

5. Referencial Teórico e Modelo Proposto

Neste capítulo será apresentada a proposta metodológica para a construção da árvore de cenários nos passos *Forward* e *Backward*. O arcabouço teórico empregado será exposto nas próximas seções.

Conforme apresentado no Capítulo 4, a geração de cenários proposta por (OLIVEIRA, 2010), através do uso da *técnica não paramétrica* de computação intensiva *Bootstrap*, apresenta-se satisfatória para o passo *Forward*, mas não na recursão *Backward*.

A seguir é mostrado um exemplo da aplicação do método proposto por (OLIVEIRA, 2010). Haja vista que não há garantias de que o valor de ENA simulado seja positivo em função da construção do modelo sem a variável Δ (equação (4.21)), a obtenção de um valor "válido" é extremamente dependente do ruído $A_{t(per, cen)}$ sorteado. Na Figura 5.1 é apresentado um caso ilustrativo em que, para um dado caminho do passo *Forward*, todos os valores de ENA da árvore são positivos, mas se considerado o passo *Backward*, em que todas as ramificações da sub árvore selecionada são consideradas, não é possível obter um valor de ENA positivo, limitando a aplicabilidade prática do método proposto.

Neste exemplo, o caminho *Forward* em que todos os valores simulados são positivos é mostrado em azul. O nó em vermelho indica que, no passo *Backward*, para as demais ramificações do estágio $t = 3$, não são obtidos valores de ENA positivos em função do resíduo anteriormente sorteado para compor a árvore de cenários.

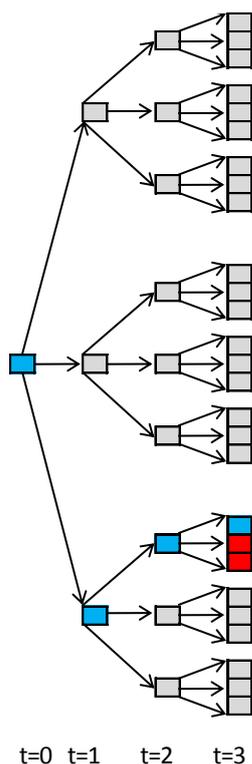


Figura 5.1 - Falha no Passo *Backward*

No modelo NEWAVE, são sorteados, a partir da distribuição Lognormal com três parâmetros ajustada aos resíduos do modelo $PAR(p)$, 20 ruídos aleatórios para cada um dos períodos do horizonte de planejamento. Seja um PMO para o mês de janeiro e cinco anos de cenários mensais simulados ($T = 60$). Neste caso, seria construída uma matriz de ruídos de dimensão 20×60 . Uma vez sorteada, esta matriz é fixa para todo o processo de geração de cenários, tanto para o passo *Forward* quanto para o *Backward*.

A partir da referida matriz, são sorteadas, de forma aleatória, 200 trajetórias, que compõem os cenários de ENA (ou caminhos *Forward*), construídos com base nos modelos $PAR(p)$ anteriormente ajustados, dos períodos $t = 1, 2, \dots, T$. No caso do passo *Backward*, a simulação é realizada no sentido contrário e, em cada um dos períodos de todos os caminhos *Forward*, o valor de ENA simulado deve ser positivo para qualquer um dos 20 resíduos anteriormente sorteados para compor a matriz de dimensão 20×60 (no caso, por exemplo, do PMO de janeiro).

De forma ilustrativa, considere o exemplo da Figura 5.2. Nesta está representada uma situação hipotética de uma matriz de ruídos (cada cédula

representa um ruído selecionado a partir de uma distribuição Lognormal com três parâmetros) com 20 aberturas e 60 estágios de planejamento (PMO de janeiro, por exemplo). Para imaginar a possível dimensão da árvore de cenários considerada, basta visualizar a cardinalidade da mesma, 20^{60} , uma vez que cada nó antecessor pode se ligar a outros 20 sucessores. No exemplo da Figura foram selecionados apenas três ramos (cenários) da árvore, representados nas cores cinza, azul e amarelo. Pode-se notar que em duas situações, o mesmo nó, destacado em verde, se repetiu para dois cenários. Em apenas um dos estágios o mesmo ruído foi selecionado pelos três cenários, destacado em vermelho na matriz. Os caminhos *Forward* selecionados estão expostos na parte inferior da matriz de ruídos.

Supondo que os ruídos selecionados fossem aplicados no modelo estocástico correspondente ao período em questão, seriam gerados três cenários de ENA, para um dado subsistema, conforme ilustra a Figura 5.3.

