida apresenta-se uma relação entre a lei exponencial e a lei  $\text{ZDGGD}$ .

**Teorema 2.5.** *Sejam  $X$  uma v.a.r. tal que  $X \sim \mathfrak{E}(-\log(q))$  e  $\alpha$  tal que  $-1 < \alpha \leq 1$ . Então a v.a.r.  $Y = [X - \alpha] \sim \text{ZDGGD}(q, \alpha)$ .*

*Prova.* Constata-se que  $S_Y = \mathbb{N}_0$ . Seja  $k \in S_Y$  então

$$\begin{aligned} P([X - \alpha] \leq k) &= P(X - \alpha < k + 1) = P(X < k + 1 + \alpha) = \int_0^{k+1+\alpha} -\log(q)e^{\log(q)x} dx \\ &= 1 - e^{\log(q)(k+1+\alpha)} = 1 - q^{k+1+\alpha}, \end{aligned}$$

o que permite concluir que  $Y \sim \text{ZDGG}(q, \alpha)$ . ■

Notemos que quando  $\alpha = 0$  recuperamos o resultado conhecido que diz que a parte inteira de uma v.a.r. que siga uma lei exponencial segue uma lei Geométrica. É este resultado, entre outros, que leva a dizer que a lei Geométrica é o análogo discreto da lei exponencial.

## 2.6 Estimação

Estudamos agora os estimadores dos parâmetros  $(q, \alpha)$  da lei  $\text{ZDGGD}$  fornecidos por três métodos de estimação: proporções, método dos momentos e método da máxima verosimilhança. Este estudo complementa o apresentado em Sastry et al [15], pois estabelecemos, para os vários estimadores, o



correspondente comportamento assintótico. Depois de apresentarmos alguns resultados preliminares sobre convergência de sucessões de v.a. reais iniciaremos o estudo com a lei Geométrica, caso particular correspondente à  $ZDGGD(q, \alpha)$  quando  $\alpha = 0$ .

### 2.6.1 Definições e resultados

**Teorema 2.6** (Teorema do Limite Central multivariado, [11], pág. 61). *Seja  $[X_n = (X_{1,n}, \dots, X_{k,n})]_{n \in \mathbb{N}}$  uma sucessão de v.a. i.i.d. sobre  $\mathbb{R}^k$ , com  $k \in \mathbb{N}$ , com média  $\theta = (\theta_1, \dots, \theta_k)$  e matriz de variâncias-covariâncias  $\Sigma = [\sigma_{i,j}]_{1 \leq i, j \leq k}$ . Seja ainda  $\bar{X}_{i,n} = \frac{1}{n} \sum_{j=1}^n X_{i,j}$ ,  $i = 1, \dots, k$ . Então*

$$\sqrt{n} \left( \begin{bmatrix} \bar{X}_{1,n} \\ \dots \\ \bar{X}_{k,n} \end{bmatrix} - \begin{bmatrix} \theta_1 \\ \dots \\ \theta_k \end{bmatrix} \right)$$

*converge em lei para um vector aleatório real  $Z$  que segue uma normal centrada, com média  $(0, \dots, 0) \in \mathbb{R}^k$ , e de matriz de variâncias-covariâncias  $\Sigma$ . <sup>(6)</sup>*

**Teorema 2.7** (Método delta multivariado, [11], pág. 61). *Seja  $[X_n = (X_{1,n}, \dots, X_{k,n})]_{n \in \mathbb{N}}$  uma sucessão de v.a. i.i.d. sobre  $\mathbb{R}^k$ ,  $k \in \mathbb{N}$ , tal que*

$$\sqrt{n} \left( \begin{bmatrix} X_{1,n} \\ \dots \\ X_{k,n} \end{bmatrix} - \begin{bmatrix} \theta_1 \\ \dots \\ \theta_k \end{bmatrix} \right) \xrightarrow{L} Z, Z \sim N(0, \Sigma).$$

*Considere-se a função  $h: \mathbb{R}^k \rightarrow \mathbb{R}^k$  continuamente diferenciável numa vizinhança  $W$  do parâmetro  $\theta = (\theta_1, \dots, \theta_k)$  e  $B = \begin{bmatrix} \frac{\partial h_i}{\partial \theta_j} \end{bmatrix}_{1 \leq i, j \leq k}$  não singular em  $W$ . Então*

$$\sqrt{n} \left( \begin{bmatrix} h_1(X_n) \\ \dots \\ h_k(X_n) \end{bmatrix} - \begin{bmatrix} h_1(\theta) \\ \dots \\ h_k(\theta) \end{bmatrix} \right) \xrightarrow{L} Z, Z \sim N(0, B\Sigma B^T).$$

O resultado seguinte estabelece a convergência quase certa e em lei dos estimadores fornecidos pelo método dos momentos.

**Teorema 2.8** ([5], pág. 147-148). *Seja  $(X_1, \dots, X_n)$  uma  $n$ - amostra de uma v.a.r.  $X$  com  $Q_\theta$  a sua lei de probabilidade sobre  $(\mathbb{R}, \mathfrak{B})$ , onde o parâmetro  $\theta = (\theta_1, \dots, \theta_k)$  é desconhecido e pertencente a um subconjunto aberto,  $\Theta$ , de  $\mathbb{R}^k$ ,  $k \in \mathbb{N}$ . Suponhamos que para  $r = 1, \dots, k$  existe o momento de ordem  $r$  de  $X$ ,  $m_r(\theta)$ , e denotemos por  $M_r = \frac{1}{n} \sum_{i=1}^n X_i^r$  o momento empírico de ordem  $r = 1, \dots, k$ . Seja  $\hat{\theta}_n = f(M_1, \dots, M_k)$  o estimador obtido pelo método dos momentos, onde  $f$  de  $\mathbb{R}^k$  em  $\mathbb{R}^k$  é diferenciável. Então*

$$\forall \theta \in \Theta, \hat{\theta}_n \xrightarrow{q.c.} \theta.$$

<sup>(6)</sup> Iremos denotar a convergência em lei por  $\xrightarrow{L}$ .

Se  $X$  admite momento de ordem  $2k$  tem-se ainda que

$$\forall \theta \in \Theta, \sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{L} Z, Z \sim N(0, D(\theta)A(\theta)D(\theta)^T),$$

$$\text{com } A(\theta) = [\text{Cov}(X^j, X^l)]_{1 \leq j, l \leq k} \text{ e } D(\theta) = \left[ \frac{\partial f_j}{\partial m_l}(\theta) \right]_{1 \leq j, l \leq k}. \quad (7)$$

Iremos introduzir agora alguma notação necessária para o Teorema seguinte onde se estabelece o comportamento limite em probabilidade e em lei dos estimadores da máxima verosimilhança.

Consideremos a sucessão  $(X_n)_{n \in \mathbb{N}}$  de v.a. reais i.i.d. com  $X$  seguindo a lei de probabilidade  $Q_\theta$  de suporte  $S_\theta$  com  $\theta \in \Theta$  e  $\Theta \subset \mathbb{R}^k$ ,  $k \in \mathbb{N}$ , e  $\theta^0 = (\theta_1^0, \dots, \theta_k^0)$  o verdadeiro valor do parâmetro. Seja  $g_\theta(y)$ ,  $y \in \mathbb{R}$ , a função de probabilidade de  $Q_\theta$  se a lei é discreta.<sup>(8)</sup> A função de verosimilhança associada a uma qualquer observação  $x^{(n)} = (x_1, \dots, x_n)$  é, para cada  $n \in \mathbb{N}$ , dada por

$$\forall \theta \in \Theta, L(\theta; x^{(n)}) = \prod_{i=1}^n g_\theta(x_i).$$

As equações de verosimilhança são

$$\frac{\partial}{\partial \theta_i} \log(L(\theta; x^{(n)})) = 0, \quad i = 1, \dots, k. \quad (2.3)$$

Seja  $\mathfrak{X}$  o menor conjunto numerável, independente de  $\theta$ , que contém  $S_\theta$ ,  $\forall \theta \in \Theta$ .<sup>(9)</sup>

**Teorema 2.9** ([11], pág. 461-465). *Considerem-se as seguintes condições,*

- A) A aplicação  $\theta \mapsto Q_\theta$  é injectiva.
- B) Existe um aberto  $W \subseteq \Theta$  contendo  $\theta_0$ , tal que, para  $\forall x \in \mathfrak{X}$ ,  $g_\theta(x)$  admite derivadas de terceira ordem  $\forall \theta \in W$ .
- C) As primeiras e segundas derivadas do logaritmo de  $g_\theta(x)$  verificam

$$1. E_\theta \left[ \frac{\partial}{\partial \theta_j} \log g_\theta(X) \right] = 0, \quad j = 1, \dots, k,$$

$$2. I_{i,j}(\theta) = E_\theta \left[ \frac{\partial}{\partial \theta_i} \log g_\theta(X) \frac{\partial}{\partial \theta_j} \log g_\theta(X) \right] = E_\theta \left[ -\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log g_\theta(X) \right],$$

onde  $I(\theta) = [I_{i,j}(\theta)]_{1 \leq i, j \leq k}$  é a designada matriz de informação de Fisher.

- D)  $I(\theta)$  é finita e definida positiva.
- E) Existem funções  $M_{i,j,l}(x)$  tais que

$$\left| \frac{\partial^3}{\partial \theta_i \partial \theta_j \partial \theta_l} \log g_\theta(x) \right| \leq M_{i,j,l}(x), \quad \forall \theta \in W \text{ com } E_{\theta^0}(M_{i,j,l}) \text{ finita, } i, j, l = 1, \dots, k.$$

<sup>(7)</sup> Iremos denotar a convergência quase certa por  $\xrightarrow{q.c.}$ .

<sup>(8)</sup>  $g_\theta(y)$  é a função densidade de  $Q$  se a lei é absolutamente contínua.

<sup>(9)</sup> Se  $Q_\theta$  é absolutamente contínua,  $\mathfrak{X}$  é o menor subconjunto de  $\mathbb{R}$ , independente de  $\theta$ , que contém  $S_\theta$ ,  $\forall \theta \in \Theta$ .

Se as equações de verosimilhança, (2.3), admitem por solução única  $\hat{\theta}_n = (\hat{\theta}_{1,n}, \dots, \hat{\theta}_{k,n})$ , então

1.  $\hat{\theta}_{i,n} \xrightarrow{P} \theta_i, i = 1, \dots, k.$
2.  $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{L} Z, Z \sim N(0, I(\theta)^{-1}).$

### 2.6.2 Lei Geométrica

Considere-se o modelo estatístico associado a uma  $n$ -amostra  $(X_1, X_2, \dots, X_n)$  de  $X \sim ZDGGD(q, 0)$  com função de probabilidade  $(1 - q)q^x \mathbb{I}_{\mathbb{N}_0}(x), x \in \mathbb{R},$

$$\left( \mathbb{R}^n, \mathfrak{B}_n, G(q)^{(n)} \right)_{q \in ]0,1[}$$

onde  $G(q)^{(n)}$  é uma lei discreta com função de probabilidade

$$g_q(x_1, x_2, \dots, x_n) = \left[ (1 - q)^n q^{\sum_{i=1}^n x_i} \right] \mathbb{I}_{\mathbb{N}_0^n}(x_1, x_2, \dots, x_n), (x_1, x_2, \dots, x_n) \in \mathbb{R}^n.$$

#### Estimador baseado na proporção de zeros

Considerando a igualdade  $P(X = 0) = 1 - q$  e tendo em conta que  $P_{0,n} = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{X_i=0}$  é um estimador de  $P(X = 0)$  obtém-se o seguinte estimador para  $q$ :  $\hat{Q}_n = 1 - P_{0,n}$ . No teorema seguinte estabelecemos algumas propriedades do estimador  $\hat{Q}_n$ .

**Teorema 2.10.** *Considere-se  $\hat{Q}_n$  o estimador de  $q$  apresentado anteriormente. Então  $\hat{Q}_n$  é um estimador cêntrico,  $\hat{Q}_n \xrightarrow{m.q.} q$ , <sup>(10)</sup>,  $\hat{Q}_n \xrightarrow{q.c.} q$  e*

$$n^{1/2}(\hat{Q}_n - q) \xrightarrow{L} Z, Z \sim N(0, q(1 - q)).$$

*Prova.* 1. Notemos que a v.a.r.  $\mathbb{I}_{X_i=0}, i \in \mathbb{N}$ , segue uma lei de Bernoulli de parâmetro  $P(X = 0) = 1 - q$ , isto é,  $X_i \sim B(1 - q)$ . Verifiquemos primeiro que  $\hat{Q}_n$  é um estimador cêntrico,

$$E(\hat{Q}_n) = E(1 - P_{0,n}) = E\left(1 - \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{X_i=0}\right) = 1 - \frac{1}{n} \sum_{i=1}^n E(\mathbb{I}_{X_i=0}) = 1 - (1 - q) = q.$$

Uma vez que  $X_1, \dots, X_n$  são independentes, temos

$$V(\hat{Q}_n) = V(1 - P_{0,n}) = V(P_{0,n}) = V\left(\frac{1}{n} \sum_{i=1}^n \mathbb{I}_{X_i=0}\right) = \frac{1}{n^2} \sum_{i=1}^n V(\mathbb{I}_{X_i=0}) = \frac{(1 - q)q}{n}.$$

Segue-se que  $\lim_{n \rightarrow +\infty} V(\hat{Q}_n) = 0$  e consequentemente  $\hat{Q}_n \xrightarrow{m.q.} q$ .

2. Como  $(\mathbb{I}_{X_i>0})_{i \in \mathbb{N}}$  é uma sucessão de variáveis aleatórias reais independentes e com a mesma lei,  $B(P(X > 0))$ , verificando  $E(\mathbb{I}_{X_i>0}) = P(X > 0) = q$ , obtemos utilizando o Teorema de Kolmogorov  $\hat{Q}_n = 1 - P_{0,n} = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{X_i>0} \xrightarrow{q.c.} E(\mathbb{I}_{X_i>0}) = q$ .

<sup>(10)</sup> Iremos denotar a convergência em média quadrática por  $\xrightarrow{m.q.}$ .

3. Aliando o que foi mencionado no ponto 2 ao facto de  $V(\mathbb{I}_{X_i>0}) = q(1-q)$  segue-se, usando o Teorema do Limite central, que

$$\frac{\sum_{i=1}^n \mathbb{I}_{X_i>0} - nq}{\sqrt{q(1-q)n}} = n^{1/2} \frac{\frac{1}{n} \sum_{i=1}^n \mathbb{I}_{X_i>0} - q}{\sqrt{q(1-q)}} \xrightarrow{L} Z, Z \sim N(0, 1). \quad \blacksquare$$

### Estimador do método dos momentos

Como temos  $E(X) = m_1 = \frac{q}{1-q}$  podemos escrever  $q = \frac{1}{1+\frac{1}{m_1}}$ . Assim, tendo em conta o momento empírico de ordem 1,  $M_1 = \frac{1}{n} \sum_{i=1}^n X_i$ , um estimador do método dos momentos para  $q$  é

$$\tilde{Q}_n = \frac{1}{1 + \frac{1}{M_1}}. \quad (2.4)$$

Podemos estabelecer o resultado seguinte.

**Teorema 2.11.** *O estimador  $\tilde{Q}_n$ , apresentado em (2.4), verifica  $E(\tilde{Q}_n) < q$ ,  $\tilde{Q}_n \xrightarrow{q.c.} q$  e*

$$n^{1/2}(\tilde{Q}_n - q) \xrightarrow{L} Z, Z \sim N(0, q(1-q)^2).$$

*Prova.* Como  $f(x) = \frac{1}{1+\frac{1}{x}}$  é uma função estritamente côncava <sup>(11)</sup> para  $x > 0$ , utilizando a desigualdade de Jensen vem

$$E(\tilde{Q}_n) = E\left(\frac{1}{1 + \frac{1}{\frac{1}{n} \sum_{i=1}^n X_i}}\right) < \frac{1}{1 + \frac{1}{E\left(\frac{1}{n} \sum_{i=1}^n X_i\right)}} = \frac{1}{1 + \frac{1-q}{q}} = q.$$

Adicionalmente, seguindo o Teorema 2.8, como  $X$  admite momentos de segunda ordem e a solução do sistema obtido pelo método dos momentos é única então  $\tilde{Q}_n \xrightarrow{q.c.} q$  e tem-se

$$n^{1/2}(\tilde{Q}_n - q) \xrightarrow{L} Z, Z \sim N(0, D(q)^2 A(q))$$

com  $A(q) = V(X) = \frac{q}{(1-q)^2}$  e  $D(q) = \frac{\frac{1}{m_1^2}}{(1+\frac{1}{m_1})^2} = (1-q)^2$  pois  $f(m_1) = \frac{1}{1+\frac{1}{m_1}}$ .

Observamos que a normalidade assintótica de  $n^{1/2}(\tilde{Q}_n - q)$ , sai de forma natural também, conjugando o Teorema 2.6, onde obtemos  $\sqrt{n}(\frac{1}{n} \sum_{i=1}^n X_i - \theta) \xrightarrow{L} Z, Z \sim N(0, \frac{q}{(1-q)^2})$  com  $\theta = \frac{q}{(1-q)}$ , e o Teorema 2.7 usando  $h(\theta) = \frac{1}{1+\frac{1}{\theta}}$ . ■

### Estimador da máxima verosimilhança

A função de verosimilhança associada à realização  $x^{(n)} = (x_1, \dots, x_n)$  da amostra é dada por

$$L(q; x^{(n)}) = \prod_{i=1}^n (1-q)q^{x_i} = (1-q)^n q^{\sum_{i=1}^n x_i}, q \in ]0, 1[.$$

<sup>(11)</sup> Tem-se  $f^{(2)}(x) = -2(x+1)^{-3} < 0, x > 0$ .

Desta forma, um estimador da máxima verosimilhança (e.m.v.) é um maximizante de  $L(q; x^{(n)})$ . Como  $L(q; x^{(n)})$  é estritamente positiva e  $\log$  é uma função estritamente crescente então um maximizante de  $L(q; x^{(n)})$  é um maximizante de  $\log(L(q; x^{(n)}))$ , isto é, de

$$l(q; x^{(n)}) = \log L(q; x^{(n)}) = \log \left\{ (1-q)^n q^{\sum_{i=1}^n x_i} \right\} = n \log(1-q) + \sum_{i=1}^n x_i \log(q).$$

$$\text{Como} \quad \frac{\partial}{\partial q} l(q; x^{(n)}) = -\frac{n}{1-q} + \frac{\sum_{i=1}^n x_i}{q} = 0 \quad \Leftrightarrow \quad q = \frac{1}{1 + \frac{1}{\frac{1}{n} \sum_{i=1}^n x_i}},$$

$$\frac{\partial^2}{\partial q^2} l(q; x^{(n)}) = -\frac{n}{(1-q)^2} - \frac{\sum_{i=1}^n x_i}{q^2} < 0,$$

então o e.m.v. de  $q$  é dado por  $\hat{Q}_n = \frac{1}{1 + \frac{1}{\bar{x}_n}}$ . Note-se que o e.m.v. e o estimador dos momentos presente em (2.4), são iguais logo verificam as mesmas propriedades, ou seja, não é cêntrico, converge quase certamente para  $q$  e  $n^{1/2}(\hat{Q}_n - q)$  é assintoticamente normal.

### Eficiência relativa assintótica

A variância da lei limite associada ao estimador das proporções,  $q(1-q)$ , é superior à correspondente ao estimador dos momentos,  $q(1-q)^2$ . Uma medida de eficiência relativa assintótica do estimador das proporções,  $\hat{Q}$ , relativamente ao estimador dos momentos,  $\tilde{Q}$ , inspirada em [10], será  $\frac{q(1-q)^2}{q(1-q)} = 1-q < 1$ . Assim, para além de ser mais eficiente assintoticamente, a eficiência relativa do estimador dos momentos aumenta quando  $q$  aumenta.

### 2.6.3 Lei ZDGGD( $q, \alpha$ )

Vamos agora estimar os parâmetros do modelo geral ZDGGD( $q, \alpha$ ). Considere-se o modelo estatístico associado a uma  $n$ -amostra  $(X_1, X_2, \dots, X_n)$  de  $X \sim \text{ZDGGD}(q, \alpha)$

$$\left( \mathbb{R}^n, \mathfrak{B}_n, \text{ZDGGD}(q, \alpha)^{(n)} \right)_{(q, \alpha) \in ]0, 1[ \times ]-1, +\infty[}$$

onde  $\text{ZDGGD}(q, \alpha)^{(n)}$  é uma lei discreta com função de probabilidade

$$\vec{g}_{q, \alpha}(x_1, x_2, \dots, x_n) = \left[ \prod_{i=1}^n [(1-q^{\alpha+1})]^{\mathbb{I}_{x_i=0}} \prod_{i=1}^n [(1-q)q^{x_i+\alpha}]^{\mathbb{I}_{x_i>0}} \right] \mathbb{I}_{\mathbb{N}_0}(x_1, \dots, x_n), (x_1, x_2, \dots, x_n) \in \mathbb{R}^n.$$

### Estimador baseado nas proporções de zeros e de uns

Como  $P(X=0) = 1 - q^{\alpha+1}$  e  $P(X=1) = (1-q)q^{1+\alpha}$  obtemos o seguinte sistema

$$\begin{cases} P(X=1) = (1-q)[1 - P(X=0)] \\ q^{\alpha+1} = 1 - P(X=0), \end{cases} \quad \text{que permite chegar a}$$

$$(q, \alpha) = \left( 1 - \frac{P(X=1)}{1 - P(X=0)}, \frac{2 \log[1 - P(X=0)] - \log[1 - P(X=0) - P(X=1)]}{\log[1 - P(X=0) - P(X=1)] - \log[1 - P(X=0)]} \right)$$

Tendo em conta que  $P_{0,n} = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{X_i=0}$  e  $P_{1,n} = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{X_i=1}$  são estimadores de  $P(X=0)$  e  $P(X=1)$  respectivamente, obtemos os seguintes estimadores para  $q$  e  $\alpha$

$$(\hat{Q}_n, \hat{A}_n) = \left( \frac{P_{1,n}}{1 - P_{0,n}}, \frac{2 \log(1 - P_{0,n}) - \log(1 - P_{0,n} - P_{1,n})}{\log(1 - P_{0,n} - P_{1,n}) - \log(1 - P_{0,n})} \right). \quad (2.5)$$

O resultado seguinte estabelece a convergência em probabilidade e em lei destes estimadores.

