



Análise da Função de Correlação de dois Pontos em Distribuições Sintéticas

Felipe Ventura Vargas ^{1,2}, Evelise Gausmann ² e Fabricio Ferrari ²

doi.org/10.20873/stmmta2021-2526-5725-58926

¹ Institution/Instituição 1

² Instituto de Matemática Estatística e Física – IMEF, Universidade Federal do Rio Grande – FURG, Rio Grande, RS, Brasil.

Resumo

A função de correlação de dois pontos é um dos principais métodos estocásticos utilizados dentro da cosmologia para a medição do grau de homogeneidade de um sistema. A metodologia consiste em comparar distribuições amostrais com uma distribuição uniforme e homogênea, identificando escalas de homogeneidade da distribuição amostral através de suas diferenças. Testamos diferentes distribuições sintéticas a fim de identificar a abrangência das informações obtidas sobre tais distribuições. Os resultados mostraram que a Função de Correlação de Dois Pontos consegue identificar escalas características nas quais distribuições podem ser consideradas homogêneas, concentradas ou apresentam vazios. Para distribuições Gaussianas, a resolução das oscilações que representam escalas características é limitado; se a distâncias entre distribuições for inferior a quatro vezes o desvio padrão das distribuições, as oscilações são sobrepostas. Mostramos também que diferentes distribuições podem gerar Funções de Correlação de dois Pontos similares, confirmando resultados encontrados na literatura.

Palavras-chave: Função de Correlação de dois pontos, Distribuições de distâncias, Escalas características

Abstract

The two-point correlation function is one of the main statistical methods used within cosmology. This function compares sample distributions with a uniform and homogeneous distribution, identifying scales of sample distribution homogeneity. We test different synthetic distributions in order to identify the completeness of information obtained about such distributions. The results showed that the Two-Point Correlation Function is able to identify characteristic scales in which distributions can be considered homogeneous, concentrated or present empty. For Gaussian distributions, the resolution of oscillations representing characteristic scales is limited: if the distances between distributions are less than four times the standard deviation of the distributions, the oscillations are superimposed. We also show that different distributions can generate Correlation Functions of two similar points, confirming results found in the literature.

Keywords: Two-point correlation function, Distributions of distances, Scales Features

Introdução

O Princípio Cosmológico é uma das hipóteses mais importantes da cosmologia, no entanto cada vez mais evidências apontam para uma falta de homogeneidade na distribuição de matéria do Universo. Em seu trabalho Shane e Wirtanen, 1954 realizaram uma contagem de nebulosas ex-

tragaláticas que apontou uma tendência de que aglomerados de galáxias ocorram em grupos.

Com evidências para a existência de aglomerados de aglomerados de galáxias surge a questão se tais estruturas seriam uma extensão das galáxias ou apenas galáxias reunidas. Caso sejam, não é possível dizer que o Universo é

homogêneo, pois galáxias são nucleadas e isto requer que uma densidade maior de matéria esteja próxima ao núcleo. Se os aglomerados de aglomerados também forem nucleados e esse padrão se repetir para escalas muito maiores então teremos pontos de grande densidade de massa e regiões sem massa alguma em todas as escalas espaciais. Existe um padrão medido de homogeneidade e isotropia em escalas acima de 100 Mpc, mas na metade do século XX, muitas evidências colocavam em cheque tais propriedades. Os super aglomerados também foram estudados em Abell, 1958 onde um catálogo é preparado com 2712 aglomerados ricos em galáxias, 1682 aglomerados que respeitam critérios estatísticos de amostras homogêneas são selecionados levando a conclusões importantes; a densidade espacial do centro dos aglomerados não varia com a distância, inferindo que tais aglomerados não eram uma extensão da distribuição de galáxias, mas sim uma forma mais homogênea de disposição, não nucleada. Abell, 1958 também afirma que existe um padrão significativo não aleatório de distribuição dos aglomerados, formando o que chamou de aglomerados de segunda ordem.

Muitas evidências apontam em favor do Princípio Cosmológico. Um dos maiores levantamentos recentemente feitos é o *Sloan Digital Sky Survey (SDSS)* Stoughton e et, al., 2002 onde, até 2001 havia detectado 14 milhões de objetos nos dois hemisférios celestes, apontando um grande grau de homogeneidade na distribuição de matéria no universo. A discussão sobre a homogeneidade do universo se estendeu até que *surveys* em escalas muito grandes fossem realizados. Mas foi necessário desenvolver ferramentas para quantificar tal homogeneidade e por fim entender se o Princípio Cosmológico é uma hipótese científica válida. A Função de Correlação é um dos principais métodos estatísticos utilizados dentro da cosmologia, popularizada na área por Groth e Peebles, 1977.

A Função de Correlação utiliza uma análise de campos estocásticos para distribuição de distâncias, partindo da hipótese de que existe um campo de distribuição de distâncias uniforme com flutuações. A distribuição amostral é então comparada com o campo teórico uniforme, qualquer diferença estatística entre a amostra e a distribuição teórica indica uma falta de homogeneidade na distribuição amostral estudada. A função nos diz o quanto uma distribuição qualquer difere de uma distribuição aleatória e uniforme, gerada por um Processo Pontual de Poisson Homogêneo. O Processo Pontual de Poisson Homogêneo é um método utilizado para gerar dados de distribuições homogêneas que sejam estatisticamente independentes. Como a distribuição gerada por esse processo é homogênea, uma comparação

é feita gerando um índice que varia entre -1 e 1. Caso o comportamento da função de correlação tenda a zero para grandes escalas espaciais, não haverá diferença entre a distribuição amostral e a distribuição teórica homogênea, indicando assim, que para a escala teórica estudada, o Princípio Cosmológico é válido.

A Função é uma interessante ferramenta de análise de distribuição estatística que pode nos dar informações sobre distribuições de pontos quaisquer. Utilizando apenas a localização espacial de objetos, pode ser aplicada em diversas escalas. Em uma das primeiras aplicações na cosmologia, Totsuji e Kihara, 1969 mostra, para dados reais, que a Função de Correlação pode ser aproximada por uma lei de potências. Quer dizer que para cada escala de distância, o grau de homogeneidade varia exponencialmente. Existem distâncias características que podem ser identificadas pela função, apontando a transição da escala em que há aglomeração para uma escala em que a distribuição começa a ser homogênea. Como visto em Einasto et al., 1986, onde função é aplicada para amostras de dados observacionais de distribuições de galáxias e vazios no universo, há um comprimento de correlação, uma escala limite até onde se pode dizer que há baixa homogeneidade. Existem duas definições para o comprimento de correlação. Para a Física Estatística, é uma distância na qual a "memória espacial" de flutuações locais persistem, ou seja, onde o efeito de pequenas perturbações no sistema é sentido. Já para a estatística cosmológica, o comprimento de correlação representa onde a amplitude de flutuações de densidade começa a se tornar pequena demais em comparação com a densidade média dos objetos ao todo (Gabrielli et al., 2005). Em Einasto et al., 1997, são analisados dados de distribuições sintéticas que estudam como a Função de Correlação de dois pontos varia em relação à geometria da distribuição gerada. Em larga escala, a função pode mostrar qual a geometria de uma distribuição de super aglomerados e em pequenas escalas a função varia com o formato e o tamanho da distribuição.

A Função de Correlação é um dos métodos mais influentes no estudo estatístico da distribuição de matéria no universo. Sendo utilizada desde a metade do século XX para determinar o grau de homogeneidade de distribuições e quantificar até que ponto o Princípio Cosmológico pode ser aplicado. Através de análise de dados numéricos de modelos com distribuições sintéticas, o objetivo do trabalho é estudar a influência de variações de parâmetros estatísticos e de arranjos de distribuições na Função de Correlação de Dois Pontos. Com esse estudo teremos um panorama mais claro de quais as características da Função de Correlação de dois pontos serão realmente úteis ao extrair infor-

mações de dados reais, assim como as limitações desta ferramenta. Este trabalho foi baseado no Trabalho de Conclusão de Curso do Autor Felipe Ventura Vargas realizado no IMEF/FURG e está organizado da seguinte forma: Na seção *Distribuições* apresentamos as leis de distribuição usadas no estudo, assim como o Processo pontual de Poisson (PPP) usado na Função de Correlação de Dois Pontos (FC2P); Na seção *Função de Correlação de Dois Pontos* apresentamos o estudo da FC2P aplicada a diversas configurações usando as distribuições apresentadas; Por fim apresentamos nossos resultados conclusões.

Distribuições

A análise da FC2P será feita com dados sintéticos controlados cujas características são conhecidas. As principais distribuições utilizadas neste trabalho são Distribuição Gaussiana, Distribuição Beta e Distribuição Homogêneas. A distribuição homogênea é obtida através de um PPP Homogêneo. A vantagem da utilização deste método é que todas as características das distribuições analisadas com a função já são conhecidas, servindo como base de comparação e para futuras análises.

O algoritmo foi gerado em Python, onde determinamos o tipo de distribuição, quantidade de pontos, afastamento entre pontos, centro (valor médio) para nuvens de pontos, afastamento entre nuvens e todas as outras possíveis características da nuvem. Assim, em um ambiente controlado, teremos certeza de que os resultados encontrados representam características das amostras. Como a FC2P pode ser utilizada em qualquer escala de distâncias e estamos usando dados sintéticos, a unidade de espaço utilizada será arbitrária ou, *u.m.* (unidade de medida).

Distribuição Gaussiana

A densidade de probabilidade Gaussiana segue a lei:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (8)$$

onde μ representa a média da distribuição e σ o desvio padrão, uma medida da dispersão dos pontos em torno da média. As nuvens gaussianas serão formadas gerando distribuições de pontos nas três dimensões (x, y, z) seguindo a lei descrita pela equação 8. Em uma distribuição Gaussiana teremos uma probabilidade maior de encontrar pontos perto do centro das nuvens (ponto médio) do que nas extremidades. O desvio padrão nos dirá o quão concentrada ou espalhada (distante do centro) estarão os pontos de uma nuvem.

Distribuição Beta

A distribuição Beta segue a lei de distribuição:

$$p(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1} \quad (9)$$

para $0 < x < 1$ e coeficientes arbitrários $\alpha, \beta > 0$. No modelo, os parâmetros α e β definem a forma da distribuição. Se $\alpha = \beta$, a distribuição é simétrica, se $\alpha > \beta$, a assimetria é negativa e, no caso de $\alpha < \beta$, sua assimetria é positiva. Já o desvio padrão σ está relacionado com os coeficientes α e β na forma:

$$\sigma(\alpha, \beta) = \frac{\alpha\beta}{(\alpha + \beta + 1)(\alpha + \beta)^2} \quad (10)$$

onde $B(\alpha, \beta)$ é uma função de normalização:

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx \quad (11)$$

A nuvem Beta (Figura 7) foi feita gerando distribuições de pontos nas três dimensões (x, y, z) seguindo a lei descrita pela equação 9 com $\alpha = \beta = 0.3$. Podemos observar que a probabilidade de termos pontos nos extremos é maior do que ter pontos no centro da distribuição, se comparado com uma distribuição Gaussiana, como mostrado na Figura 7. A distribuição Beta varia de acordo com a combinação de α e β escolhida, podendo variar do exemplo mostrado na Figura 7 até uma distribuição aproximadamente Gaussiana. Embora a distribuição Beta não tenha significado físico, sua forma nos dá uma ideia de uma distribuição com uma densidade maior de ponto nas extremidade e uma densidade menor de pontos (ou um vazio) no centro. A FC2P de vazios pode ser útil em análises de distribuições reais, portanto, se bem conhecida, pode nos dar informações importantes.

Processo Pontual de Poisson Homogêneo

O PPP Homogêneo é um processo que garante que a probabilidade de encontrar um ponto independe de outros pontos gerados anteriormente na distribuição. Neste trabalho focaremos em sua aplicação discreta para gerar os dados sintéticos. A formação do Processo Pontual consiste na geração de regiões de volume onde pontos sejam dispostos aleatoriamente e o número de pontos em cada região se dá de maneira probabilística. No caso do PPP a distribuição de

probabilidade estatística é a de Poisson dada por:

$$p(k, \lambda) = \frac{\lambda e^{-\lambda}}{k!}, \quad (12)$$

onde k é o número de eventos esperados que ocorrem no intervalo λ . No caso homogêneo, o intervalo λ é mantido constante. Um número de pontos, $p(k, \lambda) = N$, é gerado para cada região. Em seguida N pontos são dispostos aleatoriamente em uma região, ou "caixa". É calculado um novo N para cada caixa, repetindo o processo várias vezes. Diversas caixas são geradas da mesma forma, o somatório das caixas gera uma distribuição Homogênea e a probabilidade de seus pontos são independentes. Quanto mais caixas forem formadas para a distribuição, melhor essa característica é garantida. Se pegarmos mil caixas de dez pontos e somarmos, podemos ter algo como na Figura 8. Este é o método utilizado neste trabalho para gerar distribuições homogêneas de acordo com Martínez, 2001.

Função de Correlação de Dois Pontos

A FC2P é uma formulação específica de Função de Correlação em que são comparadas distribuições de frequências de distâncias entre dois pontos de dois tipos de amostra. Uma delas é uma distribuição homogênea aleatória gerada por um PPP Homogêneo e a outra é a amostra utilizada para análise, podendo ser sintética ou de dados reais de galáxias.

Para o cálculo da distribuição de frequência de distâncias devemos contabilizar todas as distâncias entre pontos sem repeti-los. Imaginemos quatro pontos dispostos no vértice de um quadrado de aresta igual a 1 u.m., como mostrado na Figura 9. Neste caso, distâncias do ponto 1 ao ponto 2 e do ponto 2 ao ponto 1 serão consideradas a mesma, sendo contabilizada apenas uma vez. No espaço euclidiano, a distância entre dois pontos é dada por $d = (x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2$, onde x, y, z são as coordenadas dos pontos e os índices i e j representam os pontos dos quais a distância está sendo calculada. Calculando as distâncias entre todos os pontos construímos a Tabela 1. Podemos ver pela tabela que há quatro distâncias de 1 u.m. (nas arestas) e duas distâncias de $\sqrt{2} \approx 1,41$ u.m. (entre os vértices), então a contagem de distâncias para esse caso será de 4 distâncias para a distância 1 u.m. e de 2 distâncias para a distância 1,41 u.m..

Estamos interessados no cálculo de frequências de distâncias que envolve uma quantidade de pontos muito maior. A Figura 8, com dez mil pontos, representa algo mais próximo do que queremos. A mesma ideia aplicada para a Figura 9 será utilizada aqui, calculando todas as distâncias en-

tre todos os pontos sem repetição. O resultado é mostrado na Figura 10, onde a abscissa representa as distâncias e a ordenada o número de vezes que essa distância aparece. A Figura 10 também é conhecida como um histograma, pois associa uma variável a uma frequência de classe - neste caso distância e frequência de distâncias. No caso de uma distribuição uniforme e homogênea como a da Figura 8, para distâncias muito pequenas temos uma baixa frequência de ocorrências. A medida que as distâncias aumentam, aumentam também a sua frequência, até que a partir de determinada distância, o centro da distribuição, os valores voltam a diminuir. Isso acontece porque para distâncias pequenas há poucos vizinhos, essa quantidade vai aumentando e com ela aumenta a contagem de distâncias. Mas ao se aproximar de distâncias maiores ainda, devido ao fato da amostra ser finita, a contagem de distâncias começa a cair novamente. A forma do histograma de distâncias depende da distribuição analisada.

A FC2P apresenta diversas fórmulas para sua representação discreta. Como é visto em Hamilton, 1993, existem formas mais complexas da função que corrigem efeitos causados pelas bordas das amostras mas que, para fins do estudo do comportamento da função, não influenciam tão fortemente em suas principais características. Portanto, a fórmula discreta aplicada será a utilizada em Einasto et al., 1986,

$$\xi_{(r)} = \frac{(DD(r))n_r}{(RR(r))n_d} - 1, \quad (13)$$

onde $(DD(r))$ é frequência de distâncias em um intervalo $r \pm dr$ entre pares de pontos da distribuição estudada e $(RR(r))$ é a frequência de distâncias para uma distribuição gerada por um PPP Homogêneo (como a mostrada na Figura 10). A razão $\frac{n_r}{n_d}$ é um coeficiente de normalização para que as amostras não precisem ter necessariamente o mesmo número de pontos. Os valores n_r e n_d representam, respectivamente, o número de distâncias na amostra homogênea e na amostra estudada.

Ao dividir a distribuição de frequência de distâncias da amostra pela distribuição de frequências gerada por uma distribuição homogênea de pontos teremos valores que nos indicam quantas vezes mais determinadas distâncias estão presentes na amostra em relação à amostra homogênea. Ao diminuir uma unidade, contagens de distâncias que são iguais se tornam nulas na FC2P (distâncias não correlacionadas). Caso haja mais distâncias na amostra do que na distribuição homogênea o valor da função de correlação será positivo (positivamente correlacionadas) indicando uma aglomeração de pontos. Caso o número de distâncias na amostra seja menor, teremos valores negativos na fun-

ção (anti correlacionadas) o que nos indicará a presença de vazios ou regiões de baixa densidade.

Em Einasto et al., 1997 os autores estudam o comportamento da FC2P variando distribuições de pontos. Neste artigo, é possível ver que os valores da FC2P estão em escala logarítmica, já que os gráficos apresentados variam muito relação às escalas de distâncias, começando por um pico muito grande depois caindo rapidamente a zero e oscilando em torno desse valor. Como há uma maior variação no primeiro pico da FC2P, normalmente a função, nesta região, é aproximada por uma lei de potências. A distância na qual o primeiro pico da FC2P chega a zero é chamada de comprimento de correlação que, segundo Gabrielli et al., 2005, é uma escala na qual as flutuações dos valores de distâncias se tornam pequenas amplitudes em relação à densidade média da distribuição de distâncias de uma distribuição homogênea de pontos. O comprimento de correlação é uma distância característica na qual as amostras tendem à homogeneidade, principalmente para dados reais. A FC2P normalmente apresenta um comportamento oscilatório tendo picos e vales. O centro dos picos representa distâncias características da FC2P. Tais distâncias normalmente são associadas ao tamanho das distribuições e ao afastamento entre distribuições. A seguir faremos uma análise das distribuições dos picos. Utilizando a Equação 13 implementaremos o cálculo discreto da FC2P na linguagem de programação Python.

FC2P das amostras sintéticas de nuvens gaussianas

Primeiramente geramos uma amostra de nuvens gaussianas centralizadas nas arestas de um cubo. Esta distribuição simples nos ajudará a analisar FC2P resultante e sua relação com as distribuições. Foram geradas oito distribuições de pontos que seguem uma lei gaussiana (Equação 8). O centro de cada distribuição é o vértice de um cubo de aresta 50 u.m. com coordenadas dos centros (x, y, z) : $(10,10,10)$, $(10,10,60)$, $(10,60,60)$, $(60,60,60)$, $(10,60,10)$, $(60,10,10)$, $(60,60,10)$ e $(60,10,60)$. O desvio padrão, σ , de todas as distribuições é 5 u.m.

No topo da Figura 11, mostramos em um gráfico tridimensional como as distribuições estão dispostas. O segundo gráfico da Figura 11 aponta de maneira mais didática as principais distâncias envolvidas na distribuição, excluindo apenas a distância que cruza o cubo, pois a figura é apenas uma projeção bidimensional das distribuições. Como explicado anteriormente, as distâncias entre todos os pontos devem ser contadas sem contagem repetida de pontos. Uma análise superficial da figura nos dá uma ideia de quais devem ser os picos esperados, pois estes es-

tão associados à distâncias características das distribuições. O diâmetro de cada nuvem está em torno de 10 u.m. (já que $\sigma = 5 \text{ u.m.}$), logo as principais distâncias envolvidas devem ser:

- 10 u.m. , escala interna das nuvens;
- 50 u.m. , distâncias (entre pontos das nuvens) que não cruzam a face do cubo nem o cubo;
- $50\sqrt{2} \text{ u.m.}$, distâncias (entre pontos das nuvens) diagonais da face do cubo;
- $50\sqrt{3} \text{ u.m.}$, distâncias (entre pontos das nuvens) nas diagonais do cubo;

A Figura 12 representa, respectivamente, o histograma de contagem de distâncias para a distribuição de oito nuvens gaussianas e a FC2P calculada para esta mesma distribuição. A Tabela 2 nos dá as posições dos picos das duas figuras. Podemos observar no histograma da Figura 12 que o primeiro pico está muito próximo de 5 u.m. que é o desvio padrão das nuvens. O segundo pico está próximo de 50 u.m. , que é a distância entre as nuvens. O terceiro pico está próximo de $50\sqrt{2} \text{ u.m.}$ que é a distância entre nuvens na diagonal da face do cubo. O quarto pico está próximo de $50\sqrt{3} \text{ u.m.}$, que é a distância na diagonal atravessando o cubo. Na FC2P da Figura 12, o segundo pico se aproxima do diâmetro das nuvens, o terceiro, quarto e quinto picos seguem a mesma lógica do histograma correspondente.

Pela definição da FC2P os picos do histograma e da FC2P não devem divergir. No entanto o primeiro pico está em um valor um pouco diferente. Como discutiremos adiante, o primeiro pico é caótico, pois pequenas variações nas condições iniciais afetam de forma significativa seu valor, podendo ser completamente diferente para distribuições estatisticamente iguais. Tal característica será estudada melhor mais adiante.

Para entender melhor o primeiro pico da FC2P foram feitas dez distribuições de pontos estatisticamente iguais. As distribuições são compostas por duas nuvens com mil pontos cada, centradas em 25 u.m. e 75 u.m. , com desvio padrão, σ , igual a 5 u.m. . A Figura 13 dá uma representação visual de como cada distribuição se dá espacialmente projetada no plano. A distribuição homogênea gerada tem vinte mil pontos (dez vezes mais pontos que a amostra testada). A FC2P para cada uma das distribuições foi calculada e em seguida uma FC2P média foi calculada para diminuir o erro estatístico, de forma similar ao procedimento em Einasto et al., 1997, onde a FC2P é a média de 10 FC2Ps. Embora as distribuições sejam estatisticamente iguais a disposição dos pontos é aleatória, logo pequenos erros estatísticos devem

surgir. O processo de calcular a média das FC2Ps foi repetido três vezes, de maneira que três FC2Ps média foram calculadas e estão dispostas na Figura 14. As FC2Ps foram suavizadas, dividindo as curvas em 51 segmentos interpolados por polinômios de terceiro grau. A suavização afeta um pouco a amplitude dos gráficos, mas não a posição dos picos, característica que estamos interessados nesse trabalho. Portanto podemos seguir com a análise da posição dos picos sem preocupações sobre erros associados à suavização.

Podemos observar que mesmo se tratando de FC2Ps médias suavizadas, o primeiro pico para cada uma delas é completamente diferente. Pequenas variações estatísticas geram grandes variações no primeiro pico da função.

A função de correlação divide o histograma de uma amostra pelo histograma de uma distribuição homogênea e uniforme. O histograma de uma distribuição homogênea e uniforme tende a zero nas extremidades (similar à Figura 10), o que deixa a razão muito sensível nessas regiões, amplificando flutuações da ordem do erro causado pelo fato da distribuição ser aleatória. Distribuições iguais, centradas na mesma posição com o mesmo desvio padrão apresentam variação maior nas extremidades do que no resto da função (Figura 14). O primeiro pico não deve ser levado em consideração ao se comparar distribuições diferentes já que diverge mesmo para distribuições iguais. De fato, a literatura não utiliza o valor do primeiro pico. Para analisar esta primeira oscilação é comum que seja feita uma aproximação por uma lei de potências. Com tal lei é encontrado o comprimento de correlação (onde o primeiro pico começa a tender a zero), que seria uma distância na qual a distribuição, para dados reais, passa a ter um comportamento mais homogêneo. No entanto as análises deste trabalho se dão em dados sintéticos e a aproximação por uma lei de potências não tem tanta representatividade.

A Função de Correlação de dois pontos depende das distâncias relativas entre os pontos comparadas com uma distribuição homogênea, então é de se esperar que ao mudar a localização das distribuições sem alterar a distância relativa entre as distribuições, não alteramos a função de correlação. Realizamos alguns testes variando a centralidade afim de testarmos essa proposição.

Foram geradas duas distribuições gaussianas com mil pontos cada e desvio padrão, $\sigma = 0,5 \text{ u.m.}$. As distribuições têm todos os parâmetros iguais, exceto a posição, pois uma está centrada nas coordenadas (2,2,2) e a outra está centrada nas coordenadas (5,5,5) (Figura 15 no topo à esquerda). Ao mover as distribuições para um novo centro, (3,3,3) e (6,6,6) (Figura 15 no topo à direita) mantemos a distância relativa entre elas constante, equivalente a 3

u.m. Caso a FC2P seja calculada para as mesmas nuvens, apenas mudando a centralização, é obtido, como esperado, exatamente a mesma FC2P, pois a função utiliza apenas a distância relativa entre as nuvens. A FC2P suavizada por 51 polinômios de terceiro grau pode ser vista na Figura 15 em baixo.

FC2P de nuvens gaussianas com sobreposição de picos

Uma característica interessante de ser observada é até onde a FC2P consegue distinguir a separação das distribuições. Para tal análise foram geradas distribuições com duas nuvens gaussianas com mil pontos cada, desvio padrão $\sigma = 5 \text{ u.m.}$, centradas em pontos diferentes (de maneira análoga às distribuições anteriores), mantendo o desvio padrão constante. As nuvens foram aproximadas e em cada passo de aproximação foi feito o cálculo da FC2P, até o limite que não era mais possível identificar o pico referente ao afastamento entre as nuvens. O centro das distribuições têm coordenadas z e y em 50 u.m. e apenas a coordenada x varia.

As Figura 16, 17 e 18 mostram as distribuições de nuvens gaussianas em aproximação, o correspondente histograma de distâncias e a FC2P respectivamente. Na Figura 16, as mostramos quatro distribuições com duas nuvens gaussianas onde a distância entre os centros das nuvens é reduzida em 10 u.m. a cada passo. Nos histogramas da Figura 17 podemos identificar os picos das distâncias características das distribuições, sendo que o segundo pico referente ao afastamento das nuvens no primeiro gráfico é facilmente distinguível, já no último os picos estão totalmente sobrepostos. No caso da FC2P observamos que na Figura 18, já no terceiro gráfico a informação sobre o segundo pico já esta totalmente perdida, quando no histograma da Figura 17 ainda podemos observá-lo. A FC2P perde informação em relação ao histograma que a forma. O limite no qual a função de correlação não perde informação sobre picos consecutivos é equivalente a quatro desvios padrões da distribuição. No caso representado, as nuvens com um desvio padrão de $\sigma = 5 \text{ u.m.}$, tiveram os picos sobrepostos na FC2P quando a distância entre as nuvens chegou a 20 u.m. , quatro vezes o desvio padrão de cada nuvem.

Para melhor entender como a FC2P varia foi feito um teste onde a densidade de nuvens. As nuvens se aproximam e o número de nuvens é aumentado em cada etapa, partimos de uma amostra com 8 nuvens gaussianas como na Figura 11, na seguinte aumentamos para 27 e por fim 64 nuvens gaussianas. Cada nuvem tem quinhentos pontos, o desvio padrão das nuvens é constante $\sigma = 5 \text{ u.m.}$ A Figura 19 mostra a comparação entre o histograma das distribuições

de 8, 27 e 64 nuvens, com um histograma gerado por uma distribuição homogênea gerado por um PPP Homogêneo. É possível perceber que ao aumentar o número de nuvens e diminuir as distâncias entre as nuvens os picos começam a se aproximar e diminuir a amplitude. Nas figuras abaixo temos os histogramas renormalizados e comparados com a distribuição normal homogênea gerada pelo PPP. No primeiro gráfico da Figura 19 podemos ver que ao aumentar o número de nuvens e diminuir as distâncias entre elas, as distâncias começam a se distribuir de maneira mais homogênea, diminuindo a amplitude dos picos e os aproximando. Os histogramas das amostras começam a tender para o histograma homogêneo.

Na Figura 20 podemos ver que a FC2P também começa a tender a zero quando é aumentada a densidade de nuvens. No limite em que a distância entre os picos é da ordem de quatro desvios padrão das nuvens, a FC2P se degenera totalmente, pois todos os picos ficam sobrepostos. Este resultado já é esperado, pois ao gerar diversas distribuições aleatórias e deixá-las próximas temos um processo muito similar ao PPP Homogêneo explicado anteriormente.

FC2P de amostras com nuvens de diferentes distribuições

Realizamos testes com diferentes distribuições para determinar as principais características das distribuições que influenciam a FC2P. Testamos diferentes distribuições de pontos em cada nuvem. Em cada teste foi alterado a lei de distribuição de uma das nuvens.

Foram mantidos constantes o número de pontos, o centro das nuvens e o desvio padrão (σ) de cada nuvem. A parte relevante em manter o centro das nuvens constantes é que a distância entre os centros se manteve constante e não o centro em si, como estudado anteriormente. Foram selecionados alguns dos testes que representam os resultados para serem mostrados nesse trabalho. Os testes apresentados foram feitos com a comparação entre uma nuvem Gaussiana, Equação 8, e uma caixa Homogênea, gerada pelo PPP. As nuvens foram reescaladas de forma que o seu desvio padrão total fosse de $\sigma = 5 \text{ u.m.}$, e foram representadas na Figura 21 à esquerda. Uma nuvem gaussiana e uma caixa formada por uma lei de distribuição Beta - de maneira análoga à caixa Homogênea, foi gerada uma distribuição Beta (Equação 9 com $\alpha = \beta = 0.3$ e então a distribuição foi reescalada para que seu desvio padrão total fosse $\sigma = 5 \text{ u.m.}$ - representadas pela Figura 21 no meio, e duas nuvens Gaussianas representadas pela Figura 21 à direita. O desvio padrão de todas as nuvens é 5 u.m. e a distância entre o centro das nuvens é 50 u.m.

A FC2P para cada tipo de distribuição foi calculada a partir da média de 10 FC2Ps, de forma análoga ao feito em Einasto et al., 1997, ou seja, as curvas de correlação são na verdade a média de várias. São esperados dois picos principais em escalas diferentes. Um pico correspondente às distâncias internas de cada nuvem e um pico correspondente às distâncias entre as nuvens. As FC2Ps encontradas foram as ilustradas na Figura 22.

A Tabela 3 indica a posição dos principais picos de cada distribuição. A função de correlação de duas gaussianas apresentou três picos principais. O que foi muito similar ao encontrado na função de correlação aplicada às distribuições compostas por uma gaussiana e um quadrado homogêneo. A escala de distâncias do terceiro pico é correspondente ao afastamento entre as nuvens.

Na Figura 22 podemos ver que a primeira oscilação é completamente diferente para as três curvas, no entanto esta região não serve para nos garantir que a FC2P consegue diferenciar distribuições diferentes. Como estudado na seção a primeira oscilação da FC2P é extremamente instável e varia muito mesmo para distribuições iguais.

Em regiões mais estáveis a análise é mais confiável. Podemos ver na Figura 22 que o comportamento das FC2P é similar na escala do segundo pico, mas a FC2P da distribuição beta apresenta oscilações. No entanto as outras duas amostras são extremamente similares. Podemos ver na Figura 23 duas curvas coloridas e uma curva pontilhada feita pela média da FC2P gerada por dez distribuições com a caixa homogênea. As linhas tracejadas superiores e inferiores são calculadas a partir do desvio padrão da média de tais curvas. É possível ver que ambas as curvas estão dentro dos limites do desvio padrão da curva média tracejada. Embora as distribuições sejam visualmente diferentes, a FC2P não consegue diferenciar leis de distribuição diferentes. Este resultado não é novo, como pode ser visto na seção 8.2.7 de Starck e Murtagh, 2005, onde distribuições foram geradas por processos totalmente diferentes e apresentaram FC2P muito similares. Como visto na Figura 22 o teste com a distribuição beta foi o mais diferente dos três resultados mostrados. Este teste apresentou nove picos (Tabela 3). O segundo pico de Beta apresentado na tabela, está bem saliente na Figura 22 e está em 11,675. Este pico é da ordem de escala do afastamento dos pontos da distribuição Beta. Ao criar o vazio no meio da distribuição, foram criadas mais distâncias da ordem de grandeza do vazio, o que aumentou a contagem na FC2P. As oscilações próximas ao terceiro pico das outras duas funções foram criadas pelo mesmo motivo, ao diminuir as distâncias no centro e aumentar nas bordas as regiões correspondentes à distância

entre as nuvens e essas distâncias modificadas foram afetadas. É possível criar distribuições que apresentam FC2P similares, no entanto se as distribuições apresentarem variações de densidades de pontos grandes o suficiente (como o vazio criado na caixa de distribuição Beta) estas podem ser detectadas pela FC2P.

Conclusão

A FC2P aplicada em sua forma espacial é uma ferramenta estatística utilizada para determinar o grau de homogeneidade do universo ao comparar a frequência de distâncias de amostras com a frequência de distâncias de uma distribuição homogênea e, assim, verificar se e em que escala o Princípio Cosmológico poderia ser aplicado. Tal ferramenta vem sendo utilizada até hoje e tem uma influência muito grande no estudo de distribuições em grandes escalas.

A FC2P consegue determinar escalas características de aglomeração através da posição dos picos. Os testes com distribuições mostraram que, utilizando apenas as distâncias relativas, a FC2P independe da posição das distribuições. Testes com diferentes nuvens mostraram que a FC2P é degenerada, pois não consegue diferenciar certas distribuições. Mas caso haja uma anomalia maior na distribuição de pontos, como o vazio gerado no centro da caixa de distribuição Beta, é possível identificar variações, que não determinam o tipo de distribuição analisado, sempre tornando possível que se crie uma distribuição de pontos que gere a mesma FC2P. Tal estudo mostra que diferentes distribuições podem produzir estatisticamente os mesmos resultados na FC2P, o que está de acordo com a literatura, como pode ser visto na seção 8.2.7 de Starck e Murtagh, 2005.

Os testes com nuvens Gaussianas nos mostraram que há uma sobreposição de picos na FC2P quando a escala das distâncias é da ordem de quatro vezes o desvio padrão, σ da distribuição. No caso mostrado, quando as nuvens chegaram a uma distância próxima equivalente a quatro vezes o desvio padrão da distribuição de pontos de cada nuvem, foi possível detectar apenas um pico na função de correlação. Este resultado é importante pois, no histograma de distribuição de distâncias das amostras é possível ainda identificar os diferentes picos que representavam as duas distribuições, enquanto a função de correlação identificava apenas um pico que poderia ser referente a uma mesma distribuição com um desvio padrão maior. Ou seja, a função de correlação perde informações em relação ao histograma. Ao variar a densidade de nuvens é possível observar como os histogramas e, também, as FC2Ps se aproximam do esperado por uma distribuição homogênea quando as nuvens começam a

ficar mais numerosas e próximas. Tal resultado é esperado, pois para um número muito grande de nuvens teríamos um processo análogo ao que gera a distribuição homogênea, o PPP Homogêneo.

Tais estudos nos mostram que a FC2P não falha ao efetuar a função para qual foi originalmente implementada: encontrar a escala para qual uma distribuição pode ser considerada homogênea. Mas ao tentar detectar estruturas mais finas e de maneira mais precisa tem suas limitações, embora ainda consiga identificar escalas de distâncias características das distribuições - se respeitados os limites nos quais os picos se sobrepõem - podendo servir para parametrizar distâncias características em métodos não paramétricos mais precisos.

Agradecimentos

E.G. agradece a CAPES - programa PNPd: Este estudo foi financiado em parte pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Referências

- Abbott, T. M. C., Allam, S., Andersen, P., Angus, C., Asorey, J., Avelino, A., Avila, S., Bassett, B. A., Bechtol, K., Bernstein, G. M., Bertin, E., Brooks, D., Brout, D., Brown, P., Burke, D. L., Calcino, J., Carnero Rosell, A., Carollo, D., Carrasco Kind, M., . . . DES Collaboration. (2019). First Cosmology Results using Type Ia Supernovae from the Dark Energy Survey: Constraints on Cosmological Parameters. *Astrophysical Journal*, 872(2), **journal article** L30, L30. <https://doi.org/10.3847/2041-8213/ab04fa>
- Abell, G. O. (1958). The Distribution of Rich Clusters of Galaxies. *apjs*, 3, 211. <https://doi.org/10.1086/190036>
- Einasto, J., Einasto, M., Frisch, P., Gottlober, S., Muller, V., Saar, V., Starobinsky, A. A. & Tucker, D. (1997). The supercluster-void network - III. The correlation function as a geometrical statistic. *mnras*, 289, 813–823. <https://doi.org/10.1093/mnras/289.4.813>
- Einasto, J., Saar, E. & Klypin, A. A. (1986). Structure of superclusters and supercluster formation. V - Spatial correlation and voids. *mnras*, 219, 457–478. <https://doi.org/10.1093/mnras/219.3.457>
- Einstein, A. (1905). The foundation of the theory of relativity. *The Principle of Relativity. Dover Books on Physics. June 1* (pp. 109–164).
- Einstein, A. (1915). The foundation of the general theory of relativity. *The Principle of Relativity. Dover Books on Physics. June 1* (pp. 129–174).
- Gabrielli, A., Sylos Labini, F., Joyce, M. & Pietronero, L. (2005). *Statistical Physics for Cosmic Structures*. Springer.
- Groth, E. J. & Peebles, P. J. E. (1977). Statistical analysis of catalogs of extragalactic objects. VII - Two- and three-point correlation functions for the high-resolution Shane-Wirtanen catalog of galaxies. *apj*, 217, 385–405. <https://doi.org/10.1086/155588>
- Hamilton, A. J. S. (1993). Toward Better Ways to Measure the Galaxy Correlation Function. *apj*, 417, 19. <https://doi.org/10.1086/173288>
- Hubble, E. P. (1926). Extragalactic nebulae. *apj*, 64. <https://doi.org/10.1086/143018>
- Hubble, E. (1929). A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae. *Proceedings of the National Academy of Science*, 15(3), 168–173. <https://doi.org/10.1073/pnas.15.3.168>
- Martínez, V. J. (2001). *Statistical of the Galaxy Distribution*. CHAPMAN; HALL/CRC.
- O’Raifeartaigh, C., O’Keeffe, M., Nahm, W. & Mitton, S. (2018). One hundred years of the cosmological constant: from “superfluous stunt” to dark energy. *European Physical Journal H*, 43(1). <https://doi.org/10.1140/epjh/e2017-80061-7>
- Shane, C. D. & Wirtanen, C. A. (1954). The distribution of extragalactic nebulae. *aj*, 59, 285–304. <https://doi.org/10.1086/107014>
- Slipher, V. M. (1913). The radial velocity of the Andromeda Nebula. *Lowell Observatory Bulletin*, 1, 56–57.
- Starck, J. & Murtagh, F. (2005). *Astronomical Image and Data Analysis* (Second). Springer.
- Stoughton, C. & et, al. (2002). Sloan Digital Sky Survey: Early Data Release. *The Astronomical Journal*, 123, 485–548. <https://doi.org/10.1086/324741>
- Totsuji, H. & Kihara, T. (1969). The Correlation Function for the Distribution of Galaxies. *pasj*, 21, 221.

Pares	Distâncias (u.m.)
1 e 2	$\frac{1}{\sqrt{2}}$
1 e 3	$\frac{1}{2}$
1 e 4	1
2 e 3	$\frac{1}{\sqrt{2}}$
2 e 4	$\frac{1}{2}$
3 e 4	1

Tabela 1

Distâncias entre pares de ponto

Picos (u.m.)	P1	P2	P3	P4	P5
Histograma	4,78	50,04	71,74	87,67	
Correlação	0,25	9,15	50,04	71,74	87,67

Tabela 2

Picos de distribuição 8 Nuvens: Histograma e Correlação

Picos (u.m.)	P1	P2	P3
Gaussianas	0,675	3,875	50,125
Homogênea	0,775	3,725	49,625
Beta	1,925	11,675	42,975

Tabela 3

Picos de distribuição diferentes: Duas nuvens gaussianas, nuvem gaussiana e nuvem cúbica homogênea, nuvem gaussiana e nuvem beta

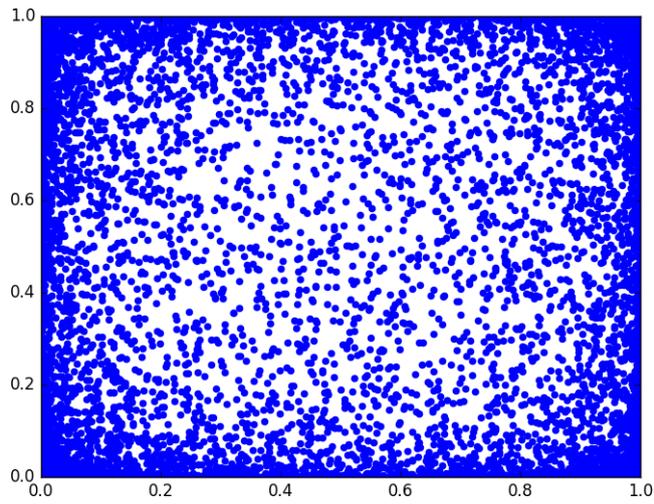


Figura 7

Exemplo Tridimensional de Nuvem Beta, com 10 mil pontos, projetada Bidimensionalmente. Neste caso podemos ver uma concentração maior de pontos nas bordas, o que pode ser utilizado para estudar propriedades de vazios.

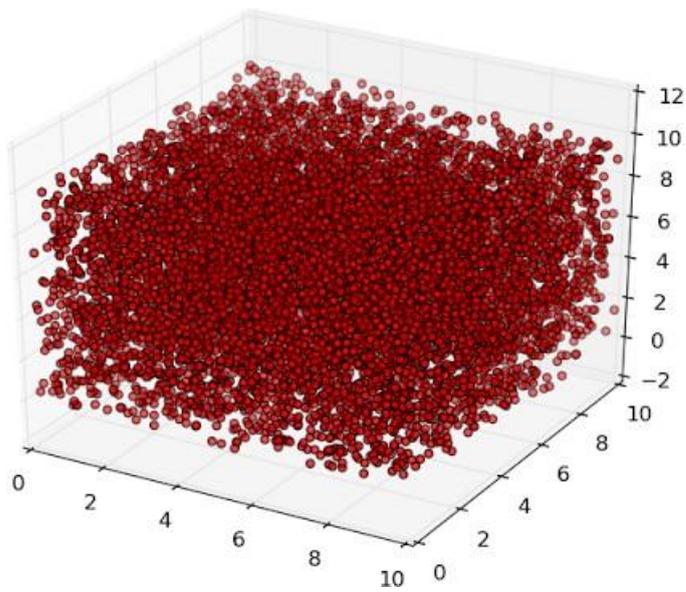


Figura 8

Amostra gerada por um Processo Pontual de Poisson Homogêneo contendo 10 mil pontos

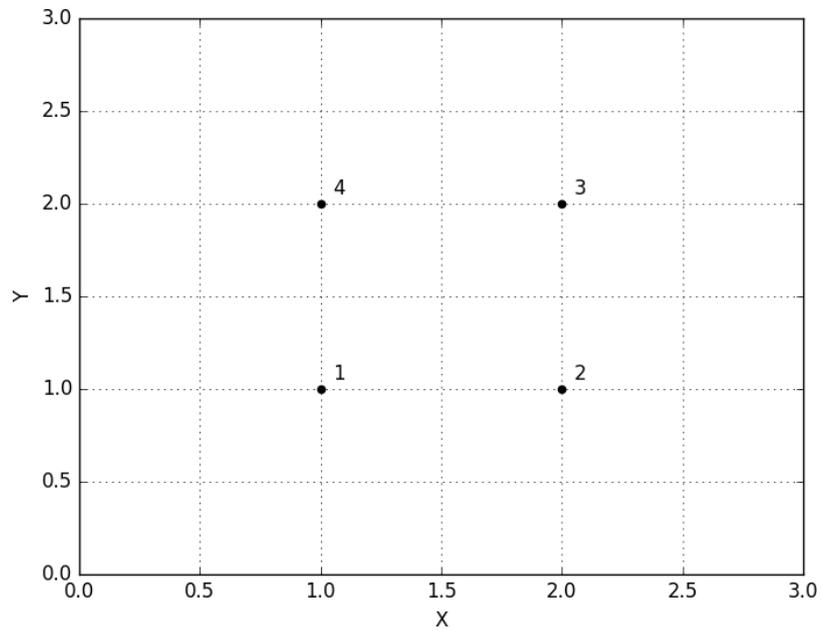


Figura 9
Exemplo de Distribuição de Pontos.

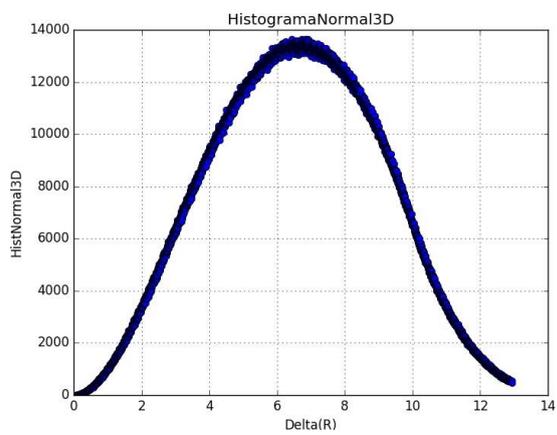


Figura 10
Exemplo de Distribuição de Frequências de Distâncias para 10 mil Pontos.

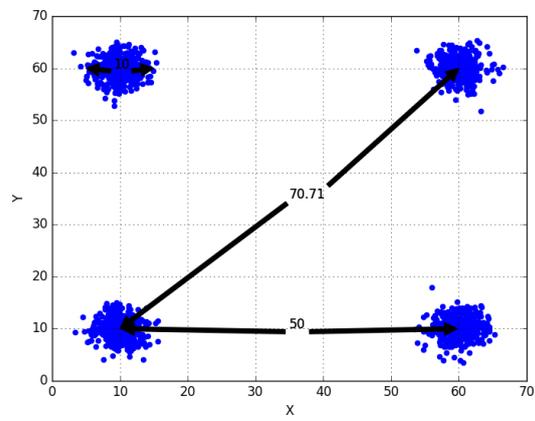
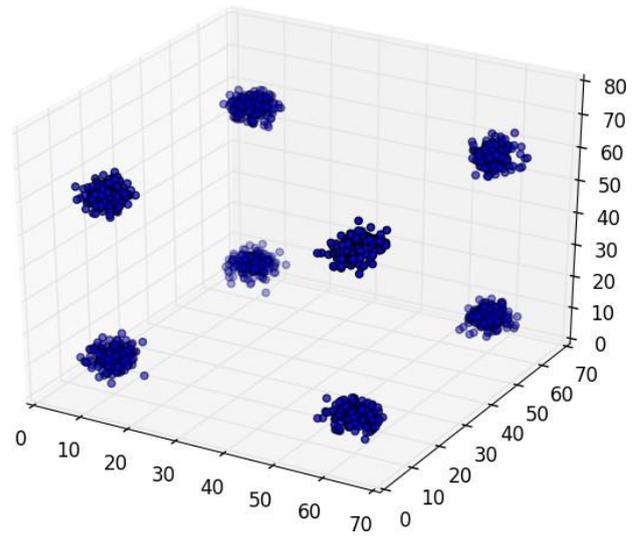


Figura 11
Gráfico tri-dimensional das oito nuvens gaussianas e sua projeção bi-dimensional, respectivamente

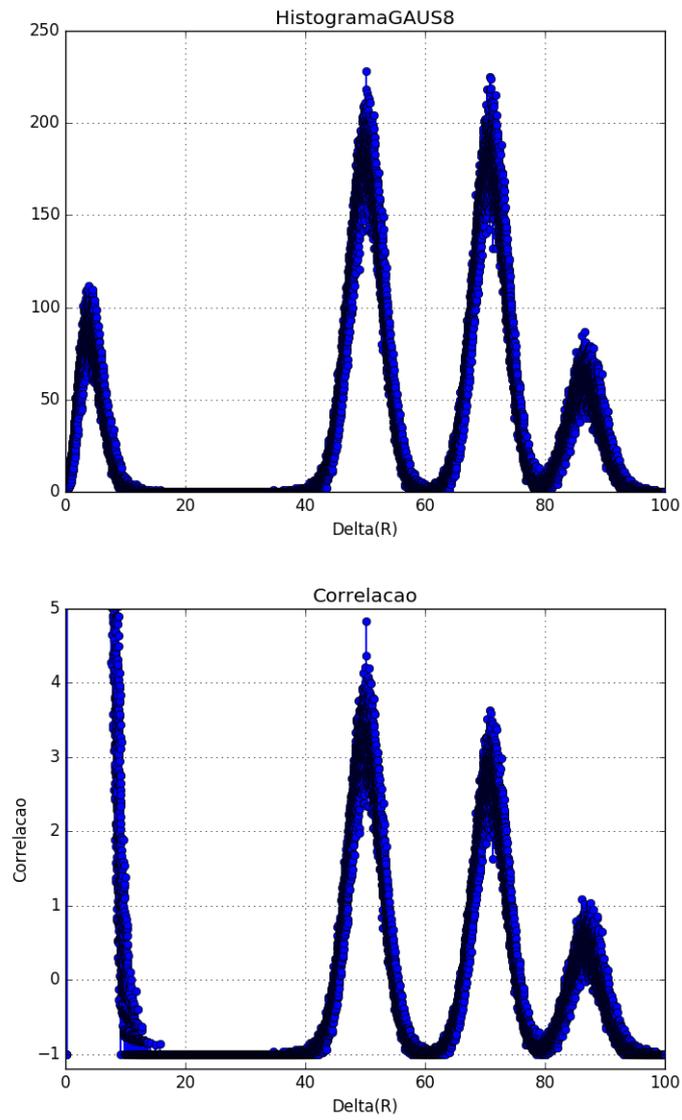


Figura 12

Histograma distribuição de distâncias das oito nuvens gaussianas e FC2P da mesma distribuição, respectivamente

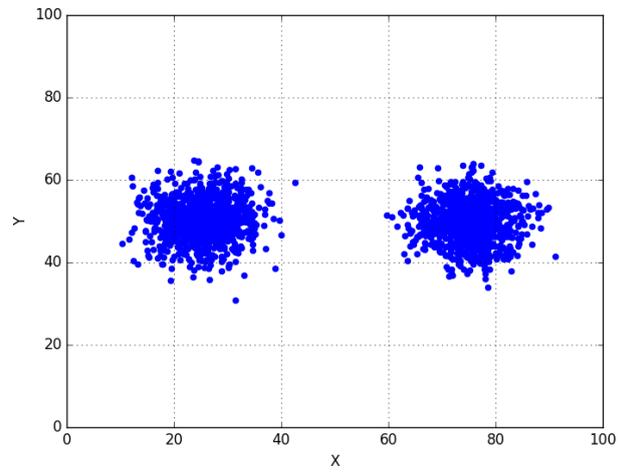


Figura 13

Visualização das distribuições Iguais. As nuvens foram projetadas bidimensionalmente para melhor visualização.

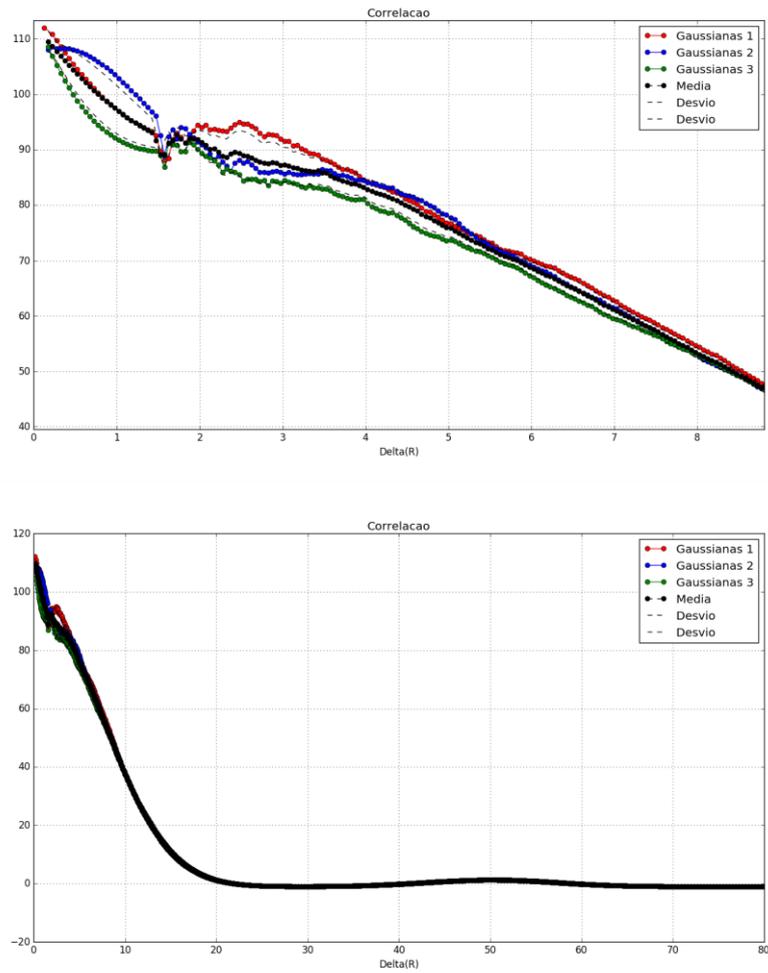


Figura 14

No topo: Correlações médias com filtro de suavidade. As curvas coloridas representam a média de 10 FC2P, a curva preta é a média das 3 curvas médias. A linhas tracejadas são calculadas a partir do desvio padrão de 30 FC2P de distribuições estatisticamente iguais. A variação do primeiro pico é grande mesmo em uma função suavizada e pode ficar fora do desvio padrão das curvas. Em baixo: Correlações médias. O primeiro pico se mostra completamente instável, mas a partir de certa escala de distância as curvas da FC2P se tornam indistinguíveis.

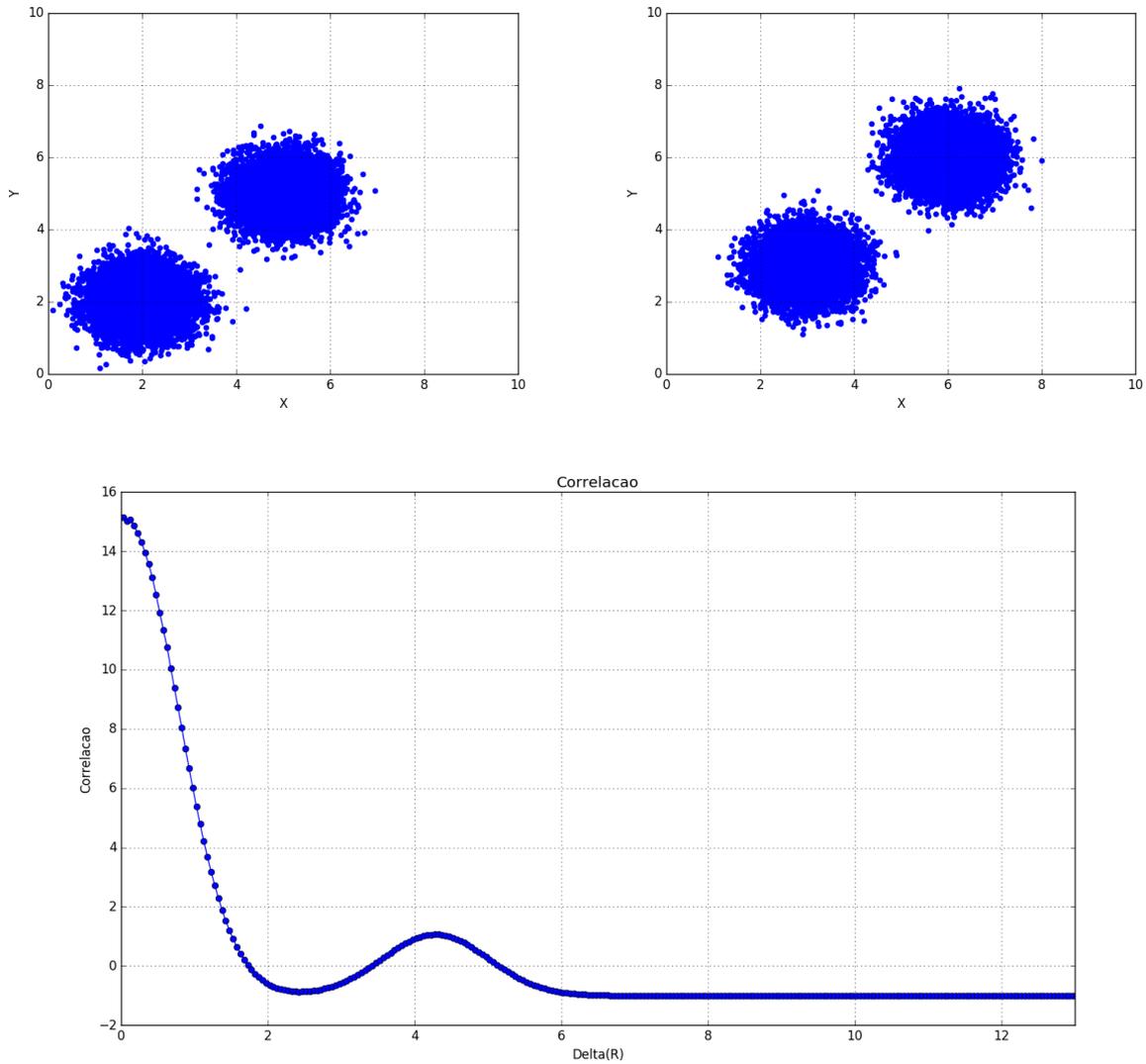


Figura 15

A nuvens estão centradas em (2,2,2) e (5,5,5) e contém mil pontos cada, $\sigma = 0,5$ u.m. em relação ao centro de cada nuvem. A nuvens estão centradas em (3,3,3) e (6,6,6) e contém mil pontos cada, $\sigma = 0,5$ u.m. em relação ao centro de cada nuvem. Função de Correlação de dois pontos suavizada encontrada para as nuvens centradas em 2 e 5 e nuvens centradas em 3 e 6. A amostra é a mesma, a única diferença é o centro deslocado.

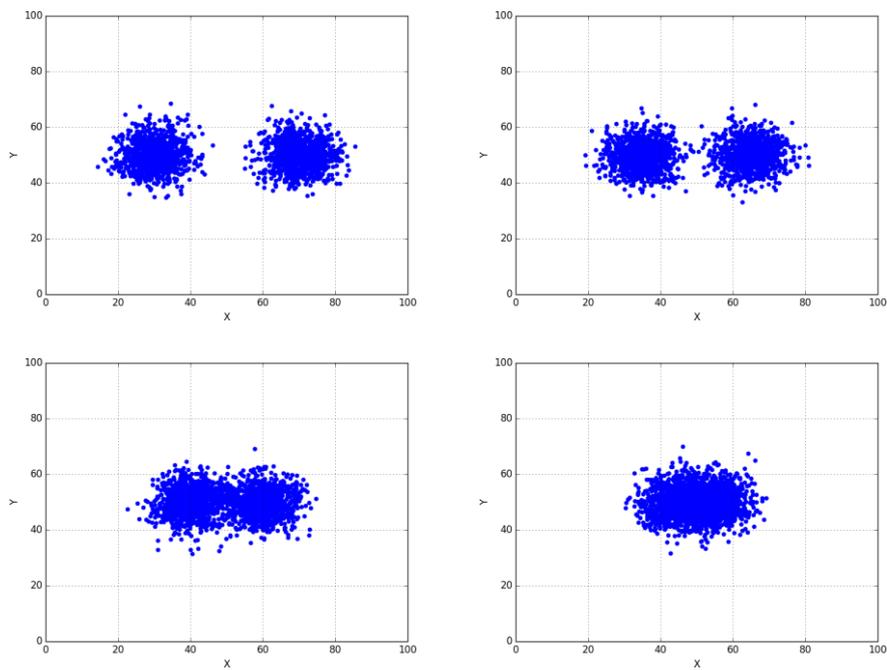


Figura 16

Distribuições centradas em $x = 30$ u.m. e $x = 70$ u.m., em $x = 35$ u.m. e $x = 65$ u.m., em $x = 40$ u.m. e $x = 60$ u.m., e em $x = 45$ u.m. e $x = 55$ u.m.

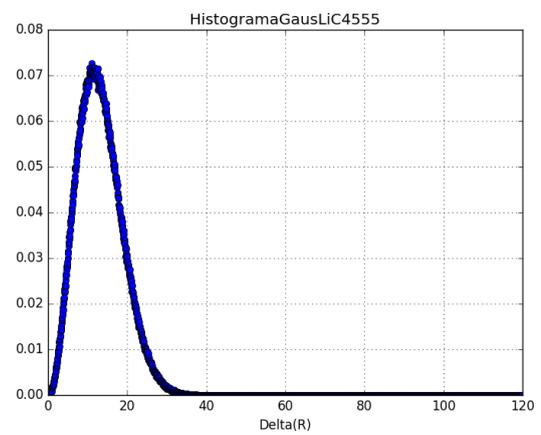
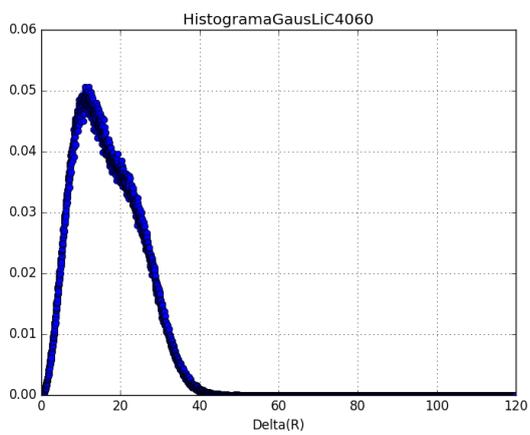
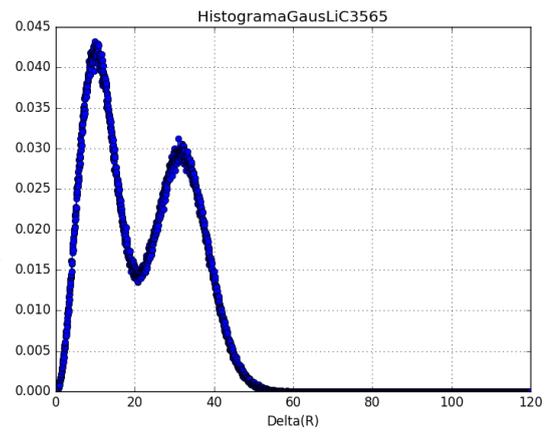
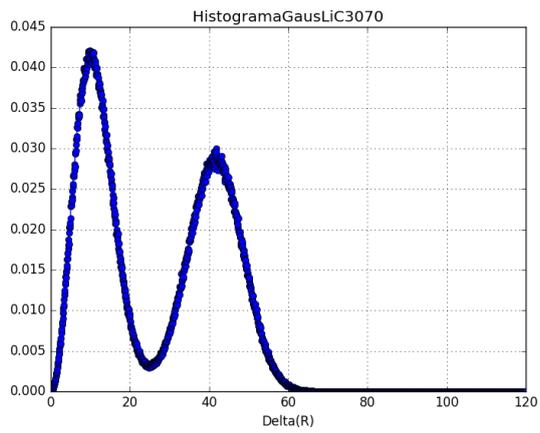


Figura 17

Histograma de nuvens em $x = 30$ u.m. e $x = 70$ u.m., em $x = 35$ u.m. e $x = 65$ u.m., em $x = 40$ u.m. e $x = 60$ u.m. e em $x = 45$ u.m. e $x = 55$ u.m.

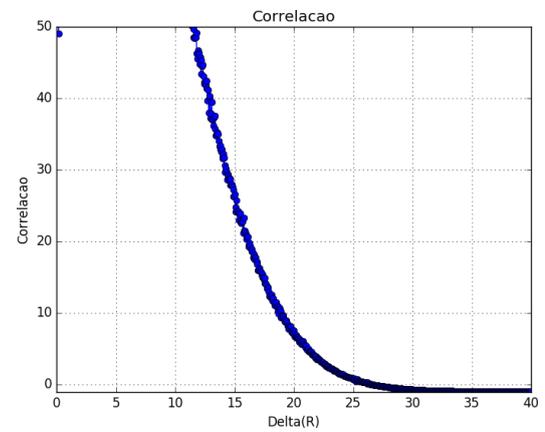
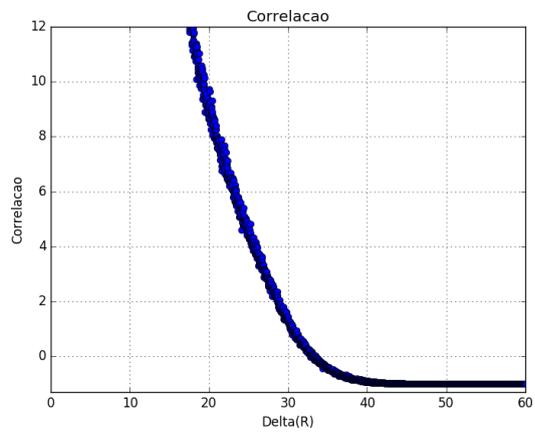
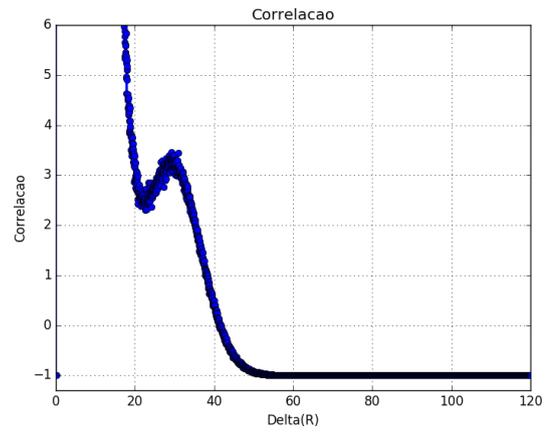
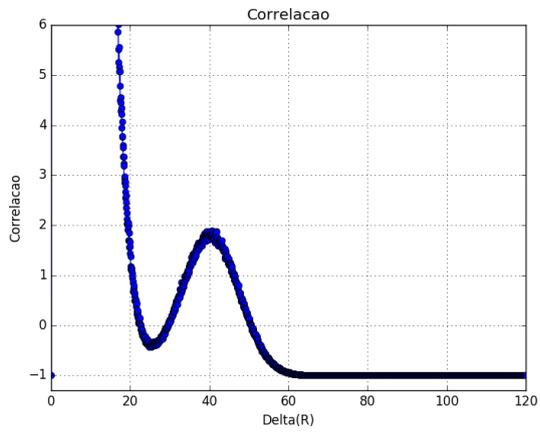


Figura 18

Correlação de nuvens em $x = 30$ u.m. e $x = 70$ u.m., em $x = 35$ u.m. e $x = 65$ u.m., em $x = 40$ u.m. e $x = 60$ u.m. e em $x = 45$ u.m. e $x = 55$ u.m.