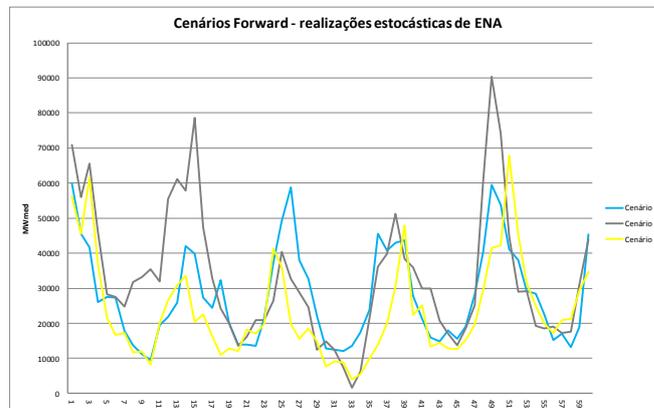


Figura 5.3 – Exemplo dos três cenários *Forward* sorteados

Em (OLIVEIRA, 2010) embora seja possível, por meio do descarte de ENAs negativas e reamostragem dos resíduos do modelo, via *Bootstrap*, a construção dos caminhos de séries sintéticas do passo *Forward*, não se consegue o mesmo no passo *Backward*. Os 20 ruídos sorteados anteriormente não geram, para todas as aberturas, valores de ENA positivos no período e cenário em questão.

Assim sendo, nas próximas seções será apresentado o arcabouço teórico empregado para a proposição da nova abordagem para a geração de séries sintéticas, por meio de um modelo que utiliza técnicas não paramétricas e de simulação estocástica. A última seção deste capítulo apresenta o modelo proposto.

5.1 A Técnica Não Paramétrica *Bootstrap*

O *Bootstrap*, introduzido por (EFRON & TIBSHIRANI, 1993), é uma técnica estatística não paramétrica computacionalmente intensiva que permite a avaliação da variabilidade de estimadores com base nos dados de uma única amostra existente.

A técnica é indicada para problemas nos quais os procedimentos estatísticos convencionais sejam de difícil aplicação. Em geral, apresenta vantagens se usado em situações de amostras pequenas ou grandes, desde que forneça resultados

próximos aos obtidos por meios assintóticos usuais em grandes amostras ou superior a amostras reduzidas.

Operacionalmente, a técnica consiste de um sorteio com reposição dos elementos de uma amostra aleatória, gerando uma “amostra *Bootstrap*”, de tamanho igual à original. Extrai-se um número suficiente de amostras a fim de se obter a “distribuição *Bootstrap*” de qualquer estatística de interesse do pesquisador. Desta forma, o conjunto de observações *Bootstrap* corresponde a uma estimativa da verdadeira distribuição amostral da estatística em questão. Como mostrado em (EFRON & TIBSHIRANI, 1993), à medida que o tamanho da amostra tende ao infinito, a distribuição *Bootstrap* converge para a distribuição verdadeira da estatística.

Considere $X = (x_1, x_2, \dots, x_{n-1}, x_n)$ a amostra original finita de tamanho n obtida de um modelo probabilístico desconhecido descrito por sua função de distribuição acumulada F e $\theta = S(X)$ a estatística de interesse. Denota-se X_i^* , $i = 1, 2, \dots, B$ como sendo a i -ésima amostra *Bootstrap* de tamanho n obtida da amostra original X por meio de sorteios com reposição. Para cada amostra tem-se a correspondente estimativa *Bootstrap* da estatística de interesse, isto é, $\theta_i^* = S(X_i^*)$.

Define-se a média, variância e erro padrão do estimador *Bootstrap* de θ por, respectivamente:

$$\theta^* = \frac{\sum_{i=1}^B \theta_i^*}{B} \tag{5.1}$$

$$\text{Var}(\theta_i^*) = \frac{\sum_{i=1}^B (\theta_i^* - \theta^*)^2}{B - 1} \tag{5.2}$$

$$SE_{boot} = \sqrt{\text{Var}(\theta_i^*)} \tag{5.3}$$

Sem maiores detalhes matemáticos, é possível provar que:

$$\text{Var}(SE_{boot}) \cong \frac{C_1}{n^2} + \frac{C_2}{nB} \quad (5.4)$$

C_1 e C_2 são constantes que dependem da distribuição populacional F , mas não dependem de n e B . Portanto, a incerteza associada ao estimador *Bootstrap* dependerá, em última análise, do tamanho da amostra original n , isto é, mesmo que se gere uma infinidade de amostras *Bootstrap*, a incerteza do estimador não vai a zero.

