



GRÁFICOS DE CONTROLE PARA PROCESSOS DE CONTAGEM E UNITÁRIOS: UMA ABORDAGEM DE CEP CLÁSSICO E BAYESIANO

Anderson Fonseca – Universidade Federal da Bahia (ANDERSONOF@UFBA.BR)

Paulo Henrique Ferreira – Universidade Federal da Bahia (PAULOHENRI@UFBA.BR)

Diego Carvalho do Nascimento – Universidad de Atacama (DIEGO.NASCIMENTO@UDA.CL)

Francisco Louzada – Universidade de São Paulo (LOUZADA@ICMC.USP.BR)

Introdução

A grande competitividade no mercado atual faz com que a busca por excelência seja altamente necessária (LOUZADA et al., 2013). Neste contexto, o Controle Estatístico de Processos (CEP) é uma alternativa muito importante e utilizada. Uma das suas técnicas de destaque é o gráfico (ou carta) de controle, no qual é possível observar se o processo está fora de controle estatístico ou não.

Objetivos

O objetivo principal deste trabalho é prover a implementação e disponibilização de funções no R, capazes de gerar gráficos de controle para o monitoramento estatístico de processos não normais (por exemplo, dados de contagem ou dados contínuos no intervalo $(0,1)$, como é o caso de taxas, índices e proporções) via inferência clássica. Adicionalmente, foi considerada a estimação de parâmetros via abordagem Bayesiana, o que permite construir um intervalo de credibilidade associado a cada limite de controle, possibilitando agregar conhecimento *a priori* do fenômeno no processo de estimação.

Material e Método

Dois tipos de dados que vêm recebendo bastante atenção na literatura, são os dados de contagem (presentes em diversas situações do cotidiano, como, por exemplo, a quantidade de peças defeituosas em uma linha de produção, os casos de COVID-19 por semana epidemiológica etc.) e os dados unitários (com aplicações em uma vasta gama de áreas, como ecologia, economia e indústria, entre outras; umidade relativa do ar e taxa de inflação são alguns exemplos de variáveis unitárias). Supor normalidade nesses cenários pode levar a conclusões errôneas, sendo importante adotar outra abordagem para monitorar esses tipos de dados.

Para dados de contagem, a Poisson é a distribuição de probabilidade mais conhecida e utilizada. Entretanto, ela supõe que a média e a variância do processo são iguais, sendo tal suposição violada em muitas ocasiões. Para contornar este problema, no âmbito de CEP,



podem ser utilizados gráficos de controle baseados em distribuições que acomodam superdispersão (ou sobredispersão) dos dados. Dentre as distribuições que possuem essa propriedade, destacam-se as distribuições resultantes de misturas uniparamétricas de Poisson, como a Poisson-Lindley (SANKARAN, 1970), a Poisson-Shanker (SHANKER et al., 2017) e a Poisson-Sujatha (SHANKER e FESSHAYE, 2016). No contexto de monitoramento de processos, os gráficos de controle desenvolvidos com base nessas distribuições, foram apresentados por Jesus et al. (2022).

O gráfico de controle se baseia na teoria geral de Shewhart, segundo a qual os limites de controle são calculados através da expressão: Média $\pm L \times$ Desvio-Padrão. Também referido como gráfico de controle de Shewhart, foi desenvolvido considerando a distribuição normal, porém é facilmente adaptado a qualquer distribuição que apresente primeiro e segundo momentos. Como caso particular, a política de controle adotada pelo Seis Sigma considera a constante L igual a 3.

Dados unitários originários de processos não Bernoulli (como taxas, índices e proporções) são frequentemente observados em situações reais (por exemplo, umidade relativa do ar e índices econômicos como taxa de juros e inflação). Adicionalmente, a assimetria geralmente está presente em boa parte dessas situações e a suposição de normalidade para dados duplamente truncados não é razoável de ser adotada. Em um contexto de CEP, poderá fazer, inclusive, com que os limites de controle excedam o intervalo (0,1).

Uma alternativa é utilizar a função quantil de distribuições unitárias (assimétricas) para calcular os limites de controle, garantindo que estes estejam dentro do intervalo (0,1). Dentre os gráficos de controle que fazem uso dessa estratégia (os quais podem ser referidos como gráficos de controle de Shewhart modificados), destacam-se as propostas baseadas nas distribuições beta (SANT'ANNA e TEN CATEN, 2012), Kumaraswamy (LIMA-FILHO e BAYER, 2019), simplex e unit-gamma (LEE HO et al., 2019), unit-Lindley (FONSECA et al., 2021).

Visto a crescente demanda de aplicações de controle de processos não normais, este trabalho buscou implementar diversas funções que geram gráficos de controle para o tipo de dados escolhido (contagem ou unitário). Todas as implementações foram feitas no *software* R versão 4.2.0 (R CORE TEAM, 2022) e estão disponíveis em: <https://github.com/CER-UFBA/cchart>.

