

MAE610 - Tópicos de Estatística / Teoria do Risco

*Agregação de portfólios de risco
correlacionados*

Seções 3 e 4

Fábio Rampazzo Mathias

Fernando Henrique Ferraz P. da Rosa

Vagner Aparecido Pedro Junior

IME-USP

13/05/2005

Introdução

Nesse trabalho descrevemos as seções 3 e 4 do artigo de Wang (1998), que trata sobre modelos de agregação de portfólios de risco correlacionados. A seção 3 descreve os modelos de perda agregada e a aplicação do método da transformação rápida de Fourier (FFT) para o cálculo de quantidades envolvendo essa perda. Também são descritos métodos de discretização de variáveis contínuas e um algoritmo para cálculo da distribuição de perda agregada.

Na seção 4 faz-se uma revisão de algumas ferramentas para distribuições de probabilidades uni- e multivariadas, como as funções característica, geradora de momentos e de probabilidade. Discute-se ainda um teorema para agregação de variáveis correlacionadas.

Modelos de agregação de perda e o método FFT

Para avaliação de perdas com seguro para um conjunto de negócios, a abordagem de frequência de ocorrências/sinistros é o método mais flexível, onde a frequência média de ocorrência e o valor médio dos sinistros são utilizados para estimar a perda média. De modo a facilitar uma análise dinâmica do risco subjacente, a distribuição de perda agregada é necessária para quantificar a variabilidade inerente ao custo da perda agregada. Essa seção descreve os modelos agregados básicos e como o método de transformações rápidas de Fourier pode ser utilizado para calcular a distribuição de perda agregada.

Distribuições para o número de sinistros

Em muitas áreas de aplicações no contexto da modelagem de processos aleatórios, a distribuição de Poisson é usualmente considerada como ponto de partida, embora a distribuição de frequência através de uma binomial negativa também seja uma alternativa considerada.

- A distribuição Poisson com média $\lambda > 0$ é definida pela função de probabilidade:

$$P(N = n) = e^{-\lambda} \frac{\lambda^n}{n!}, \quad n = 0, 1, 2, \dots$$

A função geradora de probabilidade é dada por

$$P_N(t) = \mathbf{E}[t^N] = e^{\lambda(t-1)},$$

$$\mathbf{e} \mathbf{E}[N] = \text{Var}[N] = \lambda.$$

Distribuições para o número de sinistros (cont)

- A distribuição Binomial Negativa com parâmetros $\alpha, \beta > 0$, tem função de probabilidade:

$$p_n = P(N = n) = \frac{\Gamma(\alpha + n)}{\Gamma(\alpha)n!} \left(\frac{1}{1 + \beta}\right)^\alpha \left(\frac{\beta}{1 + \beta}\right)^n, \quad n = 0, 1, 2, \dots$$

e função geradora de probabilidade dada por:

$$P_N(t) = [1 - \beta(t - 1)]^{-\alpha},$$

com $E[N] = \alpha\beta$ e $\text{Var}[N] = \alpha\beta(1 + \beta)$. Geralmente, para a distribuição binomial negativa, a variância excede a média. A razão entre a variância e a média é

$$\frac{\text{Var}[N]}{E[N]} = 1 + \beta.$$

Distribuições para o número de sinistros (cont)

Devido a essa relação é comum especificar a distribuição binomial negativa por essa razão da variância pela média; por exemplo através de um parâmetro de contágio c :

$$\text{Var}[N] = E[N](1 + cE[N]).$$

Distribuições para o valor dos sinistros

Diversas distribuições são em geral usadas para os valores dos sinistros. Entre as mais comuns temos:

Distribuição	Ranking
Pareto	1
lognormal	2
exponencial inversa / Gaussiana	3
Gaussiana inversa	4
Weibull	5
gamma	6

Distribuições para o valor dos sinistros (cont)

Se um conjunto de dados grande está disponível, uma distribuição de perda empírica pode ser também utilizada. Assim que escolhida a distribuição do valor dos sinistros, é necessário discretizá-la, afim de que a metodologia a ser apresentada possa ser aplicada de forma computacionalmente eficiente. Para essa discretização considera-se uma unidade monetária h conveniente (US\$ 1000 por exemplo), e uma variável discreta com incrementos de h em h . Consideremos então dois métodos para discretização de uma variável aleatória contínua dessa maneira.

O método do arredondamento

Suponha X uma variável aleatória contínua com distribuição acumulada $F_X(x) = P(X \leq x)$. Escolha um valor h apropriado. Seja f_j a probabilidade no ponto jh , $j = 0, 1, 2, \dots$. Então:

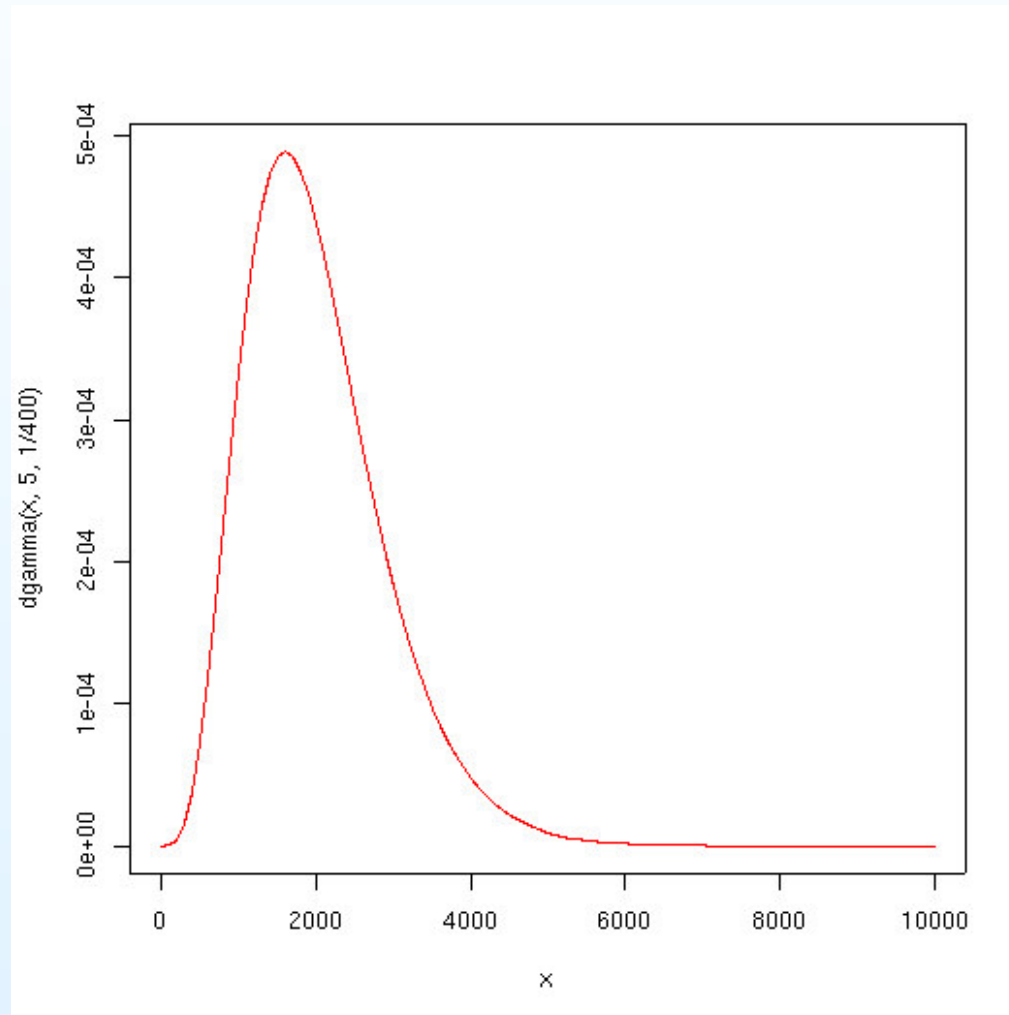
$$f_0 = F_X\left(\frac{h}{2}\right),$$

$$f_j = F_X\left(jh + \frac{h}{2}\right) - F_X\left(jh - \frac{h}{2}\right), \quad j = 1, 2, \dots$$

Enquanto a maior vantagem desse método é a simplicidade, há a desvantagem de não se preservar a média da distribuição contínua.

Exemplo: O método do arredondamento

Considere uma variável aleatória $X \sim \text{Gama}(5, 1/400)$. Na Figura abaixo temos sua densidade.

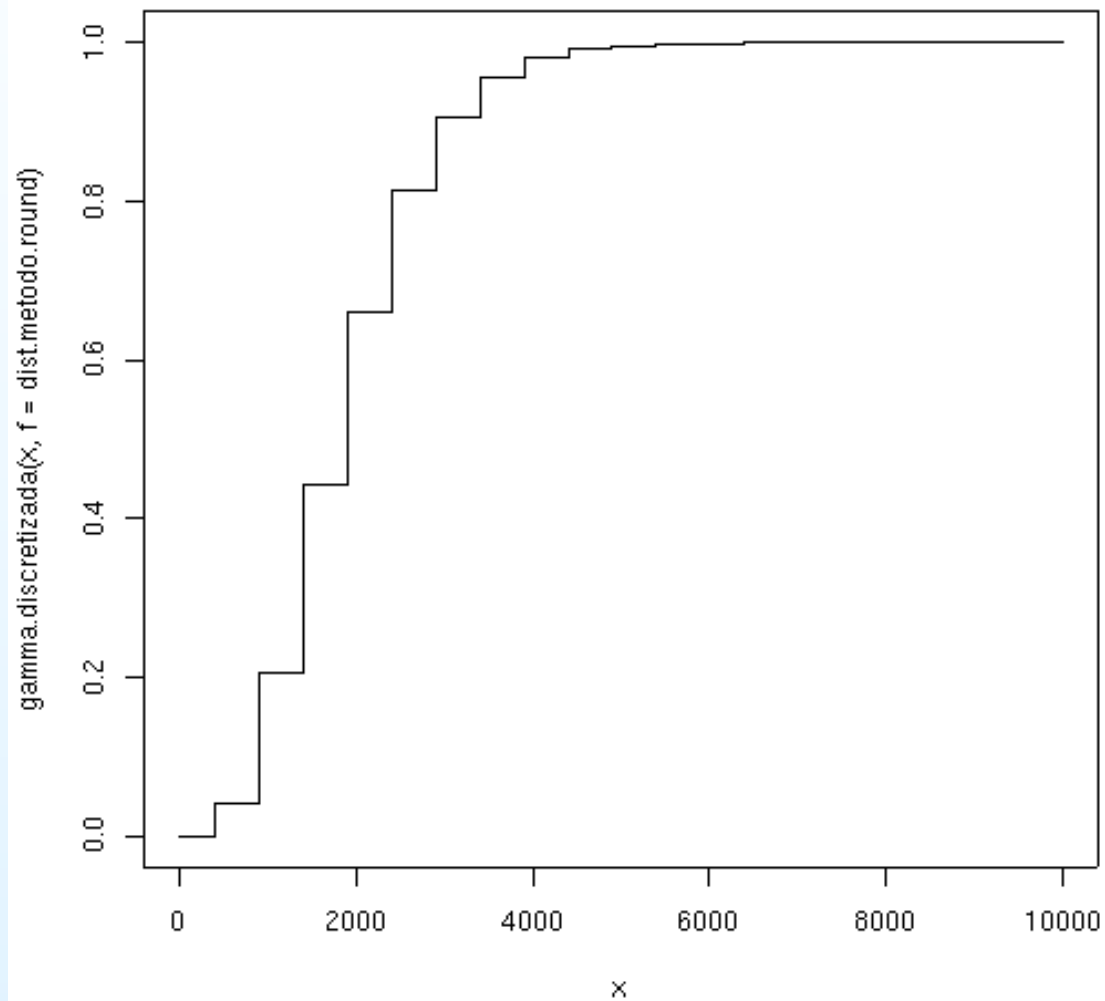


Exemplo: O método do arredondamento (cont)