**Teorema 2.12.** *Os estimadores  $\hat{Q}_n$  e  $\hat{A}_n$ , presentes em (2.5), verificam  $\hat{Q}_n \xrightarrow{P} q$ ,  $\hat{A}_n \xrightarrow{P} \alpha$  e*

$$n^{1/2} \left( \begin{bmatrix} \hat{Q}_n \\ \hat{A}_n \end{bmatrix} - \begin{bmatrix} q \\ \alpha \end{bmatrix} \right) \xrightarrow{L} Z, \quad Z \sim N(0, B(q, \alpha) \Sigma(q, \alpha) B(q, \alpha)^T) \quad (2.6)$$

com  $B(q, \alpha) = [b_{i,j}]_{1 \leq i, j \leq 2}$  e  $\Sigma(q, \alpha) = [\sigma_{i,j}]_{1 \leq i, j \leq 2}$  tais que

$$\begin{aligned} \sigma_{1,1} &= p_0(1 - p_0), \quad \sigma_{1,2} = \sigma_{2,1} = -p_0 p_1, \quad \sigma_{2,2} = p_1(1 - p_1), \\ b_{1,1} &= \frac{-p_1}{(1 - p_0)^2}, \quad b_{1,2} = \frac{-1}{1 - p_0}, \\ b_{2,1} &= \frac{(1 - p_0 - p_1) \log(1 - p_0 - p_1) + (1 - p_0) \log(1 - p_0)}{(p_0 - 1)(p_0 + p_1 - 1) (\log(1 - p_0 - p_1) - \log(1 - p_0))^2}, \\ b_{2,2} &= \frac{\log(1 - p_0)}{(1 - p_0 - p_1) (\log(1 - p_0 - p_1) - \log(1 - p_0))^2} \text{ onde } p_0 = 1 - q^{\alpha+1} \text{ e } p_1 = (1 - q)q^{\alpha+1}. \end{aligned}$$

*Prova.* 1. Como  $(\mathbb{I}_{X_n=0})$ ,  $n \in \mathbb{N}$ , são v.a. reais i.i.d. seguindo uma lei de Bernoulli de parâmetro  $P(X=0)$ ,  $B(P(X=0))$ , com  $E(\mathbb{I}_{X_n=0}) = P(X=0) = 1 - q^{\alpha+1}$  e  $(\mathbb{I}_{X_n=1})$ ,  $n \in \mathbb{N}$ , são i.i.d. seguindo uma lei de Bernoulli  $B(P(X=1))$  com  $E(\mathbb{I}_{X_n=1}) = P(X=1) = (1 - q)q^{\alpha+1}$  segue-se aplicando o teorema de Kolmogorov que  $P_{0,n} \xrightarrow{q.c.} 1 - q^{\alpha+1}$  e  $P_{1,n} \xrightarrow{q.c.} (1 - q)q^{\alpha+1}$ . Logo  $P_{0,n} \xrightarrow{P} 1 - q^{\alpha+1}$  e  $P_{1,n} \xrightarrow{P} (1 - q)q^{\alpha+1}$ . Usando as propriedades da convergência em probabilidade, vem  $\hat{Q}_n = 1 - \frac{P_{1,n}}{1 - P_{0,n}} \xrightarrow{P} 1 - \frac{(1 - q)q^{\alpha+1}}{1 - 1 + q^{\alpha+1}} = q$ . Tem-se  $1 - P_{0,n} \xrightarrow{P} q^{\alpha+1}$  e  $1 - P_{0,n} - P_{1,n} \xrightarrow{P} q^{\alpha+2}$ . Agora como  $\log(x)$  é uma função contínua em  $]0, +\infty[$  segue-se, ([12], pág.174),  $\log(1 - P_{0,n}) \xrightarrow{P} \log(q^{\alpha+1}) = (\alpha + 1) \log(q)$  e  $\log(1 - P_{0,n} - P_{1,n}) \xrightarrow{P} \log(q^{\alpha+2}) = (\alpha + 2) \log(q)$ . Usando mais uma vez as propriedades da convergência em probabilidade conclui-se  $\hat{A}_n \xrightarrow{P} \frac{2(\alpha+1)\log(q) - (\alpha+2)\log(q)}{(\alpha+2)\log(q) - (\alpha+1)\log(q)} = \alpha$ .

2. Como  $Y_n = (\mathbb{I}_{X_n=0}, \mathbb{I}_{X_n=1})$ ,  $n \in \mathbb{N}$ , são v.a. i.i.d. com média  $E(Y_n) = (p_0, p_1)$  onde  $p_0 = 1 - q^{\alpha+1}$  e  $p_1 = (1 - q)q^{\alpha+1}$  e matriz de variâncias-covariâncias

$$\Sigma = \begin{bmatrix} V(\mathbb{I}_{X_n=0}) & Cov(\mathbb{I}_{X_n=0}, \mathbb{I}_{X_n=1}) \\ Cov(\mathbb{I}_{X_n=0}, \mathbb{I}_{X_n=1}) & V(\mathbb{I}_{X_n=1}) \end{bmatrix} = \begin{bmatrix} p_0(1 - p_0) & -p_0 p_1 \\ -p_0 p_1 & p_1(1 - p_1) \end{bmatrix}$$

então usando o Teorema 2.6, obtemos  $n^{1/2} \left( \begin{bmatrix} P_{0,n} \\ P_{1,n} \end{bmatrix} - \begin{bmatrix} p_0 \\ p_1 \end{bmatrix} \right) \xrightarrow{L} Z, \quad Z \sim N(0, \Sigma)$ .

Considerando agora as funções  $h_1(p_0, p_1) = 1 - \frac{p_1}{1-p_0}$  e  $h_2(p_0, p_1) = \frac{2 \log(1-p_0) - \log(1-p_0-p_1)}{\log(1-p_0-p_1) - \log(1-p_0)}$  e usando o Teorema 2.7 obtemos

$$n^{1/2} \left( \begin{bmatrix} h_1(P_{0,n}, P_{1,n}) \\ h_2(P_{0,n}, P_{1,n}) \end{bmatrix} - \begin{bmatrix} h_1(p_0, p_1) \\ h_2(p_0, p_1) \end{bmatrix} \right) \xrightarrow{L} Z, \quad Z \sim N(0, B \Sigma B^T). \quad (2.7)$$

com  $B = \left[ \frac{\partial h_i}{\partial p_{j-1}} \right]_{1 \leq i, j \leq 2}$  onde  $\frac{\partial h_1}{\partial p_0} = \frac{-p_1}{(1-p_0)^2}$ ,  $\frac{\partial h_1}{\partial p_1} = \frac{-1}{1-p_0}$ ,

$$\begin{aligned} \frac{\partial h_2}{\partial p_0} &= \frac{(1-p_0-p_1) \log(1-p_0-p_1) + (1-p_0) \log(1-p_0)}{(p_0-1)(p_0+p_1-1) (\log(1-p_0-p_1) - \log(1-p_0))^2}, \\ \frac{\partial h_2}{\partial p_1} &= \frac{\log(1-p_0)}{(1-p_0-p_1) (\log(1-p_0-p_1) - \log(1-p_0))^2}. \end{aligned} \quad \blacksquare$$

### Método dos momentos

Tendo em conta que  $E(X) = m_1 = \frac{q^{\alpha+1}}{1-q}$  e  $E(X^2) = m_2 = \frac{q^{\alpha+1}(1+q)}{(1-q)^2}$  deduzimos

$$\begin{cases} m_2 = \frac{(1+q)}{(1-q)} m_1 \\ \alpha = \frac{\log(m_1) + \log(1-q) - \log(q)}{\log(q)} \end{cases} \Leftrightarrow \begin{cases} q = \frac{m_2 - m_1}{m_2 + m_1} \\ \alpha = \frac{\log(2m_1^2) - \log(m_2 - m_1)}{\log(m_2 - m_1) - \log(m_2 + m_1)}. \end{cases}$$

Considerando os momentos empíricos de ordem 1 e 2,  $M_1 = \frac{1}{n} \sum_{i=1}^n X_i$  e  $M_2 = \frac{1}{n} \sum_{i=1}^n X_i^2$ , um estimador para  $(q, \alpha)$  obtido pelo método dos momentos é

$$(\tilde{Q}_n, \tilde{A}_n) = \left( \frac{M_2 - M_1}{M_2 + M_1}, \frac{\log(2M_1^2) - \log(M_2 - M_1)}{\log(M_2 - M_1) - \log(M_2 + M_1)} \right). \quad (2.8)$$

Estes estimadores são quase certamente convergentes. Mais geralmente vale o seguinte resultado.

**Teorema 2.13.** *Sejam  $\tilde{Q}_n$  e  $\tilde{A}_n$  definidos em (2.8). Tem-se  $(\tilde{Q}_n, \tilde{A}_n) \xrightarrow{q.c.} (q, \alpha)$  e*

$$n^{1/2} \left( \begin{bmatrix} \tilde{Q}_n \\ \tilde{A}_n \end{bmatrix} - \begin{bmatrix} q \\ \alpha \end{bmatrix} \right) \xrightarrow{L} Z, \quad Z \sim N(0, D(q, \alpha) A(q, \alpha) D(q, \alpha)^T)$$

com  $A(q, \alpha) = [a_{i,j}(q, \alpha)]_{1 \leq i, j \leq 2}$  e  $D(q, \alpha) = [d_{i,j}(q, \alpha)]_{1 \leq i, j \leq 2}$  onde

$$\begin{aligned} a_{1,1} &= \frac{q^{\alpha+1}(1+q-q^{\alpha+1})}{(1-q)^2}, \quad a_{1,2} = a_{2,1} = \frac{q^{\alpha+1}(1+4q+q^2)}{(1-q)^3} - \frac{q^{\alpha+1}(1+q)}{(1-q)^2} \frac{q^{\alpha+1}}{(1-q)}, \\ a_{2,2} &= \frac{q^{\alpha+1}(1+11q+11q^2+q^3) - q^{2\alpha+2}(1+q)^2}{(1-q)^4}, \quad d_{1,1} = \frac{-2m_2}{(m_1+m_2)^2}, \quad d_{1,2} = \frac{2m_1}{(m_1+m_2)^2}, \\ d_{2,1} &= \frac{\frac{1}{m_2-m_1} [\log(2m_1^2) - \log(m_1+m_2)] + \frac{2}{m_1} [\log(m_2-m_1) - \log(m_1+m_2)] + \frac{1}{m_2+m_1} \log\left(\frac{2m_1^2}{m_2-m_1}\right)}{[\log(m_2-m_1) - \log(m_2+m_1)]^2} \end{aligned}$$

$$d_{2,2} = \frac{\frac{-1}{m_2-m_1} [\log(2m_1^2) - \log(m_1 + m_2)] + \frac{1}{m_2+m_1} [\log(2m_1^2) - \log(m_2 - m_1)]}{[\log(m_2 - m_1) - \log(m_2 + m_1)]^2}.$$

*Prova.* Como  $X$  admite momento de 4ª ordem e o estimador do método dos momentos verifica  $(\hat{Q}_n, \hat{A}_n) = (f_1(M_1, M_2), f_2(M_1, M_2))$  com  $f_1(m_1, m_2) = \frac{m_2-m_1}{m_2+m_1}$  e  $f_2(m_1, m_2) = \frac{\log(2m_1^2) - \log(m_2 - m_1)}{\log(m_2 - m_1) - \log(m_2 + m_1)}$ , seguindo o Teorema 2.8 conclui-se a normalidade assintótica, onde a lei normal é centrada e possui como matriz de variâncias-covariâncias  $D(q, \alpha)A(q, \alpha)D(q, \alpha)^T$ , com  $A(\theta) = [Cov(X^j, X^l)]_{1 \leq j, l \leq 2}$  isto é,  $Cov(X, X) = V(X) = \frac{q^{\alpha+1}(1+q-q^{\alpha+1})}{(1-q)^2}$ ,

$$Cov(X, X^2) = Cov(X^2, X) = E(X^3) - E(X^2)E(X) = \frac{q^{\alpha+1}(1+4q+q^2)}{(1-q)^3} - \frac{q^{\alpha+1}(1+q)}{(1-q)^2} \frac{q^{\alpha+1}}{(1-q)},$$

$$\begin{aligned} Cov(X^2, X^2) &= E(X^4) - E(X^2)E(X^2) = \frac{q^{\alpha+1}(1+11q+11q^2+q^3)}{(1-q)^4} - \left[ \frac{q^{\alpha+1}(1+q)}{(1-q)^2} \right]^2, \\ &= \frac{q^{\alpha+1}(1+11q+11q^2+q^3) - q^{2\alpha+2}(1+q)^2}{(1-q)^4} \end{aligned}$$

e  $D = \left[ \frac{\partial f_i}{\partial m_j} \right]_{1 \leq i, j \leq 2}$  ou seja,  $\frac{\partial f_1}{\partial m_1} = \frac{-2m_2}{(m_1+m_2)^2}$ ,  $\frac{\partial f_1}{\partial m_2} = \frac{2m_1}{(m_1+m_2)^2}$ ,

$$\begin{aligned} \frac{\partial f_2}{\partial m_1} &= \frac{\frac{1}{m_2-m_1} [\log(2m_1^2) - \log(m_1 + m_2)] + \frac{2}{m_1} \log\left(\frac{m_2-m_1}{m_1+m_2}\right) + \frac{1}{m_2+m_1} \log\left(\frac{2m_1^2}{m_2-m_1}\right)}{[\log(m_2 - m_1) - \log(m_2 + m_1)]^2}, \\ \frac{\partial f_2}{\partial m_2} &= \frac{\frac{-1}{m_2-m_1} [\log(2m_1^2) - \log(m_1 + m_2)] + \frac{1}{m_2+m_1} [\log(2m_1^2) - \log(m_2 - m_1)]}{[\log(m_2 - m_1) - \log(m_2 + m_1)]^2}. \quad \blacksquare \end{aligned}$$

### Estimador da Máxima Verosimilhança

Seja  $x^{(n)} = (x_1, \dots, x_n)$  uma realização da  $n$ -amostra. A função de verosimilhança é dada por

$$L(q, \alpha; x^{(n)}) = \left[ \prod_{i=1}^n [(1-q^{\alpha+1})]^{\mathbb{I}_{x_i=0}} \prod_{i=1}^n [(1-q)q^{x_i+\alpha}]^{\mathbb{I}_{x_i>0}} \right], (q, \alpha) \in ]0, 1[ \times ]-1, +\infty[$$

Consideremos as seguintes notações:  $m = m(x^{(n)}) = \sum_{i=1}^n \mathbb{I}_{x_i=0}$  = número de zeros observados em  $x^{(n)}$  e  $z = z(x^{(n)}) = \sum_{i=1}^n x_i \mathbb{I}_{x_i>0}$ .

Então  $L(q, \alpha; x^{(n)})$  pode escrever-se na forma  $L(q, \alpha; x^{(n)}) = (1-q^{\alpha+1})^m [(1-q)q^\alpha]^{(n-m)} q^z$ . A log-verosimilhança é então

$$\begin{aligned} l(q, \alpha; x^{(n)}) &= \log L(q, \alpha; x^{(n)}) = \log \left\{ (1-q^{\alpha+1})^m ((1-q)q^\alpha)^{(n-m)} q^z \right\} \\ &= m \log(1-q^{\alpha+1}) + (n-m) \log((1-q)q^\alpha) + \log(q)z. \end{aligned}$$

A solução das equações de verosimilhança

$$\begin{cases} \frac{\partial l}{\partial q} = -\frac{m(\alpha+1)q^\alpha}{1-q^{\alpha+1}} - \frac{(n-m)}{(1-q)} + \frac{\alpha(n-m)}{q} + \frac{1}{q}z = 0 \\ \frac{\partial l}{\partial \alpha} = m \frac{-\log(q)q^{\alpha+1}}{1-q^{\alpha+1}} + (n-m) \log(q) = 0 \end{cases}$$



é dada por  $(\hat{Q}_n, \hat{A}_n) = \left(1 - \frac{n-m}{z}, \frac{\log\left(1 - \frac{m}{n}\right)}{\log\left(1 - \frac{n-m}{z}\right)} - 1\right)$ . Logo o estimador da máxima verosimilhança é

$$(\hat{Q}_n, \hat{A}_n) = \left(1 - \frac{n - \sum_{i=1}^n \mathbb{I}_{X_i=0}}{\sum_{i=1}^n X_i \mathbb{I}_{X_i>0}}, \frac{\log\left(1 - \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{X_i=0}\right)}{\log\left(1 - \frac{n - \sum_{i=1}^n \mathbb{I}_{X_i=0}}{\sum_{i=1}^n \mathbb{I}_{X_i>0} X_i}\right)} - 1\right). \quad (2.9)$$

No Teorema seguinte estabelecemos propriedades destes estimadores.

**Teorema 2.14.** *Os estimadores  $\hat{Q}_n$  e  $\hat{A}_n$ , presentes em (2.9), verificam  $\hat{Q}_n \xrightarrow{P} q$ ,  $\hat{A}_n \xrightarrow{P} \alpha$  e*

$$n^{1/2} \left( \begin{bmatrix} \hat{Q}_n \\ \hat{A}_n \end{bmatrix} - \begin{bmatrix} q \\ \alpha \end{bmatrix} \right) \xrightarrow{L} Z, \quad Z \sim N(0, I((q, \alpha))^{-1}) \quad (2.10)$$

$$\text{com } I(q, \alpha) = [I_{i,j}]_{1 \leq i,j \leq 2} \text{ tal que } I_{1,1} = q^{\alpha-1} \left[ \frac{(1-q^2)(\alpha+1)^2 + (1-q^{\alpha+1})(1-(1-q)^2)}{(1-q^{\alpha+1})(1-q)^2} \right],$$

$$I_{1,2} = I_{2,1} = q^\alpha \frac{(\alpha+1)\log(q)}{1-q^{\alpha+1}}, \quad I_{2,2} = \frac{\log^2(q) q^{\alpha+1}}{1-q^{\alpha+1}}.$$

*Prova.* Para estabelecer o resultado vamos verificar a validade das condições do Teorema 2.9. Começamos por nos restringirmos ao espaço dos parâmetros  $\Theta = ]0, 1[ \times ]-1, +\infty[$ , definir  $\mathfrak{X} = \mathbb{N}_0$  e considerar  $g_{q,\alpha}(x)$  a função de probabilidade da lei ZDGGD.

- A) A aplicação  $(q, \alpha) \mapsto \text{ZDGGD}(q, \alpha)$  é injectiva.
- B) Seja  $W = ]\delta_1, 1 - \delta_1[ \times ]-1 + \delta_1, \delta_2[$  com  $0 < \delta_1 < \frac{1}{2}$  e  $\delta_2 > 0$  contendo os verdadeiros valores  $q$  e  $\alpha$ . Então,  $\forall x \in \mathbb{N}_0$ ,  $g_{q,\alpha}$  admite derivadas de terceira ordem  $\forall (q, \alpha) \in W$ .
- C) Vamos provar que  $E\left(\frac{\partial \log(g_{q,\alpha}(X))}{\partial q}\right) = 0$  e que  $E\left(\frac{\partial \log(g_{q,\alpha}(X))}{\partial \alpha}\right) = 0$ . Temos

x	0	1, 2, ...
$\frac{\partial \log(g_{q,\alpha}(x))}{\partial q}$	$-\frac{(\alpha+1)q^\alpha}{1-q^{\alpha+1}}$	$-\frac{1}{1-q} + \frac{x+\alpha}{q}$
$\frac{\partial \log(g_{q,\alpha}(x))}{\partial \alpha}$	$-\frac{q^{\alpha+1} \log(q)}{1-q^{\alpha+1}}$	$\log(q)$

Consequentemente,

$$\begin{aligned} i) E\left(\frac{\partial \log(g_{q,\alpha}(X))}{\partial q}\right) &= \sum_{x=0}^{+\infty} \frac{\partial \log(g_{q,\alpha}(x))}{\partial q} P(X=x) \\ &= -(\alpha+1)q^\alpha + \sum_{x=1}^{+\infty} \left(-\frac{1}{1-q} + \frac{x+\alpha}{q}\right) P(X=x) \\ &= -(\alpha+1)q^\alpha - \frac{1}{1-q} \sum_{x=1}^{+\infty} P(X=x) - \frac{\alpha(1-q^{\alpha+1})}{q} + \frac{\alpha(1-q^{\alpha+1})}{q} + \sum_{x=1}^{+\infty} \left(\frac{x+\alpha}{q}\right) P(X=x) \\ &= -(\alpha+1)q^\alpha - \frac{q^{\alpha+1}}{1-q} - \frac{\alpha(1-q^{\alpha+1})}{q} + \frac{q^\alpha}{1-q} + \frac{\alpha}{q} = 0, \end{aligned}$$

$$\begin{aligned} \text{ii) } E\left(\frac{\partial \log(g_{q,\alpha}(X))}{\partial \alpha}\right) &= \sum_{x=0}^{+\infty} \frac{\partial \log(g_{q,\alpha}(x))}{\partial \alpha} P(X=x) = -q^{\alpha+1} \log(q) + \sum_{x=1}^{+\infty} \log(q) P(X=x) \\ &= -q^{\alpha+1} \log(q) + q^{\alpha+1} \log(q) = 0. \end{aligned}$$

Denotando  $q = \theta_1$  e  $\alpha = \theta_2$  iremos agora verificar que

$$E\left(\frac{\partial \log(g_{\theta_1,\theta_2}(X))}{\partial \theta_i} \frac{\partial \log(g_{\theta_1,\theta_2}(X))}{\partial \theta_j}\right) = -E\left(\frac{\partial^2 \log(g_{\theta_1,\theta_2}(X))}{\partial \theta_i \partial \theta_j}\right) < +\infty, i, j = 1, 2. \quad (2.11)$$

Começemos por verificar que  $E\left(\frac{\partial^2 \log(g_{\theta_1,\theta_2}(X))}{\partial \theta_i \partial \theta_j}\right) < +\infty, i, j = 1, 2$ . O quadro seguinte permite organizar os cálculos.

x	0	1, 2, ...
$\frac{\partial^2 \log(g_{q,\alpha}(x))}{\partial q \partial q}$	$-\frac{(\alpha+1)q^{\alpha-1}(q^{\alpha+1}+\alpha)}{(1-q^{\alpha+1})^2}$	$-\frac{1}{(1-q)^2} - \frac{(x+\alpha)}{q^2}$
$\frac{\partial^2 \log(g_{q,\alpha}(x))}{\partial \alpha \partial q}$	$-\frac{(\alpha+1)q^{2\alpha+1} \log(q)}{(1-q^{\alpha+1})^2} - \frac{q^\alpha}{(1-q^{\alpha+1})} - \frac{(\alpha+1)q^\alpha \log(q)}{(1-q^{\alpha+1})}$	$\frac{1}{q}$
$\frac{\partial^2 \log(g_{q,\alpha}(x))}{\partial \alpha \partial \alpha}$	$-\frac{q^{\alpha+1} \log^2(q)}{(1-q^{\alpha+1})^2}$	0

$$\begin{aligned} \text{Obtemos, } E\left(\frac{\partial^2 \log(g_{q,\alpha}(x))}{\partial q \partial q}\right) &= \sum_{x=0}^{+\infty} \frac{\partial^2 \log(g_{q,\alpha}(x))}{\partial q \partial q} P(X=x) \\ &= -(\alpha+1) \frac{q^{\alpha-1}(\alpha+q^{\alpha+1})}{(1-q^{\alpha+1})^2} (1-q^{\alpha+1}) + \sum_{x=1}^{+\infty} \left(-\frac{1}{(1-q)^2} - \frac{(x+\alpha)}{q^2}\right) (1-q)^{q^{\alpha+x}} \\ &= -(\alpha+1) \frac{q^{\alpha-1}(\alpha+q^{\alpha+1})}{(1-q^{\alpha+1})} - \frac{1}{1-q} \sum_{x=1}^{+\infty} q^{x+\alpha} - \frac{1}{q^2} \sum_{x=1}^{+\infty} x(1-q)q^{x+\alpha} - \frac{\alpha}{q} \sum_{x=1}^{+\infty} (1-q)q^{x+\alpha} \\ &= -(\alpha+1) \frac{q^{\alpha-1}(\alpha+q^{\alpha+1})}{(1-q^{\alpha+1})} - \frac{q^{\alpha+1}}{(1-q)^2} - \frac{q^{\alpha+1}}{q^2(1-q)} - \alpha q^{\alpha+1} \\ &= -q^{\alpha-1} \left[ \frac{(1-q^2)(\alpha+1)^2 + (1-q^{\alpha+1})(1-(1-q)^2)}{(1-q^{\alpha+1})(1-q)^2} \right] < +\infty. \quad (2.12) \end{aligned}$$

Analogamente

$$\begin{aligned} E\left(\frac{\partial^2 \log(g_{q,\alpha}(x))}{\partial \alpha \partial q}\right) &= \sum_{x=0}^{+\infty} \frac{\partial^2 \log(g_{q,\alpha}(x))}{\partial \alpha \partial q} P(X=x) \\ &= -\left(\frac{(\alpha+1)q^{2\alpha+1} \log(q)}{(1-q^{\alpha+1})^2} + \frac{q^\alpha}{(1-q^{\alpha+1})} + \frac{(\alpha+1)q^\alpha \log(q)}{(1-q^{\alpha+1})}\right) (1-q^{\alpha+1}) + \frac{1}{q} \sum_{x=1}^{+\infty} P(X=x) \\ &= -\frac{(\alpha+1)q^\alpha \log(q)}{(1-q^{\alpha+1})} - q^\alpha + \frac{q^{\alpha+1}}{q} = -\frac{(\alpha+1)q^\alpha \log(q)}{(1-q^{\alpha+1})} < +\infty \quad (2.13) \end{aligned}$$

e ainda

$$E\left(\frac{\partial^2 \log(g_{q,\alpha}(x))}{\partial \alpha^2}\right) = \sum_{x=0}^{+\infty} \frac{\partial^2 \log(g_{q,\alpha}(x))}{\partial \alpha \partial q} P(X=x) = \frac{-q^{\alpha+1} \log^2(q)}{(1-q)^{\alpha+1}} < +\infty \quad (2.14)$$