Alguns métodos propostos na literatura para a construção de intervalos de confiança utilizando o *Bootstrap* são: *Non-Studentized pivotal*, *Bootstrap-t*, *Percentile*, *Bias Corrected percentile*, *Bias Corrected and accelerated percentile*, *Test-inversion* e *Studentized test-inversion*. Para detalhes sobre esses métodos, consultar (CARPENTER & BITHELL, 2000), (EFRON & TIBSHIRANI, 1993) e (SILVA, 1995).

O intervalo de confiança *Percentile* utilizando o *Bootstrap* é calculado de forma similar ao método clássico, entretanto utiliza a distribuição *Bootstrap* do parâmetro de interesse.

Para isto, adotando \hat{G} como a função de distribuição acumulada de $\hat{\theta}^*$, então o intervalo percentil $(1-2\alpha)$ é definido por α e $(1-\alpha)$ percentis de \hat{G} . Desta maneira, o ponto inferior do intervalo é dado por $\hat{G}^{-1}(\alpha)$ e o ponto superior é $\hat{G}^{-1}(1-\alpha)$. Adotando $\hat{G}^{-1}(\alpha) = \hat{\theta}^{*(\alpha)}$ então o $100.\alpha^{th}$ percentil da distribuição *Bootstrap* pode ser escrito como:

$$\left[\hat{\theta}_{\%inf}, \hat{\theta}_{\%sup} \right] = \left[\hat{\theta}^{*(\alpha)}, \hat{\theta}^{*(1-\alpha)} \right] \quad (5.5)$$

Cabe salientar que este resultado é teórico e que se refere a uma situação onde o número de amostras é infinito. Segundo (UNNY & COVER, 1985), no *Bootstrap* este número é finito e dado por (onde n é o tamanho da amostra):

$$\binom{2n-1}{n} \quad (5.6)$$

Portanto, de acordo com a equação anterior, para um tamanho de amostra igual a 10, seria possível construir $B = 92378$ amostras *Bootstrap*. Na prática, B deve ser suficientemente grande para que seja possível calcular as estatísticas de interesse de cada amostra. Essas estatísticas são ordenadas e calculam-se, desta forma, os percentis de interesse. Seja $\hat{\theta}_B^{*(\alpha)}$ o $(100.\alpha)$ i -ésimo percentil dos valores de $\hat{\theta}^*(b)$, ou seja, a $B.\alpha$ i -ésima numa lista de valores ordenado das B estatísticas oriundas das amostras *Bootstrap*. Por exemplo, caso $B = 10.000$ e $\alpha = 0,05$, então o limite inferior será o ponto que está na posição 250 e o limite superior do intervalo será o valor que está na posição 9750.

Assim, o intervalo aproximado para $(1 - 2\alpha)$ é:

$$\left[\hat{\theta}_{\%,\text{inf}}, \hat{\theta}_{\%,\text{sup}} \right] \approx \left[\hat{\theta}_B^{*(\alpha)}, \hat{\theta}_B^{*(1-\alpha)} \right] \quad (5.7)$$

Este resultado é válido quando a distribuição da estatística de interesse é aproximadamente Normal, uma vez que o intervalo percentil *Bootstrap* e o intervalo de uma distribuição Normal padrão serão equivalentes. Esta relação é válida pois, segundo o Teorema Central do Limite, quando $n \rightarrow \infty$, o histograma dos dados gerados via *Bootstrap* adquire a forma de uma distribuição Normal.

Em posse dos dois intervalos estimados, a verificação da significância deve ser feita da seguinte forma: se o valor zero estiver contido no intervalo $\left[\hat{\theta}_{\%,\text{inf}}, \hat{\theta}_{\%,\text{sup}} \right]$, entre o limite inferior e superior, então estatisticamente este parâmetro é igual a zero; caso contrário, o parâmetro estimado é estatisticamente diferente de zero.

Na próxima seção é apresentada a aplicação da técnica no contexto de simulação de variáveis estocásticas e séries temporais.

5.1.1 **Bootstrap na Simulação de Séries Temporais**

A utilização da técnica de *Bootstrap* em séries temporais surgiu em 1984, quando a metodologia foi aplicada num contexto econométrico para previsão de demanda de energia no mercado americano. Mais tarde outros trabalhos surgiram

e estudou-se a estimação do erro-padrão das estimativas dos parâmetros dos modelos de previsão. Atualmente o *Bootstrap* vem sendo aplicado em uma série de situações, com os mais variados objetivos. A técnica tem se mostrado muito robusta e os resultados encontrados bastante satisfatórios.