Temos ainda que $E[X] = 2000$ e $Var(X) = 800000$.
Consideremos $h = 500$. Através do algoritmo do arredondamento obtemos:

```
> dist.metodo.round
      0      500     1000     1500     2000
4.74e-04 4.16e-02 1.64e-01 2.38e-01 2.18e-01
      2500     3000     3500     4000     4500
1.54e-01 9.20e-02 4.91e-02 2.41e-02 1.11e-02
      5000     5500     6000     6500     7000
4.87e-03 2.05e-03 8.34e-04 3.30e-04 1.27e-04
      7500     8000     8500     9000     9500
4.82e-05 1.79e-05 6.54e-06 2.36e-06 8.40e-07
      10000
2.96e-07
```

Exemplo: O método do arredondamento (cont)



Método do emparelhamento de médias

Este método consegue fazer com que a média da distribuição contínua seja preservada.

Com a distribuição contínua com função de distribuição acumulada F_X , primeiramente achamos os valores esperados múltiplos de h :

$$(1) \quad \mathbf{E}[X; j \cdot h] = \int_0^{j \cdot h} [1 - F_X(u)] du, \quad \text{para } j = 1, 2, \dots$$

Então, calculamos o vetor de probabilidade por:

$$f_0 = \Pr\{X = 0 \cdot h\} = 1 - \mathbf{E}[X; h]/h,$$

$$f_j = \Pr\{X = j \cdot h\}$$

$$= (2\mathbf{E}[X; j \cdot h] - \mathbf{E}[X; (j - 1) \cdot h] - \mathbf{E}[X; (j + 1) \cdot h])/h, \quad j = 1$$

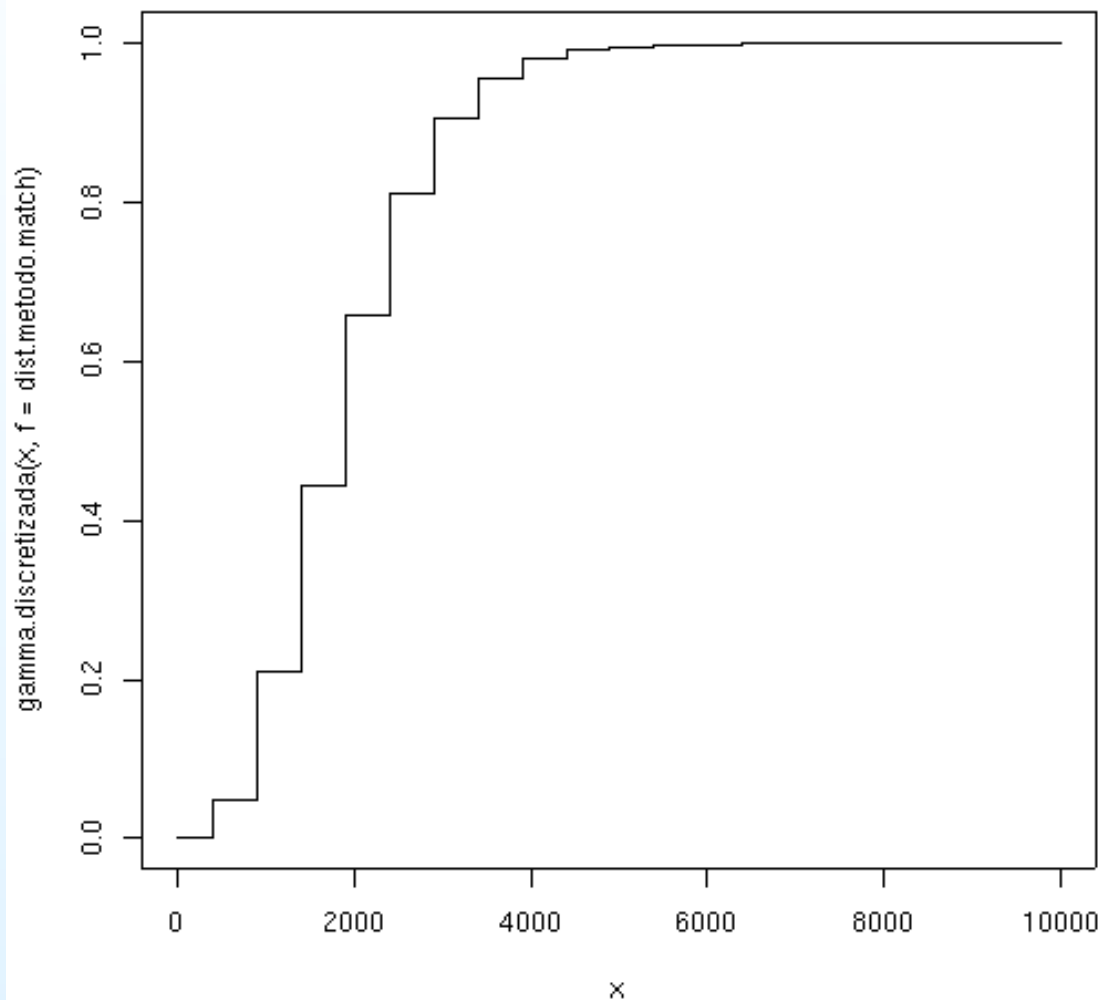
Exemplo: O método do emparelhamento de médias

(cont)

