Modelos para Séries Temporais com Base na EGEDD

Distribuição Dagum Exponencial Generalizada Exponencializada

Apresentador: Vítor Bernardo Silveira Pereira
Orientador: Cleber Bisognin
Coorientadora: Laís Helen Loose

Motivação

- A identificação de uma **distribuição** apropropriada é muito importante para a **análise estatística**;
- A distribuição Dagum, idealizada por Camilo Dagum, tem ganhado muita atenção na sua utilidade para modelagem de dados sobre a distribuição de renda e análise de confiabilidade.
- Distribuições baseadas na Dagum:
    - Distribuição exponenciada **Kumaraswamy-Dagum** (Silva et al., 2015);
    - Distribuição **Dagum Exponencial Generalizada Exponencializada** proposta em (Nasiru et al., 2019);
    - **Gompertz Generalizada**, proposta por El-Gohary et al. (2013), entre outros...

Motivação

- **Aplicações:** Principalmente na área da modelagem de dados de sobrevivência e em dados econômicos, como análise de confiabilidade, modelagem de renda pessoal e nacionais (micro e macroeconômico).

Augusto Dagum: Um dos maiores economistas e estatísticos latinos

Motivação

- Os modelos de **Regressão Linear Normal** e **Modelos Lineares Generalizados** possuem como pressuposto que a distribuição seja Normal ou pertença a família exponencial regular;

- Na prática, existem dados que não se adequam a distribuição normal ou distribuições pertecentes a familia exponencial, então nos últimos anos vêm sendo propostos **modelos alternativos**;

Motivação

- Os modelos de **Regressão Linear Normal** e **Modelos Lineares Generalizados** possuem como pressuposto que a distribuição seja Normal ou pertença a família exponencial regular;
- Na prática, existem dados que não se adequam a distribuição normal ou distribuições pertecentes a familia exponencial, então nos últimos anos vêm sendo propostos **modelos alternativos**;
- Nos novos modelos, tem-se uma reparametrização em termos da média ou quantil, para possibilitara interpretação dos parâmetros em termos de medidas de posição ou precisão;
- Na **EGEDD**, no entanto, **não há forma fechada para média e variância**, desta forma reparametrizamos a distribuição em termos dos quantis;
- Com a reparametrização, uma estrutura de regressão é incluída para **modelar a média ou o quantil**, assim tendo a mesma abordagem dos MLG'S;

Motivação

- Em Séries Temporais, uma das abordagens mais usuais é a metodologia Box e Jenkins, no entanto também estão surgindo modelos alternativos, denominados de **modelos dinâmicos**;
- Assim a proposta do projeto, é **desenvolver um novo modelo (modelo dinâmico) com suporte nos reais**, baseando-se na EGEDD e na distrubuição Dagum;

Propriedades da EGEDD

Função de distribuição acumulada

$$F_Y(y) = 1 - \left\{1 - \left[1 - \left(1 - \left(1 + \alpha y^{-\delta}\right)^{-\sigma}\right)^{\gamma}\right]^{\eta}\right\}^{\lambda}, \quad y > 0,$$

Propriedades da EGEDD

Função de distribuição acumulada

$$F_Y(y) = 1 - \left\{1 - \left[1 - \left(1 - \left(1 + \alpha y^{-\delta}\right)^{-\sigma}\right)^{\gamma}\right]^{\eta}\right\}^{\lambda}, \quad y > 0,$$

onde os paramêtros são

$$\alpha, \sigma, \lambda, \delta, \eta, \gamma > 0$$

Os parâmetros $\sigma, \lambda, \delta, \eta, \gamma$ são parâmetros de forma e α é parâmetro de escala.

Função densidade de probabilidade

$$f_Y(y) = \frac{\alpha\sigma\lambda\delta\eta\gamma\left(1+\alpha y^{-\delta}\right)^{-\sigma-1}\left[1-\left(1+\alpha y^{-\delta}\right)^{-\sigma}\right]^{\gamma-1}}{y^{\delta+1}\left\{1-\left[1-\left(1-\left(1+\alpha y^{-\delta}\right)^{-\sigma}\right)^{\gamma}\right]^{\eta}\right\}^{1-\lambda}}$$

$$*[1-(1-(1+\alpha y^{-\delta})^{-\sigma})^{\gamma}]^{\eta-1}$$

Função Quantílica

$$Q_Y(\tau) = \left\{\frac{1}{\alpha}\left[\left(1-\left(1-\left(1-(1-\tau)^{\frac{1}{\lambda}}\right)^{\frac{1}{\eta}}\right)^{\frac{1}{\gamma}}\right)^{-\frac{1}{\sigma}} - 1\right]\right\}^{-\frac{1}{\delta}}, \ \text{onde} \quad \tau \in (0,1).$$