Observamos que, para  $i, j = 1, 2$ ,

$$\frac{\partial^2 \log(g_{\theta_1, \theta_2}(x))}{\partial \theta_i \partial \theta_j} = \frac{\partial^2 g_{\theta_1, \theta_2}(x)}{\partial \theta_i \partial \theta_j} \frac{1}{g_{\theta_1, \theta_2}(x)} - \frac{\frac{\partial g_{\theta_1, \theta_2}(x)}{\partial \theta_i} \frac{\partial g_{\theta_1, \theta_2}(x)}{\partial \theta_j}}{(g_{\theta_1, \theta_2}(x))^2}.$$

Assim

$$E\left(\frac{\partial^2 \log(g_{\theta_1, \theta_2}(x))}{\partial \theta_i \partial \theta_j}\right) = E\left(\frac{\partial^2 g_{\theta_1, \theta_2}(x)}{\partial \theta_i \partial \theta_j} \frac{1}{g_{\theta_1, \theta_2}(x)} - \frac{\frac{\partial g_{\theta_1, \theta_2}(x)}{\partial \theta_i} \frac{\partial g_{\theta_1, \theta_2}(x)}{\partial \theta_j}}{(g_{\theta_1, \theta_2}(x))^2}\right), i, j = 1, 2$$

Verifiquemos agora que  $E\left(\frac{\partial^2 g_{\theta_1, \theta_2}(x)}{\partial \theta_i \partial \theta_j} \frac{1}{g_{\theta_1, \theta_2}(x)}\right) = 0$ ,  $i, j = 1, 2$ , o que permitirá concluir (2.11).

x	0	1, 2, ...
$\frac{\partial g_{q,\alpha}(x)}{\partial q}$	$-(\alpha+1)q^\alpha$	$-q^{x+\alpha} + (x+\alpha)(1-q)q^{x+\alpha-1}$
$\frac{\partial g_{q,\alpha}(x)}{\partial \alpha}$	$-q^{\alpha+1} \log(q)$	$(1-q)q^{x+\alpha} \log(q)$
$\frac{\partial^2 g_{q,\alpha}(x)}{\partial q \partial q}$	$-(\alpha+1)\alpha q^{\alpha-1}$	$-2(x+\alpha)q^{x+\alpha-1} + (x+\alpha)(x+\alpha-1)(1-q)q^{x+\alpha-2}$
$\frac{\partial^2 g_{q,\alpha}(x)}{\partial \alpha \partial q}$	$-q^\alpha [(\alpha+1) \log(q) + 1]$	$(1-q)q^{x+\alpha-1} + (1-q)(x+\alpha) \log(q)q^{x+\alpha-1} - \log(q)q^{x+\alpha}$
$\frac{\partial^2 g_{q,\alpha}(x)}{\partial \alpha^2}$	$-q^{\alpha+1} \log^2(q)$	$(1-q)q^{x+\alpha} \log^2(q)$

Tendo em conta os resultados que se encontram organizados no quadro anterior vem

$$\begin{aligned} E\left(\frac{\partial^2 g_{q,\alpha}(X)}{\partial q \partial q} \frac{1}{g_{q,\alpha}(X)}\right) &= \sum_{x=0}^{\infty} \frac{\partial^2 g_{q,\alpha}(x)}{\partial q \partial q} \frac{1}{g_{q,\alpha}(x)} g_{q,\alpha}(x) = \sum_{x=0}^{\infty} \frac{\partial^2 g_{q,\alpha}(x)}{\partial q \partial q} \\ &= -(\alpha+1)\alpha q^{\alpha-1} - 2 \sum_{x=0}^{+\infty} (x+\alpha)q^{x+\alpha-1} + \sum_{x=0}^{+\infty} (x+\alpha)(x+\alpha-1)(1-q)q^{x+\alpha-2} \\ &= -(\alpha+1)\alpha q^{\alpha-1} - \frac{2}{q(1-q)} [E(X+\alpha) - \alpha(1-q^{\alpha+1})] + \\ &+ \frac{1}{q^2} [E((X+\alpha)(X+\alpha-1)) - \alpha(\alpha-1)(1-q^{\alpha+1})] \\ &= \frac{-\alpha^2 q^{\alpha+1} - \alpha q^{\alpha+1}}{q^2} - \frac{2}{q(1-q)} [E(X) + \alpha q^{\alpha+1}] + \\ &+ \frac{1}{q^2} [E(X^2) + 2\alpha E(X) - E(X) + \alpha^2 q^{\alpha+1} - \alpha q^{\alpha+1}] \\ &= -\frac{2}{q(1-q)} E(X) - \frac{2\alpha}{q} E(X) + \frac{(1+q)E(X)}{q^2(1-q)} + \frac{2\alpha E(X) - E(X)}{q^2} - \frac{2\alpha E(X)(1-q)}{q^2} = 0 \end{aligned}$$

$$e \quad E\left(\frac{\partial^2 g_{q,\alpha}(X)}{\partial q \partial \alpha} \frac{1}{g_{q,\alpha}(X)}\right) = \sum_{x=0}^{\infty} \frac{\partial^2 g_{q,\alpha}(x)}{\partial q \partial \alpha} \frac{1}{g_{q,\alpha}(x)} g_{q,\alpha}(x) = \sum_{x=0}^{\infty} \frac{\partial^2 g_{q,\alpha}(x)}{\partial q \partial \alpha}$$

$$\begin{aligned}
&= -q^\alpha(\alpha+1)\log(q) - q^\alpha + \frac{1}{q} \sum_{x=1}^{+\infty} (1-q)q^{x+\alpha} + \\
&+ \frac{1}{q} \sum_{x=0}^{+\infty} (x+\alpha)\log(q)(1-q)q^{x+\alpha} - \frac{\log(q)}{(1-q)} \sum_{x=1}^{+\infty} (1-q)q^{x+\alpha} \\
&= -q^\alpha(\alpha+1)\log(q) + \frac{\log(q)}{q} [E(X+\alpha) - (1-q^{\alpha+1})\alpha] - \frac{\log(q)}{1-q} q^{\alpha+1} = 0
\end{aligned}$$

$$\begin{aligned}
\text{e também } E\left(\frac{\partial^2 g_{q,\alpha}(X)}{\partial\alpha\partial\alpha} \frac{1}{g_{q,\alpha}(X)}\right) &= \sum_{x=0}^{\infty} \frac{\partial^2 g_{q,\alpha}(x)}{\partial\alpha\partial\alpha} \frac{1}{g_{q,\alpha}(x)} g_{q,\alpha}(x) = \sum_{x=0}^{\infty} \frac{\partial^2 g_{q,\alpha}(x)}{\partial\alpha\partial\alpha} \\
&= -q^{\alpha+1}\log^2(q) + \log^2(q) \sum_{x=1}^{+\infty} (1-q)q^{x+\alpha} = 0.
\end{aligned}$$

- D) Agora iremos verificar que a seguinte matriz é definida positiva mostrando que os determinantes das suas submatrizes principais são positivos.

$$I(q, \alpha) = - \begin{bmatrix} E\left(\frac{\partial^2 \log(g_{q,\alpha}(x))}{\partial q^2}\right) & E\left(\frac{\partial^2 \log(g_{q,\alpha}(x))}{\partial q \partial \alpha}\right) \\ E\left(\frac{\partial^2 \log(g_{q,\alpha}(x))}{\partial \alpha \partial q}\right) & E\left(\frac{\partial^2 \log(g_{q,\alpha}(x))}{\partial \alpha^2}\right) \end{bmatrix}$$

De facto,

$$i) E\left(-\frac{\partial^2 \log(g_{q,\alpha}(x))}{\partial q \partial q}\right) = q^{\alpha-1} \left[ \frac{(1-q^2)(\alpha+1)^2 + (1-q^{\alpha+1})(1-(1-q)^2)}{(1-q^{\alpha+1})(1-q)^2} \right] > 0$$

$$\begin{aligned}
ii) E\left(-\frac{\partial^2 \log(g_{q,\alpha}(x))}{\partial q \partial q}\right) E\left(-\frac{\partial^2 \log(g_{q,\alpha}(x))}{\partial \alpha \partial \alpha}\right) &- \left[ E\left(-\frac{\partial^2 \log(g_{q,\alpha}(x))}{\partial \alpha \partial q}\right) \right]^2 \\
&= q^{\alpha-1} \left[ \frac{(1-q)^2(\alpha+1)^2 + (1-q^{\alpha+1})(1-(1-q)^2)}{(1-q^{\alpha+1})(1-q)^2} \right] \frac{\log^2(q)q^{\alpha+1}}{(1-q^{\alpha+1})} - \frac{q^{2\alpha}(\alpha+1)^2 \log^2(q)}{(1-q^{\alpha+1})^2} \\
&= \frac{q^{2\alpha} \log^2(q)}{(1-q^{\alpha+1})} \left[ \frac{(1-q)^2(\alpha+1)^2 + (1-q^{\alpha+1})(1-(1-q)^2) - (\alpha+1)^2(1-q)^2}{(1-q^{\alpha+1})(1-q)^2} \right] \\
&= \frac{q^{2\alpha} \log^2(q)}{(1-q^{\alpha+1})} \left[ \frac{(1-q^{\alpha+1})(1-(1-q)^2)}{(1-q^{\alpha+1})(1-q)^2} \right] > 0.
\end{aligned}$$

- E) Por fim iremos verificar que  $\left| \frac{\partial^3}{\partial \theta_i \partial \theta_j \partial \theta_l} \log g_{\theta_1, \theta_2}(x) \right| \leq M_{i,j,l}(x)$ ,  $\forall (q, \alpha) \in W$  e que  $E[M_{i,j,l}(X)] < +\infty$  com  $i, j, l = 1, 2$ . Recorremos ao quadro seguinte,

x	0	1, 2, ...
$\frac{\partial^3 \log(g_{q,\alpha}(x))}{\partial q \partial q \partial q}$	$m_1(q, \alpha)$	$-\frac{2}{(1-q)^3} + \frac{(k+\alpha)}{q^3}$
$\frac{\partial^3 \log(g_{q,\alpha}(x))}{\partial \alpha \partial q \partial q}$	$m_2(q, \alpha)$	$-\frac{1}{q^2}$
$\frac{\partial^3 \log(g_{q,\alpha}(x))}{\partial \alpha \partial \alpha \partial \alpha}$	$m_3(q, \alpha)$	0
$\frac{\partial^3 \log(g_{q,\alpha}(x))}{\partial q \partial \alpha \partial \alpha}$	$m_4(q, \alpha)$	0

com

$$\begin{aligned} m_1(q, \alpha) &= \frac{\partial}{\partial q} \frac{\partial^2 \log(g_{q,\alpha}(x))}{\partial q \partial q} \Big|_{x=0} = \frac{\partial}{\partial q} - \frac{(\alpha+1)q^{\alpha-1}(q^{\alpha+1} + \alpha)}{(1-q^{\alpha+1})^2} \\ &= \frac{(\alpha-1)(\alpha+1)(q^{\alpha+1} + \alpha)q^{\alpha-2}}{(1-q^{\alpha+1})^2} + \frac{(\alpha+1)^2 q^{2\alpha-1}}{(1-q^{\alpha+1})^2} + \frac{2(\alpha+1)^2 (q^{\alpha+1} + \alpha)q^{2\alpha-1}}{(1-q^{\alpha+1})^3} \end{aligned}$$

De modo análogo se obtém  $m_i(q, \alpha)$  para  $i = 2, 3, 4$ , vindo

$$\begin{cases} m_2(q, \alpha) = \frac{(q^{\alpha+1} + \alpha)q^{\alpha-1}}{(1-q^{\alpha+1})^2} + \frac{(\alpha+1)(q^{\alpha+1} + \alpha)q^{\alpha-1} \log(q)}{(1-q^{\alpha+1})^2} + \frac{(\alpha+1)q^{\alpha-1}(q^{\alpha+1} \log(q) + 1)}{(1-q^{\alpha+1})^2} + \\ \quad + \frac{2(\alpha+1)(q^{\alpha+1} + \alpha)q^{2\alpha} \log(q)}{(1-q^{\alpha+1})^3} \\ m_3(q, \alpha) = \frac{-q^{\alpha+1}(q^{\alpha+1} + 1) \log^3(q)}{(1-q^{\alpha+1})^3} \\ m_4(q, \alpha) = \frac{-2(\alpha+1)q^{2\alpha+1} \log^2(q)}{(1-q^{\alpha+1})^3} - \frac{(\alpha+1)q^\alpha \log^2(q)}{(1-q^{\alpha+1})^2} - \frac{2q^\alpha \log(q)}{(1-q^{\alpha+1})^2} \end{cases}$$

Como para  $i = 1, \dots, 4$ ,  $|m_i(q, \alpha)|$  é contínua no compacto  $[\delta_1, 1 - \delta_1] \times [-1 + \delta_1, \delta_2]$  então pelo teorema de Weierstrass  $|m_i(q, \alpha)|$  possui máximo em  $[\delta_1, 1 - \delta_1] \times [-1 + \delta_1, \delta_2]$  logo definindo  $s_i = \sup_{(q,\alpha) \in W} |m_i(q, \alpha)|$  obtemos  $|m_i(q, \alpha)| \leq s_i, \forall (q, \alpha) \in W$ . Tendo em conta ainda

que  $\left| -\frac{2}{(1-q)^3} + \frac{(k+\alpha)}{q^3} \right| \leq \frac{2}{(1-q)^3} + \frac{(k+\alpha)}{q^3} \leq \frac{2}{(\delta_1)^3} + \frac{(k+\alpha)}{\delta_1^3}$  definem-se

x	0	1, 2, ...
$M_{1,1,1}(x)$	$ s_1 $	$\frac{2}{(\delta_1)^3} + \frac{(x+\alpha)}{\delta_1^3}$
$M_{2,1,1}(x) = M_{1,2,1}(x) = M_{1,1,2}(x)$	$ s_2 $	$\frac{1}{\delta_1^2}$
$M_{2,2,2}(x)$	$ s_3 $	0
$M_{1,2,2}(x) = M_{2,1,2}(x) = M_{2,2,1}(x)$	$ s_4 $	0

que verificam  $E(M_{i,j,l}(X)) < +\infty, i, j, l = 1, 2$ . Está pois estabelecida a convergência em probabilidade dos e.m.v. e a convergência para uma lei normal centrada e de matriz de variâncias-covariâncias  $I(q, \alpha)^{-1}$  com  $I(q, \alpha) = [I_{i,j}]_{1 \leq i, j \leq 2}$  onde

$$I_{1,1} = E\left(-\frac{\partial^2 \log(g_{q,\alpha}(x))}{\partial q \partial q}\right), I_{1,2} = I_{2,1} = E\left(-\frac{\partial^2 \log(g_{q,\alpha}(x))}{\partial q \partial \alpha}\right), I_{2,2} = E\left(-\frac{\partial^2 \log(g_{q,\alpha}(x))}{\partial \alpha \partial \alpha}\right),$$

deduzidos em (2.12), (2.13) e (2.14) a menos da multiplicação por  $-1$ . ■

### Eficiência relativa assintótica

Quando o parâmetro a estimar é multidimensional, a eficiência relativa assintótica dos estimadores obtidos por dois procedimentos de estimação distintos pode ser avaliada, de acordo com [16], pelo quociente dos determinantes das matrizes variâncias-covariâncias das leis limite dos estimadores convenientemente normalizados, isto é,  $\frac{\det(\Sigma_1)}{\det(\Sigma_2)}$ . No caso em estudo limitamos a análise aos estimadores de  $(q, \alpha)$  pelo método dos momentos e das proporções e, face à complexidade das expressões, ilustramos graficamente a superfície  $\left\{ \left( \frac{\det(B(q,\alpha)\Sigma(q,\alpha)B(q,\alpha)^T)}{\det(D(q,\alpha)A(q,\alpha)D(q,\alpha)^T)} \right)^{1/2} : (q, \alpha) \in ]0.2, 0.8[ \times ]-1, 5[ \right\}$  que

se encontra na Figura 2.3. Podemos observar que quando  $\alpha$  se aproxima de  $-1$ , o estimador das proporções,  $(\hat{Q}_n, \hat{A}_n)$ , é mais eficiente que o dos momentos,  $(\tilde{Q}_n, \tilde{A}_n)$ . Para valores de  $\alpha$  afastados de  $-1$ : se  $q$  apresenta valores elevados então  $(\tilde{Q}_n, \tilde{A}_n)$  é mais eficiente que  $(\hat{Q}_n, \hat{A}_n)$ ; se  $q$  admite valores baixos os estimadores apresentam a quase a mesma eficiência.

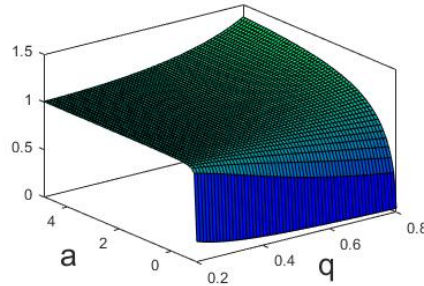


Fig. 2.3 Eficiência relativa assintótica do estimador dos momentos em relação ao das proporções.

## 2.7 Simulação

Nesta secção pretendemos avaliar por meio de estudos de simulação o comportamento dos estimadores fornecidos pelos três métodos de estimação em amostras de média e elevada dimensão. Avaliaremos os valores médios produzidos e a correspondente variabilidade, e a evolução destes resumos com a dimensão da amostra. Por fim, construímos regiões de confiança para os parâmetros, com base nas leis limites dos estimadores e comparamos o grau de confiança fixado com a denominada probabilidade de cobertura, isto é, com a proporção de amostras geradas onde a respectiva região de confiança estimada contém os verdadeiros valores dos parâmetros.

### 2.7.1 Estimação pontual

Apresentamos agora um estudo de simulação que visa comparar o desempenho dos três métodos de estimação dos parâmetros  $q$  e  $\alpha$ . Geramos uma amostra de dimensão  $n \in \{100, 500\}$  e de seguida obtemos as estimativas dos parâmetros pelos três métodos de estimação em estudo. Este procedimento foi repetido 1000 vezes, com  $q \in \{0.4, 0.6, 0.8\}$  e  $\alpha \in \{-0.7, -0.3, 0.5\}$ . As Tabelas 2.2, 2.3 e 2.4 apresentam as médias e os desvios padrão das estimativas correspondentes aos estimadores das proporções, dos momentos e máxima verosimilhança, respectivamente. Observa-se que em geral as médias das estimativas se aproximam dos valores verdadeiros com acentuado decréscimo do respectivo desvio padrão quando  $n$  aumenta. O estimador da máxima verosimilhança apresenta o melhor comportamento, com menores valores para o desvio padrão e com as médias das estimativas bastante próximas dos verdadeiros valores. Quando  $n = 100$ , as estimativas para  $\alpha$  não são muito boas, qualquer que seja o método de estimação, com desvios padrão bastante significativos. Em contrapartida para o valor de  $q$  os três métodos apresentam bons resultados.

Tabela 2.2 Resultados dos estimadores do Método das Proporções.

		$n = 100$				$n = 500$			
$\alpha$	$q$	Média( $\alpha$ )	DP( $\alpha$ )	Média( $q$ )	DP( $q$ )	Média( $\alpha$ )	DP( $\alpha$ )	Média( $q$ )	DP( $q$ )
-0.7	0.4	-0.6950	0.0818	0.3990	0.0571	-0.6987	0.0347	0.4005	0.0252
	0.6	-0.6908	0.1017	0.6002	0.0575	-0.6977	0.0442	0.6004	0.0241
	0.8	-0.6830	0.1568	0.7998	0.0425	-0.6931	0.0636	0.8007	0.0184
-0.3	0.4	-0.2822	0.1835	0.3992	0.0675	-0.2966	0.0737	0.4004	0.0298
	0.6	-0.2702	0.2046	0.6006	0.0593	-0.2931	0.0863	0.6008	0.0264
	0.8	-0.2631	0.2896	0.7976	0.0446	-0.2915	0.1169	0.7999	0.0189
0.5	0.4	0.5874	0.5073	0.4015	0.0986	0.5213	0.2035	0.4010	0.0432
	0.6	0.5795	0.4968	0.5993	0.0745	0.5143	0.1969	0.5993	0.0329
	0.8	0.6300	0.6560	0.7997	0.0488	0.5265	0.2284	0.8003	0.0220

Tabela 2.3 Resultados dos estimadores do Método dos Momentos.

		$n = 100$				$n = 500$			
$\alpha$	$q$	Média( $\alpha$ )	DP( $\alpha$ )	Média( $q$ )	DP( $q$ )	Média( $\alpha$ )	DP( $\alpha$ )	Média( $q$ )	DP( $q$ )
-0.7	0.4	-0.7014	0.1080	0.3935	0.0506	-0.7014	0.0461	0.3985	0.0230
	0.6	-0.7049	0.1808	0.5937	0.0414	-0.7022	0.0784	0.5987	0.0188
	0.8	-0.7098	0.4345	0.7957	0.0246	-0.7049	0.1875	0.7992	0.0110
-0.3	0.4	-0.3049	0.1950	0.3908	0.0598	-0.3025	0.0855	0.3981	0.0274
	0.6	-0.3026	0.2684	0.5924	0.0461	-0.3025	0.1150	0.5985	0.0208
	0.8	-0.3067	0.5167	0.7955	0.0256	-0.3056	0.2224	0.7991	0.0115
0.5	0.4	0.4771	0.4469	0.3830	0.0841	0.4928	0.2024	0.3955	0.0380
	0.6	0.4824	0.4337	0.5901	0.0547	0.4964	0.2068	0.5974	0.0247
	0.8	0.4826	0.6348	0.7955	0.0274	0.4938	0.3009	0.7988	0.0122

Tabela 2.4 Resultados dos estimadores da Máxima Verosimilhança.

		$n = 100$				$n = 500$			
$\alpha$	$q$	Média( $\alpha$ )	DP( $\alpha$ )	Média( $q$ )	DP( $q$ )	Média( $\alpha$ )	DP( $\alpha$ )	Média( $q$ )	DP( $q$ )
-0.7	0.4	-0.6984	0.0738	0.3969	0.0434	-0.6998	0.0321	0.3996	0.0195
	0.6	-0.6992	0.0853	0.5973	0.0341	-0.6993	0.0387	0.5998	0.0149
	0.8	-0.7004	0.1216	0.7982	0.0188	-0.6969	0.0556	0.7999	0.0082
-0.3	0.4	-0.2953	0.1513	0.3958	0.0519	-0.2994	0.0637	0.3995	0.0230
	0.6	-0.2914	0.1606	0.5969	0.0379	-0.2981	0.0696	0.5997	0.0168
	0.8	-0.2969	0.1986	0.7980	0.0197	-0.2980	0.0901	0.7998	0.0086
0.5	0.4	0.5252	0.3772	0.3938	0.0776	0.5062	0.1609	0.3986	0.0327
	0.6	0.5121	0.2986	0.5963	0.0454	0.5044	0.1424	0.5990	0.0202
	0.8	0.5181	0.3514	0.7986	0.0219	0.5032	0.1452	0.7996	0.0094

Observamos que quando  $\alpha = -0.7$  (isto é,  $\alpha$  próximo de  $-1$ ) e  $n = 500$  a estimação de  $q$  pelo método das proporções parece ser melhor do que a obtida pelo método dos momentos. Em suma os resultados da simulação apoiam os resultados teóricos no que toca à consistência dos estimadores, sendo, aparentemente, a velocidade de convergência do estimador de  $\alpha$  menor do que a do estimador de  $q$ . Fixamos agora o parâmetro  $(q, \alpha) = (0.4, -0.7)$ . Comparamos os valores  $nV(\hat{q}_n)$ ,  $nV(\hat{\alpha}_n)$  e  $nCov(\hat{q}_n, \hat{\alpha}_n)$  relativos ao método das proporções com os seus valores teóricos, isto é, com os elementos da matriz de variâncias-covariâncias da lei assintótica obtida no Teorema 2.12 calculada nos pontos  $(q, \alpha) = (0.4, -0.7)$ . Repetimos o processo para os estimadores dos momentos e da máxima verosimilhança, usando os Teoremas 2.13 e 2.14 respectivamente, e resumimos os valores obtidos nas Tabelas 2.5, 2.6 e 2.7.