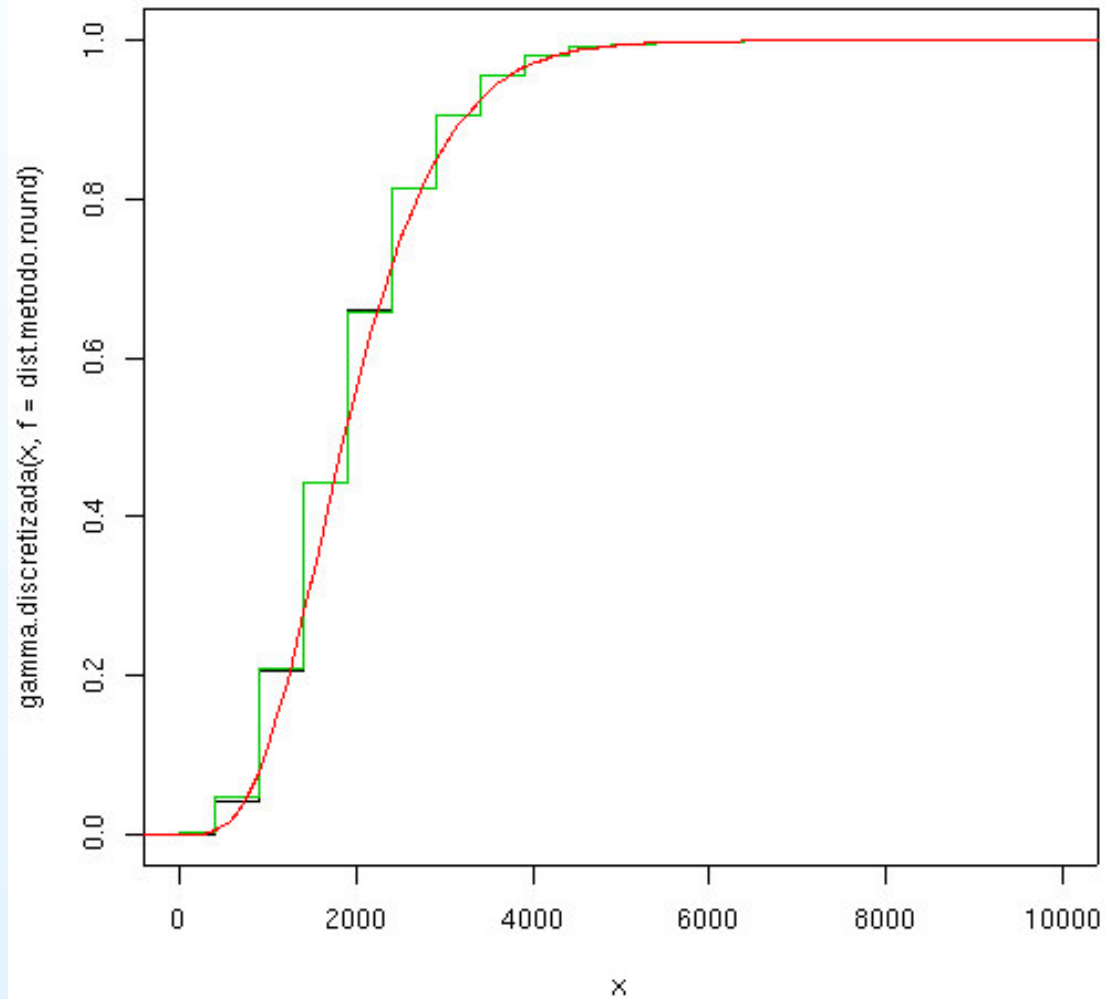
Considerando a mesma X do exercício anterior e $h = 500$, obtemos através desse método a função de distribuição:

```
> dist.metodo.match
      0      500     1000     1500     2000
1.77e-03 4.60e-02 1.62e-01 2.34e-01 2.16e-01
      2500     3000     3500     4000     4500
1.54e-01 9.28e-02 4.99e-02 2.46e-02 1.14e-02
      5000     5500     6000     6500     7000
5.01e-03 2.11e-03 8.62e-04 3.42e-04 1.32e-04
      7500     8000     8500     9000     9500
5.01e-05 1.86e-05 6.82e-06 2.46e-06 8.77e-07
      10000
0.00e+00
```

Exemplo: O método do emparelhamento de médias (cont)



Exemplo: Os dois métodos e a distribuição original



Agregação de freqüência e sinistros

A *perda agregada* é representada por uma soma de aleatória de variáveis aleatórias, Z :

$$Z = X_1 + X_2 + \dots + X_N,$$

onde X_i é o i ésimo sinistro e N é o número (aleatório) de sinistros. Assumindo os sinistros *iid* temos que essa quantidade tem a distribuição de probabilidade:

$$\begin{aligned} f_Z(x) &= P(Z = x) \\ &= \sum_{n=0}^{\infty} P(N = n)P(Z = x|N = n) \\ &= \sum_{n=0}^{\infty} P(N = n)f_X^{*n}(x). \end{aligned}$$

Cálculo da distribuição de perda agregada por FFT

Calculando a função característica de Z , temos:

$$\begin{aligned}\phi_Z(t) &= E[e^{it(Z)}] = E_N[E[e^{it(X_1+\dots+X_N)} | N]] \\ &= E_N[\phi_X(t)^N] = P_N(\phi_X(t)),\end{aligned}$$

onde P_N é a função geradora de probabilidade de N . A partir dessa relação propõe-se o seguinte algoritmo para calcular a distribuição da perda agregada:

Passo 1 Escolha $n = 2^r$ para algum inteiro r . n é o número de pontos desejados no suporte de $f_Z(x)$, ou seja, a distribuição de probabilidade da perda agregada deve ter massa de probabilidade negligível fora do intervalo $[0, n]$. Essa quantidade deve ser escolhida antes de obter a distribuição de perda.