Reparametrizações

$$\mu = \left\{ \frac{1}{\alpha} \left[\left(1 - \left(1 - \left(1 - (1-\tau)^{\frac{1}{\lambda}} \right)^{\frac{1}{\eta}} \right)^{\frac{1}{\gamma}} \right)^{-\frac{1}{\sigma}} - 1 \right] \right\}^{-\frac{1}{\delta}}$$

Reparametrizações

$$\mu = \left\{ \frac{1}{\alpha} \left[\left(1 - \left(1 - \left(1 - (1 - \tau)^{\frac{1}{\lambda}} \right)^{\frac{1}{\eta}} \right)^{\frac{1}{\gamma}} \right)^{-\frac{1}{\sigma}} - 1 \right] \right\}^{-\frac{1}{\delta}}$$

Reparametrização por δ

$$\delta = -\frac{1}{\log(\mu)} \log \left\{ \frac{1}{\alpha} \left[\left(1 - \left(1 - \left(1 - (1 - \tau)^{\frac{1}{\lambda}} \right)^{\frac{1}{\eta}} \right)^{\frac{1}{\gamma}} \right)^{-\frac{1}{\sigma}} - 1 \right] \right\}$$

Em que μ_t é o quantil, também utilizamos a reparametrização por α, queremos testar juntamente com a reparametrização por δ, o viés, consistência e normalidade assintótica dos estimadores, para verificar qual reparametrização seria melhor.

Submodelos da EGEDD

Distribuição	α	λ	σ	δ	η	γ
EGDD	α	1	σ	δ	η	γ
DD	α	1	σ	δ	1	1
EGEBD	1	λ	σ	δ	η	γ
DB	1	1	σ	δ	1	1
EGBD	1	1	σ	δ	η	γ
EGEFD	α	λ	1	δ	η	γ
EGFD	α	1	1	δ	η	γ
FD	α	1	1	δ	1	1

Notamos que existes parâmetros melhores para fazer as reparametrizações, sem perder a generalidade da EGEDD como o δ.

Regressão Dagum

Incluindo a estrutura de regressão pelo quantil, por:

$$g(\mu_t) = x_t^\top \boldsymbol{\beta} = \sum_{j=1}^{k} x_{tj}\boldsymbol{\beta}, \quad t = 1, \ldots, n,$$

em que $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)^\top$ é o vetor de parâmetros desconhecidos e $\boldsymbol{\beta} \in \mathbb{R}^k$ e $\boldsymbol{x}_t = (x_{t1}, \cdots, x_{tk})^\top$ são observações de k covariáveis $k < n$, as quais são supostamente fixas e conhecidas. Caso o intercepto β_0 seja incluído no modelo, temos que $x_{t1} = 1$, para todo $t = 1, \ldots, n$

Regressão Dagum

Incluindo a estrutura de regressão pelo quantil, por:

$$g(\mu_t) = x_t^\top \boldsymbol{\beta} = \sum_{j=1}^{k} x_{tj}\boldsymbol{\beta}, \quad t = 1, \ldots, n,$$

em que $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)^\top$ é o vetor de parâmetros desconhecidos e $\boldsymbol{\beta} \in \mathbb{R}^k$ e $\boldsymbol{x}_t = (x_{t1}, \cdots, x_{tk})^\top$ são observações de k covariáveis $k < n$, as quais são supostamente fixas e conhecidas. Caso o intercepto β_0 seja incluído no modelo, temos que $x_{t1} = 1$, para todo $t = 1, \ldots, n$

Função de ligação

Necessita ser contínua e duas vezes diferenciável, logo a nossa função de ligação é $g(\mu_t) = \log(\mu_t)$.

Assim, ficaremos com o vetor de parâmetros (com a reparametrização por δ), sendo:

$$\boldsymbol{\Theta} = (\alpha, \sigma, \lambda, \eta, \gamma, \boldsymbol{\beta}^\top)^\top,$$

em que μ_t é o quantil da observação y_t no tempo t.

Modelo DARMA(p,q)

Consideramos o quantil condicional com a adição de uma componente adicional, ε_t, na qual é um modelo ARMA(p,q). Assim, um modelo mais geral para μ_t:

$$g(\mu_t) = x_t^\top \boldsymbol{\beta} + u_t$$

Modelo DARMA(p,q)

Consideramos o quantil condicional com a adição de uma componente adicional, ε_t, na qual é um modelo ARMA(p,q). Assim, um modelo mais geral para μ_t:

$$g(\mu_t) = x_t^\top \boldsymbol{\beta} + u_t$$

onde $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)^\top$ é o vetor de parâmetros desconhecidos e $u_t = g(y_t) - \boldsymbol{x}_t^\top \boldsymbol{\beta}$ possui estrutura ARMA (p,q), ou seja,

$$\phi(B)(g(y_t) - \boldsymbol{x}_t^\top \boldsymbol{\beta}) = \theta(B)\epsilon_t$$