Tabela 2.5 Comparação dos resultados empíricos com os teóricos pelo método das proporções.

	n=100	n=500	Teóricos
$E(\hat{q}_n)$	0.3990	0.4005	0.4
$E(\hat{\alpha}_n)$	-0.6950	-0.6987	-0.7
$nV(\hat{q}_n)$	0.3264	0.3179	0.3159
$nV(\hat{\alpha}_n)$	0.6690	0.6022	0.5885
$nCov(\hat{q}_n, \hat{\alpha}_n)$	0.2806	0.2655	0.2586

Tabela 2.6 Comparação dos resultados empíricos com os teóricos pelo método dos momentos.

	n=100	n=500	Teóricos
$E(\tilde{q}_n)$	0.3935	0.3985	0.4
$E(\tilde{\alpha}_n)$	-0.7014	-0.7014	-0.7
$nV(\tilde{q}_n)$	0.2556	0.2641	0.2654
$nV(\tilde{\alpha}_n)$	1.1673	1.0648	1.0313
$nCov(\tilde{q}_n, \tilde{\alpha}_n)$	0.3667	0.3619	0.3551

Tabela 2.7 Comparação dos resultados empíricos com os teóricos pelo método da máxima verosimilhança.

	n=100	n=500	Teóricos
$E(\hat{q}_n)$	0.3969	0.3996	0.4
$E(\hat{\alpha}_n)$	-0.6984	-0.6998	-0.7
$nV(\hat{q}_n)$	0.1883	0.1899	0.1896
$nV(\hat{\alpha}_n)$	0.5450	0.5150	0.5038
$nCov(\hat{q}_n, \hat{\alpha}_n)$	0.1624	0.1617	0.1552

Os valores das Tabelas 2.5, 2.6 e 2.7 mostram que para  $n = 100$  as estimativas dos elementos das matrizes de variâncias-covariâncias da lei assintótica estão próximas dos valores teóricos excepto as relativas à variância dos estimadores de  $\alpha$ . Para  $n = 500$  observamos uma melhoria em todas



as estimativas em relação ao caso  $n = 100$ , com todas as estimativas bastante próximas dos valores teóricos.

### 2.7.2 Regiões de Confiança

Seja  $(\hat{Q}_n, \hat{A}_n)$  o estimador da máxima verosimilhança, dado em (2.9). No seguimento do Teorema 2.14 podemos aceitar, quando  $n$  é suficientemente grande, que

$$n^{1/2} \left( \begin{bmatrix} \hat{Q}_n \\ \hat{A}_n \end{bmatrix} - \begin{bmatrix} q \\ \alpha \end{bmatrix} \right) \sim N(0, I((q, \alpha))^{-1}). \quad (2.15)$$

Tendo em conta agora que se  $Y$  é um vector aleatório de dimensão 2, com  $Y \sim N(m, \Sigma)$ , então  $(Y - m)\Sigma^{-1}(Y - m)^T \sim \chi_2^2$ , ([19], pág.103-105), podemos obter uma região de confiança para  $(q, \alpha)$  com grau de confiança assintótico  $a \in ]0, 1[$ , a partir, de (2.15) considerando a seguinte região  $I = \left\{ (q, \alpha) : \begin{bmatrix} \hat{Q}_n - q & \hat{A}_n - \alpha \end{bmatrix} nI(q, \alpha) \begin{bmatrix} \hat{Q}_n - q & \hat{A}_n - \alpha \end{bmatrix}^T \leq z \right\}$  onde  $z$  representa o quantil  $a$  da lei  $\chi_2^2$ . A interpretação frequentista de probabilidade permite afirmar que se obtivermos muitas realizações da amostra de dimensão  $n$ , e se em cada realização verificarmos se a respectiva região de confiança estimada inclui  $(q, \alpha)$ , a proporção de vezes que essas regiões contêm  $(q, \alpha)$  é aproximadamente  $a$ . Vamos pois avaliar a eficácia desta região, com  $n \in \{100, 500, 1000\}$ , para estimar a verdadeira probabilidade de cobertura  $P((q, \alpha) \in I)$  com  $q, \alpha$  os verdadeiros valores, considerando  $a = 0.95$ . Para cada  $(q, \alpha)$  e cada  $n$  calculamos as estimativas de  $q$  e  $\alpha$  e a região e verificamos se o verdadeiro par  $(q, \alpha)$  pertence à região, repetindo este procedimento 10 000 vezes. A proporção de sucessos é uma estimativa da verdadeira probabilidade de cobertura ( $\simeq 0.95$ ). Usamos ainda a mesma metodologia para o método das proporções e o método dos momentos utilizando os Teoremas 2.12 e 2.13 respectivamente. Na Tabela 2.8 encontram-se os resultados obtidos, onde  $P$ ,  $M$  e  $V$  representam as referidas proporções tendo em conta o método das proporções, o método dos momentos e máxima verosimilhança, respectivamente.

Tabela 2.8 Estimativas das probabilidades de cobertura da região de confiança com confiança assintótica 0.95.

		$n = 100$			$n = 500$			$n = 1000$		
$\alpha$	$q$	P	M	V	P	M	V	P	M	V
-0.7	0.4	0.9395	0.9380	0.9439	0.9505	0.9465	0.9499	0.9513	0.9490	0.9512
	0.6	0.9413	0.9377	0.9494	0.9457	0.9437	0.9479	0.9534	0.9460	0.9507
	0.8	0.9349	0.9368	0.9480	0.9460	0.9436	0.9462	0.9517	0.9448	0.9492
-0.3	0.4	0.9348	0.9117	0.9396	0.9432	0.9340	0.9458	0.9483	0.9408	0.9489
	0.6	0.9312	0.9259	0.9428	0.9451	0.9396	0.9476	0.9513	0.9419	0.9492
	0.8	0.9283	0.9319	0.9463	0.9419	0.9407	0.9492	0.9484	0.9442	0.9513
0.5	0.4	0.9004	0.8596	0.9224	0.9308	0.8950	0.9414	0.9421	0.9145	0.9464
	0.6	0.8991	0.8866	0.9334	0.9353	0.9271	0.9431	0.9457	0.9368	0.9497
	0.8	0.8968	0.9178	0.9407	0.9368	0.9378	0.9454	0.9443	0.9414	0.9503

Observamos na Tabela 2.8 que quando  $n = 100$  é para o método da máxima verosimilhança que se observa mais concordância entre as proporções de vezes que o valor de  $(q, \alpha)$  pertence à região de confiança estimada e o valor 0.95. As diferenças entre 0.95 e as proporções observadas com os três métodos diminuem quando  $n$  aumenta, como esperado.

## 2.8 Aplicação a dados reais

Começamos por recordar três modelos clássicos habitualmente usados quando pretendemos modelar dados de contagem, nomeadamente as lei de Poisson, negativa binomial e binomial. Em seguida destacamos a nova distribuição discreta e a distribuição Poisson-Lindley. Iremos apresentar características teóricas de todas as lei mencionadas, nomeadamente a esperança, variância, inflação ou deflação em zero e o índice de dispersão com o objectivo de identificar modelos que possam reproduzir as características empíricas de dados reais. Por fim analisaremos três conjuntos de dados reais.

### 2.8.1 Modelos clássicos de contagem

Diremos que a v.a.r.  $X$ , com  $S_X = \mathbb{N}_0$ , segue uma lei de Poisson de parâmetro  $\lambda > 0$ ,  $X \sim P(\lambda)$ , se

$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}, \quad k \in \mathbb{N}_0.$$

O modelo admite  $E(X) = V(X) = \lambda$  e consequentemente  $ID = 1$ . É inflacionado em zero se e só se  $\lambda < 1$ .

Dizemos que a v.a.r.  $X$ , com  $S_X = \mathbb{N}_0$ , segue uma lei negativa binomial de parâmetros  $r$  e  $p$  com  $r > 0$  e  $0 < p < 1$ ,  $X \sim NB(r, p)$ , se

$$P(X = k) = \frac{\Gamma(r+k)}{\Gamma(r)\Gamma(k+1)} (1-p)^r p^k, \quad k \in \mathbb{N}_0.$$

Temos  $E(X) = \frac{pr}{1-p}$  e  $V(X) = \frac{pr}{(1-p)^2}$  e consequentemente  $ID = \frac{1}{1-p} > 1$ , ou seja a lei NB apresenta sobredispersão para todos os valores dos parâmetros. Notemos ainda que no caso  $r = 1$  obtemos a lei Geométrica. É inflacionada em zero se e só se  $p < \frac{1}{r}$ .

Por fim, a v.a.r.  $X$  segue uma lei binomial de parâmetros  $n$  e  $p$ , com  $S_X = \{0, 1, \dots, n\}$ , verificando  $n \in \mathbb{N}$  e  $0 < p < 1$ ,  $X \sim Bi(n, p)$ , se

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n.$$

Neste caso tem-se  $E(X) = np$  e  $V(X) = np(1-p)$  e consequentemente  $ID = (1-p) < 1$ , isto é, a lei é subdispersa. É inflacionada em zero se e só se  $(1-p)^n > np(1-p)^{n-1}$  isto é se e só se  $p < \frac{1}{1+n}$ .

### 2.8.2 Nova Distribuição Discreta

A v.a.r.  $N$  segue uma nova distribuição discreta, com parâmetros  $\alpha$  e  $\theta$  tais que  $\alpha \neq 0$ ,  $\alpha < 1$  e  $0 < \theta < 1$ , se

$$P(N = n) = \frac{\log(1 - \alpha\theta^n) - \log(1 - \alpha\theta^{n+1})}{\log(1 - \alpha)}, \quad n \in S_N = \mathbb{N}_0.$$

Esta lei foi introduzida em [8], e escrevemos  $N \sim ND(\alpha, \theta)$ . De facto trata-se de uma lei de probabilidade já que

$$\lim_{m \rightarrow +\infty} \sum_{n=0}^m P(N = n) = \lim_{m \rightarrow +\infty} \frac{\log(1 - \alpha) - \log(1 - \alpha\theta^{m+1})}{\log(1 - \alpha)} = 1$$

e  $P(N = n) > 0$ ,  $\forall n \in \mathbb{N}_0$ , pois para  $0 < \theta < 1$ , se  $0 < \alpha < 1$  então  $\log(1 - \alpha) < 0$  e  $\log(\frac{1 - \alpha\theta^n}{1 - \alpha\theta^{n+1}}) < 0$ , se  $\alpha < 0$ ,  $\log(1 - \alpha) > 0$  e  $\log(\frac{1 - \alpha\theta^n}{1 - \alpha\theta^{n+1}}) > 0$ .

A função de distribuição é dada por

$$\forall x \in \mathbb{R}, F_N(x) = P(N \leq x) = \begin{cases} 0, & \text{se } x < 0 \\ 1 - \frac{\log(1 - \alpha\theta^{\lfloor x \rfloor + 1})}{\log(1 - \alpha)} & \text{se } x \geq 0. \end{cases}$$

Esta distribuição coincide com a lei geométrica quando  $\alpha \rightarrow 0$ . De facto,  $\lim_{\alpha \rightarrow 0} \frac{\frac{\partial}{\partial \alpha} \log(1 - \alpha\theta^{n+1})}{\frac{\partial}{\partial \alpha} \log(1 - \alpha)} = \lim_{\alpha \rightarrow 0} \frac{(1 - \alpha)\theta^{n+1}}{(1 - \alpha\theta^{n+1})} = \theta^{n+1}$ ; assim, usando a regra de l'Hôpital obtemos  $\lim_{\alpha \rightarrow 0} P(N \leq n) = 1 - \theta^{n+1}$ ,  $n \in \mathbb{N}_0$ . Tem-se

$$E(N) = \frac{1}{\log(1 - \alpha)} \sum_{n=1}^{+\infty} \log(1 - \alpha\theta^n),$$

$$V(N) = \frac{1}{\log(1 - \alpha)} \sum_{n=1}^{+\infty} (2n - 1) \log(1 - \alpha\theta^n) - \left[ \frac{1}{\log(1 - \alpha)} \sum_{n=1}^{+\infty} \log(1 - \alpha\theta^n) \right]^2.$$

Devido à complexidade destas expressões torna-se difícil analisar o valor do índice de dispersão. Mas tomando como exemplo dois casos para os parâmetros, se  $(\alpha, \theta) = (-0.1, 0.25)$  então  $V(N) = 0.459 > E(N) = 0.346$  e se  $(\alpha, \theta) = (0.1, 0.1)$  tem-se  $V(N) = 0.118 > E(N) = 0.105$ , ou seja, verificamos que a  $ND$  admite sobredispersão para alguns valores dos parâmetros. Apresenta também subdispersão pois considerando por exemplo  $(\alpha, \theta) = (-1, 0.10)$  obtemos  $V(N) = 0.165 < E(N) = 0.0.257$ . É uma lei com inflação em zero pois é unimodal sendo a moda zero. De facto considerando a função  $f(x) = \frac{\log(1 - \alpha\theta^x) - \log(1 - \alpha\theta^{x+1})}{\log(1 - \alpha)}$  para  $x \geq 0$  vem

$$\begin{aligned} f'(x) &= \left[ \frac{\log(1 - \alpha\theta^x) - \log(1 - \alpha\theta^{x+1})}{\log(1 - \alpha)} \right]' = \frac{1}{\log(1 - \alpha)} \left[ \frac{-\alpha\theta^x \log(\theta)}{\log(1 - \alpha\theta^x)} + \frac{-\alpha\theta^{x+1} \log(\theta)}{1 - \alpha\theta^{x+1}} \right] \\ &= -\alpha\theta^x \frac{\log(\theta)}{\log(1 - \alpha)} \left[ \frac{1 - \theta}{(1 - \alpha\theta^{x+1})(1 - \alpha\theta^x)} \right] < 0, \quad \forall \alpha, \theta \text{ tais que } \alpha \neq 0 \wedge \alpha < 1, 0 < \theta < 1. \end{aligned}$$

Assim  $f(x)$  é uma função estritamente decrescente para  $x \geq 0$  logo  $(P(N = n), n \in \mathbb{N}_0)$  é também estritamente decrescente.

Na Figura 2.4 apresentam-se histogramas de realizações das leis  $ND(-0.1, 0.25)$  e  $ND(0.1, 0.1)$  respectivamente, destacando-se a inflação em zero nos dois casos como foi anteriormente referido.

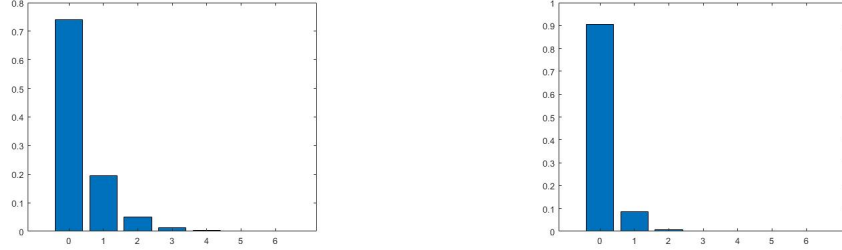


Fig. 2.4 Histogramas de uma realização da lei  $ND(-0.1, 0.25)$  (esquerda) e da lei  $ND(0.1, 0.1)$  (direita).

### 2.8.3 Lei de Poisson-Lindley

Uma v.a.r.  $X$ , com  $S_X = \mathbb{N}_0$ , segue a lei de Poisson-Lindley de parâmetro  $\theta > 0$ , que abreviamos para  $X \sim PL(\theta)$ , se

$$P(X = k) = \frac{\theta^2(k + \theta + 2)}{(\theta + 1)^{k+3}}, \quad k \in \mathbb{N}_0.$$

Trata-se de uma lei de probabilidade, pois

$$\begin{aligned} \sum_{k=0}^{+\infty} P(X = k) &= \sum_{k=0}^{+\infty} \frac{\theta^2(k + \theta + 2)}{(\theta + 1)^{k+3}} = \frac{\theta^2}{(\theta + 1)^3} \left[ \frac{1}{(\theta + 1)} \sum_{k=0}^{+\infty} k \left( \frac{1}{\theta + 1} \right)^{k-1} + (\theta + 2) \sum_{k=0}^{+\infty} \left( \frac{1}{\theta + 1} \right)^k \right] \\ &= \frac{\theta^2}{(\theta + 1)^3} \left[ \frac{1}{(\theta + 1)} \frac{(\theta + 1)^2}{\theta^2} + (\theta + 2) \frac{\theta + 1}{\theta} \right] = \frac{\theta^2}{(\theta + 1)^3} \frac{(\theta + 1)^3}{\theta^2} = 1. \end{aligned}$$

O nome da lei vem do facto de ter sido criada a partir de uma lei de Poisson quando o parâmetro  $\lambda$  segue uma lei de Lindley de parâmetro  $\theta$ ,  $L(\theta)$ , isto é, com função densidade dada pela expressão  $f_\Lambda(\lambda) = \frac{\theta^2}{(\theta + 1)}(1 + \lambda)e^{-\theta\lambda}\mathbb{I}_{\lambda > 0}$ . De facto, considerando  $X|\Lambda = \lambda \sim P(\lambda)$  e  $\Lambda \sim L(\theta)$  então

$$\begin{aligned} P(X = k) &= \int_{-\infty}^{+\infty} P(X = k|\Lambda = \lambda)f_\Lambda(\lambda)d\lambda = \int_0^{+\infty} \frac{e^{-\lambda}\lambda^k}{k!} \frac{\theta^2}{(\theta + 1)}(1 + \lambda)e^{-\theta\lambda}d\lambda \\ &= \frac{\theta^2}{k!(\theta + 1)} \int_0^{+\infty} (\lambda^k + \lambda^{k+1})e^{-(\theta+1)\lambda}d\lambda = \frac{\theta^2}{k!(\theta + 1)} \left[ \frac{\Gamma(k+1)}{(\theta + 1)^{k+1}} + \frac{\Gamma(k+2)}{(\theta + 1)^{k+2}} \right] \\ &= \frac{\theta^2(k + \theta + 2)}{(\theta + 1)^{k+3}}. \end{aligned}$$

Temos ainda

$$\begin{aligned} E(X) &= E(E(X|\Lambda = \lambda)) = \int_0^{+\infty} \lambda \frac{e^{-\lambda}\lambda^k}{k!} \frac{\theta^2}{(\theta + 1)}(1 + \lambda)e^{-\theta\lambda}d\lambda \\ &= \frac{\theta^2}{(\theta + 1)} \left[ \frac{\Gamma(2)}{\theta^2} + \frac{\Gamma(3)}{\theta^3} \right] = \frac{\theta + 2}{(\theta + 1)\theta} \end{aligned}$$

e  $V(X) = \frac{\theta^3 + 4\theta^2 + 6\theta + 2}{\theta^2(\theta + 1)^2}$ , pois

$$\begin{aligned} E(X^2) &= E(E(X^2|\Lambda = \lambda)) = \int_0^{+\infty} (\lambda + \lambda^2) \frac{e^{-\lambda} \lambda^k}{k!} \frac{\theta^2}{(\theta + 1)} (1 + \lambda) e^{-\theta\lambda} d\lambda \\ &= \frac{\theta + 2}{(\theta + 1)\theta} + \frac{\theta^2}{(\theta + 1)} \left[ \frac{\Gamma(3)}{\theta^3} + \frac{\Gamma(4)}{\theta^4} \right] = \frac{2\theta + 6}{\theta^2(\theta + 1)}. \end{aligned}$$

Deduzimos, como foi notado em [4], que  $ID = 1 + \frac{\theta^2 + 4\theta + 2}{\theta(\theta + 1)(\theta + 2)} > 1$ , ou seja a lei é sobredispersa. Por fim, podemos notar que se  $\theta > \sqrt{2} - 1$  então  $P(X = 0) > P(X = 1)$  e que se  $0 < \theta < \sqrt{2} - 1$  obtemos  $P(X = 0) < P(X = 1)$ , que ilustramos na Figura 2.5.

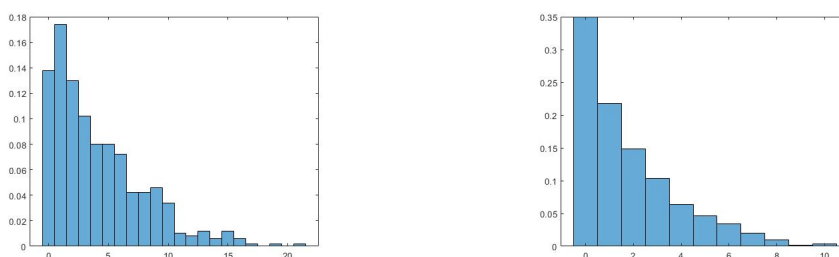


Fig. 2.5 Histogramas de uma realização da lei  $PL(0.4)$  (esquerda) e da lei  $PL(0.8)$  (direita).

Resumimos na Tabela 2.9 alguma da informação atrás mencionada que nos pode ajudar a seleccionar um modelo para um conjunto de dados reais. Verificamos que a lei  $ZDGGD$  é a única que acomoda simultaneamente deflação e inflação em zero tal como subdispersão e sobredispersão.

Tabela 2.9 Resumo de algumas características das leis Poisson, negativa binomial, binomial, nova distribuição discreta, Poisson-Lindley e  $ZDGGD$ .

	Inflação	Deflação	Sobredispersão	Subdispersão
P	✓	✓	×	×
NB	✓	✓	✓	×
Bi	✓	✓	×	✓
ND	✓	×	✓	✓
PL	✓	✓	✓	×
$ZDGGD$	✓	✓	✓	✓

### 2.8.4 Dados de Actuariado

Uma questão relevante em ciências actuárias é a fixação do prémio de cada apólice. No processo de fixação do prémio de uma apólice é necessário obter uma boa modelação para o número de pedidos de indemnização por apólice e para a indemnização pedida. Ora, o número de pedidos de indemnização por apólice possui inflação em zero e sobre-dispersão como factos estilizados. Para ilustrar estes traços apresentam-se de seguida dois gráficos provenientes de dados retirados de [20] onde estão apresentados vários conjuntos de dados relativos a pedidos de indemnizações por apólice num período

fixo de tempo. A inflação em zero é clara nas Figuras 2.6 e 2.7 e a sobredispersão revela-se, embora de forma suave, na Tabela 2.10 onde surgem alguns resumos do conjunto 2. Consequentemente é

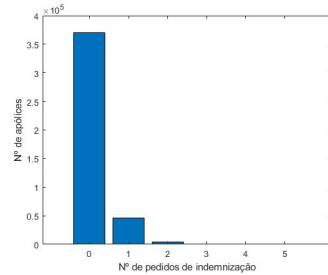


Fig. 2.6 Conjunto 2.