Algoritmo (continuação)

- Passo 2 Discretize a distribuição de probabilidade dos sinistros, usando o método da média emparelhada ou método do arredondamento. Seja $(f_0, f_1, \dots, f_{m-1})$ a função de densidade de probabilidade discreta obtida. Acrescente zeros no final desse vetor até que ele tenha n elementos, obtendo f_X .
- Passo 3 Aplique uma transformação rápida de Fourier (FFT) em f_X , obtendo $\tilde{f}_X = FFT(f_X)$.
- Passo 4 Aplique a função geradora de probabilidade do número de sinistros, elemento a elemento, em $\tilde{f}_X = FFT(f_X)$, obtendo $\tilde{f}_Z = P_N(\tilde{f}_X)$.
- Passo 5 Aplique uma transformação rápida de Fourier inversa (IFFT) em \tilde{f}_Z , obtendo $f_Z = IFFT(\tilde{f}_Z)$.

Exemplo de aplicação

Considere o primeiro exercício da segunda lista de exercícios de MAE610. Temos que $N \sim Po(\lambda = 2)$ e X_i segue a distribuição de probabilidades:

X_i	1	2	3	4
f_k	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{3}{8}$

Pede-se para calcular a distribuição da soma Z , para k de 0 até 20. Na lista, otivemos esses valores pela fórmula de Panjer. O mesmo pode ser feito pelo algoritmo acima, através de transformações de Fourier (FFT).

Passo 1 Escolhemos $n = 20$, desconsiderando a recomendação de $n = 2^r$ para algum r inteiro, pois essa recomendação não afeta os resultados, e é indicada apenas para otimização dos cálculos.

Aplicação (continuação)

Passo 2 Nesse caso a distribuição dos sinistros já é discreta, não é necessário discretizá-la. Acrescentamos somente o número de zeros necessários no final do vetor (os comandos indicados foram executados no R):

```
> Y <- c(0,1/8,1/4,1/4,3/8)
> Y <- c(Y,rep(0,15))
> names(Y) <- 0:19
> Y
      0      1      2      3      4      5      6      7
0.00 0.12 0.25 0.25 0.38 0.00 0.00 0.00
      8      9     10     11     12     13     14     15
0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
     16     17     18     19
0.00 0.00 0.00 0.00
```

Aplicação (continuação)

Passo 3 Aplicamos então a transformação de Fourier a esse vetor:

0	1	2
1.000+0.00i	0.584-0.74i	-0.202-0.77i
3	4	5
-0.545-0.20i	-0.250+0.24i	0.125+0.12i
6	7	8
0.077-0.18i	-0.216-0.16i	-0.250+0.15i
9	10	11
0.052+0.26i	0.250+0.00i	0.052-0.26i
12	13	14
-0.250-0.15i	-0.216+0.16i	0.077+0.18i
15	16	17
0.125-0.12i	-0.250-0.24i	-0.545+0.20i
18	19	
-0.202+0.77i	0.584+0.74i	

Aplicação (continuação)

Passo 4 Notemos que a função geradora de probabilidade de uma variável aleatória Poisson com parâmetro λ é dada por:

$$g_X(s) = e^{-\lambda(1-s)}$$

Aplicando essa função ao vetor transformado com $\lambda = 2$, temos:

```
> fgt.poi <- function(x, lambda) {  
    exp(-lambda*(1-x))  
}  
> ztil <- fgt.poi(fft(Y), lambda=2)
```

Aplicação (continuação)

```
> ztil
      0          1          2
1.000+0.00i 0.036-0.43i 0.003-0.09i
      3          4          5
0.042-0.02i 0.073+0.04i 0.168+0.04i
      6          7          8
0.148-0.06i 0.083-0.03i 0.079+0.02i
      9         10         11
0.130+0.08i 0.223+0.00i 0.130-0.08i
     12         13         14
0.079-0.02i 0.083+0.03i 0.148+0.06i
     15         16         17
0.168-0.04i 0.073-0.04i 0.042+0.02i
     18         19
0.003+0.09i 0.036+0.43i
```

Aplicação (continuação)

Passo 5 Por fim aplicando a transformação inversa em \tilde{f}_Z , temos:

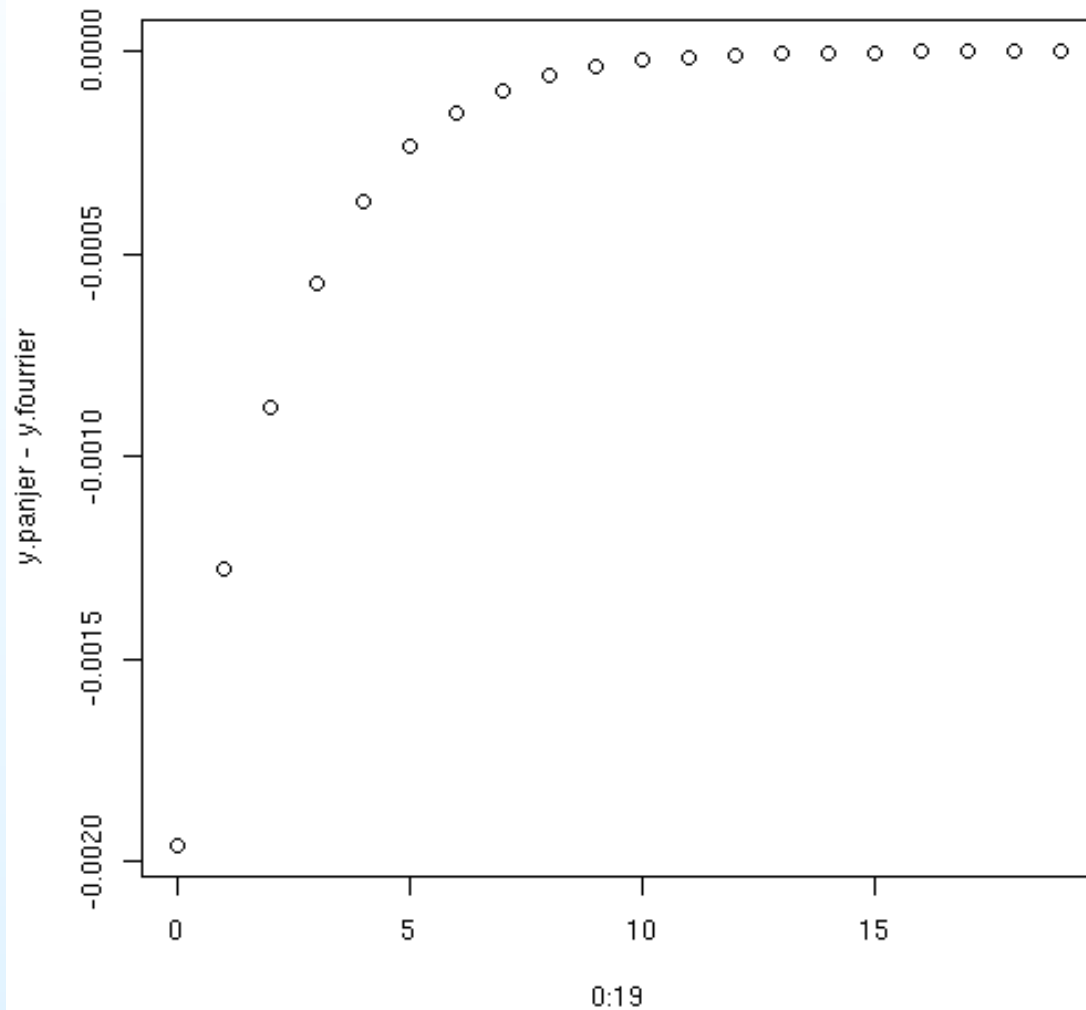
```
> Re(fft(ztil,inverse=T)/length(Y))
      0      1      2      3      4
0.1373 0.0351 0.0728 0.0855 0.1378
      5      6      7      8      9
0.0660 0.0830 0.0783 0.0767 0.0466
     10     11     12     13     14
0.0454 0.0371 0.0295 0.0194 0.0162
     15     16     17     18     19
0.0119 0.0086 0.0057 0.0042 0.0029
```

que é a densidade de probabilidade de Z .

Comparação: Panjer vs FFT

k	panjer	fourier	k	panjer	fourier
0	0.14	0.14	10	0.05	0.05
1	0.03	0.04	11	0.04	0.04
2	0.07	0.07	12	0.03	0.03
3	0.08	0.09	13	0.02	0.02
4	0.14	0.14	14	0.02	0.02
5	0.07	0.07	15	0.01	0.01
6	0.08	0.08	16	0.01	0.01
7	0.08	0.08	17	0.01	0.01
8	0.08	0.08	18	0.00	0.00
9	0.05	0.05	19	0.00	0.00

Comparação: Panjer vs FFT



Técnicas para combinar várias linhas de negócios

Suponhamos que tenhamos interesse em comparar duas linhas de negócio:

- A linha 1 tem número de sinistros N e valor de sinistros X .
- A linha 2 tem número de sinistros K e valor de sinistros Y .
- N , K , X e Y são mutuamente independentes.

O interesse está em descobrir a distribuição de probabilidades da perda agregada dos portfólios combinados:

$$Z = (X_1 + \dots + X_N) + (Y_1 + \dots + Y_K).$$

Sob as suposições acima, temos que:

$$\phi_Z(t) = P_N(\phi_X(t))P_K(\phi_Y(t))$$

Técnicas para combinar várias linhas de negócios (cont)

Sejam \tilde{g} e \tilde{h} as transformadas de Fourier das distribuições de perda agregada da Linha 1 e 2 respectivamente:

$$\tilde{g} = P_N(\tilde{f}_X) \text{ e } \tilde{h} = P_K(\tilde{f}_Y)$$

Basta aplicar o produto complexo (elemento a elemento) de \tilde{g} e \tilde{h} e aplicarmos a transformada de Fourier inversa para obtermos a distribuição desejada:

$$f_Z = IFFT(\tilde{g}\tilde{h})$$

O modelo Poisson

Temos aqui um modelo Poisson básico para combinar diferentes linhas de negócios:

Assuma que estamos combinando k linhas de negócios. Para $j = 1, 2, \dots, k$ assumamos que a Linha j tem frequência Poisson com média λ_j e distribuição de risco F_j . Nós assumimos que a perda proveniente de diferentes linhas de negócios são *independentes*. Em termos da função característica teremos

$$\begin{aligned}\phi_Z(t) &= \prod_{j=1}^k P_{N_j}(\phi_{X_j}(t)) \\ &= \prod_{j=1}^k e^{\lambda_j(\phi_{X_j}(t)-1)} \\ &= e^{\lambda(\phi_X(t)-1)},\end{aligned}$$

O modelo Poisson (cont)

onde $\lambda = \lambda_1 + \dots + \lambda_k$, e

$$\phi_X(t) = \frac{\lambda_1}{\lambda} \phi_{X_1}(t) + \dots + \frac{\lambda_k}{\lambda} \phi_{X_k}(t).$$

Assim, as perdas agregadas para as k linhas de negócio combinadas têm uma frequência Poisson com média

$$\lambda = \lambda_1 + \lambda_2 + \dots + \lambda_k$$

e a distribuição dos sinistros é uma média ponderada de cada distribuição individual:

$$F(x) = \frac{\lambda_1}{\lambda} F_1(x) + \frac{\lambda_2}{\lambda} F_2(x) + \dots + \frac{\lambda_k}{\lambda} F_k(x).$$

O modelo Poisson (cont)

Em resumo, temos que sob a suposição de independência mútua entre as linhas e um processo de Poisson para cada linha, a distribuição agregada pode ser calculada como se fosse uma única linha, usando as equações encontradas para ajustar os parâmetros de cada distribuição.

As suposições de independência são muito restritivas. Modelos mais realistas devem levar em conta a presença de uma incerteza sobre os parâmetros do modelo, e uma possível correlação entre as linhas de negócio. Como alternativa para o modelo Poisson, o modelo binomial negativo é comumente usado para ajustar esses dois pontos levantados. Basta notar que uma distribuição binomial negativa pode ser obtida a partir da suposição de uma distribuição de Poisson com parâmetro de média desconhecido e seguindo uma gama.

O modelo com distribuição Binomial Negativa

Considere k linhas de negócio com distribuição de frequência e de sinistro como indicadas na Tabela abaixo.

Linha	$E(N_i)$	$\frac{\text{Var}(N_i)}{E(N_i)}$	Dist. de Risco
1	$E(N_1)$	$1 + \beta_1$	F_1
2	$E(N_2)$	$1 + \beta_2$	F_2
\vdots	\vdots	\vdots	\vdots
k	$E(N_k)$	$1 + \beta_k$	F_k

- A média da frequência agregada é a soma de cada média individual

$$E[N_{agg}] = E[N_1] + E[N_2] + \cdots + E[N_k].$$

O modelo com distribuição Binomial Negativa (cont)

- A variância total da frequência agregada pode ser calculada por:

$$\text{Var}[N_{agg}] = \text{Var} \left[\sum_{i=1}^k N_i \right] = \sum_{i=1}^k \text{Var}[N_i] + 2 \sum_{i < j} \text{Cov}[N_i, N_j].$$

Uma abordagem simples e direta para levar em conta a incerteza sobre os parâmetros das distribuições de frequência é assumir uma distribuição binomial negativa para todas as linhas combinadas. Dessa forma, os parâmetros da binomial negativa podem ser prontamente estimados a partir de $E[N_{agg}]$ e $\text{Var}[N_{agg}]$. A distribuição dos sinistros para todas as linhas combinadas pode ser calculada como uma média ponderada de cada distribuição individual dos sinistros:

O modelo com distribuição Binomial Negativa (cont)

$$F(x) = \frac{\mathbf{E}[N_1]}{\mathbf{E}[N_{agg}]} F_1(x) + \frac{\mathbf{E}[N_2]}{\mathbf{E}[N_{agg}]} F_2(x) + \dots + \frac{\mathbf{E}[N_k]}{\mathbf{E}[N_{agg}]} F_k(x).$$

Com essa abordagem leva-se em conta a incerteza sobre os parâmetros da distribuição de freqüência, e, ao mesmo tempo, é possível ajustar para correlações entre as freqüências em cada linha de negócio. Suponha por exemplo que as freqüências N_j tenham a seguinte matriz de correlação:

$$\begin{pmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1k} \\ \rho_{21} & \rho_{22} & \cdots & \rho_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{k1} & \rho_{k2} & \cdots & \rho_{kk} \end{pmatrix},$$

O modelo com distribuição Binomial Negativa (cont)

nós podemos então calcular a razão global de média por variância usando a expressão para $\text{Var}[N_{agg}]$ que encontramos, e ajustar as correlações usando:

$$\text{Cov}[N_i, N_j] = \rho_{ij} \sqrt{\text{Var}[N_i]} \sqrt{\text{Var}[N_j]}.$$

Entre os outros possíveis métodos conhecidos na literatura nesse tema, temos o método de Panjer, que usa uma relação recursiva. O método de Heckman-Meyers, que utiliza a inversão direta de funções características. O método de Robertson, usando o método de FFT com funções *piecewise* ao invés de uma distribuição discreta para os sinistros.

Algumas ferramentas para distribuições multivariadas

- O caso univariado

Seja X uma variável aleatória não-negativa discreta, contínua ou mista. Seja $f_X(x)$ sua função de (densidade de) probabilidade:

$$f_X(x) = \begin{cases} P(X = x), & \text{se } X \text{ discreta} \\ \frac{d}{dx}F_X(x), & \text{se } X \text{ contínua} \end{cases}$$

- A função geradora de probabilidade (f.g.p) é definida por:

$$P_X(t) = E[t^X] = \begin{cases} \sum f_X(x)t^x & \text{se } X \text{ discreta} \\ \int f_X(x)t^x dx & \text{se } X \text{ contínua.} \end{cases}$$

O caso univariado (cont)

- A função geradora de momentos (f.g.m) é definida por:

$$M_X(t) = E[e^{tX}] = P_X(e^t)$$

- A função característica (f.c), também chamada de transformada de Fourier, é dada por:

$$\phi_X(t) = E[e^{itX}] = P_X(e^{it}) = M_X(it),$$

- É válido que: $P_X(1) = M_X(0) = \phi_X(0) = 1$, e

$$E[X] = \left[\frac{d}{dx} P_X(t) \right]_{t=1} = \left[\frac{d}{dx} M_X(t) \right]_{t=0} = -i \left[\frac{d}{dx} \phi_X(t) \right]_{t=0}$$

O caso multivariado

Para um conjunto de variáveis (X_1, \dots, X_k) , seja f_{X_1, \dots, X_k} a sua função de (densidade de) probabilidade:

$$f_{X_1, \dots, X_k} =$$

$$\begin{cases} Pr\{X_1 = x_1, \dots, X_k = x_k\}, & \text{se } X_j \text{ discretas} \\ \frac{\delta^k}{\delta x_1 \dots \delta x_k} F_{X_1, \dots, X_k}(x_1, \dots, x_k), & \text{se } X_j \text{ contínuas.} \end{cases}$$

Para qualquer subconjunto de $\{X_1, X_2, \dots, X_k\}$, sua distribuição de probabilidade conjunta é chamada de distribuição marginal de f_{X_1, \dots, X_k} . Como um caso especial, f_{X_1} é uma distribuição marginal univariada de f_{X_1, \dots, X_k} .