A função que gera os gráficos de controle foi nomeada de *plot.control* e apresenta a seguinte estrutura: *plot.control(x, type = c('unit', 'count'), fase1 = 0.7, alpha = .1, L = 3, iter = 1000, xlab = 'a', ylab = 'b', tit = NULL, size = 10)*. Seus argumentos são: x : vetor com dados;



type: tipo de dados - unitário ('*unit*') ou contagem ('*count*'); *fase1*: proporção de dados para a fase 1 (estimação); *alpha*: probabilidade de alarme falso (α , ou erro do tipo I); *L*: constante do gráfico de controle de Shewhart; *iter*: número de iterações (algoritmo Bayesiano); *xlab*: legenda do eixo x; *ylob*: legenda do eixo y; *tit*: título do gráfico; *size*: tamanho da fonte dos elementos textuais do gráfico.

Para dados de contagem, foram implementados os gráficos de controle baseados nas distribuições de Poisson, Poisson-Lindley, Poisson-Shanker e Poisson-Sujatha. Já para dados unitários, considerou-se as distribuições beta, Kumaraswamy, unit-Lindley e unit-half-normal (BAKOUCH et al., 2021), sendo que o gráfico de controle baseado nesta última é inédito na literatura sobre CEP (e com a diferenciação dela ser parametrizada em torno da mediana).

Para estimar os parâmetros das distribuições, também foi adotada uma abordagem Bayesiana (MIGON et al., 2013), com estimação da *posteriori* via método de Monte Carlo Hamiltoniano (HMC), mediante a integração com o STAN via pacote "rstan" (STAN DEVELOPMENT TEAM, 2022). Os dados utilizados para fazer essa estimação são chamados de dados da fase 1, enquanto os restantes são dados da fase 2 (monitoramento *online* do processo). O paradigma Bayesiano permite encontrar os quantis da distribuição *via posteriori* diretamente dos parâmetros e calcular os intervalos de credibilidade dos limites de controle. Este trabalho considerou *prioris* vagas (não informativas), isto é, cada parâmetro adotado teve distribuição normal centrada em zero e com variância grande (6,25), para fins de ilustração.

Resultados e Discussão

Por meio das funções desenvolvidas, foi feita uma aplicação aos dados sobre umidade relativa do ar na cidade de Copiapó - Chile. Considerou-se uma amostra de registros mínimos da umidade relativa do ar em cada parte do dia (parte 1 - da meia-noite às 5h59 UTC, parte 2 - das 6h às 11h59 UTC, parte 3 - do meio-dia às 17h59 UTC, e parte 4 - das 18h às 23h59 UTC), no período de dezembro de 2016 a dezembro de 2020, contendo 200 observações escolhidas aleatoriamente.

Foram plotados os quatro gráficos de controle (beta, Kumaraswamy, unit-Lindley e unit-half-normal *control charts*) para os dados unitários selecionados. Através da estimação Bayesiana (testes e gráficos de convergência respaldam positivamente a robustez das cadeias obtidas), com as primeiras 160 observações (dados da fase 1), foram calculados os limites de controle e seus respectivos intervalos de credibilidade, utilizando os percentis 2,5% e 97,5% da estimativa da distribuição *a posteriori* dos parâmetros para calcular esses intervalos. A política de controle adotada foi de $\alpha = 10\%$. Vale observar que o mesmo conjunto de dados foi utilizado em Fonseca et al. (2021), porém, naquela ocasião, as 200

observações foram tidas como dados da fase 1 e outras observações foram usadas para a fase 2. Por conta da limitação de espaço, esta será a única aplicação apresentada no texto, mas no repositório do GitHub encontram-se outras aplicações (inclusive, para dados de contagem).

Na Figura 1, é possível visualizar os gráficos construídos. As bandas vermelhas são os intervalos de credibilidade de 95% para os limites de controle, a linha azul representa a linha central (a média do processo) e a linha vertical tracejada marca o fim da fase 1. Os pontos que sinalizam que o processo está sob controle estatístico estão na cor preta, enquanto que os pontos na cor verde indicam que o processo está fora de controle estatístico. Por fim, em azul estão os pontos dentro dos intervalos dos limites. Observa-se, dentre outros, que o gráfico de controle unit-Lindley sinaliza mais pontos (fora de controle) acima do limite superior, enquanto os demais apresentam a maioria dos pontos (fora de controle) abaixo do limite inferior.

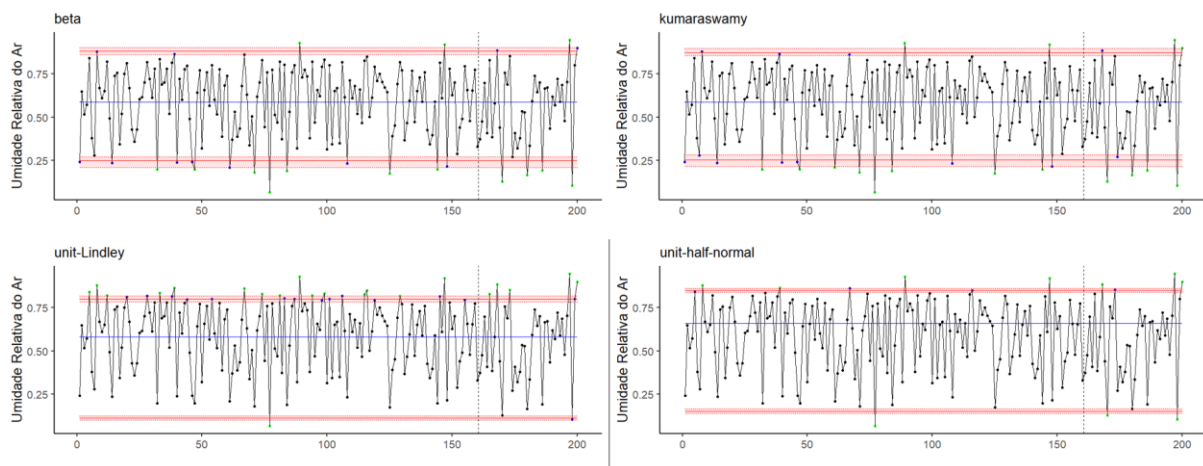


Figura 1 - Gráficos de controle gerados.

Uma alternativa para saber quais distribuições se adequam aos dados é fazer o teste de Kolmogorov-Smirnov (inserido nas funções construídas). Assim, tem-se um indicativo de que, ao nível de significância de 10%, todas as distribuições se ajustam bem aos dados analisados, com maior p-valor obtido para a distribuição unit-Lindley (beta: 0,1210; Kumaraswamy: 0,1935; unit-Lindley: 0,6047; unit-half-normal: 0,3375).

Conclusão

Neste trabalho foram desenvolvidas e disponibilizadas funções R que permitem a fácil construção de gráficos de controle para o monitoramento estatístico de processos pelo usuário. Além de agregar a abordagem de gráficos de controle via inferência Bayesiana para distribuições já existentes na literatura e para variáveis de contagem ou unitárias.



Referências

- Bakouch, Hassan S., Ali Saadati Nik, Akbar Asgharzadeh, e Hugo S. Salinas. **A flexible probability model for proportion data:** Unit-half-normal distribution. *Communications in Statistics: Case Studies, Data Analysis and Applications* 7.2 (2021): 271-288.
- Fonseca, Anderson, Paulo Henrique Ferreira, Diego Carvalho do Nascimento, Rosemeire Fiaccone, Christopher Ulloa-Correa, Ayón García-Piña, e Francisco Louzada. **Water particles monitoring in the Atacama desert:** SPC approach based on proportional data. *Axioms* 10.3 (2021): 154.
- H. W. Borchers, **pracma:** Practical Numerical Math Functions, 2022, URL: <https://CRAN.R-project.org/package=pracma>, r package version 2.3.8.
- H. Wickham. **ggplot2:** Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.
- Jesus, Bruno D., Paulo H. Ferreira, Laion L. Boaventura, Rosemeire L. Fiaccone, Wesley Bertoli, Pedro L. Ramos, e Francisco Louzada. **Statistical process control of overdispersed count data based on one-parameter Poisson mixture models.** *Quality and Reliability Engineering International* 38.5 (2022): 2324-2344.
- Lee Ho, Linda, Fidel Henrique Fernandes, e Marcelo Bourguignon. **Control charts to monitor rates and proportions.** *Quality and Reliability Engineering International* 35.1 (2019): 74-83.
- Lima-Filho, L.M.d.A., e Fabio Mariano Bayer. **Kumaraswamy control chart for monitoring double bounded environmental data.** *Communications in Statistics - Simulation and Computation*. 50.9 (2019): 2513–2528.
- Louzada, Francisco, Carlos A.R. Diniz, Paulo H. Ferreira, e Edil L. Ferreira. **Controle Estatístico de Processos:** uma abordagem prática para cursos de Engenharia e Administração. Grupo Gen-LTC, 2013.
- Migon, Helio S., Dani Gamerman, e Francisco Louzada. **Statistical inference:** an integrated approach. CRC press, 2014.
- R Core Team, **R:** A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2022, URL: <https://www.R-project.org/>.
- Sankaran, Munuswamy. **275. note:** The discrete Poisson-Lindley distribution. *Biometrics* 26.1 (1970): 145-149.
- Sant'Anna, Ângelo Márcio Oliveira, e Carla Schwengber ten Caten. **Beta control charts for monitoring fraction data.** *Expert Systems with Applications* 39.11 (2012): 10236-10243.
- Shanker, Rama, e Hagos Fesshaye. **On discrete Poisson-Shanker distribution and its applications.** *Biometrics & Biostatistics International Journal* 5.1 (2017): 00121.
- Shanker, Rama, Kamlesh Kumar Shukla, e Hagos Fesshaye. **A generalization of Sujatha distribution and its applications with real lifetime data.** *Journal of Institute of Science and Technology* 22.1 (2017): 66-83.
- Stan Development Team, **RStan:** the R interface to Stan, 2022, URL: <https://mc-stan.org/>.
- Thomas W. Yee (2022). **VGAM:** Vector Generalized Linear and Additive Models. R package version 1.1-6. URL <https://CRAN.R-project.org/package=VGAM>