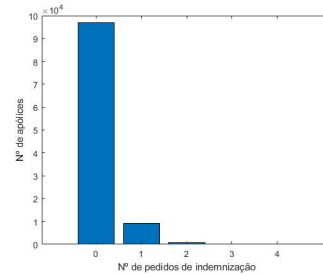


Fig. 2.7 Conjunto 5.

fundamental procurar modelos que possuam as características teóricas correspondentes aos factos estilizados. Ora como foi verificado anteriormente, a lei *ZDGGD* permite modelar a inflação em zero e a sobredispersão. Assim, começaremos por verificar se é adequado modelar o número de pedidos de indemnização por apólice referente ao conjunto 2 de [20] por meio desta família de leis, comparando o ajustamento obtido com outros modelos. Iremos usar o teste de adequação do qui-quadrado a uma

Tabela 2.10 Estatísticas relativas aos pedidos de indemnização por apólice (conjunto 2).

Média	Variância	Assimetria	Curtose	ID
0.1317	0.1385	2.9805	12.8563	1.0515

família de leis para testar a hipótese da lei do número de pedidos de indemnização por apólice,  $Q$ , pertencer à família das leis *ZDGGD*. Iremos repetir também o teste para as leis *ND* e *NB*, casos tratados nos artigos [8] e [20] respectivamente, e por fim, iremos também fazer o teste para a lei de Poisson pois estes dados verificam  $ID \simeq 1$ .

Tabela 2.11 Testes de adequação para o número de pedidos de indemnização por apólice.

N	O	Valores Esperados			
		ZDGGD	ND	NB	Poisson
0	370412	370388.3107	370688.4284	370 481.4204	369269.6423
1	46545	46575.0622	46311.123	46 416.1612	48631.6266
2	3935	3916.9627	3899.7695	4023.9579	3202.3926
3	317	329.4166	313.3176	297.0835	140.5850
4	28	27.7039	25.0740	20.0224	4.6288
$\geq 5$	3	2.5439	2.2875	1.3546	0.1247
Total	421 240	421 240	421 240	421 240	421 240
		$\hat{q} = 0.0841$	$\hat{\alpha} = -1.349$	$\hat{r} = 2.6047$	$\hat{\lambda} = 0.1317$
		$\hat{\alpha} = -0.1460$	$\hat{\theta} = 0.08$	$\hat{p} = 0.0481$	
$\chi^2$		0.5907	2.2326	8.0043	543.3113
$df$		2	2	2	2
p-valor		0.7443	0.3275	0.0183	0

Na Tabela 2.11 apresentam-se valores esperados sob a hipótese de a lei  $Q$  pertencer à família de leis  $ZDGGD$ ,  $ND$ ,  $NB$  ou Poisson.  $N$  representa as classes do número de pedidos de indemnização por apólice e  $O$  as frequências absolutas observadas de cada classe. Apresentam-se ainda as estimativas da máxima verosimilhança.  $\chi^2$  é a estatística de teste e  $df$  é o número de graus de liberdade da lei da estatística de teste. Ao nível de significância 0.01 somos conduzidos a aceitar como possíveis modelos, a  $ZDGGD$ ,  $ND$  e  $NB$ . Adicionalmente, usando como referência o valor  $\chi^2$  verificamos que obtemos um melhor ajuste aos dados quando consideramos a hipótese de  $Q$  seguir uma lei da família das  $ZDGGD$ . Curiosamente, embora os dados apresentem  $ID \simeq 1$  e inflação em zero, a lei de Poisson é a que apresenta pior ajustamento, sendo rejeitada para qualquer nível de significância.

### 2.8.5 Outros Dados: greves e lesões

Iremos agora abordar um conjunto de dados analisado em [13] referente ao número de surtos de greves na indústria mineira de carvão no Reino Unido em períodos de 4 semanas sucessivas, entre os anos 1948 e 1959. Na Tabela 2.12 apresentam-se algumas estatísticas referentes a estes dados. Observa-se que  $ID < 1$ , e portanto vamos tentar modelar os dados pelas leis  $ZDGGD(q, \alpha)$  e  $Bi(4, p)$ . Não vamos incluir  $ND$  nesta análise porque embora  $ID < 1$  ela é unimodal em zero o que não é compatível com a deflação em zero destes dados que se deduz da Tabela 2.13.

Tabela 2.12 Estatísticas relativas aos surtos de greve.

Média	Variância	Assimetria	Curtose	ID
0.9936	0.7419	0.8023	3.5038	0.7467

Tabela 2.13 Testes de adequação para o número de surtos de greve.

N	O	Valores Esperados	
		ZDGGD	Bi
0	46	46.0035	49.7819
1	76	78.0645	65.8107
2	24	22.6621	32.6252
3	9	6.5788	7.1883
$\geq 4$	1	2.6911	0.5939
Total	156	156	156
		$\hat{q} = 0.2903$	$n = 4$
		$\hat{\alpha} = -0.7175$	$\hat{p} = 0.2484$
$\chi^2$		0.1911	4.7772
$df$		1	2
p-valor		0.6620	0.0918

Verificamos, observando a Tabela 2.13, que o ajuste feito usando a lei  $ZDGGD$  é bom e bastante melhor que no caso da binomial. Em particular, a amostra observada não é compatível com a lei  $Bi(4, p)$  para níveis de significância superiores a 0.10. Observamos, que tal como a amostra, a lei  $ZDGGD(0.2903, -0.7175)$  usada na construção deste teste admite  $ID < 1$ .

Finalmente iremos modelar um conjunto de dados estudados em [17] referente ao número de lesões de determinado tipo em mamíferos. O objectivo principal é comparar os resultados que surgem usando a lei *ZDGGD* com os obtidos no artigo referido anteriormente usando a lei *PL*. Na Tabela 2.14 apresentam-se alguns resumos da amostra.

Tabela 2.14 Estatísticas relativas ao número de lesões.

Média	Variância	Assimetria	Curtose	ID
0.4742	0.7398	2.4158	10.6954	1.5600

Tabela 2.15 Testes de adequação para o número de lesões.

N	O	Valores Esperados			
		ZDGGD	PL	G	NB
0	413	412.9736	405.6757	407.6583	412.7421
1	124	124.0222	133.5740	131.1437	124.9650
2	42	42.2172	42.6203	42.1889	41.5113
3	15	14.3707	13.2949	13.5722	14.19246
4	5	4.8918	4.0769	3.3662	4.9246
5	0	1.6652	1.2336	1.4046	1.7228
$\geq 6$	2	0.8593	0.5246	1.6661	0.9378
Total	601	601	601	601	601
		$\hat{q} = 0.3404$	$\hat{\theta} = 2.6853$	$\hat{q} = 0.3217$	$\hat{r} = 0.8373$
		$\hat{\alpha} = 0.0783$			$\hat{p} = 0.3616$
$\chi^2$		0.0520	1.2787	0.6594	0.1040
$df$		2	3	3	2
p-valor		0.9743	0.7342	0.8827	0.9493

Ao calcular os estimadores da máxima verosimilhança sob a hipótese da lei dos dados em estudo pertencer à família de leis *ZDGGD* obtemos  $\hat{\alpha} = 0.0783$  <sup>(12)</sup>. Para verificar se este valor é significativo vamos fazer o teste  $H_0 : \alpha = 0$  vs  $H_1 : \alpha \neq 0$  usando  $Z = \frac{n^{1/2}(\hat{\alpha}_n - \alpha)}{\sqrt{I((q, \alpha))_{2,2}^{-1}}} \sim N(0, 1)$ . Obtemos assim  $Z_{obs} = 0.8670$  e logo  $|Z_{obs}| < 1.96$ , pelo que aceitamos ao nível de significância 0.05 a hipótese  $H_0 : \alpha = 0$ . Desta forma fomos conduzidos à modelação dos dados por uma lei Geométrica. Como podemos verificar, as quatro leis, *ZDGGD*, *PL*, Geométrica e *NB* apresentam bons resultados sendo a lei *ZDGGD* a que se ajustou melhor. De facto, apesar do teste apoiar a decisão de que  $\alpha = 0$ , o valor  $\alpha = 0.0783$  permitiu que a lei *ZDGGD* captasse as frequências observadas da classe 0 e 1 melhor do que a lei Geométrica.

<sup>(12)</sup> Observamos que ao tentar incluir neste estudo a lei *ND* na forma geral houve dificuldade em maximizar a função log verosimilhança. Como a lei Geométrica se ajusta bem aos dados, o que indica que o parâmetro  $\alpha$  da lei *ND* está perto de 0 e tendo em conta a expressão da função de probabilidade da lei *ND*, o facto de  $\alpha$  estar perto de 0 pode ter complicado o processo de maximização.



## Capítulo 3

# Modelos INARCH com lei condicional ZDGGD

Motivados pelo modelo INARCH proposto em [3] para um processo estocástico  $X = (X_t, t \in \mathbb{Z})$  de valores inteiros não negativos em que a lei de  $X_t$  condicional ao passado é de Poisson, vamos neste Capítulo abordar o caso em que tal lei condicional é a lei ZDGGD. Deste modo, na Secção 3.1, começaremos por apresentar o modelo ZDGGD – INARCH. Nas Secções 3.2 e 3.3, estudamos os modelos Geométricos INARCH e Geométricos inflacionados em zero INARCH, respectivamente, tendo como objectivo principal estabelecer condições de estacionaridade de primeira e segunda ordem. Na Secção 3.4 abordamos o método de estimação da máxima verosimilhança condicional, que iremos ilustrar com simulações na Secção 3.5. Por fim, na Secção 3.6, consideramos uma série de contagem observada e avaliamos qual o modelo, de entre os estudados, que conduz ao melhor ajustamento face a alguns critérios. Os códigos utilizados ao longo deste Capítulo foram feitos com recurso aos softwares Matlab e R.

### 3.1 ZDGGD-INARCH

**Definição 3.1.** Um processo  $X = (X_t, t \in \mathbb{Z})$  verifica um modelo Geométrico generalizado com perturbação em zero ARCH de valores inteiros de ordem  $p \in \mathbb{N}$ , ou de modo abreviado ZDGGD-INARCH( $p$ ), se para todo  $t \in \mathbb{Z}$

$$\begin{cases} X_t | \underline{X}_{t-1} \sim \text{ZDGGD}(q_t, \alpha_t), \\ E(X_t | \underline{X}_{t-1}) = \lambda_t = \frac{q_t^{\alpha_t+1}}{1 - q_t} = a_0 + \sum_{i=1}^p a_i X_{t-i} \end{cases} \quad (3.1)$$

de parâmetros  $q_t \in ]0, 1[$  e  $\alpha_t \in [-1, +\infty[$  com  $a_0 > 0$  e  $a_i \geq 0$  para  $i = 1, \dots, p$ . Para todo  $t \in \mathbb{Z}$ ,  $\underline{X}_{t-1}$  representa a  $\sigma$ -álgebra gerada por  $\{X_{t-k}, k \in \mathbb{N}\}$ .

O resultado seguinte será útil no estudo de estacionaridade de primeira ordem.

**Lema 3.1.** As raízes de  $p(z) = 1 - a_1 z - \dots - a_p z^p$ , onde  $a_i \geq 0$  para  $i = 1, \dots, p$ , estão no exterior do círculo unitário se e só se  $\sum_{i=1}^p a_i < 1$ .

*Prova.* ( $\Rightarrow$ ) Como  $p(0) = 1 > 0$  e  $p$  é uma função contínua então usando a hipótese vem  $p(1) > 0$  e conseqüentemente  $\sum_{i=1}^p a_i < 1$ . ( $\Leftarrow$ ) Suponhamos que existe uma raiz de  $p(z)$ ,  $z_*$ , tal que  $|z_*| \leq 1$ . Assim

$$1 = \sum_{i=1}^p a_i z_*^i = \left| \sum_{i=1}^p a_i z_*^i \right| \leq \sum_{i=1}^p a_i |z_*|^i \leq \sum_{i=1}^p a_i < 1,$$

que é um absurdo, logo as raízes estão no exterior do círculo unitário. ■

Assim, estamos em condições de estabelecer o resultado seguinte.

**Teorema 3.1.** *Um processo  $X = (X_t, t \in \mathbb{Z})$  que verifica o modelo ZDGGD-INARCH( $p$ ), dado em (3.1), é estacionário de primeira ordem se e só se*

$$\sum_{i=1}^p a_i < 1.$$

*Prova.* Para provarmos que o processo  $X$  é estacionário de primeira ordem temos de verificar que  $E(X_t) = \mu < \infty$  para todo o  $t \in \mathbb{Z}$ . Tendo em conta que estamos a lidar com funções mensuráveis positivas podemos escrever

$$\mu_t = E(X_t) = E(E(X_t | X_{-t-1})) = E\left(a_0 + \sum_{i=1}^p a_i X_{t-i}\right) = a_0 + \sum_{i=1}^p a_i \mu_{t-i}.$$

Obtemos a seguinte equação de recorrência não homogénea para  $\mu_t$

$$\mu_t = a_0 + \sum_{i=1}^p a_i \mu_{t-i}.$$

Ora, esta equação admite uma solução independente de  $t$  e finita se e só se as raízes do polinómio característico da equação de recorrência homogénea associada se encontram fora do círculo unitário, o que por sua vez acontece, usando o Lema 3.1, se e só se

$$\sum_{i=1}^p a_i < 1. \quad \blacksquare$$

**Nota.** *Deduzimos, desta prova, que um processo  $X = (X_t, t \in \mathbb{Z})$  que verifica o modelo ZDGGD-INARCH( $p$ ) e é estacionário de primeira ordem possui como função média*

$$E(X_t) = \mu = \frac{a_0}{1 - \sum_{i=1}^p a_i}, \quad \forall t \in \mathbb{Z}.$$

O estudo da estacionaridade de segunda ordem, que desenvolvemos a seguir, é apresentado apenas para a sub-família destes modelos em que  $\alpha_t = 0, \forall t \in \mathbb{Z}$ .

### 3.2 Modelo Geométrico INARCH(p)

Relembramos agora que se a v.a.r.  $Y \sim ZDGGD(q, \alpha)$  e se  $\alpha = 0$  então  $Y \sim G(q)$ . Deste modo se no modelo (3.1) considerarmos  $\alpha = 0$  e  $p_t = 1 - q_t$  para todo o  $t \in \mathbb{Z}$  e se definirmos o processo  $\lambda_t = \frac{1-p_t}{p_t}$  para todo o  $t \in \mathbb{Z}$  obtemos o modelo Geométrico INARCH, caso particular do modelo Binomial Negativo INGARCH, abreviadamente NB-INGARCH, introduzido em (Zhu,[21]).

**Definição 3.2.** Um processo  $X = (X_t, t \in \mathbb{Z})$  verifica um modelo Geométrico ARCH de valores inteiros de ordem  $p \in \mathbb{N}$ , ou de modo abreviado G-INARCH(p), se para todo  $t \in \mathbb{Z}$

$$\begin{cases} X_t | \underline{X}_{t-1} \sim G(1 - p_t) \\ E(X_t | \underline{X}_{t-1}) = \frac{1 - p_t}{p_t} = \lambda_t = a_0 + \sum_{i=1}^p a_i X_{t-i} \end{cases} \quad (3.2)$$

com  $a_0 > 0$  e  $a_i \geq 0$  para  $i = 1, \dots, p$ .

Podemos afirmar, tendo em conta o Teorema 3.1, que  $X$  é estacionário de primeira ordem se e só se  $\sum_{i=1}^p a_i < 1$  e que sob esta condição se tem  $E(X_t) = \mu = \frac{a_0}{1 - \sum_{i=1}^p a_i}$ .

Temos  $V(X | \underline{X}_{t-1}) > E(X_t | \underline{X}_{t-1})$  uma vez que  $V(X | \underline{X}_{t-1}) = \frac{1-p_t}{p_t} \frac{1}{p_t} = \lambda_t(\lambda_t + 1)$ . Podemos ainda observar que um processo  $X$  de segunda ordem que verifique tal modelo apresenta sobredispersão. De facto, considerando a variância de  $X_t$  não condicionada obtemos

$$V(X_t) = E\{V(X_t | \underline{X}_{t-1})\} + V\{E(X_t | \underline{X}_{t-1})\} = E\{\lambda_t(1 + \lambda_t)\} + V(\lambda_t) \quad (3.3)$$

e assim

$$V(X_t) = E(\lambda_t) + E(\lambda_t^2) - (E(\lambda_t))^2 + (E(\lambda_t))^2 + V(\lambda_t) = E(\lambda_t) + (E(\lambda_t))^2 + 2V(\lambda_t) > E(X_t),$$

o que permite concluir que  $\frac{V(X_t)}{E(X_t)} > 1$  para todo o  $t \in \mathbb{Z}$ .

Relativamente à estacionaridade de segunda ordem, começamos por mostrar o seguinte resultado, seguindo o trabalho desenvolvido em [18].

**Teorema 3.2.** Seja  $X = (X_t, t \in \mathbb{Z})$  um processo que verifica o modelo G-INARCH(p) estacionário de primeira ordem. O vector

$$W_t = \left[ E(X_t^2) \quad E(X_t X_{t-1}) \quad \dots \quad E(X_t X_{t-(p-1)}) \right]^T, \quad t \in \mathbb{Z},$$

verifica a seguinte equação de recorrência vectorial de ordem  $p$ ,  $W_t = B_0 + \sum_{k=1}^p B_k W_{t-k}$ , onde  $B_0$  é um vector de dimensão  $p$  e  $B_k$  para  $k = 1, \dots, p$  é uma matriz quadrada de ordem  $p$  <sup>(1)</sup>.

*Prova.* Começamos por notar que  $E(X_t^2)$  e  $E(X_t X_{t-h})$  podem não ser finitas mas como as funções são mensuráveis e positivas, podemos aplicar o operador esperança matemática. Temos, para todo o

<sup>(1)</sup>Deduzidas ao longo da prova.

$t \in \mathbb{Z}$ ,

$$\begin{aligned} E(X_t^2) &= E(E(X_t^2 | \underline{X}_{t-1})) = E[V(X_t | \underline{X}_{t-1}) + (E(X_t | \underline{X}_{t-1}))^2] = E[\lambda_t(1 + \lambda_t) + \lambda_t^2] \\ &= E[\lambda_t + 2\lambda_t^2] = E[\lambda_t] + 2E[\lambda_t^2]. \end{aligned}$$

Tendo em conta que  $X$  é estacionário de primeira ordem, obtemos  $E(\lambda_t) = E(X_t) = \mu$ . Podemos ainda escrever

$$\begin{aligned} E(\lambda_t^2) &= E\left[\left(a_0 + \sum_{i=1}^p a_i X_{t-i}\right)^2\right] = E\left[a_0^2 + 2a_0 \sum_{i=1}^p a_i X_{t-i} + \sum_{i=1}^p \sum_{j=1}^p a_i a_j X_{t-i} X_{t-j}\right] \\ &= E\left[a_0^2 + 2a_0(\lambda_t - a_0) + \sum_{i=1}^p a_i^2 X_{t-i}^2 + 2 \sum_{i=1}^p \sum_{j=i+1}^p a_i a_j X_{t-i} X_{t-j}\right] \\ &= a_0^2 + 2a_0(\mu - a_0) + \sum_{i=1}^p a_i^2 E(X_{t-i}^2) + 2 \sum_{i=1}^p \sum_{j=i+1}^p a_i a_j E(X_{t-i} X_{t-j}), \end{aligned}$$

logo

$$E(X_t^2) = \mu + 2a_0^2 + 4a_0(\mu - a_0) + 2 \sum_{i=1}^p a_i^2 E(X_{t-i}^2) + 4 \sum_{i=1}^p \sum_{j=i+1}^p a_i a_j E(X_{t-i} X_{t-j}), \quad (3.4)$$

com  $0 < \mu + 2a_0^2 + 4a_0(\mu - a_0) = \mu + 4a_0\mu - 2a_0^2$ . Para fazer aparecer a equação de recorrência vectorial desejada observamos que para  $h \geq 1$

$$\begin{aligned} E(X_t X_{t-h}) &= E(E(X_t X_{t-h} | \underline{X}_{t-1})) = E(X_{t-h} E(X_t | \underline{X}_{t-1})) = E(X_{t-h} \lambda_t) \\ &= E\left[X_{t-h} \left(a_0 + \sum_{i=1}^p a_i X_{t-i}\right)\right] = a_0 E(X_{t-h}) + \sum_{i=1}^p a_i E(X_{t-h} X_{t-i}) \\ &= a_0 \mu + \sum_{i=1}^{h-1} a_i E(X_{t-i} X_{t-h}) + a_h E(X_{t-h}^2) + \sum_{i=h+1}^p a_i E(X_{t-h} X_{t-i}). \end{aligned} \quad (3.5)$$

Assim, considerando (3.4) e (3.5) para  $h = 1, \dots, p-1$ , obtemos o sistema

$$\left\{ \begin{aligned} E(X_t^2) &= \mu + 4a_0\mu - 2a_0^2 + 2 \sum_{i=1}^p a_i^2 \underbrace{E(X_{t-i}^2)}_{W_{t-i}} + 4 \sum_{i=1}^p \sum_{j=i+1}^p a_i a_j \underbrace{E(X_{t-i} X_{t-j})}_{W_{t-i}} \\ E(X_t X_{t-1}) &= a_0 \mu + a_1 \underbrace{E(X_{t-1}^2)}_{W_{t-1}} + a_2 \underbrace{E(X_{t-1} X_{t-2})}_{W_{t-1}} + \dots + a_p \underbrace{E(X_{t-1} X_{t-p})}_{W_{t-1}} \\ \dots & \\ E(X_t X_{t-h}) &= a_0 \mu + \sum_{i=1}^{h-1} a_i \underbrace{E(X_{t-i} X_{t-h})}_{W_{t-i}} + a_h \underbrace{E(X_{t-h}^2)}_{W_{t-h}} + \sum_{i=h+1}^p a_i \underbrace{E(X_{t-h} X_{t-i})}_{W_{t-h}} \\ \dots & \\ E(X_t X_{t-(p-1)}) &= a_0 \mu + \sum_{i=1}^{p-2} a_i \underbrace{E(X_{t-i} X_{t-(p-1)})}_{W_{t-i}} + a_{p-1} \underbrace{E(X_{t-(p-1)}^2)}_{W_{t-(p-1)}} + a_p \underbrace{E(X_{t-(p-1)} X_{t-p})}_{W_{t-(p-1)}} \end{aligned} \right.$$

que permite concluir que o vector  $\left[ E(X_t^2) \quad E(X_t X_{t-1}) \quad \dots \quad E(X_t X_{t-(p-1)}) \right]^T$  verifica a equação  $W_t = B_0 + \sum_{k=1}^p B_k W_{t-k}$  com

$$B_0 = \left[ \mu + 4a_0\mu - 2a_0^2 \quad \alpha_0\mu \quad \dots \quad \alpha_0\mu \right]^T$$

e  $B_k = \left[ b_{i,j}^{(k)} \right]_{1 \leq i,j \leq p}$  para  $k = 1, \dots, p$  uma matriz quadrada de ordem  $p$  com elementos

$$b_{i,j}^{(k)} = \begin{cases} 2a_k^2 & \text{se } j = 1 \\ 4a_k a_{k-1+j} & \text{se } j = 2, \dots, p \end{cases} \quad b_{i,j}^{(k)} = \begin{cases} a_{k-1+j} & \text{se } i = k+1, j = 1, \dots, p \\ a_k & \text{se } i = k+j, j = 2, \dots, p \\ 0 & \text{caso contrário} \end{cases}$$

com  $a_j = 0, j > p$ . ■

Reencontramos deste modo resultados presentes em [18] e [21]. Enunciamos de seguida uma condição necessária e suficiente para a estacionaridade de segunda ordem de um processo estocástico que verifique o modelo  $G - INARCH(p)$  cuja prova se encontra em [18].

**Teorema 3.3.** *Seja  $X = (X_t, t \in \mathbb{Z})$  um processo que verifica o modelo  $G - INARCH(p)$  estacionário de primeira ordem. O processo é estacionário de segunda ordem se e só se*

$$P(L) = I_p - \sum_{k=1}^p B_k L^k \quad (3.6)$$

é uma matriz polinomial cujo determinante,  $\det P(z)$ , possui todas as raízes fora do círculo unitário.  $I_p$  é a matriz identidade de ordem  $p$ .

**Exemplo 3.1.** *Consideremos um processo estacionário de primeira ordem,  $X$ , seguindo um modelo  $G - INARCH(2)$  com  $a_1 + a_2 < 1$ ,  $a_1 \geq 0$  e  $a_2 > 0$ . Usando o Teorema 3.2 obtemos a seguinte equação vectorial de recorrência*

$$\begin{bmatrix} E(X_t^2) \\ E(X_t X_{t-1}) \end{bmatrix} = \begin{bmatrix} \mu + 4a_0\mu - 2a_0^2 \\ a_0\mu \end{bmatrix} + \begin{bmatrix} 2a_1^2 & 4a_1 a_2 \\ a_1 & a_2 \end{bmatrix} \begin{bmatrix} E(X_{t-1}^2) \\ E(X_{t-1} X_{t-2}) \end{bmatrix} + \begin{bmatrix} 2a_2^2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} E(X_{t-2}^2) \\ E(X_{t-2} X_{t-3}) \end{bmatrix}.$$

Assim, usando o Teorema 3.3 temos que  $X$  é um processo estacionário de segunda ordem se e só se

$$\begin{aligned} \det(P(z)) &= \det \left( \begin{bmatrix} 1 - 2a_1^2 z - 2a_2^2 z^2 & -4a_1 a_2 z \\ -a_1 z & 1 - a_2 z \end{bmatrix} \right) \\ &= 1 - 2a_1^2 z - 2a_2^2 z^2 - a_2 z - 2a_1^2 a_2 z^2 + 2a_2^3 z^3 \end{aligned}$$

possui todas as raízes fora do círculo unitário. Se  $a_1 = 0$ , por exemplo,  $a_2$  deve verificar  $a_2 < 1$  e  $\det P(z) = (a_2 z - 1)(\sqrt{2}a_2 z - 1)(\sqrt{2}a_2 z + 1)$  deve possuir as raízes fora do círculo unitário que por sua vez acontece se e só se  $a_2 < \frac{\sqrt{2}}{2}$ . Notamos que não podemos usar o Lema 3.1 pois um coeficiente é negativo. Assim, iremos apenas mostrar uma condição necessária de estacionaridade de segunda

ordem. Se  $\det P(z)$  possui todas as raízes fora do círculo unitário então

$$\begin{aligned} \det P(1) = 1 - 2a_1^2 - 2a_2^2 - a_2 - 2a_1^2a_2 + 2a_2^3 > 0 &\Leftrightarrow (1 - a_2) > 2a_1^2 + 2a_2^2 + 2a_1^2a_2 - 2a_2^3 \\ &\Leftrightarrow 1 > 2 \left( a_1^2 + \frac{2a_1^2a_2}{1 - a_2} \right) + 2a_2^2. \end{aligned} \quad (3.7)$$

Apresentamos agora um resultado que permite obter a função de autocovariância de um processo estacionário de segunda ordem  $X$  que verifique o modelo G-INARCH(p).

**Teorema 3.4.** *Seja  $X = (X_t, t \in \mathbb{Z})$  um processo que verifica o modelo G-INARCH(p) estacionário de segunda ordem. A função de autocovariância de  $X$ ,  $\gamma_X(k) = \text{Cov}(X_t, X_{t-k})$ ,  $k \geq 0$ , verifica as equações*

$$\gamma_X(0) = \mu + \mu^2 + 2 \sum_{i=1}^p a_i \gamma_X(i) \quad (3.8)$$

$$\gamma_X(k) = \sum_{i=1}^p a_i \gamma_X(|k - i|). \quad (3.9)$$

*Prova.* Para  $k > 0$  obtemos

$$\begin{aligned} \gamma_X(k) &= \text{Cov}(X_t, X_{t-k}) = E[(X_t - \mu)(X_{t-k} - \mu)] = E[E((X_t - \mu)(X_{t-k} - \mu) | \underline{X}_{t-1})] \\ &= E[(X_{t-k} - \mu)(E(X_t | \underline{X}_{t-1}) - \mu)] = \text{Cov}(X_{t-k}, \lambda_t) \\ &= \text{Cov}\left(X_{t-k}, a_0 + \sum_{i=1}^p a_i X_{t-i}\right) = \sum_{i=1}^p a_i \text{Cov}(X_{t-k}, X_{t-i}) = \sum_{i=1}^p a_i \gamma_X(|k - i|). \end{aligned}$$

Considerando agora o caso  $k = 0$ , obtemos de (3.3)

$$\gamma_X(0) = V(X_t) = \mu + \mu^2 + 2V(\lambda_t) = \mu + \mu^2 + 2 \sum_{i=1}^p a_i \gamma_X(i),$$

pois, utilizando o facto de que  $\gamma_X(k) = \sum_{i=1}^p a_i \gamma_X(|k - i|)$  temos

$$\begin{aligned} V(\lambda_t) &= V\left(a_0 + \sum_{i=1}^p a_i X_{t-i}\right) = \text{Cov}\left(\sum_{i=1}^p a_i X_{t-i}, \sum_{j=1}^p a_j X_{t-j}\right) = \sum_{i=1}^p \sum_{j=1}^p a_i a_j \gamma_X(|i - j|) \\ &= \sum_{i=1}^p a_i \sum_{j=1}^p a_j \gamma_X(|i - j|) = \sum_{i=1}^p a_i \gamma_X(i). \quad \blacksquare \end{aligned}$$

A partir de (3.9) deduzimos que a função de autocorrelação de  $X$ ,  $\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)}$ ,  $h \geq 1$ , verifica a equação de recorrência seguinte

$$\rho_X(k) = \sum_{i=1}^p a_i \rho_X(|k - i|), k \geq 1. \quad (3.10)$$

Para determinar  $\rho_X(h)$  para  $h = 1, \dots, p$  dados  $a_1, a_2, \dots, a_p$  resolvemos o seguinte sistema, denominado sistema de Yule-Walker,  $\rho = Ra$ , onde  $\rho = \begin{bmatrix} \rho_X(1) & \rho_X(2) & \dots & \rho_X(p) \end{bmatrix}^T$ ,

Fig. 3.1 Trajectória de um processo que verifica o modelo G-INARCH(1) com  $a_0 = 1$  e  $a_1 = 0.5$ .

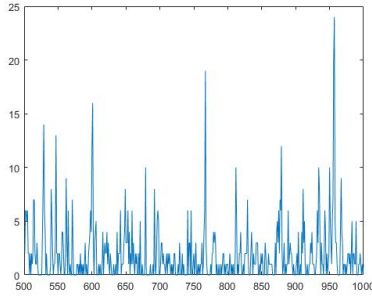


Tabela 3.1 Valores teóricos e empíricos de uma trajectória de um processo G-INARCH(1),  $a_0 = 1$  e  $a_1 = 0.5$ .

Valores	Teóricos	Empíricos
$E(X_t)$	2	2.103
$V(X_t)$	12	14.449
$a_0$	1	0.938
$a_1 = \rho(1)$	0.5	0.554

$R = [\rho_X(|j-i|)]_{i \leq i, j \leq p}$  e  $a = [a_1 \ a_2 \ \dots \ a_p]^T$ . Depois, para obter  $\rho_X(h)$  para  $h \geq p+1$  utilizamos (3.10). Podemos ainda obter  $\gamma_X(0)$  da seguinte forma a partir de (3.8)

$$\gamma_X(0) = \mu + \mu^2 + 2 \sum_{i=1}^p a_i \gamma_X(i) \Leftrightarrow \gamma_X(0) - 2 \sum_{i=1}^p a_i \gamma_X(i) = \mu + \mu^2 \Leftrightarrow \gamma_X(0) = \frac{\mu + \mu^2}{1 - 2 \sum_{i=1}^p a_i \rho_X(i)}.$$

Observamos que o sistema anterior serve também de base para um método de estimação, denominado estimação de Yule-Walker, de  $a_0, a_1, \dots, a_p$ . De facto, substituindo as autocorrelações pelas autocorrelações empíricas e resolvendo em seguida o sistema, obtemos estimativas para  $a_1, \dots, a_p$  que depois permitem obter uma estimativa para  $a_0$  já que  $E(X_t) = \frac{a_0}{1 - \sum_{i=1}^p a_i}$ .

**Exemplo 3.2.** Seja  $X = (X_t, t \in \mathbb{Z})$  um processo que verifica o modelo G-INARCH(1) estacionário de primeira ordem verificando  $a_1 < 1$ . Utilizando o Teorema 3.2 obtemos  $E(X_t^2) = \mu + 2a_0^2 + 4a_0(\mu - a_0) + 2\alpha_1^2 E(X_{t-1}^2)$  com  $\mu = \frac{a_0}{1-a_1}$ . Do Teorema 3.3 sabemos que  $X$  é estacionário de segunda ordem se e só se  $2a_1^2 < 1$ . Deste modo, consideremos  $2a_1^2 < 1$ . A função de autocorrelação de  $X$  é então, a partir de (3.10), dada por

$$\rho_X(k) = a_1^k, k \geq 1$$

e a função de autocovariância é

$$\gamma_X(k) = a_1^k \frac{\mu + \mu^2}{1 - 2a_1^2}, k \geq 0.$$

Na Figura 3.1 encontra-se uma trajectória de  $X$  verificando um modelo G-INARCH(1) com  $a_0 = 1$  e  $a_1 = 0.5$ . Na Tabela 3.1 encontram-se valores teóricos e empíricos, onde os valores empíricos de  $a_0$  e  $a_1$  foram obtidos a partir do método de estimação Yule-Walker acima referido. Como se pode observar a série apresenta sobredispersão e os valores empíricos estão relativamente próximos dos teóricos.

### 3.3 Modelo Geométrico inflacionado em zero INARCH(p)

A estacionaridade de segunda ordem foi apenas estudada, na secção anterior, para o caso particular dos modelos G-INARCH(p). Desta forma não conseguimos tirar proveito da flexibilidade que a lei

ZDGGD possui para a probabilidade de ocorrer zero. Para contornar este problema, começamos por ter em conta a definição seguinte.

**Definição 3.3.** A v.a.r.  $Y$ , com  $S_Y = \mathbb{N}_0$ , segue uma lei Geométrica inflacionada<sup>(2)</sup> em zero de parâmetros  $q \in ]0, 1[$  e  $w \in ]0, 1[$ ,  $ZIG(q, w)$ , se

$$P(Y = k) = w\delta_{0,k} + (1-w)q(1-q)^k, \quad k \in \mathbb{N}_0. \quad (3.11)$$

Notemos que  $E(Y) = (1-w)\frac{1-q}{q}$ . Como  $E(Y^2) = (1-w)\frac{(2-q)(1-q)}{q^2}$  obtemos ainda  $V(Y) = (1-w)\frac{1-q}{q} \left(1 + \frac{(1+w)(1-q)}{q}\right)$ .<sup>(3)</sup> Desta forma, podemos generalizar o modelo G-INARCH considerando o modelo apresentado de seguida que é caso particular do modelo binomial negativo inflacionado em zero INARCH (ZINB-INGARCH) introduzido em [22].

**Definição 3.4.** Um processo  $X = (X_t, t \in \mathbb{Z})$  verifica um modelo Geométrico inflacionado em zero ARCH de valores inteiros de ordem  $p \in \mathbb{N}$ , de modo abreviado  $ZIG-INARCH(p)$ , se para todo  $t \in \mathbb{Z}$

$$\begin{cases} X_t | \underline{X}_{t-1} \sim ZIG(1-p_t, w) \\ E(X_t | \underline{X}_{t-1}) = (1-w)\frac{1-p_t}{p_t} = (1-w)\lambda_t \\ \lambda_t = a_0 + \sum_{i=1}^p a_i X_{t-i} \end{cases} \quad (3.12)$$

com  $w \in ]0, 1[$ ,  $a_0 > 0$  e  $a_i \geq 0$  para  $i = 1, \dots, p$ . Se  $w = 0$  obtemos o modelo  $G-INARCH(p)$ .

Uma vez que  $V(X | \underline{X}_{t-1}) = (1-w)\lambda_t \{1 + (1+w)\lambda_t\}$  também se tem, neste modelo,  $V(X | \underline{X}_{t-1}) > E(X | \underline{X}_{t-1})$ . Além disso se  $X$  for um processo de segunda ordem verifica-se ainda  $V(X_t) > E(X_t)$  já que

$$\begin{aligned} V(X_t) &= E\{V(X_t | \underline{X}_{t-1})\} + V\{E(X_t | \underline{X}_{t-1})\} = (1-w)E(\lambda_t) + (1-w)(1+w)E(\lambda_t^2) + (1-w)^2V(\lambda_t) \\ &= (1-w)E(\lambda_t) + (1-w)(1+w)(E(\lambda_t))^2 + 2(1-w)V(\lambda_t). \end{aligned}$$

**Teorema 3.5.** Seja  $X = (X_t, t \in \mathbb{Z})$  um processo que verifica um modelo  $ZIG-INARCH(p)$ .  $X$  é estacionário de primeira ordem se e só se

$$\sum_{i=1}^p (1-w)a_i < 1.$$

*Prova.* A prova usa os mesmos argumentos do Teorema 3.1, mas a equação de recorrência obtida, para  $\mu_t = E(X_t)$ , é agora

$$\mu_t = E(E(X_t | \underline{X}_{t-1})) = E\left((1-w)a_0 + \sum_{i=1}^p (1-w)a_i X_{t-i}\right) = (1-w)a_0 + \sum_{i=1}^p (1-w)a_i \mu_{t-i}. \quad \blacksquare$$

<sup>(2)</sup> Notemos que aqui a noção de inflação em zero difere da apresentada no Capítulo 2.

<sup>(3)</sup>  $\delta_{0,k} = 1$  se  $k = 0$  e  $\delta_{0,k} = 0$  caso contrário.



Decorre do Teorema anterior, que se  $X = (X_t, t \in \mathbb{Z})$  é um processo que verifica o modelo ZIG-INARCH(p) e é estacionário de primeira ordem então possui média dada por

$$\mu = E(X_t) = \frac{(1-w)a_0}{1 - \sum_{i=1}^p (1-w)a_i}.$$

Tal como na secção anterior, começamos por mostrar um resultado que nos ajuda a obter uma condição necessária e suficiente para a estacionaridade de segunda ordem.

**Teorema 3.6.** *Seja  $X = (X_t, t \in \mathbb{Z})$  um processo que verifica o modelo ZIG- INARCH(p) estacionário de primeira ordem. O vector*

$$W_t = \left[ E(X_t^2) \quad E(X_t X_{t-1}) \quad \dots \quad E(X_t X_{t-(p-1)}) \right]^T, \quad t \in \mathbb{Z}$$

*verifica a seguinte equação de recorrência vectorial de ordem  $p$ ,  $W_t = B_0 + \sum_{k=1}^p B_k W_{t-k}$ , onde  $B_0$  é um vector de dimensão  $p$  e  $B_k$  para  $k = 1, \dots, p$  é uma matriz quadrada de ordem  $p$ .*

*Prova.* Iremos apenas indicar os elementos do vector  $B_0$  e das matrizes  $B_k$ ,  $k = 1, \dots, p$  já que a prova segue os mesmos passos usados na prova do Teorema 3.2. Tendo em conta que, para todo o  $t \in \mathbb{Z}$ ,

$$\begin{aligned} E(X_t^2) &= \mu + 2(1-w)\alpha_0^2 + 4\alpha_0(\mu - (1-w)\alpha_0) + 2 \sum_{i=1}^p (1-w)\alpha_i^2 E(X_{t-i}^2) + \\ &+ 4 \sum_{i=1}^p \sum_{j=i+1}^p (1-w)\alpha_i \alpha_j E(X_{t-i} X_{t-j}) \end{aligned}$$

e para  $h > 0$

$$\begin{aligned} E(X_t X_{t-h}) &= (1-w)\alpha_0 \mu + \sum_{i=1}^{h-1} (1-w)\alpha_i E(X_{t-i} X_{t-h}) + (1-w)\alpha_h E(X_{t-h}^2) + \\ &+ \sum_{i=h+1}^p (1-w)\alpha_i E(X_{t-h} X_{t-i}), \end{aligned}$$

temos  $B_0 = \left[ \mu + 2(1-w)a_0^2 + 4a_0(\mu - (1-w)a_0) \quad (1-w)a_0\mu \quad \dots \quad (1-w)a_0\mu \right]^T$  e

$B_k = \left[ b_{i,j}^{(k)} \right]_{1 \leq i,j \leq p}$ ,  $k = 1, \dots, p$ , uma matriz quadrada de ordem  $p$  com elementos

$$b_{1,j}^{(k)} = \begin{cases} 2(1-w)a_k^2, & j = 1 \\ 4(1-w)a_k a_{k-1+j}, & j = 2, \dots, p \end{cases} \quad b_{i,j}^{(k)} = \begin{cases} (1-w)a_{k-1+j}, & i = k+1, j = 1, \dots, p \\ (1-w)a_k, & i = k+j, j = 2, \dots, p \\ 0, & \text{caso contrário} \end{cases}$$

com  $a_j = 0$ , para  $j > p$ . ■

Desta forma obtemos resultados estabelecidos em [18] e [21]. No próximo Teorema, cuja prova se encontra em [18], podemos encontrar uma condição necessária e suficiente para a estacionaridade de segunda ordem.

**Teorema 3.7.** *Seja  $X = (X_t, t \in \mathbb{Z})$  um processo que verifica o modelo ZIG- INARCH( $p$ ) estacionário de primeira ordem. O processo é estacionário de segunda ordem se e só se*

$$P(L) = I_p - \sum_{k=1}^p B_k L^k \quad (3.13)$$

*é uma matriz polinomial cujo determinante,  $\det P(z)$ , possui todas as raízes fora do círculo unitário.*

Assim, se um processo for estacionário de segunda ordem e verificar o modelo ZIG-INARCH podemos encontrar no próximo Teorema como obter a sua função de autocovariância.

**Teorema 3.8.** *Seja  $X = (X_t, t \in \mathbb{Z})$  um processo que verifica o modelo ZIG- INARCH( $p$ ) estacionário de segunda ordem. A função de autocovariância de  $X$ ,  $\gamma_X(k) = \text{Cov}(X_t, X_{t-k})$ ,  $k \geq 0$ , verifica as equações*

$$\gamma_X(0) = \mu + \frac{1+w}{1-w} \mu^2 + 2 \sum_{i=1}^p a_i \gamma_X(i) \quad (3.14)$$

$$\gamma_X(k) = \sum_{i=1}^p (1-w) a_i \gamma_X(|k-i|) \quad (3.15)$$

*Prova.* Os resultados surgem de forma natural, adaptando a prova do Teorema 3.4. ■

**Exemplo 3.3.** *Seja  $X = (X_t, t \in \mathbb{Z})$  um processo que verifica o modelo ZIG- INARCH(1) estacionário de primeira ordem verificando  $(1-w)a_1 < 1$ . Utilizando o Teorema 3.6 obtemos  $E(X_t^2) = \mu + 2(1-w)a_0^2 + 4a_0(\mu - (1-w)a_0) + 2(1-w)a_1^2 E(X_{t-1}^2)$ . O Teorema 3.7 permite estabelecer que  $X$  é estacionário de segunda ordem se e só se  $2(1-w)a_1^2 < 1$  já que  $\det P(L) = 1 - 2(1-w)a_1^2 L$ . Sob esta condição a função de autocorrelação é*

$$\rho_X(k) = ((1-w)a_1)^k, k \geq 1$$

*e a função de autocovariância*

$$\gamma_X(k) = ((1-w)a_1)^k \frac{\mu + \frac{1+w}{1-w} \mu^2}{1 - 2(1-w)a_1^2}, k \geq 0,$$

onde  $\mu = \frac{(1-w)a_0}{1-(1-w)a_1}$ . Na Figura 3.2 encontra-se uma trajectória considerando  $a_0 = 1$ ,  $a_1 = 1$  e  $w = 0.4$  e na Tabela 3.2 encontram-se os valores teóricos e empíricos de  $E(X_t)$ ,  $V(X_t)$  e de  $\rho(1)$  correspondentes. Podemos observar na Tabela 3.2 que a série observada apresenta sobredispersão e valores empíricos próximos dos teóricos.

Fig. 3.2 Trajectória de um processo que verifica o modelo ZIG-INARCH(1) com  $a_0 = 1$ ,  $a_1 = 0.5$  e  $w = 0.4$ .

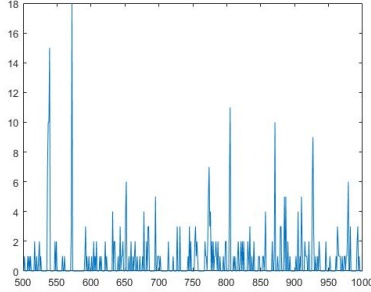


Tabela 3.2 Valores teóricos e empíricos de uma trajectória de um processo ZIG-INARCH(1),  $a_0 = 1$ ,  $a_1 = 0.5$  e  $w = 0.4$ .

Valores	Teóricos	Empíricos
$E(X_t)$	0.857	0.758
$V(X_t)$	3.673	3.254
$\rho(1)$	0.3	0.293

### 3.4 Modelo G-INARCH(1): Máxima verosimilhança condicional

Iremos agora abordar um método de estimação dos parâmetros  $\alpha = (a_0, a_1)$  do modelo  $G-INARCH(1)$ . Suponhamos que  $x^{(T)} = (x_1, \dots, x_T)$ ,  $n \geq 2$ , é uma trajectória de um processo  $(X_t, t \in \mathbb{Z})$  que segue um modelo  $G-INARCH(1)$ . Como a lei de  $X_t | \underline{X}_{t-1}$  é conhecida,  $P(X_t = k | \underline{X}_{t-1}) = p_t(1 - p_t)^k$ ,  $k \in \mathbb{N}_0$ , com  $p_t = \frac{1}{1 + \lambda_t}$  e  $\lambda_t = a_0 + a_1 X_{t-1}$ , podemos estimar os seus parâmetros usando o método de máxima verosimilhança condicional. A função de verosimilhança condicional é dada por

$$L(\alpha; x^{(T)}) = P(X_T = x_T, \dots, X_2 = x_2 | \underline{X}_1),$$

que ainda pode ser reescrita da seguinte forma

$$L(\alpha; x^{(T)}) = \prod_{t=2}^T P(X_t = x_t | X_{t-1} = x_{t-1}) = \prod_{t=2}^T \left( \frac{1}{1 + \lambda_t} \right) \left( \frac{\lambda_t}{1 + \lambda_t} \right)^{x_t}$$

já que  $(X_t, t \in \mathbb{Z})$  é uma cadeia de Markov de ordem 1. Assim a log-verosimilhança condicional é

$$\begin{aligned} \log L(\alpha; x^{(T)}) &= \sum_{t=2}^T l_t(\alpha; x^{(n)}) = \sum_{t=2}^T \log \left\{ \left( \frac{1}{1 + \lambda_t} \right) \left( \frac{\lambda_t}{1 + \lambda_t} \right)^{x_t} \right\} \\ &= \sum_{t=2}^T x_t \log(\lambda_t) - (1 + x_t) \log(1 + \lambda_t). \end{aligned} \quad (3.16)$$

As primeiras derivadas de  $l_t(\alpha; x^{(n)})$  são

$$\frac{\partial l_t}{\partial a_i} = \frac{x_t}{\lambda_t} \frac{\partial \lambda_t}{\partial a_i} - \frac{1 + x_t}{1 + \lambda_t} \frac{\partial \lambda_t}{\partial a_i} = \left[ \frac{x_t}{\lambda_t} - \frac{1 + x_t}{1 + \lambda_t} \right] \frac{\partial \lambda_t}{\partial a_i}, \quad i = 0, 1,$$

e as segundas derivadas

$$\frac{\partial^2 l_t}{\partial a_i \partial a_j} = \left[ -\frac{x_t}{\lambda_t^2} \frac{\partial \lambda_t}{\partial a_i} + \frac{1 + x_t}{(1 + \lambda_t)^2} \frac{\partial \lambda_t}{\partial a_i} \right] \frac{\partial \lambda_t}{\partial a_j} + \left[ \frac{x_t}{\lambda_t} - \frac{1 + x_t}{1 + \lambda_t} \right] \frac{\partial^2 \lambda_t}{\partial a_i \partial a_j}, \quad i, j = 0, 1,$$

com  $\frac{\partial \lambda_t}{\partial a_0} = 1$ ,  $\frac{\partial \lambda_t}{\partial a_1} = x_{t-1}$  e  $\frac{\partial^2 \lambda_t}{\partial a_i \partial a_j} = 0$  para  $i, j = 0, 1$ . Como não parece ser possível obter a forma explícita dos estimadores da máxima verosimilhança condicional revela-se necessário recorrer a métodos de optimização numérica para a estimação indicada.

### 3.5 Simulação

Foi feito um estudo de simulação com o objectivo de avaliar o método da máxima verosimilhança condicional para estimar os parâmetros  $a_0$  e  $a_1$  do modelo  $G - INARCH(1)$ . Gerámos uma trajectória de um processo estacionário de segunda ordem seguindo um modelo  $G - INARCH(1)$ ,  $2a_1^2 < 1$ , com tamanho  $n \in \{100, 500\}$ <sup>(4)</sup> e de seguida maximizámos a função log-verosimilhança condicional dada em (3.16). Este último procedimento foi repetido 1000 vezes para cada par  $(a_0, a_1) \in \{0.5, 1\} \times \{0.2, 0.5, 0.7\}$ . Na Tabela 3.3 encontram-se as médias e os desvios padrão das estimativas obtidas. Observamos que as médias das estimativas se aproximam dos verdadeiros valores de  $a_0$  e  $a_1$  com os respectivos desvios padrão a diminuírem quando  $n$  aumenta.

Com o objectivo de estudar a velocidade de convergência dos estimadores da máxima verosimilhança condicional para uma lei normal iremos, por fim, repetir o procedimento descrito anteriormente 2000 vezes mas considerando agora  $n = 1500$ , com  $a_0 = 0.5$  e  $a_1 = 0.2$ . Estimamos os parâmetros  $m$  e  $\sigma$  de uma lei normal através do método da máxima verosimilhança para modelar  $\sqrt{n}(\hat{a}_0 - a_0)$  e  $\sqrt{n}(\hat{a}_1 - a_1)$  e apresentamos os resultados obtidos na Figuras 3.3. Efectuamos ainda dois testes do qui quadrado para testar as hipóteses de  $\sqrt{n}(\hat{a}_0 - a_0)$  e  $\sqrt{n}(\hat{a}_1 - a_1)$  pertencerem à família das leis normais, tendo obtido  $\chi^2 = 3.9247$ ,  $df = 7$  e  $p - \text{valor} = 0.7884$  e  $\chi^2 = 0.8828$ ,  $df = 6$  e  $p - \text{valor} = 0.9897$  respectivamente, o que nos leva a aceitar a normalidade em ambos os casos. De facto, todos os resultados obtidos estão de acordo com a teoria abordada em [2], onde é apresentado a convergência quase certa dos estimadores da máxima verosimilhança condicional para os verdadeiros valores dos parâmetros e a normalidade assintótica desses mesmos estimadores para os parâmetros do modelo geométrico  $INGARCH(1, 1)$ .

Tabela 3.3 Resultados dos estimadores da máxima verosimilhança condicional para o modelo  $G - INARCH(1)$ .

		$n = 100$				$n = 500$			
$a_0$	$a_1$	$M(a_0)$	$DP(a_0)$	$M(a_1)$	$DP(a_1)$	$M(a_0)$	$DP(a_0)$	$M(a_1)$	$DP(a_1)$
0.5	0.2	0.5042	0.1078	0.1839	0.1272	0.5004	0.0475	0.1965	0.0570
	0.5	0.5110	0.1176	0.4661	0.1668	0.5009	0.0502	0.4944	0.0713
	0.7	0.5140	0.1243	0.6599	0.1758	0.5016	0.0527	0.6927	0.0785
1	0.2	1.0141	0.1891	0.1808	0.1226	1.0020	0.0853	0.1964	0.0557
	0.5	1.0286	0.2191	0.4671	0.1578	1.0041	0.0944	0.4939	0.0667
	0.7	1.0369	0.2392	0.6609	0.1667	1.0054	0.1020	0.6932	0.0723

<sup>(4)</sup>De facto as trajectórias geradas possuem tamanho 600 e 1000, mas retiramos os primeiros 500 valores.

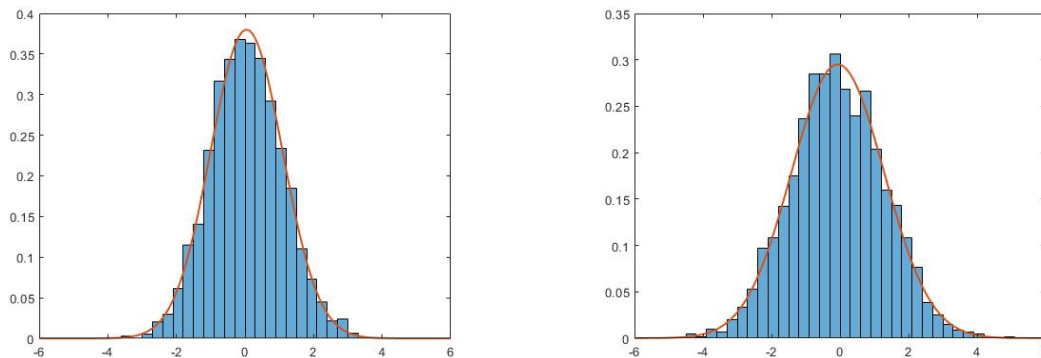


Fig. 3.3 Histograma de  $\sqrt{n}(\hat{a}_0 - a_0)$  (esquerda) e Histograma de  $\sqrt{n}(\hat{a}_1 - a_1)$  (direita).

### 3.6 Dados Reais

Nesta secção, iremos abordar a modelação de um conjunto de dados reais relativos ao número de novos casos de infecção pelo Hantavirus por semana no estado federal Eslévico-Holsácia da Alemanha entre 2005 e 2018 totalizando 742 observações. Estes dados foram retirados da base de dados do Instituto Robert-Koch (<https://survstat.rki.de>).

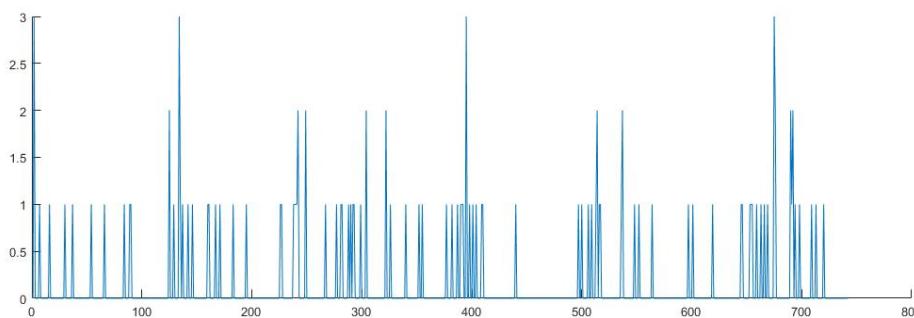
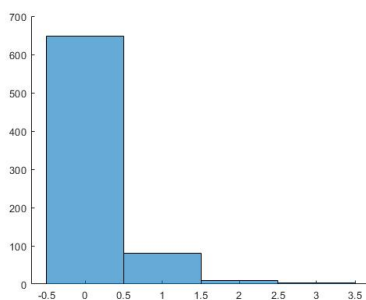


Fig. 3.4 Trajectória do número de novos casos de infecção por Hantavirus por semana.



Estatística	Valor
Média	0.1509
Variância	0.1877
Proporção de 0	0.8733

Tabela 3.4 Histograma do número de novos casos de infecção por Hantavirus por semana e algumas estatísticas.

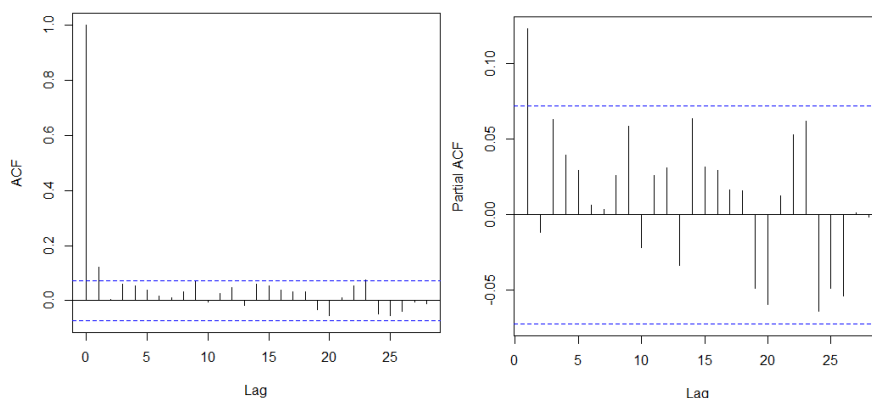


Fig. 3.5 Autocorrelações (esquerda) e autocorrelações parciais (direita) empíricas.

No processo de identificação do modelo que melhor se ajusta aos dados passámos por duas fases. A primeira consistiu em identificar, dentro dos modelos mencionados neste trabalho, quais os mais compatíveis com o valor das autocorrelações parciais da série observada. Na segunda fase, usámos os critérios log-verosimilhança ( $-Log$ ), informação de Akaike ( $AIC$ ) e informação Bayesiana ( $BIC$ ), a comparação entre valores de resumos teóricos dos modelos e da série observada e por fim a análise de resíduos para decidir qual o modelo considerado que se ajusta melhor à série em estudo. Como as autocorrelações parciais deixam de ser significativas a partir da segunda (Figura 3.5) optámos por usar os modelos  $G - INARCH(1)$  e  $INARCH(1)$ . Tendo em conta o número elevado de zeros que a série possui (Figura 3.4), constituindo 87.33% dos dados, decidimos ainda considerar os modelos  $INARCH(1)$  Geométrico e de Poisson inflacionados em zero ( $ZIG-INARCH$  e  $ZIP-INARCH$  respectivamente). Tendo em consideração a informação contida na Tabela 3.5, observamos que o modelo  $ZIG-INARCH(1)$  apresenta o melhor valor para o critério  $-Log$  enquanto que o modelo  $G-INARCH(1)$  apresenta os melhores valores para os critérios  $AIC$  e  $BIC$  seguido pelo  $ZIG - INARCH(1)$ .

Tabela 3.5 Estimativas dos parâmetros dos modelos e valores dos critérios  $-Log$ ,  $AIC$  e  $BIC$ .

Modelo	Estimativas			Critérios		
	$\alpha_0$	$\alpha_1$	$w$	$-Log$	$AIC$	$BIC$
$G-INARCH(1)$	0.1320	0.1269		327.1165	658.2330	667.4517
$INARCH(1)$	0.1321	0.1259		332.2810	668.5619	677.7806
$ZIG-INARCH(1)$	0.1508	0.1441	0.1238	326.9672	659.9343	673.7624
$ZIP-INARCH(1)$	0.2800	0.2480	0.5259	327.6763	661.3525	675.1806

Na Tabela 3.6 encontram-se valores de resumos teóricos associados a cada modelo e a valores empíricos relativos à amostra. Observamos que o modelo  $ZIG - INARCH(1)$  é o que capta melhor a sobredispersão da série apresentando média e variância muito próximas dos respectivos valores empíricos. Relativamente à autocorrelação de primeira ordem também o modelo  $ZIG - INARCH(1)$  possui um valor teórico bastante próximo do observado.

Tabela 3.6 Média, variância e autocorrelação de primeira ordem para cada modelo e da amostra. Média (Mr) e variância (Vr) dos resíduos de Pearson para cada modelo.

Modelo	G-INARCH(1)	INARCH(1)	ZIG-INARCH(1)	ZIP-INARCH(1)	Amostra
Média	0.1512	0.1511	0.1512	0.1505	0.1509
Variância	0.1799	0.1536	0.1873	0.1808	0.1877
$\rho(1)$	0.1269	0.1259	0.1262	0.1176	0.1227
Mr	0.0000	0.001	0.0001	0.0006	0
Vr	1.0858	1.2450	1.0482	1.0699	1

Por fim, tivemos em conta a análise dos resíduos de Pearson.

**Definição 3.5.** *Considere-se um modelo, dependente de um parâmetro  $\theta \in \mathbb{R}^n$ ,  $n \in \mathbb{N}$ , com média condicional  $E[X_t | \underline{X}_{t-1}; \theta]$  e variância condicional  $V[X_t | \underline{X}_{t-1}; \theta]$ . Se  $\hat{\theta}$  é uma estimativa de  $\theta$  então os resíduos de Pearson são dados por  $R_t := R_t(\hat{\theta}) = \frac{X_t - E[X_t | \underline{X}_{t-1}; \hat{\theta}]}{\sqrt{V[X_t | \underline{X}_{t-1}; \hat{\theta}]}}$ .*

Se o modelo correspondente se ajustar bem aos dados então estes resíduos deveriam comportar-se como ruídos brancos reduzidos. Na Tabela 3.6 encontram-se a média e a variância dos resíduos de Pearson (Mr e Vr) correspondentes a cada modelo. Observamos que o modelo *ZIG – INARCH(1)* apresenta o melhor resultado com média e variância mais próximas de zero e um. Apresentamos na Figura 3.6 as autocorrelações e autocorrelações parciais dos resíduos de Pearson correspondentes a este modelo. Observamos que não há autocorrelações nem autocorrelações parciais significativas, o que está de acordo com a hipótese do resíduo ser um ruído branco.

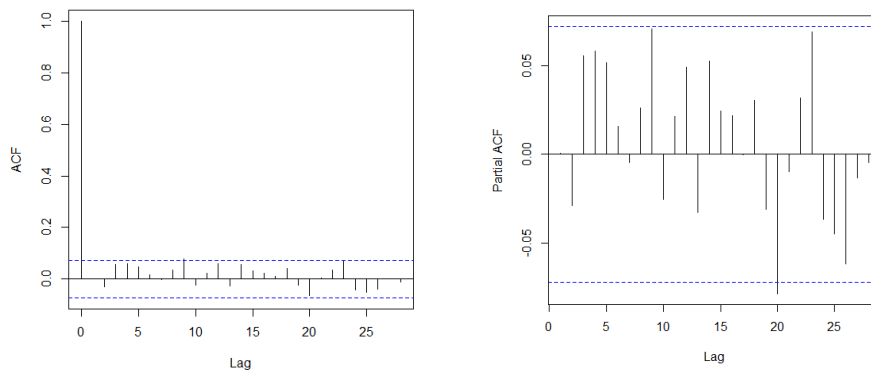


Fig. 3.6 Autocorrelações (esquerda) e autocorrelações parciais (direita) dos resíduos de Pearson correspondentes ao modelo *ZIG – INARCH(1)*.

Denotando a série relativa aos novos casos de infecção pelo Hantavirus por semana no estado federal Eslésvico-Holsácia da Alemanha por  $X$ , tendo em conta que o modelo *ZIG – INARCH(1)* apresentou em termos gerais os melhores resultados, um modelo para a série  $X$  é então

$$\begin{cases} X_t | \underline{X}_{t-1} \sim ZIG(1 - \hat{p}_t, 0.1238) \\ E(X_t | \underline{X}_{t-1}) = (1 - 0.1238)\hat{\lambda}_t \\ \hat{\lambda}_t = \frac{1 - \hat{p}_t}{\hat{p}_t} = 0.1508 + 0.1441X_{t-1}. \end{cases}$$





# Bibliografia

- [1] Chandra, N. K., Roy, D., and Ghosh, T. (2013). A generalized Poisson distribution. *Communications in Statistics - Theory and Methods*, 42(15):2786–2797.
- [2] Davis, R. A. and Liu, H. (2016). Theory and inference for a class of nonlinear models with application to time series of counts. *Statistica Sinica*, pages 1673–1707.
- [3] Ferland, R., Latour, A., and Oraichi, D. (2006). Integer-valued garch process. *Journal of Time Series Analysis*, 27(6):923–942.
- [4] Ghitany, M. E. and Al-Mutairi, D. K. (2009). Estimation methods for the discrete Poisson–Lindley distribution. *Journal of Statistical Computation and Simulation*, 79(1):1–9.
- [5] Gonçalves, E. and Mendes-Lopes, N. (2003). *Estatística - Teoria Matemática e Aplicações*. Escolar Editora, Lisboa.
- [6] Gonçalves, E. and Mendes-Lopes, N. (2008). *Séries temporais: modelações lineares e não lineares*. Sociedade Portuguesa de Estatística.
- [7] Gómez-Déniz, E. (2010). Another generalization of the geometric distribution. *TEST: An Official Journal of the Spanish Society of Statistics and Operations Research*, 19(2):399–415.
- [8] Gómez-Déniz, E., Sarabia, J., and Calderín-Ojeda, E. (2011). A new discrete distribution with actuarial applications. *Insurance: Mathematics and Economics*, 48:406–412.
- [9] Klugman, S., Panjer, H., and Willmot, G. (2004). *Loss models*. Wiley series in probability and statistics. Wiley Interscience, Hoboken, NJ, 2. ed edition.
- [10] Lehmann, E. (2004). *Elements of Large-Sample Theory*. Springer Texts in Statistics. Springer New York.
- [11] Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation*. Springer-Verlag, New York, NY, USA, second edition.
- [12] Resnick, S. (2014). *A probability path. Reprint of the 2005 edition*.
- [13] Ridout, M. S. and Besbeas, P. (2004). An empirical model for underdispersed count data. *Statistical Modelling*, 4(1):77–89.
- [14] Ross, S. M. (1990). *A Course in Simulation*. Prentice Hall PTR, USA.
- [15] Sastry, D. V. S., Bhati, D., Rattihalli, R. N., and Gómez-Déniz, E. (2016). On zero-distorted generalized geometric distribution. *Communications in Statistics - Theory and Methods*, 45(18):5427–5442.
- [16] Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley Series in Probability and Statistics - Applied Probability and Statistics Section Series. Wiley.

- 
- [17] Shanker, R. (2015). On Poisson-Lindley distribution and its applications to biological sciences. *Biometrics Biostatistics International Journal*, 2:103–107.
- [18] Silva, F. A. C. d. (2016). *Compound Poisson integer-valued GARCH processes*. PhD thesis, 00500:: Universidade de Coimbra.
- [19] Wilks, S. (1943). *Mathematical Statistics*. Mathematical Statistics. Princeton University Press.
- [20] Willmot, G. E. (1987). The Poisson-inverse gaussian distribution as an alternative to the negative binomial. *Scandinavian Actuarial Journal*, 1987(3-4):113–127.
- [21] Zhu, F. (2010). A negative binomial integer-valued garch model. *Journal of Time Series Analysis*, 32:54–67.
- [22] Zhu, F. (2012). Zero-inflated Poisson and negative binomial integer-valued garch models. *Journal of Statistical Planning and Inference*, 142(4):826–839.

## Anexo A

### Cálculo de Momentos

Esta secção tem como objectivo mostrar os cálculos que levam à obtenção dos momentos,  $E(X^r)$ ,  $r = 1, \dots, 4$ , e dos momentos centrados  $\mu'_r = E[(X - E(X))^r]$ ,  $r = 2, 3, 4$  de uma v.a.r.  $X \sim ZDGGD(q, \alpha)$ . Desta forma começamos por relembrar que para obter  $E(X^r)$  utilizamos  $\Phi_X^{(r)}(t)|_{t=0} = i^r E(X^r)$  e que  $\Phi_X(t) = 1 - q^{\alpha+1} + \frac{e^{it}(1-q)q^{\alpha+1}}{1-e^{it}q}$ ,  $t \in \mathbb{R}$ . Assim temos que

$$\begin{aligned}
 \Phi_X^{(1)}(t) &= \frac{ie^{it}(1-q)q^{\alpha+1}(1-e^{it}q) + e^{it}(1-q)q^{\alpha+1}ie^{it}q}{(1-e^{it}q)^2} \\
 &= i \frac{e^{it}(1-q)q^{\alpha+1}}{(1-e^{it}q)^2}, \\
 \Phi_X^{(2)}(t) &= i \frac{ie^{it}(1-q)q^{\alpha+1}(1-e^{it}q)^2 + e^{it}(1-q)q^{\alpha+1}2ie^{it}q(1-e^{it}q)}{(1-e^{it}q)^4} \\
 &= i^2 \frac{e^{it}(1-q)q^{\alpha+1}(1-e^{it}q + 2e^{it}q)}{(1-e^{it}q)^3} \\
 &= i^2 \frac{e^{it}(1-q)q^{\alpha+1}(1+e^{it}q)}{(1-e^{it}q)^3}, \\
 \Phi_X^{(3)}(t) &= i^2 \frac{[(1-q)q^{\alpha+1}ie^{it}(1+qe^{it}) + (1-q)q^{\alpha+1}e^{it}iqe^{it}](1-qe^{it})^3}{(1-e^{it}q)^6} + \\
 &\quad + \frac{(1-q)q^{\alpha+1}e^{it}(1+qe^{it})3iqe^{it}(1-qe^{it})^2}{(1-e^{it}q)^6} \\
 &= i^3 \frac{(1-qe^{it})^2(1-q)q^{\alpha+1}[(1-qe^{it})(e^{it} + 2qe^{2it}) + e^{it}(1+qe^{it})3qe^{it}]}{(1-e^{it}q)^6} \\
 &= i^3 \frac{(1-q)q^{\alpha+1}[(1-qe^{it})(e^{it} + 2qe^{2it}) + e^{it}(1+qe^{it})3qe^{it}]}{(1-e^{it}q)^4} \\
 &= i^3 \frac{(1-q)q^{\alpha+1}[e^{it} + 4qe^{2it} + q^2e^{3it}]}{(1-e^{it}q)^4},
 \end{aligned}$$

$$\begin{aligned}
\Phi_X^{(4)}(t) &= i^3(1-q)q^{\alpha+1} \frac{(ie^{it} + 8iqe^{2it} + 3iq^2e^{3it})(1-qe^{it})^4 + (e^{it} + 4qe^{2it} + q^2e^{3it})4i(1-qe^{it})^3qe^{it}}{(1-e^{it}q)^8} \\
&= i^4(1-q)q^{\alpha+1} \frac{(1-qe^{it})^3 [(e^{it} + 8qe^{2it} + 3q^2e^{3it})(1-qe^{it}) + (e^{it} + 4qe^{2it} + q^2e^{3it})4qe^{it}]}{(1-e^{it}q)^8} \\
&= i^4 \frac{(1-q)q^{\alpha+1} [e^{it} + 11qe^{2it} + 11q^2e^{3it} + q^3e^{4it}]}{(1-e^{it}q)^5}.
\end{aligned}$$

O que permite concluir

$$\begin{aligned}
\Phi_X^{(1)}(t)|_{t=0} &= i \frac{q^{\alpha+1}}{1-q} \Rightarrow E(X) = \frac{q^{\alpha+1}}{1-q}, \\
\Phi_X^{(2)}(t)|_{t=0} &= i^2 \frac{q^{\alpha+1}(1+q)}{(1-q)^2} \Rightarrow E(X^2) = \frac{q^{\alpha+1}(1+q)}{(1-q)^2}, \\
\Phi_X^{(3)}(t)|_{t=0} &= i^3 \frac{q^{\alpha+1}(1+4q+q^2)}{(1-q)^3} \Rightarrow E(X^3) = \frac{q^{\alpha+1}(1+4q+q^2)}{(1-q)^3}, \\
\Phi_X^{(4)}(t)|_{t=0} &= i^4 \frac{q^{\alpha+1}(1+11q+11q^2+q^3)}{(1-q)^4} \Rightarrow E(X^4) = \frac{q^{\alpha+1}(1+11q+11q^2+q^3)}{(1-q)^4}.
\end{aligned}$$