Consideremos os análogos das funções geradoras de probabilidade, momento e característica no caso multivariado:

O caso multivariado (cont)

$$P_{X_1, \dots, X_k}(t_1, \dots, t_k) = E[t_1^{X_1} \dots t_k^{X_k}];$$

$$M_{X_1, \dots, X_k}(t_1, \dots, t_k) = E[e^{t_1 X_1 + \dots + t_k X_k}] = P_{X_1, \dots, X_k}(e^{t_1}, \dots, e^{t_k});$$

$$\phi_{X_1, \dots, X_k}(t_1, \dots, t_k) = E[e^{i(t_1 X_1 + \dots + t_k X_k)}] = P_{X_1, \dots, X_k}(e^{it_1}, \dots, e^{it_k}).$$

Note que nos termos da função de (densidade de) probabilidade temos:

$$P_{X_1, \dots, X_k}(t_1, \dots, t_k) = \begin{cases} \sum_{(x_1, \dots, x_k)} f_{X_1, \dots, X_k}(x_1, \dots, x_k) t_1^{x_1} \dots t_k^{x_k}, \\ \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_{X_1, \dots, X_k}(u_1, \dots, u_k) t_1^{u_1}, \dots, t_k^{u_k} du_1 \dots du_k \end{cases}$$

Cada uma dessas funções determina univocamente uma distribuição multivariada de variáveis aleatórias. Temos ainda as seguintes propriedades:

O caso multivariado (cont)

- A função geradora de probabilidade ou a função característica univariada pode ser obtida para uma distribuição marginal F_{X_j} da seguinte forma:

$$P_{X_j}(t_j) = P_{X_1, \dots, X_j, \dots, X_k}(1, \dots, 1, t_j, 1, \dots, 1),$$

$$\phi_{X_j}(t_j) = \phi_{X_1, \dots, X_j, \dots, X_k}(0, \dots, 0, t_j, 0, \dots, 0)$$

- Se as variáveis X_1, \dots, X_k são mutuamente independentes temos:

$$P_{X_1, \dots, X_k}(t_1, \dots, t_k) = \prod_{j=1}^k P_{X_j}(t_j),$$

O caso multivariado (cont)

- Se dois conjuntos de variáveis X_1, \dots, X_m e Y_1, \dots, Y_n são independentes, então:

$$P_{X_1, \dots, X_m, Y_1, \dots, Y_n}(t_1, \dots, t_m, s_1, \dots, s_n) = P_{X_1, \dots, X_m}(t_1, \dots, t_m)P(Y_1, \dots, Y_n)$$

- As covariâncias podem ser obtidas através da relação $Cov[X_i, X_j] = E[X_i X_j] - E[X_i]E[X_j]$, com:

$$E[X_i X_j] = \frac{\delta^2}{\delta t_i \delta t_j} P_{X_1, \dots, X_m}(1, \dots, 1) = -\frac{\delta^2}{\delta t_i \delta t_j} \phi_{X_1, \dots, X_m}(0, \dots, 0).$$

O caso multivariado (cont)

- Para uma distribuição multivariada discreta, a função de probabilidade conjunta é dada por:

$$f_{X_1, \dots, X_k}(x_1, \dots, x_k) = \frac{\delta^{x_1 + \dots + x_k}}{(\delta t_1)^{x_1} \dots (\delta t_k)^{x_k}} P_{X_1, \dots, X_k}(0, \dots, 0) \prod_{i=1}^k \frac{1}{x_i!}.$$

Agregação de variáveis correlacionadas

Teorema 1 *Para quaisquer k variáveis correlacionadas X_1, \dots, X_k , com função geradora de probabilidades conjunta P_{X_1, \dots, X_k} e função característica conjunta dada por ϕ_{X_1, \dots, X_k} , a soma $Z = X_1 + \dots + X_k$ tem uma função geradora de probabilidade e uma função característica:*

$$P_Z(t) = P_{X_1, \dots, X_k}(t, \dots, t), \quad \phi_Z(t) = \phi_{X_1, \dots, X_k}(t, \dots, t).$$

Esse teorema permite que obtenhamos a função característica da soma Z através da função característica conjunta, de maneira simples e direta. A partir daí basta usar uma transformação de Fourier inversa para obter a distribuição de probabilidades de Z .

Agregação de portfólios de risco com freqüências correlacionadas

Considere a agregação de dois portfólios de risco correlacionados:

$$Z = (X_1 + \dots + X_N) + (Y_1 + \dots + Y_K),$$

com N e K correlacionados, e o par (N, K) independente de X e Y , e X_i s e Y_j s mutuamente independentes. Temos que:

$$\begin{aligned} P_Z(t) &= E[t^Z] = E[t^{(X_1 + \dots + X_N) + (Y_1 + \dots + Y_K)}] = \\ &= E_{N,K} E[t^{(X_1 + \dots + X_k) + (Y_1 + \dots + Y_m)} \mid N = n, K = m] \\ &= E_{N,K} [P_X(t)^N P_Y(t)^K] \\ &= P_{N,K}(P_X(t), P_Y(t)). \end{aligned}$$

Agregação de portfólios de risco com freqüências correlacionadas (cont)

e em termos da função característica, obtemos:

$$\phi_Z = P_{N,K}(\phi_X(t), \phi_Y(t)).$$

Conclusões

- Diversas abordagens diferentes permitem a modelagem das distribuições de frequências e sinistros, com diferentes suposições e resultados possíveis.
- O cálculo da distribuição agregada tanto por FFT quanto por Panjer é equivalente.
- O uso de estruturas de correlação complexas e informações sobre a incerteza dos parâmetros podem ser acomodados com o uso de um modelo de distribuição Binomial Negativa.

Sobre

A versão eletrônica desse arquivo pode ser obtida em
<http://www.ime.usp.br/~feferraz>

Copyright (c) 1999-2005 Fernando Henrique Ferraz
É dada permissão para copiar, distribuir e/ou modificar
sob os termos da Licença de Documentação Livre GNU
publicada pela Free Software Foundation;
Uma cópia da licença em está inclusa na seção intitulada
"Sobre / Licença de Uso".