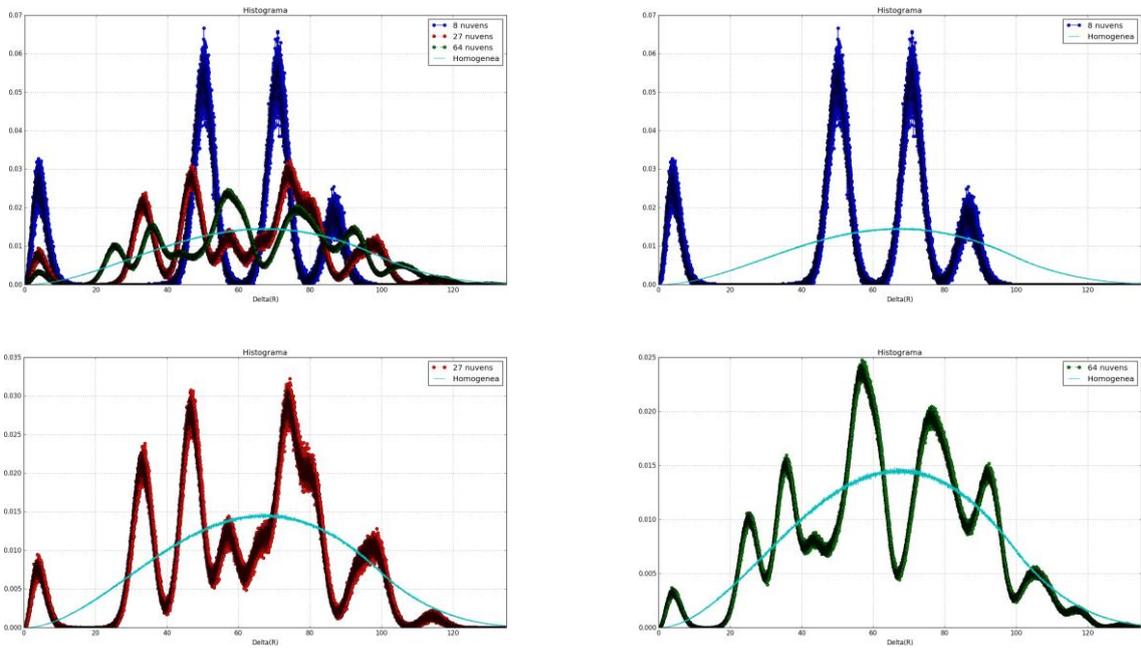


Figura 19
Histograma renormalizado e comparado com histograma homogêneo: 8 nuvens, com histograma homogêneo: 27 nuvens e com histograma homogêneo: 64 nuvens

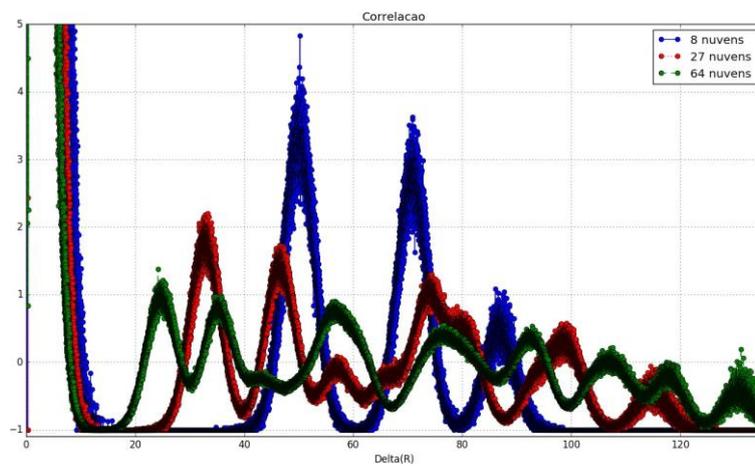


Figura 20
Composição de FC2P de 8 nuvens, 27 nuvens e 64 nuvens

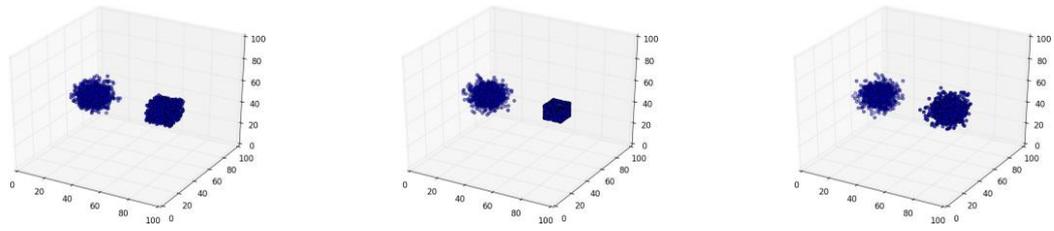


Figura 21

Nuvens Gaussiana e Caixa Homogênea; Nuvens Gaussiana e Caixa Beta; Nuvens Gaussianas

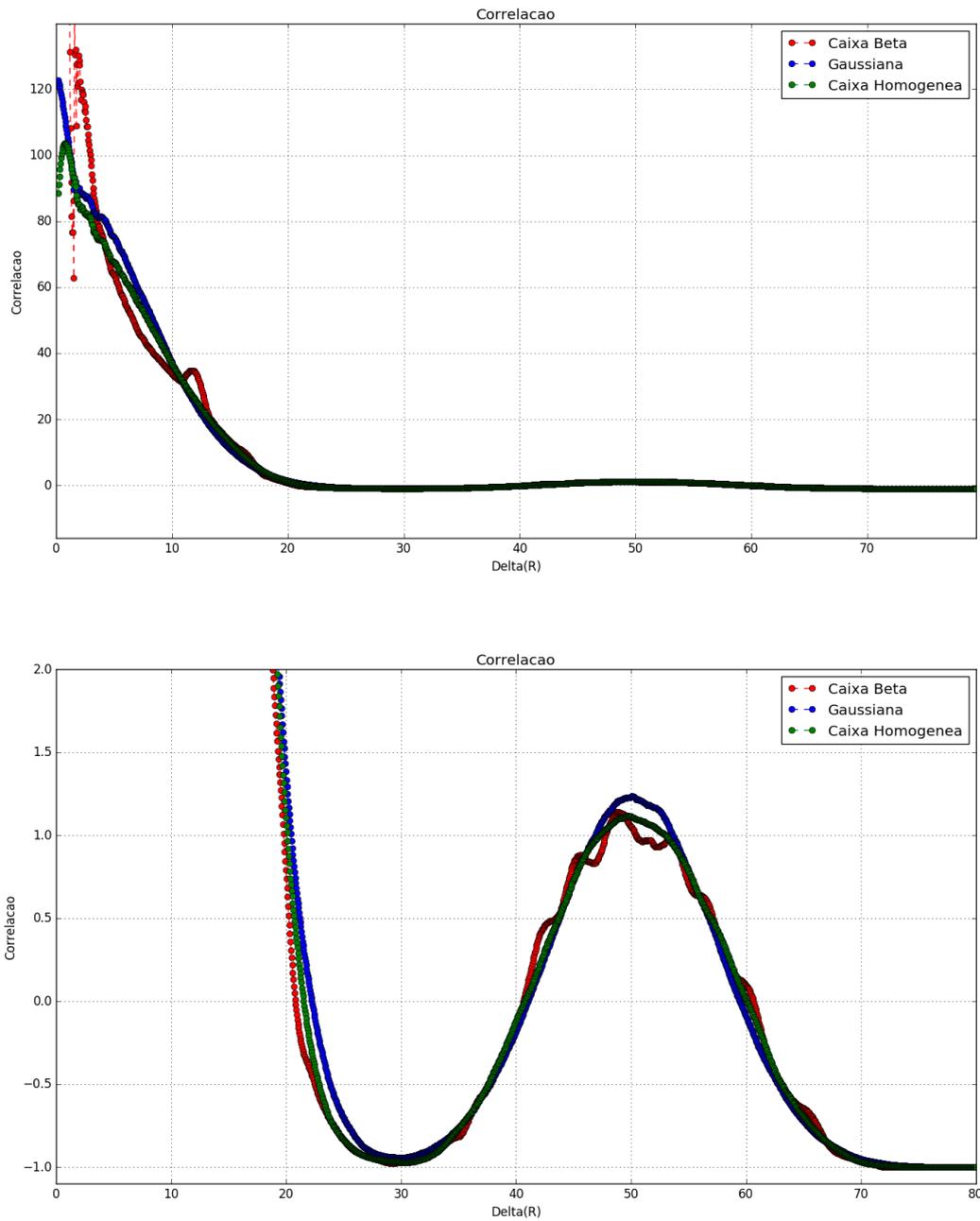


Figura 22
Comparação Correlação Distribuições Distintas; Comparação Correlação Distribuições: Escala do Segundo pico.

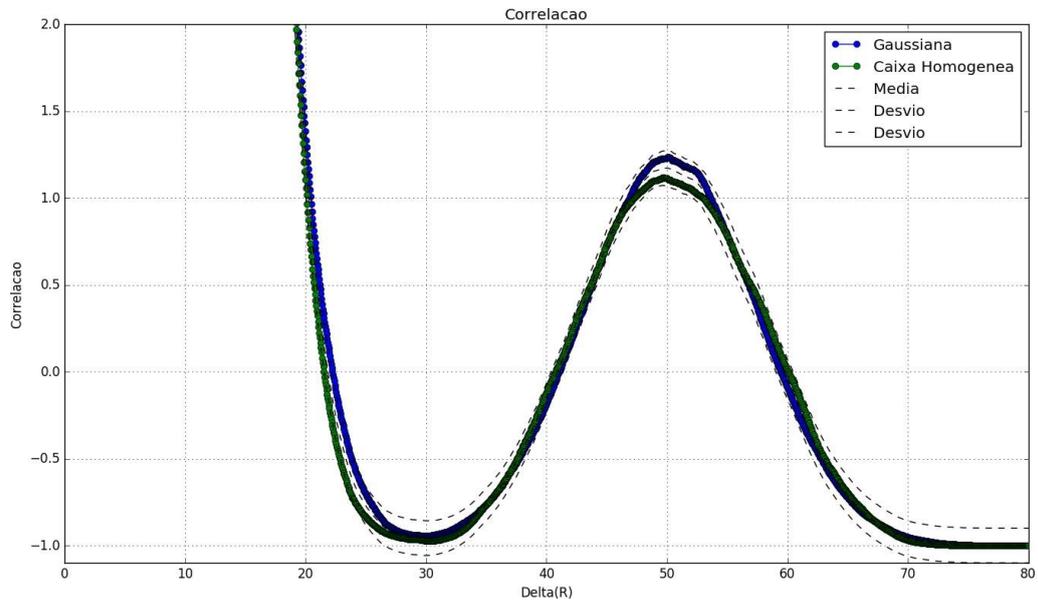


Figura 23
Comparação Correlação Distribuições: Com desvios no terceiro pico