Por fim, para obter os momentos centrados

$$\begin{aligned}
\mu'_2 &= E(X^2) - (E(X))^2 = \frac{q^{\alpha+1}(1+q)}{(1-q)^2} - \left[ \frac{q^{\alpha+1}}{1-q} \right]^2 = \frac{q^{\alpha+1}(1+q-q^{\alpha+1})}{(1-q)^2}, \\
\mu'_3 &= E[X^3 - 3X^2\mu + 3X\mu^2 - \mu^3] = E(X^3) - 3E(X^2)\mu + 3E(X)\mu^2 - \mu^3 \\
&= \frac{q^{\alpha+1}(1+4q+q^2) - 3q^{\alpha+1}(1+q)q^{\alpha+1} + 2q^{3(\alpha+1)}}{(1-q)^3}, \\
\mu'_4 &= E[X^4 - 4X^3\mu + 6X^2\mu^2 - 4X\mu^3 + \mu^4] = E(X^4) - 4E(X^3)\mu + 6E(X^2)\mu^2 - 3\mu^4 \\
&= \frac{q^{\alpha+1}(1+11q+11q^2+q^3) - 4q^{2(\alpha+1)}(1+4q+q^2) + 6q^{3(\alpha+1)}(1+q) - 3q^{4(\alpha+1)}}{(1-q)^4}.
\end{aligned}$$

## Anexo B

# Códigos

Encontram-se nesta secção os códigos utilizados com recurso ao software Matlab e R.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%funcao para gerar n valores provenientes de uma
%ZDGGD(q,a)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [x] = zdggd(q,a,n)

x=rand(1,n);
  for i=1:n
    if(x(i)<=1 q^(a+1))
      x(i)=0;

    else
      x(i)=ceil(log(1-x(i))/log(q) (a+1));
    end
  end
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Regiao Assimetria negativa
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

x = 0.0001:0.0001:0.15;
y = 0.9999:0.0001: 0.8;
[x,y] = meshgrid(x,y);
cond1 = 1+4*x+x.^2+2*x.^(2*(y+1))<3*x.^(y+1).*(1+x);
cond1 = double(cond1);
cond1(cond1 == 0) = NaN;
cond = cond1;
surf(x,y,cond)
view(0,90)

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Regiao de deflacao em zero
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%x = 0.0001:0.001:1;
%y = 0.9999:0.001:0;
%[x,y] = meshgrid(x,y);
%cond1 = y<log(x.*(2-x))./log(x);
%cond1 = double(cond1);
%cond1(cond1 == 0) = NaN;
%cond = cond1;
```



---

```

function e = em(x)

m1=mean(x);
m2=mean(x.^2);

a=(log(2*m1^2)-log(m2-m1))/(log(m2-m1)-log(m1+m2));
q=(m2-m1)/(m1+m2);
e=[q a];

end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Estimador da maxima verosimilhanca
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function e = emv(x)

n=length(x);
m=sum(x==0);
s=sum(x(x>0));

a=log(1-m/n)/log(1-(n-m)/s)-1;
q=1-(n-m)/s;

e=[q a];

end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Estudo de simulacao-estimacao pontual
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
n=100;
q=0.4;
a=-0.3;

p=[];
m=[];
v=[];

for i=1:1000

    rng(i);
    %usar quando a=0.5
    %rng(1000+i);
    x=zdggd(q,a,n);

    p=[p transpose(ep(x))];
    m=[m transpose(em(x))];
    v=[v transpose(emv(x))];
end

%alfa q
n*cov(m(2,:),m(1,:))
n*cov(v(2,:),v(1,:))
fprintf('proporcoes \n')
[mean(p(2,:)) std(p(2,:)) mean(p(1,:)) std(p(1,:))]
fprintf('momentos \n')
[mean(m(2,:)) std(m(2,:)) mean(m(1,:)) std(m(1,:))]
fprintf('maxima verosimilhanca \n')
[mean(v(2,:)) std(v(2,:)) mean(v(1,:)) std(v(1,:))]

```

```

%Valores teóricos - tabelas 2.6, 2.7, 2.8
a = -0.7;
q = 0.4;
a_11 = q^(a+1) * (1+q-q^(a+1)) / (1-q)^2;
a_12 = q^(a+1) / (1-q)^3 * (1+4*q+q^2-q^(a+1)*(1+q));
a_21 = a_12;
a_22 = (q^(a+1) * (1+11*q+11*q^2+q^3) - (q^(a+1))^2 * (1+2*q+q^2)) / (1-q)^4;
A = [a_11 a_12;
     a_21 a_22];
m1 = q^(a+1) / (1-q);
m2 = q^(a+1) * (1+q) / (1-q)^2;
d_11 = -2*m2 / (m1+m2)^2;
d_12 = 2*m1 / (m1+m2)^2;
d_21 = 1 / (m2-m1) * (log(2*m1^2) - log(m1+m2)) + 2 / m1 * (log(m2-m1) - log(m1+m2)) + 1 / (m1+m2) * (log(2*m1^2) - log(m2-m1));
d_22 = -1 / (m2-m1) * (log(2*m1^2) - log(m1+m2)) + 1 / (m1+m2) * (log(2*m1^2) - log(m2-m1));
D = [d_11 d_12;
     d_21 d_22];
fprintf('momentos \n')
V = D * A * transpose(D)

i_11 = (a+1) * q^(a-1) * (a+q^(a+1)) / (1-q^(a+1)) +
q^(a+1) / (1-q)^2 + q^(a+1) / (q^2*(1-q)) + a * q^(a-1);
i_12 = q^a * (a+1) * log(q) / (1-q^(a+1));
i_22 = log(q)^2 * q^(a+1) / (1-q^(a+1));
fprintf('maxima verosimilhanca \n')
inv([i_11 i_12; i_12 i_22])
s_11 = (1-q^(a+1)) * q^(a+1);
s_12 = -(1-q^(a+1)) * (1-q) * q^(a+1);
s_21 = s_12;
s_22 = (1-q) * q^(a+1) * (1-(1-q) * q^(a+1));
S = [s_11 s_12;
     s_21 s_22];
p0 = 1-q^(a+1);
p1 = (1-q) * q^(a+1);
b_11 = -p1 / (1-p0)^2;
b_12 = -1 / (1-p0);
b_21 = ((p0+p1-1) * log(1-p0-p1) + (1-p0) * log(1-p0)) / ((p0-1) * (p0+p1-1) * (log(1-p0-p1) - log(1-p0))^2);
b_22 = log(1-p0) / ((1-p0-p1) * (log(1-p0-p1) - log(1-p0))^2);
B = [b_11 b_12;
     b_21 b_22];
fprintf('proporcoes \n')

```



```
C=B*S*transpose(B)
```

```
%Estudo de simulacao-probabilidade de cobertura
tamanho=[100 500 1000];

a_0=0.5;
q_0=0.8;

%quantil de ordem 0.95
sig=chi2inv(0.95,2);

for k=1:3

n=tamanho(k)
count_q_v=0;
count_q_m=0;
count_q_p=0;

for i=1:10000
rng(1000+i);
x=zdgdd(q_0,a_0,n);

v=emv(x);
q=v(1,1);
a=v(1,2);

i_11=n*(a_0+1)*q_0^(a_0-1)*(a_0+q_0^(a_0+1))/(1-q_0^(a_0+1))+
n*q_0^(a_0+1)/(1-q_0)^2+n*q_0^(a_0+1)/
(q_0^2*(1-q_0))+n*a_0*q_0^(a_0-1);
i_12=n*q_0^a_0*(a_0+1)*log(q_0)/(1-q_0^(a_0+1));
i_22=n*log(q_0)^2*q_0^(a_0+1)/(1-q_0^(a_0+1));
I=[i_11 i_12;
i_12 i_22];

z=[q-q_0 a-a_0]*I*transpose([q-q_0 a-a_0]);

if(z<sig)
count_q_v=count_q_v+1;
end

m=em(x);
q=m(1,1);
a=m(1,2);

a_11=q_0^(a_0+1)*(1+q_0-q_0^(a_0+1))/(1-q_0)^2;
a_12=q_0^(a_0+1)/(1-q_0)^3*(1+4*q_0+q_0^2-q_0^(a_0+1)*(1+q_0));
a_21=a_12;
a_22=(q_0^(a_0+1)*(1+11*q_0+11*q_0^2+q_0^3)-
(q_0^(a_0+1))^2*(1+2*q_0+q_0^2))/(1-q_0)^4;
A=[a_11 a_12;
a_21 a_22];

m1=q_0^(a_0+1)/(1-q_0);
m2=q_0^(a_0+1)*(1+q_0)/(1-q_0)^2;

d_11=-2*m2/(m1+m2)^2;
```

```

d_12=2*m1/(m1+m2)^2;

d_21=1/(m2*m1)*(log(2*m1^2)-log(m1+m2))+2/m1
*(log(m2*m1)-log(m1+m2))+1/(m1+m2)*(log(2*m1^2)-log(m2*m1));

d_21=d_21/(log(m2*m1)-log(m2+m1))^2;

d_22= 1/(m2*m1)*(log(2*m1^2)-log(m1+m2))
+1/(m1+m2)*(log(2*m1^2)-log(m2*m1));
d_22=d_22/(log(m2*m1)-log(m2+m1))^2;

D=[d_11 d_12;
    d_21 d_22];
V=D*A*transpose(D);
V=inv(V);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

z=n*[q q_0 a a_0]*V*transpose([q q_0 a a_0]);

if (z<sig)
    count_q_m=count_q_m+1;
end

s_11=(1-q_0^(a_0+1))*q_0^(a_0+1);
s_12=(1-q_0^(a_0+1))*(1-q_0)*q_0^(a_0+1);
s_21=s_12;
s_22=(1-q_0)*q_0^(a_0+1)*(1-(1-q_0)*q_0^(a_0+1));
S=[s_11 s_12;
    s_21 s_22];
p0=1-q_0^(a_0+1);
p1=(1-q_0)*q_0^(a_0+1);
b_11=p1/(1-p0)^2;
b_12=1/(1-p0);
b_21=((p0+p1-1)*log(1-p0*p1)+(1-p0)
*log(1-p0))/((p0-1)*(p0+p1-1)*(log(1-p0*p1)-log(1-p0))^2);
b_22=log(1-p0)/((1-p0*p1)*(log(1-p0*p1)-log(1-p0))^2);
B=[b_11 b_12;
    b_21 b_22];
C=B*S*transpose(B);
C=inv(C);
p=ep(x);
q=p(1,1);
a=p(1,2);
z=n*[q q_0 a a_0]*C*transpose([q q_0 a a_0]);
if (z<sig)
    count_q_p=count_q_p+1;
end

end

[ count_q_p/10000 count_q_m/10000 count_q_v/10000 ]

end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Estimadores/Dados reais
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
format short
d1=[370412 46545 3935 317 28 3];
d2=[413 124 42 15 5 0 2];
d3=[46 76 24 9 1];

```

```

dados=d2;

d=[];
for j=1:length(dados)

    d=[d ones(1,dados(j))*(j-1)];
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Emv
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
x=d;
n=length(x);
m=sum(x==0);
s=sum(x(x>0));

a=log(1-m/n)/log(1-(n-m)/s)-1
q=1-(n-m)/s

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%negativa binomial
%parmhat = nbinfit(d)

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%ND
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
dados=[370412 46545 3935 317 28 3];
d=[];
for j=1:length(dados)

    d=[d ones(1,dados(j))*(j-1)];
end

LL=@(x)(-sum(log((log(1-x(1)).*x(2).^d)
-log(1-x(1)).*x(2).^(d+1))./log(1-x(1)))));

format short
options = optimset('fmincon');
options = optimset(options, 'TolFun', 1e-006);
options = optimset(options, 'LargeScale', 'off');
options = optimset(options, 'MaxFunEvals', 1000);
options = optimset(options, 'MaxIter', 800);
options = optimset('maxfunevals',20000);
options = optimoptions('fmincon','display','none');

theta0 = [-1;0.5];
lb = [-5;-5];
ub = [5;1];

[thetahat] = fmincon(LL,theta0,[],[],[],[],lb,ub,[],options)

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Significancia alfa
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
a=0;
q=0.3404;
n=601;
format short

I=[(a+1)*q^(a-1)*(a+q^(a+1))/(1-q^(a+1))+

```

```

q^(a+1)/(1-q)^2+q^(a+1)/(q^2*(1-q))+
a*q^(a-1) q^a*(a+1)*log(q)/(1-q^(a+1));
    q^a*(a+1)*log(q)/(1-q^(a+1)) log(q)^2*q^(a+1)/(1-q^(a+1) ]);
i=inv(I);

z=n^(1/2)*0.0783/(i(2,2))^(1/2)

pd = makedist('Normal ');
    2*(1-cdf(pd, z))

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Trajectory G-INARCH(1)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
a_0=1;
a_1=0.5;

t=[a_0/(1-a_0-a_1)];
rng(1);
for i=2:1000

    lambda=a_0+a_1*t(i-1);
    t=[t nbinrnd(1,1/(1+lambda))];
end

plot([501:1000],t(501:1000))

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Trajectory ZIG-INARCH
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
a_0=1;
a_1=0.5;
w=0.4;

t=[(1-w)*a_0/(1-(1-w)*a_1)];

for i=2:1000
    rng(7000+i);
    r=rand;
    if(r<(1-w))
        lambda=a_0+a_1*t(i-1);
        t=[t nbinrnd(1,1/(1+lambda))];
    else
        t=[t 0];
    end
end

plot([501:1000],t(501:1000));
aux=t(501:1000);
m1=mean(aux)
m=(1-w)*a_0/(1-(1-w)*a_1)
var(aux)
(m+(1+w)/(1-w)*m^2)/(1-2*(1-w)*a_1^2)

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Estudo Simulacao G-INARCH(1)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [a_0 a_1] =CLS(d,n)

s1=d(2:n);

```

```

s2=d(1:n-1);

a_1 = (sum(s1*s2) 1/(n-1)*sum(s1
)*sum(s2))/(sum(s2.^2) 1/(n-1)*(sum(s2))^2);
a_0 = 1/(n-1)*(sum(s1) a_1*sum(s2));
end

a_0=0.5;
a_1=0.2;

cls=[];
n=1500;
N=500+n;

for j=1:2000

rng(j);
s=[a_0/(1-a_1)];
l=[a_0/(1-a_1)];

for i=2:N

lambda=a_0+a_1*s(i-1);
l=[l lambda];
s=[s nbinrnd(1,1/(1+lambda))];

end

dados=s(501:N);
[a,b]=CLS(dados,n);
dados1=s(502:N);
dados2=s(501:N-1);

LL=@(x)( sum(log((1./(1+x(1)+x(2)*dados2))
.*(1-1./(1+x(1)+x(2)*dados2)).^dados1))) ;

options = optimset('fmincon');
options = optimset(options, 'TolFun', 1e-006);
options = optimset(options, 'LargeScale', 'off');
options = optimset(options, 'MaxFunEvals', 1000);
options = optimset(options, 'MaxIter', 800);
options = optimset('maxfunvals',20000);
options = optimoptions('fmincon','display','none');

theta0 = [a;b];
lb = [0;0];
ub = [4;1];
[thetahat] = fmincon(LL,theta0,[],[],[],[],lb,ub,[],options);
cls=[cls thetahat];

end

mu=mean(sqrt(n)*(cls(1,:) a_0));
sigma=sqrt(var(sqrt(n)*(cls(1,:) a_0)));
figure(1)
histogram(sqrt(n)*(cls(1,:) a_0),'Normalization','pdf')
hold on
y = -6:0.1:6;
f = exp(-(y-mu).^2./(2*sigma^2))./(sigma*sqrt(2*pi));
plot(y,f,'LineWidth',1.5)

```

```

figure(2)
ecdf(sqrt(n)*(cls(1,:)-a_0))
hold on
x = -6:0.1:6;
y=normcdf(x,mu,sigma);
plot(x,y)
hold off
legend('Impirica ','Teorica ')

mu=mean(sqrt(n)*(cls(2,:)-a_1));
sigma=sqrt(var(sqrt(n)*(cls(2,:)-a_1)));

figure(3)
histogram(sqrt(n)*(cls(2,:)-a_1),'Normalization ','pdf')
hold on
y = -6:0.1:6;
f = exp(-(y-mu).^2./(2*sigma^2))./(sigma*sqrt(2*pi));
plot(y,f,'LineWidth ',1.5)

figure(4)
ecdf(sqrt(n)*(cls(2,:)-a_1))
hold on
x = -6:0.1:6;
y=normcdf(x,mu,sigma);
plot(x,y)
hold off
legend('Impirica ','Teorica ')

[h,p,st]=chi2gof(sqrt(n)*(cls(1,:)-a_0))
[h,p,st]=chi2gof(sqrt(n)*(cls(2,:)-a_1))

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Funcao para estimar G-INARCH(1)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [a_0,a_1] = G(dados)
N=length(dados);

dados=dados(1:N);
[a,b]=CLS(dados,N);

dados1=dados(2:N);
dados2=dados(1:N-1);

LL=@(x)(-sum(log((1./(1+x(1)+x(2)*dados2))
.*(1-1./(1+x(1)+x(2)*dados2)).^dados1))) ;

options = optimset('fmincon');
options = optimset(options,'TolFun',1e-006);
options = optimset(options,'LargeScale','off');
options = optimset(options,'MaxFunEvals',1000);
options = optimset(options,'MaxIter',800);
options = optimset('maxfunevals',20000);
options = optimoptions('fmincon','display','none');

theta0 = [a;b];
lb = [0;0];
ub = [10;1];
[thetahat] = fmincon(LL,theta0,[],[],[],[],lb,ub,[],options);
a_0=thetahat(1,1);

```

```

a_1=thetahat(2,1);
l=LL([a_0 a_1])

AIC=2*l+4
BIC=2*l+2*log(N)

m=a_0/(1-a_1)
r_1=a_1
v=(m+m^2)/(1-2*r_1*a_1)
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Funcao para estimar INARCH(1)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [a_0,a_1] = P(dados)
N=length(dados);

dados=dados(1:N);
[a,b]=CLS(dados,N);

dados1=dados(2:N);
dados2=dados(1:N-1);
dados3=factorial(dados1);
LL=@(x)(-sum(log((x(1)+x(2)).*dados2).^dados1
.*exp(1).^(-x(1)-x(2)).*dados2)./dados3));

options = optimset('fmincon');
options = optimset(options, 'TolFun', 1e-006);
options = optimset(options, 'LargeScale', 'off');
options = optimset(options, 'MaxFunEvals', 1000);
options = optimset(options, 'MaxIter', 800);
options = optimset('maxfunvals',20000);
options = optimoptions('fmincon','display','none');

theta0 = [a;b];
lb = [0;0];
ub = [10;1];
[thetahat] = fmincon(LL,theta0,[],[],[],[],lb,ub,[],options);
a_0=thetahat(1,1);
a_1=thetahat(2,1);
l=LL([a_0 a_1])
AIC=2*l+4
BIC=2*l+2*log(N)

m=a_0/(1-a_1)
r_1=a_1
v=a_0*(1-a_1)/((1-a_1^2)*(1-a_1)^2)
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Funcao para estimar ZIG-INARCH(1)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [a_0,a_1,w] = ZIG(x)
N=length(x);
dados1=x(2:N);
dados2=x(1:N-1);
dados3=[];

for i=1:length(dados1)

```

```

    dados3=[dados3 (dados1(i)==0)];

end

LL=@(x)( sum(log((x(3)+(1-x(3))*1./(1+x(1)+x(2)*dados2))
.^dados3*((1-x(3)).*(1./(1+x(1)+x(2)*dados2))
.*(1-1./(1+x(1)+x(2)*dados2)).^dados1).^1 dados3)))));

options = optimset('fmincon');
options = optimset(options, 'TolFun', 1e-006);
options = optimset(options, 'LargeScale', 'off');
options = optimset(options, 'MaxFunEvals', 1000);
options = optimset(options, 'MaxIter', 800);
options = optimset('maxfunvals',20000);
options = optimoptions('fmincon','display','none');

theta0 = [0.5;0.5;0.5];
lb = [0;0;0];
ub = [10;1;1];
[thetahat] = fmincon(LL,theta0,[],[],[],[],lb,ub,[],options);
a_0=thetahat(1,1);
a_1=thetahat(2,1);
w=thetahat(3,1);

l=LL([a_0 a_1 w])
AIC=2*l+6
BIC=2*l+3*log(N)

m=(1-w)*a_0/(1-(1-w)*a_1)
r_1=(1-w)*a_1
v=(m+(1+w)/(1-w)*m^2)/(1-2*r_1*a_1)

end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Funcao para estimar ZIP INARCH(1)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function [a_0,a_1,w] = ZIP(x)
N=length(x);
dados1=x(2:N);
dados2=x(1:N-1);
dados3=[];
dados4=factorial(dados1);

for i=1:length(dados1)
    dados3=[dados3 (dados1(i)==0)];
end

LL=@(x)( sum(log((x(3)+(1-x(3))*exp(1).^(x(1)-x(2)).*dados2))
.^dados3*((1-x(3)).*(x(1)+x(2)).*dados2).^dados1
.*exp(1).^(x(1)-x(2)).*dados2)./dados4).^1 dados3)))));

options = optimset('fmincon');
options = optimset(options, 'TolFun', 1e-006);
options = optimset(options, 'LargeScale', 'off');
options = optimset(options, 'MaxFunEvals', 1000);
options = optimset(options, 'MaxIter', 800);
options = optimset('maxfunvals',20000);

```



```

options = optimoptions('fmincon','display','none');

theta0 = [0.5;0.5;0.5];
lb = [0;0;0];
ub = [10;1;1];
[thetahat] = fmincon(LL,theta0,[],[],[],[],lb,ub,[],options);
a_0=thetahat(1,1);
a_1=thetahat(2,1);
w=thetahat(3,1);
l=LL([a_0 a_1 w])
AIC=2*l+6
BIC=2*l+3*log(N)

m=(1-w)*a_0/(1-(1-w)*a_1)
r_1=(1-w)*a_1
v=(1-w)*a_0*(1+w*a_0-(1-w)*a_1)/((1-(1-w)*a_1^2)*(1-(1-w)*a_1)^2)
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Leitura_dados
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
T = readtable('C:/Users/User/Desktop/hant.csv');
A = table2array(T);
A=transpose(A);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Estatistica
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
length(A)
mean(A)
var(A)
sum(A==0)/length(A)
var(A)/mean(A)

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Estimacao
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
[a_0,a_1]=G(A);
%[a_0,a_1]=P(A);
%[a_0,a_1,w]=ZIG(A);
%[a_0,a_1,w]=ZIP(A);

a_0
a_1
%w

%r-residuos de person
r=[];
e=[];
for i=2:length(A)

    lambda=a_0+a_1*A(i-1);

    r=[r (A(i)-(1-w)*lambda)/sqrt((1-w)*lambda*(1+w*lambda+lambda))];

%para modelos com lei poisson
%r=[r (A(i)-(1-w)*lambda)/sqrt((1-w)*lambda*(1+w*lambda))];

    e=[e lambda*(1-w)];
end

```

```

mr=mean(r)
vr=var(r)
r=transpose(r);

figure(1)
hold on
plot([4:length(A)],A(4:length(A)))
plot([1:length(e)],e,'r')
hold off

figure(2)
plot([1:length(r)],r,'r')

%cria ficheiro com residuos de pearson
xlswrite('residuos.xls',r)

figure(3)
hold on
histogram(A)
hold off

figure(4)
hold on
plot([1:length(A)],A(1:length(A)))
hold off

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Codigos em software R para obter ACF E PACF da serie e residuos
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
Dados1<-read.csv(file='C:/Users/User/Desktop/hant.csv',header = TRUE, sep = ";", dec = ",")
p<-Dados1[,1]
acf(p)
pacf(p)
pacf(p,1, pl = FALSE)

residuos<-read.csv(file='C:/Users/User/Desktop/residuos.csv',header = FALSE, sep = ";", dec = ",")
p<-residuos[,1]
acf(p)
pacf(p)

```