Modelo DARMA(p,q)

Consideramos o quantil condicional com a adição de uma componente adicional, ε_t, na qual é um modelo ARMA(p,q). Assim, um modelo mais geral para μ_t:

$$g(\mu_t) = x_t^\top \boldsymbol{\beta} + u_t$$

onde $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)^\top$ é o vetor de parâmetros desconhecidos e $u_t = g(y_t) - \boldsymbol{x_t}^\top \boldsymbol{\beta}$ possui estrutura ARMA (p,q), ou seja,

$$\phi(B)(g(y_t) - \boldsymbol{x}_t^\top \boldsymbol{\beta}) = \theta(B)\epsilon_t$$

$$\phi(z) = -\sum_{i=0}^{p} \phi_i z^i, \quad \theta(z) = -\sum_{j=0}^{q} \theta_i z^i$$

onde L é operador *backward*, $\phi_0 = -1 = \theta_0$ e assumimos que $\phi(\cdot)$ e $\alpha(\cdot)$ não possuem raízes em comum e ϵ_t é um erro aleatório.

Modelo final ARMA:

$$g(\mu_t) = x_t^\top \beta + \sum_{j=1}^{n} \phi_j (g(\mu_t) - x_t^\top \boldsymbol{\beta}) + \phi(B)\epsilon_t$$

Estimador de Máxima Verossimilhança

A estimação do vetor de parâmetros $\boldsymbol{\theta}$ através do método da máxima verossimilhança, denotado por EMV, é dado por

$$\widehat{\boldsymbol{\theta}} = argmax_{\boldsymbol{\theta} \in \Theta} \{\ell(\boldsymbol{\theta})\}$$

Estimador de Máxima Verossimilhança

A estimação do vetor de parâmetros $\boldsymbol{\theta}$ através do método da máxima verossimilhança, denotado por EMV, é dado por

$$\widehat{\boldsymbol{\theta}} = argmax_{\boldsymbol{\theta} \in \Theta}\{\ell(\boldsymbol{\theta})\}$$

em que $\boldsymbol{\Theta} \subseteq \{\mathbb{R}_+^5 \times \mathbb{R}^{k+1+p+q}\}$ é o espaço de parâmetros e $\ell(\boldsymbol{\theta})$ é a função de log-verossimilhança de $\boldsymbol{y}$ é dada por

$$\ell(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}; \boldsymbol{y}) = \sum_{t=1}^{n} \ell_t(\boldsymbol{\theta}, y_t),$$

Estimador de Máxima Verossimilhança

A estimação do vetor de parâmetros $\boldsymbol{\theta}$ através do método da máxima verossimilhança, denotado por EMV, é dado por

$$\widehat{\boldsymbol{\theta}} = argmax_{\boldsymbol{\theta} \in \Theta}\{\ell(\boldsymbol{\theta})\}$$

em que $\boldsymbol{\Theta} \subseteq \{\mathbb{R}_{+}^{5} \times \mathbb{R}^{k+1+p+q}\}$ é o espaço de parâmetros e $\ell(\boldsymbol{\theta})$ é a função de log-verossimilhança de $\boldsymbol{y}$ é dada por

$$\ell(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}; \boldsymbol{y}) = \sum_{t=1}^{n} \ell_t(\boldsymbol{\theta}, y_t),$$

em que

$$\ell_t(\boldsymbol{\theta}, y_t) = \log(\alpha\delta\sigma\gamma\eta\lambda) - (\delta + 1)\log(y_t) - (\sigma + 1)\log(z_t)$$
$$+(\gamma - 1)\log\left(1 - z_t^{-\sigma}\right) + (\eta - 1)\log\left[1 - \left(1 - z_t^{-\sigma}\right)^{\gamma}\right]$$
$$+(\lambda - 1)\log\left\{1 - \left[1 - \left(1 - z_t^{-\sigma}\right)^{\gamma}\right]^{\eta}\right\},$$