No contexto de séries temporais, há basicamente duas maneiras para aplicação: *Bootstrap* nos resíduos e o chamado método “*moving blocks*”, (SOUZA & CAMARGO, 2004)

No primeiro caso, garantindo a hipótese de independência dos resíduos, estes são usados para gerar novas séries. Inicialmente ajusta-se um modelo para a série em questão e, a partir dos resíduos obtidos, são geradas B amostras *Bootstrap*, sorteando B vezes com reposição dos resíduos. Por fim, geram-se B séries *Bootstrap* utilizando os parâmetros do modelo original e os resíduos *Bootstrap*. Portanto, têm-se B séries temporais *Bootstrap*.

No caso do método “*moving blocks*”, nenhum modelo é ajustado a priori. A partir dos dados, são construídos blocos, de tamanho M , da série original. Sorteiam-se k blocos amostrando com reposição, agregando-os para formar a amostra *Bootstrap*. Repete-se este passo B vezes, gerando B novas séries *Bootstrap*. Neste método, a escolha do tamanho M dos blocos é um problema não totalmente resolvido e fonte de diversas críticas na literatura. O procedimento requer, evidentemente, estacionariedade (de segunda ordem) da série original.

Desta forma, neste trabalho, a técnica não paramétrica será empregada nos resíduos do modelo $PAR(p)$ para as simulações das realizações da variável ENA no passo *Forward*. O algoritmo da aplicação será definido na seção 5.4.

No caso da recursão *Backward*, os resíduos utilizados como parcela probabilística dos modelos de simulação da *Forward* serão avaliados e uma distribuição de probabilidade contínua será ajustada aos mesmos. A partir desta distribuição, novos ruídos serão amostrados, via Monte Carlo, para compor as realizações de ENA da *Backward*. Desta forma, nas próximas seções são apresentados, respectivamente, a orientação do arcabouço teórico para os passos do ajuste, estimação paramétrica e seleção das distribuições de probabilidades. Finalmente é apresentado a método de simulação de Monte Carlo.

5.2 Distribuições de Probabilidade: Ajuste, Estimação Paramétrica e Seleção

A partir da amostra de resíduos dos modelos $PAR(p)$ ajustados para cada um dos estágios do horizonte de planejamento (ver detalhes no Apêndice A), serão ajustadas distribuições de probabilidade contínuas. Inicialmente é importante definir quais as distribuições passíveis de ajuste, com base no seu suporte. Amostras de resíduos apresentam-se no domínio dos números reais, o que restringe as possibilidades de distribuições, uma vez que as mesmas devem ter suporte no intervalo $-\infty \leq x \leq \infty$. Desta maneira, as seguintes possibilidades foram testadas: *Normal*, *Logística*, *Gumbel*, *Generalizada de Valores Extremos (GEV)* e *t Location Scale*.

Todas as características destas distribuições, como parâmetros, momentos e Função Densidade de Probabilidade, estão descritas no Apêndice B.

Haja vista que as distribuições possíveis foram caracterizadas, o próximo passo é a estimação paramétrica e seleção daquela mais aderente aos resíduos de cada estágio.

Usualmente, os métodos tradicionais para *Goodness of Fit* são os testes não paramétricos de Qui-Quadrado, Kolmogorov-Smirnoff e Anderson-Darling, (GIBBONS & CHAKRABORTI, 2003). Entretanto, estes métodos podem não ser adequados em todas as situações e devem ter sua utilização ponderada. Os testes de Qui-Quadrado, Kolmogorov-Smirnoff são apresentados em detalhes no Apêndice C.

Estudos recentes mostram alguns inconvenientes dos métodos não paramétricos. De acordo com (VOSE, 2008):

- ✓ O teste do Qui-Quadrado depende da especificação do número de classes do histograma em que os dados serão agrupados, e não há nenhuma "regra de ouro" que forneça o valor mais adequado. O teste também faz algumas suposições muito fortes que só são atendidas quando se tem uma base de dados muito grande;
- ✓ Os testes de Kolmogorov-Smirnoff e Anderson-Darling foram projetados para testar a qualidade do ajuste de distribuições com os valores dos parâmetros definidos, não aqueles em que os parâmetros são estimados a

partir dos dados. Correções são possíveis apenas para poucos tipos de distribuições;

- ✓ Nenhum destes testes penaliza distribuições para o número de parâmetros utilizado (parcimônia). Assim, a distribuição, com quatro parâmetros pode também ajustar os dados melhor porque tem muito mais flexibilidade na forma do que uma distribuição com dois parâmetros, mas a melhoria aparente é espúria;
- ✓ Nenhum destes testes pode ser empregado com dados truncados, censurados ou binários;
- ✓ O método de montagem (geralmente máxima verossimilhança, ou método de momentos) é incompatível com a medida do grau de ajuste.
- ✓ Nenhum destes testes é capaz de fornecer uma ponderação adequada para a plausibilidade de cada distribuição candidata.

Considerando o exposto, faz-se necessário buscar técnicas mais efetivas para a análise da aderência dos dados às distribuições de probabilidades teóricas apresentadas anteriormente. A literatura recente mostra que os meios mais eficazes para testar o *goodness of fit* de uma dada distribuição são os *Crítérios de Informação*. Estes critérios consideram o número de observações da amostra, número de parâmetros da distribuição avaliada e o valor da máxima verossimilhança.

Isto posto, nesta Tese serão utilizados os Critérios de Informação como critério de *ranqueamento* e seleção das distribuições de probabilidade ajustadas aos dados. O arcabouço teórico destes critérios está descrito em detalhes no Apêndice C.

Por fim, haja vista que as distribuições de probabilidades dos resíduos do modelo $PAR(p)$ foram selecionadas, é possível amostrar, via técnicas de Simulação Estocástica, valores aleatórios destas distribuições que servirão como parcela probabilística no modelo de geração de ENAs no passo *Backward*. Na próxima seção é apresentada a técnica de simulação de Monte Carlo e sua aplicação na proposta para construção da árvore de cenários.

5.3 Simulação de Monte Carlo

Os métodos de Simulação Estocástica consistem no processo de gerar amostras de variáveis aleatórias num ambiente computacional e usar essas amostras com o objetivo de explorar o espaço de incerteza ou campo de possibilidades de um dado fenômeno físico ou qualquer outro tipo de variável de estudo cujo comportamento possa ser quantificado matematicamente. A Simulação Estocástica replica o comportamento de sistemas complexos explorando a sua aleatoriedade para obter cenários das possíveis saídas desses sistemas

Os métodos de simulação fazem parte da ciência de Processos Estocásticos e utilizam o chamado *Método de Monte-Carlo*, proposto por (METROPOLIS, 1949), em seus algoritmos. De acordo com (LOPES, 2012), os dois principais requisitos para os métodos de simulação de Monte Carlo são: possuir o conhecimento das distribuições de probabilidade das variáveis de entrada do sistema e possuir um gerador de números aleatórios para gerar cenários das variáveis de entrada do sistema. Simulando um grande número de cenários, a distribuição de probabilidade de todas as saídas da simulação pode ser aproximada com precisão, que aumenta à medida que o número de cenários aumenta. Assim, a geração de números aleatórios é o alicerce de qualquer sistema de Simulação Estocástica.

Por conseguinte, é necessário definir uma rotina capaz de gerar os ruídos – valores pseudoaleatórios amostrados das distribuições de probabilidade. Deve-se atentar às propriedades desejáveis de um bom gerador de números pseudoaleatórios, conforme aponta (LOPES, 2012):

- ✓ Possuir um padrão aleatório: o gerador deve passar por testes estatísticos de aleatoriedade;
- ✓ Possuir um período longo;
- ✓ Ser eficiente: deve ser executado rapidamente e requerer um baixo armazenamento;
- ✓ Ser de fácil reprodução: a partir de determinadas condições iniciais, deve produzir sempre a mesma sequência;

- ✓ Ser portátil: a partir de determinadas condições iniciais, a sequência gerada deve ser a mesma em qualquer computador.

Formalmente, segundo (ROBERT & CASELLA, 2004), a distribuição Uniforme, $U_{[0,1]}$, fornece a representação probabilística básica de aleatoriedade e todas as demais distribuições requerem uma sequência de variáveis uniformes para serem simuladas. Desta forma, pode-se definir um *Gerador de Números Pseudoaleatórios Uniforme* como um algoritmo que, iniciado com um valor u_0 e uma transformação D , produz uma sequência $(u_i) = (D^i(u_0))$ de valores no intervalo $[0,1]$. Para todos os n valores, (u_1, \dots, u_n) reproduz o comportamento de uma amostra *iid* (V_1, \dots, V_n) de variáveis aleatórias uniformes quando comparadas através de testes de aleatoriedade.

Por conseguinte, todo o algoritmo de simulação de variáveis aleatórias contínuas, o caso daquelas empregadas neste trabalho, parte da geração de variáveis uniformes, transformadas através do método da *Função Inversa Generalizada*, (ROBERT & CASELLA, 2004). De acordo com (HUYNH, et al., 2008):

Teorema 5.1: seja F uma função de distribuição de probabilidade em \mathfrak{R} com a Função Inversa Generalizada definida por $F^{-1}(u) = \inf \{x : F(x) = u, 0 \leq u \leq 1\}$.

1. Se u é uma variável aleatória uniformemente distribuída no intervalo $[0,1]$, então a variável aleatória $F^{-1}(u)$ possui distribuição F .
2. Se X tem distribuição F , então $F(X)$ é uniformemente distribuída no intervalo $[0,1]$.

Com vistas à exemplificação, considere a geração de números aleatórios de uma distribuição de Cauchy, (CASELA & BERGER, 2001), cujas Funções Densidade de Probabilidade e Distribuição Acumulada são dadas, respectivamente, pelas equações (5.8) e (5.9):

$$f_x(x) = \frac{1}{\pi(x^2 + 1)} \tag{5.8}$$

$$F_X(x) = \frac{1}{\pi} \int_{-\infty}^x \frac{du}{u^2 + 1} = \frac{1}{\pi} \left[\arctan(x) - \frac{\pi}{2} \right] = u \quad (5.9)$$

Em consequência:

$$U = F_X(x) \Rightarrow X = F_X^{-1}(U) = \tan\left(\pi U + \frac{\pi}{2}\right) \quad (5.10)$$

Portanto, para gerar X , valores aleatórios provenientes de uma distribuição de Cauchy, usa-se uma das relações da equação (5.11) (uma vez que a função Tangente é periódica com período π):

$$\begin{aligned} X &= \tan\left(\pi\left(U + \frac{1}{2}\right)\right) \\ X &= \tan\left(\pi\left(U - \frac{1}{2}\right)\right) \end{aligned} \quad (5.11)$$

Estão disponíveis na literatura diversas extensões e sofisticações da geração clássica do método, como os métodos de *Quasi Monte Carlo*, (ROBERT & CASELLA, 2004) e *Markov Chain Monte Carlo*, (GIBBS, et al., 1998) e (GAMERMAN, 1997).

Nesta Tese, a implementação de todas as rotinas computacionais que compõem a estruturação do modelo de simulação da árvore de cenários foi desenvolvida na linguagem *Matlab*. Neste software, o principal gerador de números aleatórios disponível é o chamado *Mersenne Twister*, proposto por (MATSUMOTO & NISHIMURA, 1998). O mesmo baseia-se em profundos fundamentos teóricos da matemática discreta e da álgebra abstrata. O funcionamento deste gerador se dá por meio de uma recorrência linear matricial ao longo de campos binários finitos. De acordo com (MAZZOTTI, 2010), este gerador corrige deficiências de seus precedentes, uma vez que gera uma sequência pseudoaleatória de alta qualidade e de forma muito eficiente, tendo sido aprovado em diversos testes de aleatoriedade, incluindo o *Teste de Diehard*, desenvolvido

por (MARSAGLIA, 1995) que, segundo (RUKHIN, et al., 2010), é um dos testes mais rigorosos para avaliar os geradores de números pseudoaleatórios. Os detalhes matemáticos do algoritmo do *Mersenne Twister* não serão expostos aqui e podem ser consultados em (MATSUMOTO & NISHIMURA, 1998). Uma vez caracterizado e suas vantagens apresentadas, este será o algoritmo usado para a realização das simulações estocásticas por meio do método de Monte Carlo empregado no trabalho. Testes realizados não mostraram diferenças significativas a partir da aplicação do método *Quasi Monte Carlo*.

Haja vista o exposto, no contexto do problema proposto, de posse do conhecimento das distribuições de probabilidades – e seus parâmetros – mais adequadas aos resíduos de cada uma das configurações do modelo $PAR(p)$ de um dado PMO, para a execução do passo *Backward*, faz-se necessário simular ruídos aleatórios destas distribuições para compor a parcela estocástica do modelo definido anteriormente com vistas à construção da árvore de cenários.

Finalmente, na próxima seção será apresentado o modelo para construção de árvores de cenários via técnicas não paramétricas e de simulação de Monte Carlo proposto nesta Tese.

5.4 Modelo Proposto

No modelo NEWAVE, os ruídos da amostra *Forward* são representados de uma forma muito simplificada, sorteados a partir de uma população reduzida (normalmente a amostra *Backward* possui apenas 20 representantes). Segundo (PENNA, 2009), essa opção foi escolhida por se aproximar melhor dos pressupostos que garantem a convergência assintótica do algoritmo de solução. Portanto, pode-se interpretar a amostra *Backward* como o conjunto de todas as possíveis realizações, e a amostragem da *Forward* como um subconjunto selecionado a partir das aberturas selecionadas.

Contudo, considerando um conjunto contínuo de possíveis realizações do processo estocástico, para efeito de convergência pode-se supor que os cenários visitados nos passos *Forward* e *Backward* não precisam ser necessariamente os mesmos, mas sim que devam visitar as mesmas regiões da distribuição de

probabilidades do processo estocástico. Esta premissa conduz à proposta descrita neste documento.

Deste modo, a ideia deste trabalho é a definição de um modelo que conjuga as técnicas *Bootstrap* e de simulação estocástica na construção da árvore de cenários utilizada pela PDDE.

A proposta da junção das duas técnicas tem sua motivação na preservação das características das séries históricas de ENA. No contexto do SIN, uma das características desejáveis dos cenários simulados é a preservação da correlação entre os valores de ENA dos subsistemas, haja vista a dependência espacial dos mesmos. No caso do modelo que emprega a estrutura Lognormal, a preservação desta característica se dá em função do uso da matriz $D_m^{R \times S}$, conforme mostrado na equação (4.19). No caso da utilização do *Bootstrap* na geração dos caminhos *Forward*, a alternativa proposta por (OLIVEIRA, 2010) foi o sorteio, no vetor de resíduos de cada estágio, daqueles de mesma posição para todos os subsistemas. Esta alternativa mostrou-se eficiente na geração dos cenários *Forward* para a maior parte dos casos.

A justificativa para o não emprego do *Bootstrap* na fase *Backward* já foi apresentada anteriormente. Por sua vez, a simples amostragem de ruídos, via Monte Carlo, em ambas as recursões, não seria capaz de preservar a correlação espacial entre os valores de ENA dos quatro subsistemas. Haja vista que no passo *Backward* são usados os valores passados da variável simulada na fase *Forward*, com a correlação preservada, espera-se que na recursão inversa o mesmo aconteça. Desta maneira, em função da especificidade do problema, a conjugação das técnicas de simulação propostas está, portanto, fundamentada. Todavia, vale ressaltar que, por definição, nenhuma das técnicas garante a geração de realizações apenas positivas de ENA. Nos casos em que forem gerados valores negativos, o nó é descartado e novos elementos aleatórios são amostrados at.

É importante ressaltar que a proposta é estabelecer uma metodologia genérica que possa ser utilizada na construção de árvores de cenários a partir da utilização de um modelo estocástico para modelagem da base de dados. Toda a metodologia é construída a partir da matriz de resíduos do modelo probabilístico gerador, com base nos passos apresentados a seguir, com foco na aplicação aos dados e ENA e ao modelo $PAR(p)$.

1. *Passo Forward:*

- 1.1. Amostrar, através da técnica não paramétrica *Bootstrap*, resíduos dos modelos estocásticos definidos para cada um dos períodos do horizonte de planejamento. Nesta amostragem, a posição do resíduo sorteado, no período em questão, deve ser a mesma para todos os subsistemas, com vistas à preservação da correlação entre os mesmos.
- 1.2. Compor, a partir dos resíduos sorteados, os caminhos *Forward*, respeitando a restrição de que os valores de ENA, obtidos a partir do modelo estocástico, sejam positivos.
- 1.3. Verificar que, em caso de valores de ENA simulados negativos, tais elementos devem ser descartados (para todos os subsistemas) e novos sorteios devem ser realizados.
- 1.4. Selecionar os N caminhos *Forward*, simulados dos períodos ($t = 1, 2, \dots, T$).

Para ilustrar, consideremos o *subsistema 1* e imaginemos que, para o *quinto período do terceiro caminho Forward*, tenha sido selecionado um modelo $PAR(1)$ e o resíduo de posição *45* tenha sido amostrado da matriz de resíduos do modelo estocástico gerador. Portanto, a ENA simulada neste instante seria dada a partir da seguinte componente estocástica: ${}^F A_{t(5,3)}^{45}$, em que, o índice do lado esquerdo indica a amostra *Forward* (F). Ao lado direito, os subíndices indicam o período do horizonte de planejamento (5) e o número do caminho *Forward* em questão (3). O superíndice indica a posição do resíduo amostrado (45) na matriz de resíduos mensais gerados pelo modelo PAR .

Considerando o exemplo do PMO de janeiro e cinco anos de séries simuladas, o "produto" deste passo será uma matriz de resíduos, de dimensão 200×60 , que garantem a positividade das ENAs e a preservação das características da série histórica.

2. *Passo Backward:*

- 2.1. Ajustar a matriz de resíduos mensais gerados pelo modelo PAR para cada período do horizonte de planejamento às distribuições de probabilidade selecionadas no Apêndice B.

- 2.2. *Rankear* as famílias de distribuições de probabilidade.
- 2.3. Estimar os parâmetros da distribuição de probabilidade identificada como a mais aderente no *ranking* do item anterior.
- 2.4. Iniciar o processo do passo *Backward*, com início em $(t = T, T - 1, \dots, 1)$, simulando, via Monte Carlo, 20 elementos aleatórios a partir das distribuições de probabilidades que foram definidas.
- 2.5. Realizar teste (se positivo ou negativo) para verificar e garantir que, para qualquer um dos 20 ruídos amostrados para o período e cenário em questão, o valor da realização da variável estocástica simulada seja positivo.
- 2.6. Para a determinação dos valores passados dos cenários, utilizar a matriz de ENAs simuladas no passo *Forward*.

Deste modo, o produto deste passo é um cubo com dimensões, no caso do PMO de janeiro e cinco anos de simulações, $(200 \times 60 \times 20)$, em que, associados a cada um dos 60 elementos de cada um dos 200 cenários, está disponível um vetor de 20 ruídos simulados, via Monte Carlo, e já testados, garantindo que o valor da ENA simulado no passo *Backward* seja positivo.

Portanto, as matrizes de ruídos utilizadas nos passos *Forward* e *Backward* são diferentes, contudo conservam, em ambos os casos, a distribuição de probabilidade dos resíduos do modelo $PAR(p)$. No primeiro caso, em função de serem estes próprios resíduos, amostrados via *Bootstrap*, utilizados como parcela estocástica na simulação das ENAs. No segundo, os ruídos amostrados, através de simulação estocástica, são oriundos das distribuições de probabilidades contínuas que melhor se ajustam aos dados.

Para ilustrar, seja o mesmo exemplo do passo *Forward*: imaginemos que, para o *quinto período do terceiro caminho Forward*, tenha sido selecionado um modelo $PAR(1)$ e o resíduo de posição 1 (lembrando que estão disponíveis 20 elementos) tenha sido amostrado a partir da distribuição de probabilidade ajustada aos resíduos históricos. Portanto, a ENA simulada neste instante seria dada a partir da seguinte componente estocástica: ${}^B A_{t(5,3)}^1$.

O índice do lado esquerdo indica a amostra *Backward* (**B**). Ao lado direito, os subíndices indicam o período do horizonte de planejamento (**5**) e o número do caminho *Forward* em questão (**3**). O superíndice denota a posição do resíduo amostrado (**I**) no vetor de 20 elementos aleatórios provenientes da distribuição de probabilidade selecionada. Assim, tem-se então uma estrutura, para cada subsistema, semelhante ao cubo mostrado na Figura 5.4. A seta vertical indica a dimensão referente ao número de cenários (caminhos) gerados. A seta horizontal a dimensão do número de estágios considerados no horizonte de simulação e, finalmente, a seta diagonal relaciona-se à dimensão do número de ruídos disponíveis para cada elemento da matriz (o *default* é 20 em função das considerações acerca da convergência do algoritmo mostradas anteriormente).

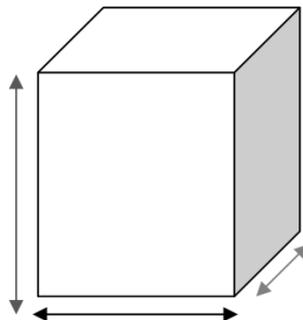


Figura 5.4 - Estrutura do Passo *Backward*

Finalmente, ainda de maneira ilustrativa, os resultados do modelo estocástico que são fornecidos à PDDE, para cada subsistema, são:

- i. Matriz com as ordens dos modelos $PAR(p)$.
- ii. Matriz com os parâmetros dos modelos $PAR(p)$.
- iii. Matriz de resíduos (amostrados via *Bootstrap*) para as simulações *Forward*.
- iv. Matriz de ruídos (simulados via Monte Carlo) para as recursões *Backward*.

Vale ressaltar que, para a PDDE, todos os resíduos do item (iii) e os ruídos do item (iv) já são aplicados ao modelo estocástico do período selecionado, tendo sido realizado o teste que verifica a positividade da realização da variável ENA.

No próximo capítulo, serão apresentados resultados da implementação desenvolvida. Os mesmos estão divididos em seções que tratam do ajuste dos modelos, simulação dos cenários da árvore e impacto na política de planejamento da operação do despacho hidrotérmico no contexto do SEB.