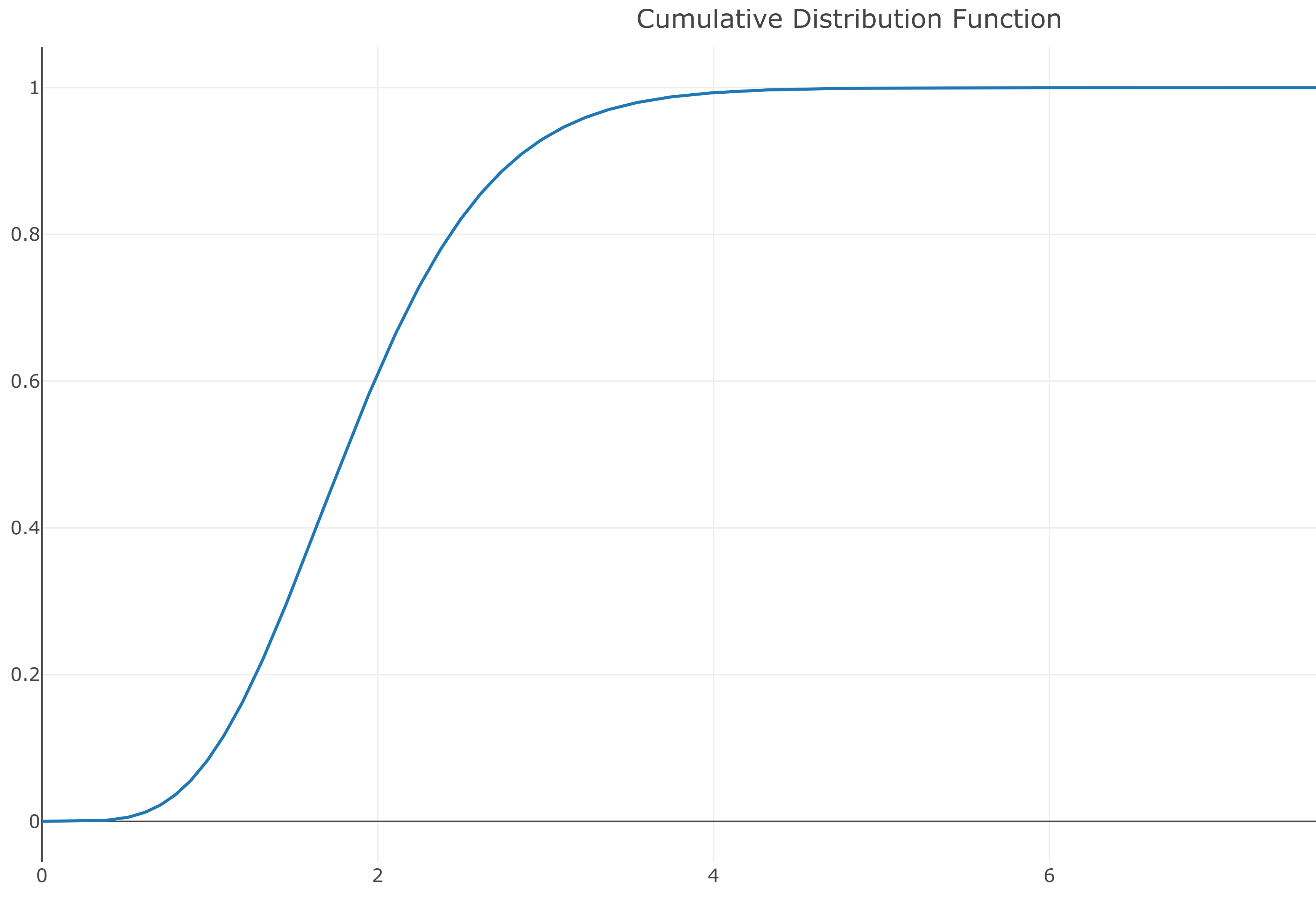
Estimador de Máxima Verossimilhança

Considerando a reparametrização por δ, o vetor de parâmetros do modelo é dado por:

$$\boldsymbol{\theta} = (\alpha, \sigma, \gamma, \eta, \lambda, \boldsymbol{\beta}^\top, \boldsymbol{\phi}^\top, \boldsymbol{\theta}^\top)^\top$$

Outros Procedimentos Estatísticos:

- Construção de testes de hipótese para verificar distribuição dos dados, verificando se é para a EGEDD ou um submodelo;
- Utilização de métodos númericos para execução de simulações (Newton-Rapson, Gosolnp com os métodos CG, BFGS e L-BFGS-B);
- Utilização de Simulações de Monte Carlo para estimação de parâmetros e análise de diagnósticos dos modelos;
- Desenvolvimento de Intervalos de Confiança e testes de hipóteses baseados na distribuição assintótica dos EMV's;
- Utilização dos estimadores de Entropia Extensiva;
- Aplicação a dados reais.



Vítor Bernardo Silveira Pereira - Email: vitorpereira3115@gmail.com

Referências Bibliográficas

- Dagum, C. (1977). El modelo log-logístico y la distribución del ingreso en la argentina. El Trimestre Económico, 44(176 (4):837–864
- Nasiru, S., Mwita, P. N., e Ngesa, O. (2019). Exponentiated generalized exponential dagum distribution. Journal of King Saud University-Science, 31(3):362–371.
- R Core Team (2021). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria