MÉTODOS DE GERAÇÃO DE NÚMEROS PSEUDO-ALEATÓRIOS

• Introdução

Começámos por abordar o problema "A Barragem de Pós-Boa" e desenvolvemos um modelo de simulação simplificado que nos permitiu testar diferentes políticas de gestão da barragem. Ao efectuarmos uma **simulação com 'dados históricos'** relativos aos anos de 1992, 1993 e 1994 estivemos a avaliar o que teria acontecido nesses anos se a barragem tivesse sido gerida com uma determinada política. Referimos, então, que essa abordagem pode ser importante, especialmente para avaliar o desempenho do sistema numa dada situação gravosa do passado, mas que não nos deixa muito tranquilos relativamente ao futuro e à sua imprevisibilidade/aleatoriedade.

Para uma avaliação mais cuidadosa da adequabilidade de uma dada política de gestão da barragem gostaríamos de dispor de uma amostra maior (e não apenas relativa a três anos). Mas, ainda que dispuséssemos de uma amostra maior, uma simulação com 'dados históricos' só permitiria avaliar o que teria acontecido nos anos relativos à amostra ... Assim, seria certamente desejável que, em vez de uma amostra de dados relativos à precipitação e outra relativa às necessidades de água a jusante, pudessemos dispor das correspondentes distribuições estatísticas.

A partir das distribuições estatísticas que descrevem o comportamento das variáveis referidas, poderíamos levar a cabo uma **simulação com 'dados aleatórios'** (*com comportamento estatístico aceitável face às distribuições...*) **correspondentes a um período muito mais longo** (por exemplo, 100 ou 200 anos). Para tal, no entanto, precisaríamos de **saber gerar os referidos 'dados aleatórios'**.

Esta secção dos apontamentos visa exactamente abordar os **métodos de geração de números pseudo-aleatórios** (NPA). "Números **pseudo**-aleatórios ?" - esta é a primeira dúvida que o leitor levantará. Na realidade, vamos aprender a conceber **algoritmos perfeitamente determinísticos no seu funcionamento, que a partir de um primeiro número (a *semente*), geram uma sequência de números que globalmente apresentam o comportamento estatístico desejável**. A cada semente estará associada uma sequência de NPA. Assim, temos que reconhecer que *rigorosamente* não estaremos a falar de números aleatórios ... mas de números *pseudo*-aleatórios ...

Uma questão clássica surge agora: "Se dispusesse de dois métodos de geração - um que lhe permitisse gerar **números realmente aleatórios** e outro que originasse **números pseudo-aleatórios** (com a mesma qualidade estatística) - qual o método que preferiria adoptar ?"

Claro que a resposta óbvia é "O método que origina números realmente aleatórios !"... Só que nem sempre a resposta óbvia é a resposta mais adequada ... Com efeito, se utilizarmos números realmente aleatórios na simulação d' "A Barragem de Pós-Boa" para compararmos duas políticas de gestão distintas, as ocorrências geradas na simulação associada à primeira política serão distintas das ocorrências geradas na simulação associada à segunda política ... o que não sendo propriamente um *problema*, ainda assim não nos permite comparar *exactamente* as duas políticas sujeitas à *mesma* aleatoriedade ...

É evidente que este problema desaparece com os números pseudo-aleatórios... Basta usar a mesma *semente* nos processos de geração para estarmos a gerir exactamente a mesma realidade (em termos de precipitação e necessidades de água) sujeita, num caso, à primeira política de gestão e, no outro caso, à segunda política de gestão ... Assim, nestas circunstâncias, as diferenças de resultados observados nas simulações correspondentes a essas duas políticas dever-se-iam exclusivamente às políticas e não à aleatoriedade da simulação.

Começaremos por aprender a gerar NPA com distribuição Uniforme [0 ; 1], e, em seguida, veremos métodos de *transformação* noutras distribuições.

• Geração de NPA com Distribuição Uniforme [0 ; 1]

Para gerarmos NPA com distribuição Uniforme [0 ; 1] utilizaremos o chamado **Método Congruencial Misto**, que consiste na utilização de uma relação recursiva do tipo:

$$\mathbf{x_{i+1} = (a . x_i + c) \bmod m}$$

com **a**, **c**, **m** inteiros positivos e **a**, **c** < **m**

A operação 'y **mod** w' traduz-se por '**resto da divisão inteira de** y **por** w'. [Por exemplo, 2 mod 3 = 2; 3 mod 2 = 1; 8 mod 3 = 2.]

De notar que a sequência de números obtidos dependerá da escolha das constantes **a**, **c** e **m**, bem como do número *semente*, $\mathbf{x_0}$, que inicia o processo.

Ilustremos a utilização do **Método Congruencial Misto** para **a = 4**, **c = 3** e **m = 5**, a partir da *semente* $\mathbf{x_0 = 3}$:

i	x_i	$4 . x_i + 3$	$x_{i+1} = (4 . x_i + 3) \bmod 5$
0	$\mathbf{x_0 = 3}$	15	$0 = x_1$
1	$x_1 = 0$	3	$3 = x_2$
2	$x_2 = 3$	15	$0 = x_3$
3	$x_3 = 0$	3	$3 = x_4$
4	$x_4 = 3$	15	$0 = x_5$
5	$x_5 = 0$	3	$3 = x_6$

Como se pode ver, a sequência obtida foi 3, 0, 3, 0, 3, 0, 3, ... Ou seja, de aleatório não tem nada ... e tem muito pouca variação ... Moral da história: os valores de **a**, **c**, **m** e $\mathbf{x_0}$ adoptados deram origem a uma sequência totalmente inútil, para efeitos de geração de número pseudo-aleatórios ...

Tentemos uma nova sequência, agora a partir de **a = 6**, **c = 4**, **m = 10** e $\mathbf{x_0 = 7}$:

i	x_i	$6 . x_i + 4$	$x_{i+1} = (6 . x_i + 4) \bmod 10$
0	$\mathbf{x_0 = 7}$	46	$6 = x_1$
1	$x_1 = 6$	40	$0 = x_2$
2	$x_2 = 0$	4	$4 = x_3$
3	$x_3 = 4$	28	$8 = x_4$
4	$x_4 = 8$	52	$2 = x_5$
5	$x_5 = 2$	16	$6 = x_6$
6	$x_6 = 6$	40	$0 = x_7$
7	$x_7 = 0$	4	$4 = x_8$
8	$x_8 = 4$	28	$8 = x_9$
9	$x_9 = 8$	52	$2 = x_{10}$

Como se vê a sequência obtida já foi um pouco mais interessante: 7, 6, 0, 4, 8, 2, 6, 0, 4, 8, 2, ... Obtivemos uma sequência de cinco números, que se repete indefinidamente ...

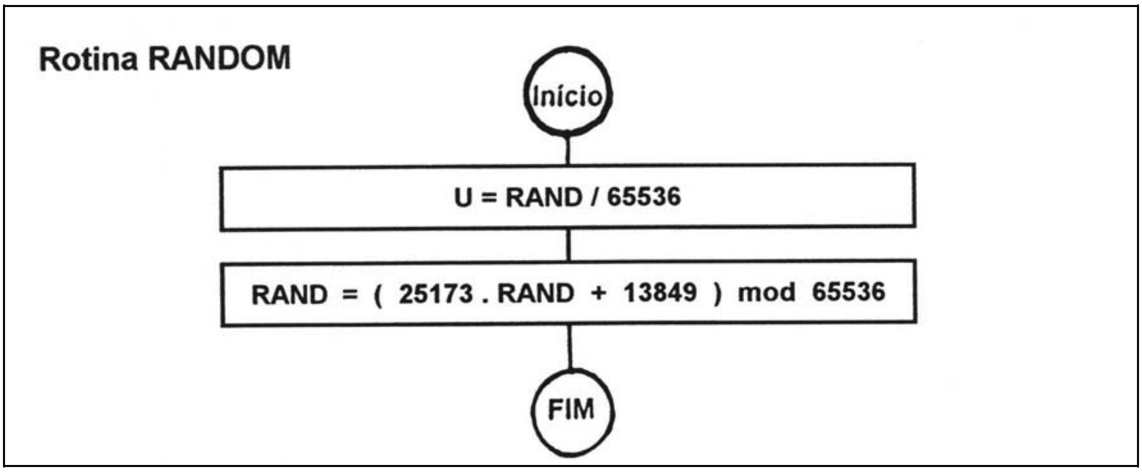
Gostaríamos de descobrir valores para as constantes **a**, **c** e **m** que, a cada *semente*, fizesse corresponder um *ciclo* tão grande quanto possível (isto é, gostaríamos de ter gerado muitos valores antes de se iniciar a repetição).

A utilização do Método Congruencial Misto, com **a = 25173**, **c = 13849** e **m = 65536**, permite-nos obter sequências de 65536 valores inteiros (entre 0 e 65535) para cada semente. Adicionalmente, verifica-se que as sequências são uniformemente distribuídas (entre 0 e 65535) e que se pode admitir a independência entre valores gerados consecutivos.

$$\boxed{x_{i+1} = (25173 . x_i + 13849) \bmod 65536}$$

Se dividirmos cada valor gerado por 65536, **obteremos** um valor pertencente a [0 ; 1 [. Para cada *semente*, poder-se-á, assim, obter uma ***grande*** **sequência** de valores (65536) que tem **comportamento Uniforme [0 ; 1]** e que exibe **independência**, sob ponto de vista estatístico, **entre valores gerados consecutivos**. O gerador assim constituído é **muito simples** e de **fácil implementação informática**, cobrindo com **grande *densidade***, o intervalo [0 ; 1] (o que possibilita a sua utilização em aplicações de maior rigor).

Apresentaremos, em seguida, o fluxograma da **rotina 'RANDOM'** (correspondente ao gerador apresentado), que afecta à variável U o NPA U[0;1] gerado :



Antes da primeira **invocação da 'RANDOM'** a variável RAND deverá ser inicializada com o valor da *semente*. Após a primeira invocação da rotina 'RANDOM' é calculado o primeiro NPA U[0;1], que será igual ao quociente da *semente* por 65536. Em seguida, a variável RAND é 'actualizada', de acordo com a relação recursiva correspondente ao Método Congruencial Misto. Após 65536 invocações da rotina 'RANDOM', a partir de uma cada *semente*, RAND volta a ficar igual à *semente*, pelo que, a partir daí, se repetirá a sequência de NPA gerados.

Assumamos a *semente* RAND = 10000 e determinemos os primeiros valores gerados:

$RAND_{inicial}$	U	$RAND_{final}$
10000	**0,1526**	20073
20073	**0,3063**	28918
28918	**0,4413**	58311

Utilizando a *semente* **10000** e o gerador descrito na rotina 'RANDOM' podemos obter a seguinte **TABELA DE NÚMEROS PSEUDO-ALEATÓRIOS COM DISTRIBUIÇÃO UNIFORME [0 , 1]**:

0,1526	0,3063	0,4413	0,8898	0,0202	0,7723	0,1453	0,7020	0,4005	0,3174
0,0730	0,0229	0,5279	0,8635	0,5836	0,3996	0,8867	0,5824	0,2867	0,6288
0,2114	0,8961	0,0415	0,7574	0,4862	0,4763	0,9575	0,0823	0,9456	0,2451
0,3940	0,1503	0,0740	0,3742	0,9603	0,7584	0,6057	0,2855	0,5159	0,9534
0,5250	0,6509	0,1701	0,0322	0,0665	0,3923	0,2824	0,5417	0,7144	0,6815
0,4620	0,2787	0,2031	0,9405	0,5684	0,2901	0,5170	0,8411	0,7109	0,8728
0,8904	0,3051	0,2489	0,5737	0,8079	0,5045	0,7918	0,1896	0,0036	0,8612
0,1990	0,7674	0,4620	0,0469	0,5798	0,6040	0,4175	0,9836	0,2936	0,2463
0,3545	0,8432	0,9498	0,4911	0,0073	0,0474	0,4077	0,3853	0,3602	0,2683
0,7768	0,2258	0,6478	0,4277	0,4164	0,5595	0,3542	0,6974	0,9359	0,1832
0,2141	0,4993	0,1946	0,1438	0,2115	0,5054	0,3020	0,7383	0,5811	0,6374
0,6173	0,5134	0,8071	0,0673	0,7501	0,2660	0,6240	0,2169	0,5596	0,0430
0,0160	0,7581	0,1552	0,1413	0,2176	0,6127	0,8964	0,1078	0,7131	0,9528
0,3924	0,7394	0,2370	0,2000	0,5027	0,0828	0,7736	0,0264	0,3366	0,4350
0,5572	0,3537	0,2536	0,3429	0,0723	0,3539	0,8629	0,6034	0,0532	0,4488
0,0240	0,2632	0,4843	0,3103	0,8462	0,6194	0,5999	0,8603	0,6889	0,2187
0,8845	0,2706	0,1615	0,8580	0,0724	0,9632	0,1233	0,5843	0,1481	0,6099
0,6838	0,6943	0,3458	0,1325	0,2023	0,7350	0,1499	0,7030	0,2882	0,5033
0,2949	0,7432	0,8008	0,0459	0,7650	0,9251	0,8496	0,6599	0,7947	0,1704
0,7939	0,8920	0,8684	0,6510	0,2430	0,5392	0,7204	0,7889	0,0564	0,6482
0,3351	0,2559	0,3435	0,5161	0,9468	0,9738	0,4635	0,6896	0,0400	0,1146
0,0260	0,9657	0,3491	0,1002	0,8902	0,3912	0,8583	0,6023	0,1655	0,2629
0,8021	0,5429	0,2113	0,1279	0,6648	0,0939	0,6371	0,7828	0,1808	0,6771
0,3020	0,2747	0,3346	0,9644	0,3156	0,9004	0,3604	0,8776	0,0373	0,2067
0,7427	0,5888	0,0763	0,9906	0,2158	0,5196	0,2920	0,2989	0,7640	0,3421
0,7939	0,4285	0,6219	0,9780	0,9413	0,9262	0,2737	0,5993	0,3436	0,5890
0,4538	0,6278	0,8601	0,4637	0,1466	0,7354	0,6003	0,8472	0,5865	0,8439
0,9235	0,2863	0,2578	0,1254	0,4391	0,5780	0,8950	0,0015	0,0064	0,7691
0,4823	0,1443	0,7348	0,1565	0,2542	0,4755	0,9838	0,2870	0,6951	0,1671
0,3626	0,9576	0,5392	0,6409	0,7307	0,2151	0,7711	0,5689	0,1975	0,3564
0,6251	0,8728	0,1222	0,9284	0,5854	0,7245	0,1143	0,7279	0,3867	0,5460
0,0332	0,8018	0,0131	0,0090	0,9883	0,4470	0,6990	0,0356	0,3388	0,2104
0,9290	0,7975	0,6942	0,8888	0,4372	0,7166	0,9138	0,5291	0,2081	0,4649
0,1071	0,4282	0,4761	0,9641	0,6335	0,1330	0,7256	0,3861	0,1021	0,4403
0,6909	0,1529	0,3726	0,3970	0,3563	0,9363	0,5542	0,3000	0,9564	0,6582
0,0063	0,6962	0,9753	0,4904	0,3382	0,3824	0,1476	0,8548	0,4095	0,4056
0,4576	0,4234	0,3293	0,0625	0,1397	0,1179	0,4572	0,9001	0,6782	0,1102
0,4024	0,7152	0,8076	0,8223	0,0244	0,5547	0,5121	0,9260	0,4326	0,7156
0,0262	0,3432	0,9862	0,7446	0,8342	0,2457	0,2948	0,4386	0,6706	0,0555
0,2176	0,8445	0,5197	0,4444	0,8640	0,2593	0,6158	0,3342	0,5934	0,2297
0,4434	0,8969	0,0152	0,5526	0,7368	0,5543	0,9888	0,2749	0,4803	0,9783
0,6235	0,6936	0,9085	0,0908	0,2787	0,3554	0,5056	0,0636	0,5638	0,0572
0,0060	0,3187	0,3382	0,8460	0,3942	0,5278	0,7112	0,0185	0,9044	0,6130
0,0421	0,1217	0,0210	0,7460	0,9405	0,9525	0,4785	0,3487	0,2657	0,5577
0,2611	0,0930	0,1269	0,2341	0,6033	0,9008	0,8837	0,5287	0,9895	0,9440
0,1454	0,2384	0,1540	0,6444	0,7713	0,4100	0,0812	0,6738	0,8708	0,3403
0,0054	0,9543	0,8034	0,5763	0,4111	0,6576	0,1783	0,9149	0,0327	0,2056
0,8550	0,4029	0,8544	0,2703	0,9428	0,3371	0,1107	0,7734	0,8012	0,2646
0,2859	0,8869	0,0394	0,9820	0,1144	0,0325	0,5169	0,5364	0,5807	0,8823
0,3430	0,2246	0,9190	0,3581	0,8771	0,5934	0,6138	0,6317	0,7286	0,4398

De notar que os três primeiros NPA (indicados na primeira linha da Tabela) são os valores que tínhamos determinado 'manualmente' (ver o quadro anterior).

É importante que não nos esqueçamos que, para além das qualidades já referidas da rotina 'RANDOM' apresentada, há uma importante 'limitação': ao fim de 65536 invocações consecutivas desta rotina 'fecha-se o ciclo' e retorna-se ao primeiro NPA gerado ! Assim, em aplicações em que seja previsível a necessidade de gerar mais do que 65536 NPA U[0;1], dever-se-á acrescentar um **'mecanismo de mudança de semente'** ao fim de 65536 invocações da rotina. A falta de tal 'mecanismo' levará a uma repetição da sequência de NPA gerados, que poderá perturbar a qualidade da simulação levada a cabo.

Antes de passarmos aos métodos de 'transformação' dos NPA U[0;1] em NPA com outras distribuições, é importante deixar um reparo à eventual **influência da(s) *semente(s)* nos resultados de uma simulação**. Em alguns casos, os resultados poderão ser influenciados pela *semente* usada para iniciar o processo de geração de NPA. Para averiguar desta eventual influência, é aconselhável efectuar várias simulações de uma mesma 'política/solução' com diferentes *sementes* - um resultado 'global' pode ser obtido a partir dos resultados 'obtidos com as diferentes *sementes*, podendo então avaliar-se da real influência das *sementes* nos resultados da simulação. Refira-se ainda que, em simulações que pela sua duração obriguem à utilização de várias *sementes*, é natural que esta influência seja atenuada ...

E agora abordemos os métodos de 'transformação' dos NPA U[0;1] em NPA com outras distribuições.

• O Método da Inversão

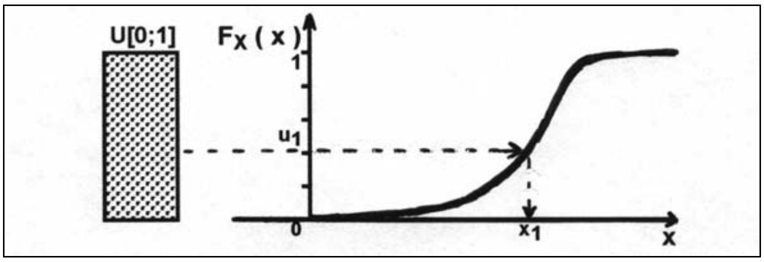
O Método da Inversão é o mais eficiente dos vários métodos que estudaremos. No entanto, como se verá de seguida, este método nem sempre é utilizável ...

Recordemos algumas características da função de distribuição acumulada $\mathbf{F_X (x)}$ relativa à variável aleatória **X** : trata-se de uma função monótona crescente, contínua à direita e que toma valores no intervalo [0 ; 1].

Se a função $\mathbf{F_X (x)}$ for representada por uma expressão analítica [o que nem sempre acontece ... basta lembrarmo-nos da distribuição Normal !] **e** se essa função for invertível [o que, obviamente, está longe de ser 'automático' ...], então poderemos igualar **u**, um NPA U[0;1] a $\mathbf{F_X (x)}$ e, recorrendo à função inversa, determinar **x** - o correspondente NPA da 'família' X:

$$\boxed{\mathbf{u = F_X (x) \quad \Leftrightarrow \quad x = F_X^{-1} (u)}}$$

Na figura seguinte esquematiza-se a aplicação do Método da Inversão:



Ao primeiro NPA U[0;1], $\mathbf{u_1}$, corresponderá o primeiro NPA X, $\mathbf{x_1}$.

Uma repetição do processo esquematizado na figura anterior vai originar uma 'colecção' de valores x_i, que, no seu conjunto, apresentarão a distribuição **X**.

- Exemplifiquemos a aplicação do Método da Inversão, com a distribuição **X ~ Uniforme [a , b]** (com a < b):

$$f_X(x) = \begin{cases} 0 & ;\ x \notin [a, b] \\ 1/(b-a) & ;\ x \in [a, b] \end{cases}$$

$$F_X(x) = \begin{cases} 0 & ;\ x < a \\ \dfrac{(x-a)}{(b-a)} & ;\ x \in [a, b] \\ 1 & ;\ x > b \end{cases}$$

$$\boxed{u = \frac{(x-a)}{(b-a)} \Leftrightarrow x = (b-a).u + a}$$

Assim, por cada NPA U[0;1] gerado, **u**, obter-se-á um correspondente NPA U[a;b], $\mathbf{x = (b-a).u + a}$.

- Exemplifiquemos a aplicação do Método da Inversão, com a distribuição **X ~ Exponencial (λ)** (com $\lambda > 0$; $\mu_X = 1/\lambda$):

$$f_X(x) = \begin{cases} 0 & ;\ x < 0 \\ \lambda . e^{-\lambda . x} & ;\ x \geq 0 \end{cases}$$

$$F_X(x) = \begin{cases} 0 & ;\ x < 0 \\ 1 - e^{-\lambda . x} & ;\ x \geq 0 \end{cases}$$

$$\boxed{u = 1 - e^{-\lambda . x} \Leftrightarrow x = -(1/\lambda) . \ln(1-u)}$$

Como U ~ Uniforme [0 ; 1], U ~ (1 - U) [Porque será ?], pelo que a expressão anterior pode ser substituída pela seguinte:

$$\boxed{x = -(1/\lambda) . \ln(u)}$$

Assim, por cada NPA U[0;1] gerado, **u**, obter-se-á um correspondente NPA Exp (λ), $\mathbf{x = -(1/\lambda) . \ln(u)}$. [<u>Alternativamente</u> poder-se-ia utilizar $x = -(1/\lambda) . \ln(1-u)$]

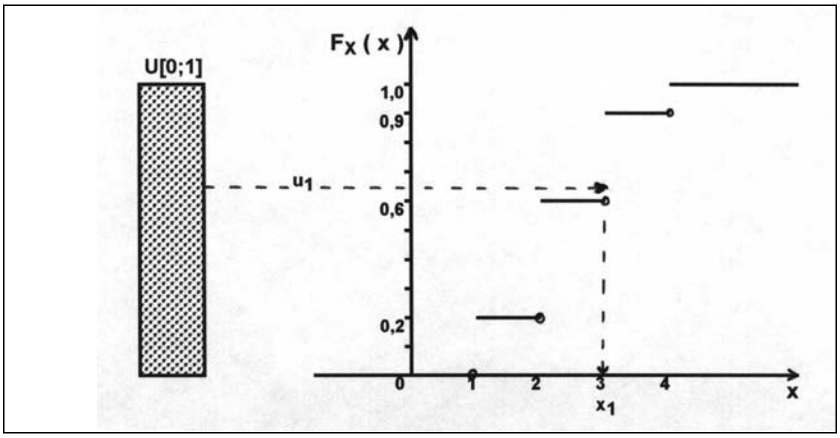
A filosofia subjacente ao Método da Inversão pode ser utilizada para gerar NPA correspondentes a **distribuições discretas**.

- Consideremos a variável aleatória discreta X com função de probabilidade dada pelo quadro seguinte:

x	1	2	3	4
P (X = x)	0,2	0,4	0,3	0,1

A função de distribuição acumulada **F_X (x)** pode representar-se no quadro seguinte:

x	< 1	$\in$ [1 ; 2 [	$\in$ [2 ; 3 [	$\in$ [3 ; 4 [	$\geq$ 4
F_X (x)	0	0,2	0,6	0,9	1,0



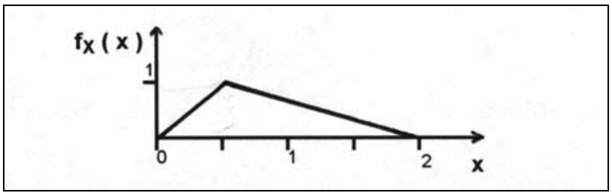
Assim, seguindo o processo esquematizado na figura anterior, começa-se por gerar um NPA U[0;1], **u** e determina-se o correspondente NPA X, **x**, do modo seguinte:

$$\begin{aligned} u \leq 0{,}2 \quad &\Rightarrow \quad x = 1 \\ 0{,}2 < u \leq 0{,}6 \quad &\Rightarrow \quad x = 2 \\ 0{,}6 < u \leq 0{,}9 \quad &\Rightarrow \quad x = 3 \\ u > 0{,}9 \quad &\Rightarrow \quad x = 4 \end{aligned}$$

Como se vê, é muito simples gerar NPA provenientes de distribuições discretas.

Apresentemos um último exemplo de aplicação do Método da Inversão:

- Seja X a variável aleatória contínua, com função de densidade de probabilidade, **f_X (x)**, que se esboça na figura seguinte:



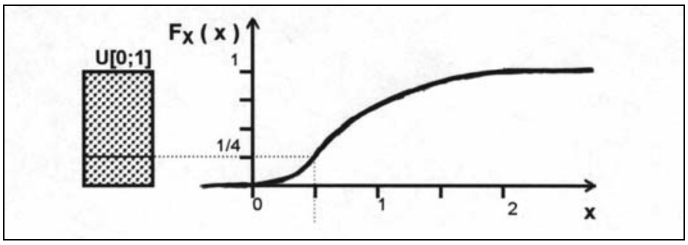
É fácil de obter a expressão analítica da correspondente função de distribuição acumulada, **F_X (x)** [Será que para si é mesmo fácil ? ... Se não é, devia ser ...] . Comecemos por escrever a função de densidade de probabilidade **f_X (x)**:

$$f_X(x) = \begin{cases} 0 & ; \; x \notin [\,0 \,;\, 2\,] \\ 2 \,.\, x & ; \; x \in [\,0 \,;\, 0{,}5\,] \\ (4 - 2\,.\,x)\,/\,3 & ; \; x \in [\,0{,}5 \,;\, 2\,] \end{cases}$$

Poderemos agora obter a função de distribuição acumulada, $F_X(x)$:

$$F_X(x) = \begin{cases} 0 & ; \; x < 0 \\ x^2 & ; \; x \in [\,0 \,;\, 0{,}5\,] \\ (4\,x - x^2 - 1)\,/\,3 & ; \; x \in [\,0{,}5 \,;\, 2\,] \\ 1 & ; \; x \geq 2 \end{cases}$$

Na figura seguinte esquematiza-se a aplicação do Método da Inversão, para a geração de NPA X:



Nota importante: Quando se utiliza o Método da Inversão e a função de distribuição acumulada, $F_X(x)$ é definida por mais do que uma expressão analítica (como acontece no exemplo que se está a considerar), é preciso ter o cuidado de 'inverter' as correspondentes expressões cuidadosamente nos correspondentes domínios de valores de **u**.

Assim, em relação ao exemplo, basta olhar para o esboço da função de densidade de probabilidade para determinar a área do triângulo ' de base [0 ; 0,5] ' (área igual a **1/4**). Este valor é obviamente igual a $F_X(0{,}5)$. Assim, só se 'inverterá' a $u = x^2$ se $u \leq 1/4$. Se $u > 1/4$, dever-se-á 'inverter' $u = (4\,x - x^2 - 1)\,/\,3$.

A inversão de $u = x^2$ não parece ser 'perturbadora'. $x = \pm \sqrt{u}$ parece ser a solução ... Mas, é preciso ter-se o cuidado de reconhecer que, neste caso, a inversão da expressão $u = x^2$ deverá originar valores de **x** pertencentes ao intervalo [0 ; 0,5] ! Assim, a expressão da 'inversa' desejada é $x = +\sqrt{u}$. [Aproveitemos para verificar que, com a expressão anterior, se conclui que se u = 0, então x = 0 e que se u = 1/4, então x = 0,5 , como desejávamos. ☺ É sempre bom verificar-se os resultados ...]

A inversão de $u = (4\,x - x^2 - 1)\,/\,3$ já parece 'dar alguma luta'. $x = 2 \pm \sqrt{3 - 3\,.\,u}$ é a solução 'matemática'... Mas como esta inversão deve originar valores de **x** pertencentes ao intervalo [0,5 ; 2], facilmente concluímos que a expressão da 'inversa' desejada é $x = 2 - \sqrt{3 - 3\,.\,u}$. [Verifiquemos que, com a expressão anterior, se conclui que se u = 1/4 então x = 0,5 e que se u = 1, então x = 2 , o que nos deixa todos satisfeitos ! ☺]

Poderemos, finalmente, esquematizar a rotina de geração de NPA X, baseada no Método da Inversão:

- gerar NPA U[0;1], **u**.
- se **u** ≤ **1/4**, então o NPA X será igual a $\mathbf{x = +\sqrt{u}}$;
  se **u > 1/ 4**, então o NPA X será igual a $\mathbf{x = 2 - \sqrt{3 - 3 \cdot u}}$.

Antes de iniciarmos a apresentação de outro método de geração de NPA, é útil recordarmo-nos que o **Método da Inversão só pode ser utilizado quando existe uma expressão analítica invertível para a função de distribuição acumulada** correspondente. Nessas situações a utilização deste **método de geração é muito eficiente: por cada NPA U[0;1] gerado obtém-se um NPA da distribuição desejada !**

• O Método da Rejeição

Se não existir uma expressão analítica invertível para a função de distribuição acumulada correspondente à distribuição que se pretende gerar não poderemos utilizar o Método da Inversão. No entanto, bastará que se disponha da **expressão analítica da função de densidade de probabilidade** correspondente à distribuição (de domínio limitado) que se pretende gerar para podermos utilizar o **Método da Rejeição**.

Seja X uma distribuição contínua com domínio limitado ao intervalo [a , b]. A utilização do Método da Rejeição para geração de NPA X pode sistematizar-se do modo seguinte:

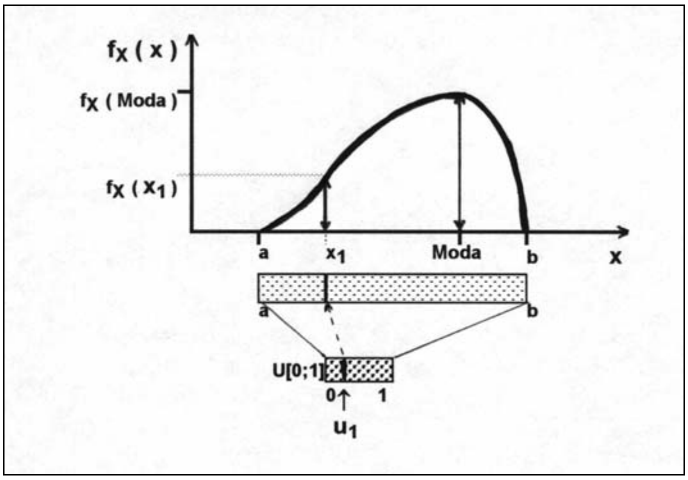
1 - geração de um NPA U[0;1], $\mathbf{u_1}$,

2 - $\mathbf{x_1 = a + (b - a) . u_1}$, (**a** e **b** representam, respectivamente, os limites inferior e superior do domínio de variação de X)

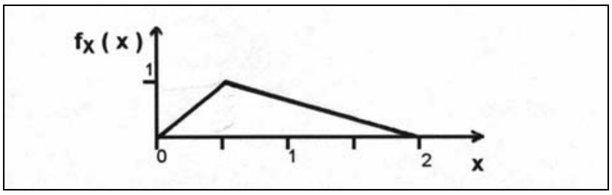
3 - $\mathbf{P_a} = \mathbf{f_X (x_1) \ / \ f_X (moda)}$

4 - geração de um NPA U[0;1], $\mathbf{u_2}$,

5 - se $\mathbf{P_a} < \mathbf{u_2}$, rejeita-se $\mathbf{x_1}$ e retorna-se a **1**;
caso contrário, assume-se $\mathbf{x_1}$ como um NPA X



- Exemplifiquemos a aplicação do Método da Rejeição, com a distribuição **X** cuja função de densidade de probabilidade, **$f_X(x)$**, se esboça na figura seguinte:



$$f_X(x) = \begin{cases} 0 & ;\ x \notin [0\,;2] \\ 2\,.\,x & ;\ x \in [0\,;0{,}5] \\ (4-2\,.\,x)/3 & ;\ x \in [0{,}5\,;2] \end{cases}$$

Relativamente a esta distribuição, podemos indicar que: a = 0 ; b = 2 ; Moda = 0,5 e f_X (Moda) = 1, pelo que a aplicação do Método da Rejeição se traduz nos seguintes passos:

1 - geração de um NPA U[0;1], **u_1**,

2 - **$x_1 = 2\,.\,u_1$** ,

3 - **P_a** $= \begin{cases} \mathbf{2\,.\,x_1} & ;\ \mathbf{x_1 \in [0\,;0{,}5]} \\ \mathbf{(4-2\,.\,x_1)/3} & ;\ \mathbf{x_1 \in [0{,}5\,;2]} \end{cases}$

4 - geração de um NPA U[0;1], **u_2**,

5 - se **P_a** < **u_2**, rejeita-se **x_1** e retorna-se a **1**;
caso contrário, assume-se **x_1** como um NPA X

Utilizemos os vinte primeiros NPA U[0;1] da Tabela apresentada anteriormente e vejamos quantos NPA X obtemos:

u_1	x_1	P_a	u_2	NPA X
0,1526	0,3052	0,6104	**0,3063**	**0,3052**
0,4413	0,8826	0,7449	**0,8898**	—
0,0202	0,0404	0,0808	**0,7723**	—
0,1453	0,2906	0,5812	**0,7020**	—
0,4005	0,8010	0,7993	**0,3174**	**0,8010**
0,0730	0,1460	0,2920	**0,0229**	**0,1460**
0,5279	1,0558	0,6295	**0,8635**	—
0,5836	1,1672	0,5552	**0,3996**	**1,1672**
0,8867	1,7734	0,1511	**0,5824**	—
0,2867	0,5734	0,9511	**0,6288**	**0,5734**

Como se pode observar, o Método da Rejeição é, na melhor das situações, '50 % menos eficiente' do que o Método da Inversão, já que necessita de, pelo menos, dois NPA U[0;1] para gerar um NPA X. Em relação ao exemplo que acabamos de apresentar, **utilizamos vinte NPA U[0;1] para obter apenas cinco NPA X !!!** Assim, sempre que tal for possível, optaremos pelo Método da Inversão, em detrimento do Método da Rejeição !

- Utilizemos o Método da Rejeição para gerar NPA com a distribuição **'Normal'**:

A função de densidade de probabilidade de uma variável aleatória **X ~ Normal (μ ; σ)** é dada por:

$$\mathbf{f_X(x) = 1/(\sqrt{2.\pi}\,.\,\sigma)\ .\ e^{-(1/2)\,.\,[(x-\mu)/\sigma]^2}\ ;\ x \in \Re}\ .$$

A moda de uma variável aleatória **X ~ Normal (μ ; σ)** coincide com o valor médio μ, correspondendo-lhe $\mathbf{f_X(\mu) = 1 / (\sqrt{2.\pi}\ .\ \sigma)}$.

Dado que a distribuição Normal tem domínio ilimitado, não poderemos utilizar o Método da Rejeição para gerar *rigorosamente* a distribuição Normal. No entanto, poderemos truncar as 'caudas' da distribuição em zonas de muito baixa probabilidade de ocorrência (por exemplo, $\mu_X - 3 . \sigma_X$ e $\mu_X + 3 . \sigma_X$) e gerar a distribuição truncada (o que, para muitas aplicações, não será *problemático* ...).

Tomemos, por exemplo a distribuição **X ~ Normal (μ = 10 ; σ = 2)**, cuja função de densidade de probabilidade é dada por:

$$\mathbf{f_X(x) = 1 / (\sqrt{2.\pi}\ .\ 2)\ .\ e^{-(1/2)\,.\,[(x-10)/2]^2}\ ;\ x \in \Re}\ .$$

Se considerarmos as truncaturas em $\mu \pm 3 . \sigma$, ter-se-á: a = 4 ; b = 16 ; Moda = 10 e f_X (Moda) = $1 / (\sqrt{2.\pi}\ .\ 2)$, pelo que a aplicação do Método da Rejeição se traduz nos seguintes passos:

1 - geração de um NPA U[0;1], $\mathbf{u_1}$,

2 - $\mathbf{x_1 = 4 + 12 . u_1}$,

3 - $\mathbf{P_a = e^{-(1/2)\,.\,[(x-10)/2]^2}}$

4 - geração de um NPA U[0;1], $\mathbf{u_2}$,

5 - se $\mathbf{P_a}$ < $\mathbf{u_2}$, rejeita-se $\mathbf{x_1}$ e retorna-se a **1**;
caso contrário, assume-se $\mathbf{x_1}$ como um NPA X

Utilizemos os vinte primeiros NPA U[0;1] da Tabela apresentada anteriormente e vejamos quantos NPA X obtemos:

u_1	x_1	P_a	u_2	NPA X
0,1526	5,8312	0,1139	**0,3063**	—
0,4413	9,2956	0,9399	**0,8898**	9,2956
0,0202	4,2424	0,0159	**0,7723**	—
0,1453	5,7436	0,1039	**0,7020**	—
0,4005	8,8060	0,8368	**0,3174**	8,8060
0,0730	4,8760	0,0376	**0,0229**	4,8760
0,5279	10,3348	0,9861	**0,8635**	10,3348
0,5836	11,0032	0,8818	**0,3996**	11,0032
0,8867	14,6404	0,0678	**0,5824**	—
0,2867	7,4404	0,4409	**0,6288**	—

Em relação ao exemplo que acabamos de apresentar, **utilizamos <u>vinte</u> NPA U[0;1] para obter apenas <u>cinco</u> NPA X ~ Normal (μ = 10 ; σ = 2) !!!**

☞ Imagine que, para determinada aplicação é muito importante o comportamento das 'caudas' da distribuição Normal. Poderíamos 'ampliar' as truncaturas efectuadas para, por exemplo, $\mu \pm \mathbf{4} . \sigma$.

Qual a influência de tal 'ampliação', na eficiência do método de geração ?

☞ Considere a variável aleatória X ~ Exponencial (λ).

Sabe-se que

$$f_X(x) = \begin{cases} 0 & ;\ x < 0 \\ \lambda \cdot e^{-\lambda \cdot x} & ;\ x \geq 0 \end{cases} .$$

Para poder utilizar o Método da Rejeição para gerar NPA X ~ 'Exponencial' (λ) é necessário truncar o domínio de variação de X. Se se adoptar como limite superior do domínio de variação de X o valor $\mu_X + 3 \cdot \sigma_X$, que é igual a ($^4/_\lambda$) [Concorda ?], está a ignorar-se uma pequena parte da 'cauda' da distribuição. Qual a probabilidade correspondente ?

Se pretender diminuir essa probabilidade, poderá 'ampliar' esse limite superior para μ_X + **4** . σ_X , isto é, para ($^5/_\lambda$) - qual a probabilidade correspondente à parte da 'cauda' da distribuição que é truncada ? E qual a influência de tal 'ampliação', na eficiência do método de geração ?

Utilize o Método da Rejeição para gerar cinco NPA X ~ 'Exponencial' (λ = 1).

• O Teorema do Limite Central - Geração de NPA com Distribuição Normal

Como se referiu, não é possível gerar NPA com distribuição Normal recorrendo ao Método da Inversão. Por outro lado, o recurso ao Método da Rejeição não nos permite gerar *rigorosamente* a distribuição Normal, mas apenas *parte* dela, já que há sempre necessidade de definir limites (inferior e superior) para os valores a gerar, que se traduzem na truncatura das 'caudas' da distribuição (e tudo isto a par de uma *baixa* eficiência do método ...).

Recordemo-nos do **Teorema do Limite Central**, que pode ser enunciado de modo muito simples: "A soma de um número elevado de variáveis aleatórias independentes e identicamente distribuídas tende para a distribuição Normal", ou seja:

Sejam $X_1, X_2, \ldots, X_n$ v.a. i.i.d. com $E[X_i] = \mu$ e $Var[X_i] = \sigma^2$,

$S_n \sim X_1 + X_2 + \ldots + X_n \sim$ Normal ($n \cdot \mu$; $n \cdot \sigma^2$)

Aproveitemos para referir que, independentemente da distribuição 'base' das variáveis aleatórias X_i , S_n tende sempre para a distribuição Normal, desde que o número de 'parcelas' seja *suficientemente elevado* ! Claro que se a distribuição 'base' for simétrica, o número de 'parcelas' necessário para se 'atingir' a distribuição Normal tenderá a ser menor do que o número correspondente se a distribuição 'base' for assimétrica ... (Por exemplo, distribuição Uniforme *versus* distribuição Exponencial.)

Dado que conseguimos gerar com facilidade NPA com distribuição Uniforme [0 ; 1], é natural que invoquemos o Teorema do Limite Central utilizando essa distribuição como distribuição 'base':

Sejam $U_1, U_2, \ldots, U_n$ v.a. independentes com distribuição Uniforme [0 ; 1] , isto é, $E[U_i] = {}^1/_2$ e $Var[U_i] = {}^1/_{12}$,

$S_n \sim U_1 + U_2 + \ldots + U_n \sim$ Normal ($^n/_2$; $^n/_{12}$)

Para n igual a 12, já obtemos uma aproximação muito boa à distribuição Normal e, além disso, esse valor de n proporcionar-nos-á valores 'simpáticos' dos parâmetros:

Sejam $U_1, U_2, \dots, U_n$ v.a. independentes com distribuição Uniforme [0 ; 1] .

$S_{12} \sim U_1 + U_2 + \dots + U_{12} \sim$ Normal ($\mu = 6$; $\sigma = 1$)

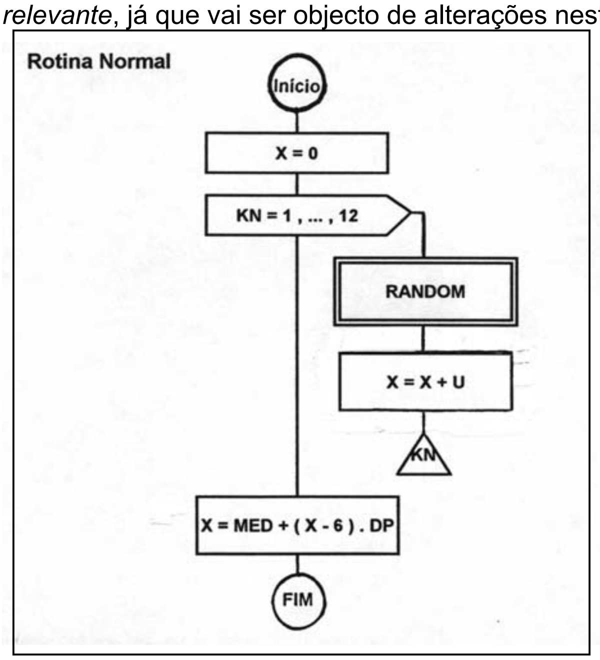
Assim, acabamos de apresentar um método de geração de NPA Normal ($\mu = 6$; $\sigma = 1$): geramos doze NPA U [0 ; 1] e somamos esses valores, obtendo um valor que pode ser considerado um NPA Normal ($\mu = 6$; $\sigma = 1$) !

Claro que se nos interessar gerar NPA com distribuição X ~ Normal (μ_X ; σ^2_X), bastará lembrarmo-nos que:

$$(S_{12} - 6) / 1 \sim (X - \mu_X) / \sigma_X \quad \Leftrightarrow \quad X \sim \mu_X + (S_{12} - 6) \cdot \sigma_X$$

Assim, para gerar um NPA Normal (μ_X ; σ^2_X) começamos por gerar doze NPA U [0 ; 1], somamos esses valores, multiplicamos o valor obtido por σ_X e, finalmente, adicionamos μ_X a esse valor.

Elaboremos, em seguida a **rotina 'Normal'**, que procederá à geração de um NPA Normal (μ_X = MED ; σ_X = DP), afectando o valor gerado à variável **X**. Estamos a assumir que quando esta rotina for invocada, já terão sido previamente afectados às variáveis MED e DP os correspondentes valor médio e desvio padrão da distribuição Normal. Recorde-se que quando se invocar a **rotina 'RANDOM'** é gerado um NPA U[0;1] que é afectado à variável **U**. Para efectuar a soma dos doze NPA U[0;1] utilizaremos um ciclo controlado pela variável **KN**. Para evitar *contratempos* num programa de simulação (que invoque esta rotina) deverá reservar-se a variável KN apenas para utilização *local* nesta rotina. Quando esta rotina for invocada, a variável X não deverá estar a servir para qualquer outro fim *relevante*, já que vai ser objecto de alterações nesta rotina.



E assim temos a **rotina 'Normal'**, que poderemos invocar para proceder à geração de um **NPA Normal (μ = MED ; σ = DP)**.

E pronto ! Agora que já sabemos gerar NPA de várias distribuições, vamos às aplicações da Simulação!

A L G U M A S   A P L I C A Ç Õ E S   D A   S I M U L A Ç Ã O

Nesta secção apresentaremos algumas aplicações da Simulação. Começaremos por abordar a "**Simulação Aplicada à Gestão de Recursos Hídricos**" (retomando o problema "A Barragem de Pós-Boa" numa versão mais geral), ilustraremos o Método de Monte Carlo com a "**Estimação de uma Distribuição de Mínimos**", apresentaremos um exercício de "**Simulação Aplicada à Gestão de Projectos**" e um outro de "**Simulação Aplicada à Gestão de Stocks**", introduziremos a "**Simulação de Processos de Poisson**" e terminaremos com uma "**Introdução à Simulação de Filas de Espera**".

• Simulação Aplicada à Gestão de Recursos Hídricos

Comecemos por considerar o **problema "A Barragem de Pós-Boa (II)" :**

A Barragem de Pós-Boa é relativamente pequena, com uma capacidade máxima de retenção, **VMAX**, de 100 u.vol. de água.

A equação simplificada de "balanço hidrológico mensal" é

$$V_f = V_0 + P - T - N - DS,$$

sendo

V_f — Volume no final do mês (u.vol.)
V_0 — Volume no início do mês (u.vol.)
P — Volume de precipitação mensal (u.vol.)
T — Volume turbinado durante o mês (u.vol.)
N — Volume para satisfação de necessidades de abastecimento mensal (u.vol.)
DS — Volume de descarga de superfície durante o mês (u.vol.)

Por razões de equilíbrio ecológico e ambiental nunca se pode permitir que o nível da barragem desça abaixo de dez por cento da capacidade máxima de retenção.

As necessidades mensais de água a jusante, **Nec** (u.vol.), têm distribuição Normal, com valor médio e desvio padrão variáveis com o mês, de acordo com a tabela seguinte:

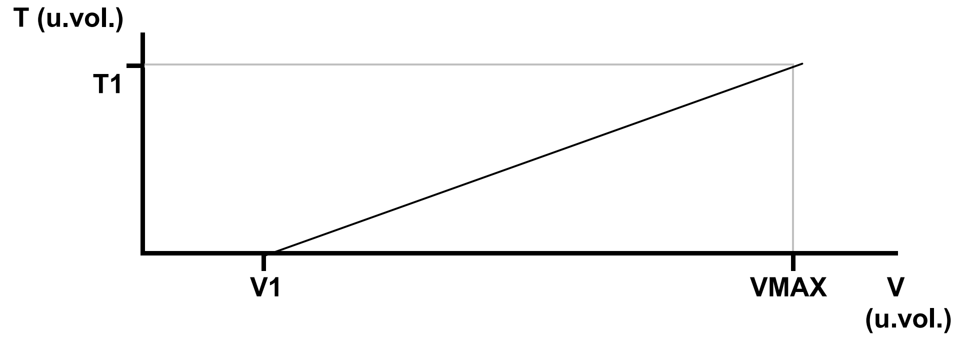
Mês	J	F	M	A	M	J	J	A	S	O	N	D
μ_{Nec} (u.vol.)	8	8	8	9	10	13	15	15	15	13	9	8
σ_{Nec} (u.vol.)	1,0	0,8	0,7	0,8	0,9	1,1	1,2	1,3	1,1	0,7	0,8	0,9

No <u>início do mês</u>, o gestor da barragem decide qual o volume de água que vai destinar à satisfação das necessidades previstas para esse mês, com base no volume de água então disponível na barragem. Sempre que possível (isto é, sempre que não seja violada a condição de equilíbrio ambiental) o volume de água destinado à satisfação das necessidades será igual ao valor previsto (indicado na tabela anterior).

Em função do volume, **V**, de água na barragem no início do mês após a afectação do volume de água destinado à satisfação das necessidades o gestor da barragem determina o volume de água a turbinar durante esse mês, de acordo com o gráfico seguinte:

(continua)

(continuação)



O gestor da barragem pretende definir o valor dos parâmetros **V1** e **T1** da política de gestão de água a turbinar.

Por razões de segurança exige-se que V1 não seja inferior a vinte por cento da capacidade máxima de retenção.

O lucro mensal L (u.m.) associado ao volume turbinado T (u.vol.) é dado pela função $\mathbf{L = T^2}$.

Sabe-se que uma descarga total (T + DS) superior ou igual a 30 u.vol. origina situações de cheia a jusante.

Para decidir quais os valores a adoptar para os parâmetros **V1** e **T1** o gestor consultou os seus registos com os valores das precipitações dos últimos 50 anos, a partir dos quais determinou as seguintes distribuições mensais (em u.vol.):

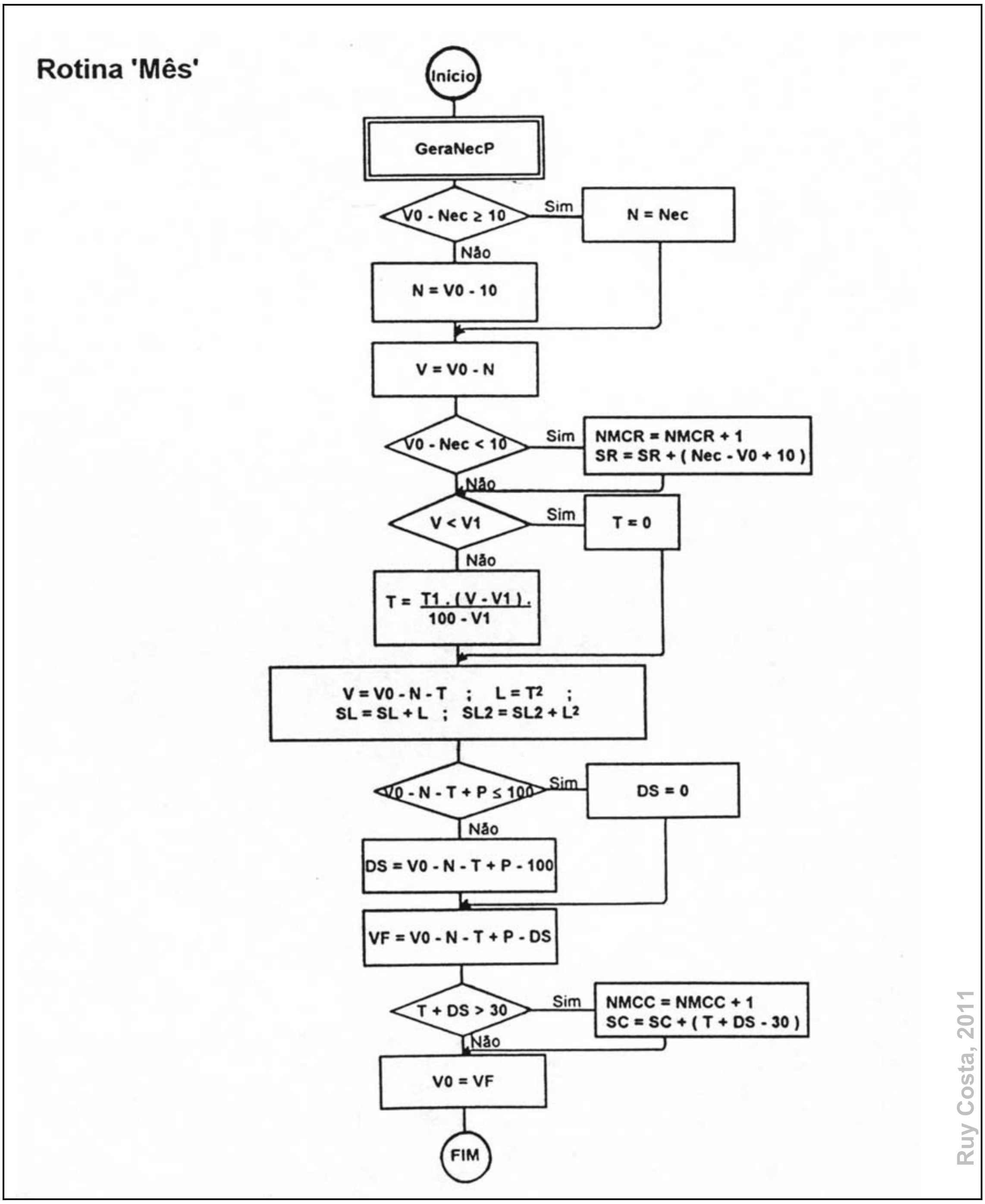
Mês	Distribuição da Precipitação Mensal (u.vol.)
J	Normal (μ = 30 ; σ = 2,5)
F	Normal (μ = 40 ; σ = 3,0)
M	Normal (μ = 40 ; σ = 3,5)
A	Normal (μ = 30 ; σ = 2,0)
M	Uniforme [20 ; 30]
J	Uniforme [5 ; 15]
J	Uniforme [0 ; 10]
A	Uniforme [0 ; 5]
S	Uniforme [6 ; 18]
O	Normal (μ = 30 ; σ = 2,5)
N	Normal (μ = 30 ; σ = 2,5)
D	Normal (μ = 30 ; σ = 2,5)

Elabore um modelo de simulação que lhe permita aconselhar o gestor da barragem, relativamente aos valores a adoptar para os parâmetros **V1** e **T1** da política de gestão de águas referida.

Como se pode observar, esta segunda versão do problema "A Barragem de Pos--Boa" pouco difere da primeira. Acrescentou-se algum 'realismo', relativamente à primeira versão, admitindo que quer as necessidades de água a jusante, quer a precipitação são variáveis aleatórias, com distribuições variáveis com o mês do ano. Assim, deixamos de estar num cenário de **simulação com dados históricos** (como acontecera na primeira versão), para passarmos a estar perante uma ***verdadeira*** **simulação** em que os dados estão modelados com distribuições estatísticas. Poderemos, assim, simular um horizonte temporal tão grande quanto desejarmos, bastando para tal, proceder à correspondente geração de valores pseudo-aleatórios para as necessidades e para a precipitação.

Se nos recordarmos do fluxograma apresentado na secção **II** destes apontamentos e dos comentários então feitos, bastará substituir a primeira instrução após o início do ciclo 'Mês' "**Nec = Nec (Mês,Ano) ; P = P(Mês,Ano)**" por "**Gerar P e Nec, de acordo com as correspondentes distribuições estatísticas**". É ainda conveniente acumularmos o quadrado dos valores mensais de lucro (em SL2, previamente inicializada a zer) para uma posterior avaliação da variabilidade desta medida de desempenho. É claro que será necessário alterar o ciclo 'Ano' (que deixará de decorrer entre 1998 e 1994) e, correspondentemente, o apuramento dos valores médios associados às medidas de desempenho.[Recorda-se que havia divisões por 36, já que eram 36 os meses então simulados].

Comecemos, então, por apresentar o fluxograma da **rotina 'Mês'**, o 'coração' do nosso modelo:



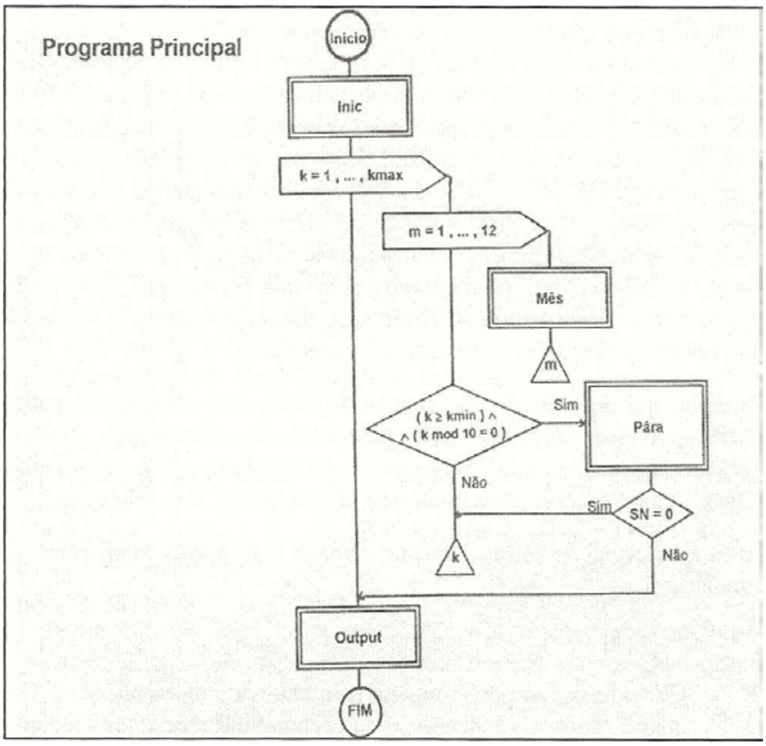
Uma **rotina** é um conjunto de instruções que se 'destaca' num dado **programa** (designação dada ao total das instruções). Figurativamente, se compararmos um programa ao corpo humano, poderemos imaginar as rotinas como importantes orgãos desse corpo, que são comandadas pelo '**programa principal**' (o cérebro). De notar que uma rotina pode ser invocada por outra rotina ! Assim, um dado programa é constituído por um 'programa principal' e por várias 'rotinas' (que desempenham algumas tarefas específicas). O 'programa principal' é um conjunto de instruções, sendo algumas delas invocações (ou chamadas) das 'rotinas'. **Um modelo bem concebido está, geralmente, associado a um 'programa principal' relativamente curto**, de fácil leitura e compreensão.

No nosso modelo, a **rotina 'Mês'** constitui o conjunto de instruções que deve ser executado todos os meses. Assim, podemos considerar esta rotina como o verdadeiro 'coração' do nosso modelo. De realçar que a **rotina 'Mês'** começa por invocar a **rotina 'GeraNecP'**, que procederá à geração de números pseudo-aleatórios que são afectados às variáveis **Nec** e **P**, correspondendo, respectivamente, aos valores gerados das necessidades de água a jusante e da precipitação para esse mês. (Cá está um exemplo de uma rotina que invoca uma outra rotina !). Deixaremos para breve a representação do fluxograma desta rotina.

E, por falar em rotinas, que outras rotinas poderemos conceber para simplificar ao máximo o 'programa principal' ?

Poderemos conceber a **rotina 'Inic'** destinada a proceder às inicializações das variáveis. Por outro lado, poderemos criar a **rotina 'Output'** destinada a proceder ao cálculo dos valores correspondentes às Medidas de Desempenho do Sistema e sua posterior impressão. Poderemos ainda criar a **rotina 'Pára'** que, quando invocada, verifica se estão cumpridas as condições para se terminar a simulação.

Esbocemos, agora, o fluxograma do nosso **programa principal** :



Comecemos por realçar a extrema **simplicidade** do **programa principal** apresentado, cuja leitura se faz com grande facilidade:

Começa-se por invocar a **rotina 'Inic'**, que assegurará a inicialização das variáveis. Em seguida, dá-se inicio ao **ciclo 'Ano'**, controlado pela variável k, que toma valores inteiros de 1 a kmax (Atenção: Não nos poderemos esquecer de inicializar kmax !). Cada ano corresponde a 12 meses - o ciclo 'Mês' é controlado pela variável m (m = 1, 2, ..., 12) e limita-se a invocar a **rotina 'Mês'** (o 'coração' do nosso modelo).

Terminada a simulação de um ano (genericamente, do k-ésimo ano), investiga-se se se deve (ou não) avaliar a precisão dos resultados com vista à paragem da simulação: pergunta-se se já se atingiu ou ultrapassou a duração mínima da simulação (kmin - mais uma variável a ser inicializada !) **e** se, simultaneamente, k é múltiplo de 10 (far-se-ão avaliações da precisão de 10 em 10 anos). [Perguntar "k mod 10 = 0 ?" é equivalente a perguntar "O resto da divisão inteira de k por 10 é 0 ?", ou seja, "k é múltiplo de 10 ?"]. Se a resposta for afirmativa, invoca-se a **rotina 'Pára'** que vai investigar se a amplitude do intervalo de confiança para o valor médio do Lucro é menor ou igual a um valor pré-fixado $\mathbf{Amp_{max}}$ (mais uma variável a inicializar !). A resposta (**S**im = 1 ; **N**ão = 0) será codificada na variável **SN**. Se a resposta for negativa (isto é, se SN = 0), a simulação deve prosseguir, pelo que se muda de 'Ano', incrementando a variável k (final do ciclo controlado pela variável k); se a resposta for afirmativa (isto é, se SN = 1, ou seja, se a amplitude do intervalo de confiança para o valor médio do Lucro for menor ou igual a $\mathbf{Amp_{max}}$, tendo-se já simulado um múltiplo de 10 anos igual ou superior a kmin anos), pode-se parar a simulação, pelo que se sái do ciclo 'Ano' directamente para a **rotina 'Output'**, que tratará do cálculo e impressão dos valores das medidas de desempenho do sistema.

De notar que se tiverem sido simulados kmax anos (sem que a **rotina 'Pára'** tivesse ordenado a paragem da simulação) termina obrigatoriamente o ciclo 'Ano', sendo de imediato invocada a **rotina 'Output'** para apuramento e impressão de resultados.

Após a execução da **rotina 'Output'** termina a execução do programa de simulação.

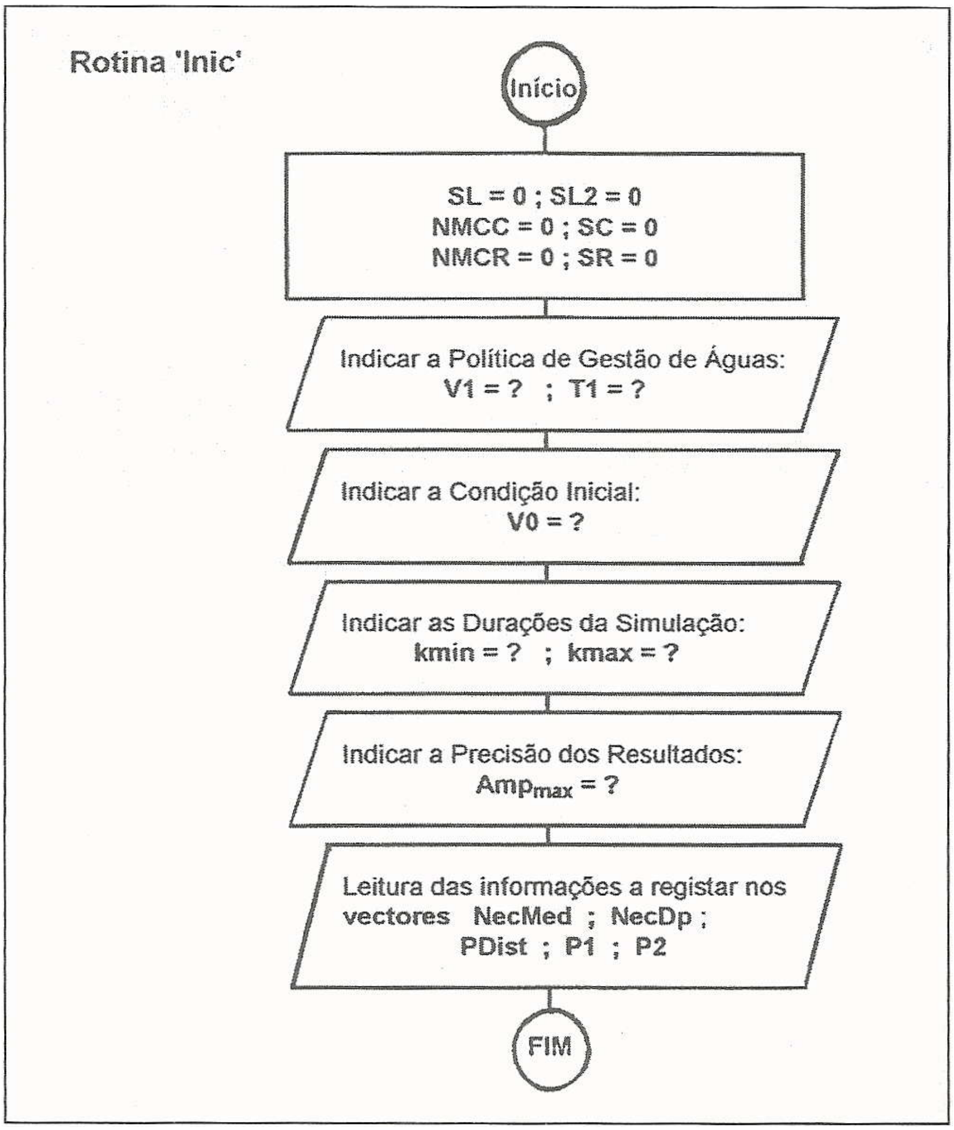
☞ De notar que, no programa apresentado não comtemplamos a 'fase inicial' destinada à atenuação da eventual influência das condições iniciais. A consideração desse 'período inicial, obrigaria à definição (inicialização) de uma variável k0 e, na rotina 'Mês', antes de se acumular valores em SL, SL2, SR, SC, NMCC e NMCR dever-se-ia investigar se k (o ano a ser simulado) era superior a k0. Só em caso afirmativo se procederia à acumulação respectiva. Seria também necessário garantir um cálculo correcto (nas rotinas 'Pára' e 'Output') das grandezas estatísticas relativas ao Lucro, já que o número de anos *efectivos* simulados seria igual ao número de anos simulados diminuído de k0 !

☞ E que tal proceder às alterações referidas nos fluxogramas ?

Para terminarmos o nosso modelo de simulação d' "A Barragem de Pós-Boa" só nos falta explicitarmos as **rotinas 'Inic'**, **'Pára'**, **'Output'** e **'GeraNecP'**.

Recordemos as variáveis que devem ser inicializadas na **rotina 'Inic'**. **SL**, **SL2**, **NMCC**, **SC**, **NMCR** e **SR** são variáveis às quais deve ser atribuído inicialmente o valor 0. Deve pedir-se ao utilizador para indicar os valores relativos à política de gestão de águas a testar (**V1** e **T1**), o valor relativo ao volume inicial de água na albufeira (**V0**), as durações mínima e máxima da simulação (**kmin** e **kmax**) e, finalmente, a amplitude máxima permitida para o intervalo de confiança a 95 % para o valor médio do Lucro ($\mathbf{Amp_{max}}$). Para facilitar a geração dos valores pseudo-aleatórios correspondentes às necessidades de água mensais a jusante e às precipitações mensais é conveniente criar alguns **vectores** (cada um deles **com 12 posições: 1 = Janº ; ... ; 12 = Dezº**): os vectores **NecMed** e **NecDp** permitem um registo mais fácil, respectivamente, do valor médio e do desvio padrão das necessidades mensais (que, recorda-se, seguem distribuições Normais); no vector **PDist** registaremos 1 quando a distribuição correspondente à precipitação de dado mês for Normal e 2 quando essa distribuição for Uniforme; nos vectores **P1** e **P2** registar-se-ão os valores dos dois parâmetros intervenientes nessas distribuições - se se tratar da distribuição **Normal (μ ; σ)**, registaremos, para um dado mês, o valor médio μ correspondente no vector P1 e o desvio padrão σ correspondente no vector P2; se se tratar da distribuição **Uniforme [a ; b]**, registaremos, para um dado mês, o limite mínimo **a** correspondente no vector P1 e o limite máximo **b** correspondente no vector P2.

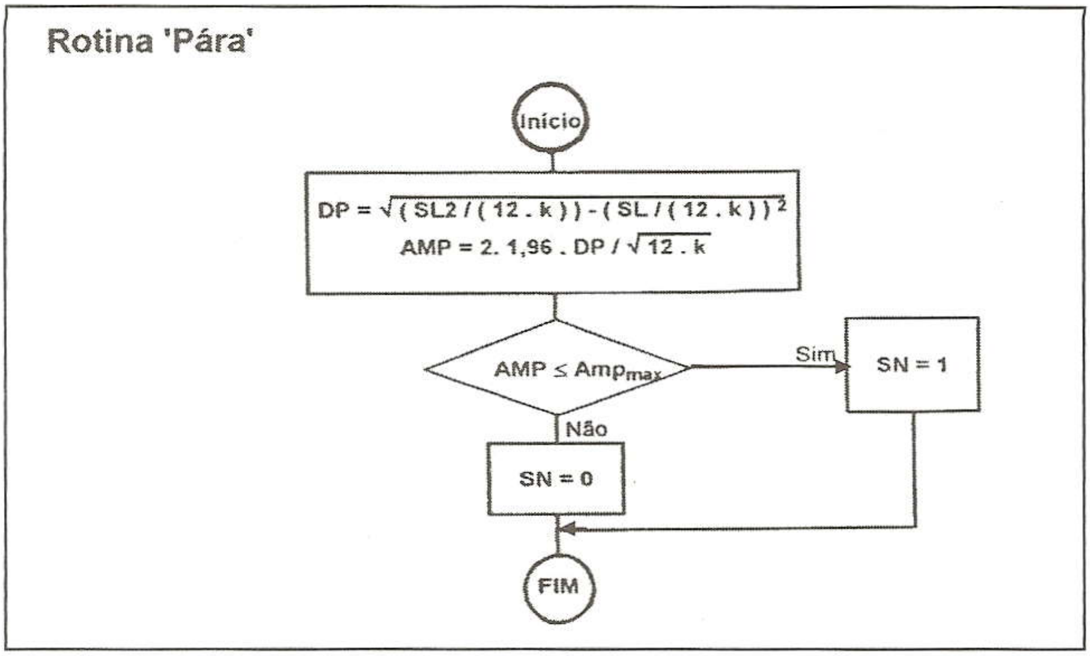
Esbocemos, então, o correspondente fluxograma:



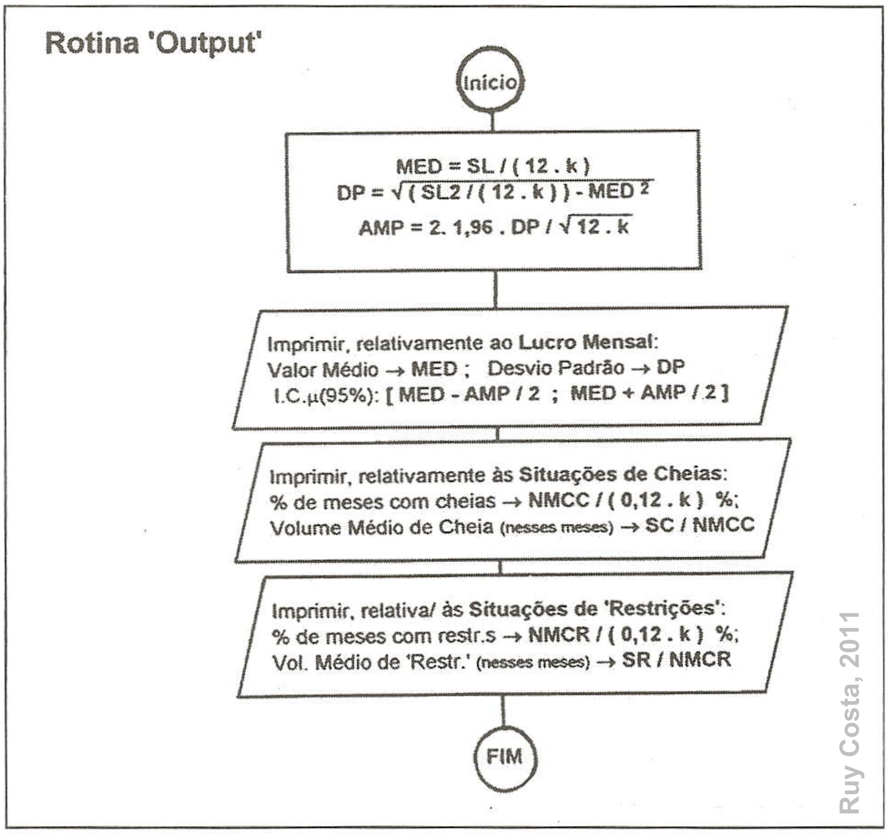
Para que não fique qualquer dúvida, indica-se, em seguida, os valores a registar nos 5 vectores referidos:

Mês	J (1)	F (2)	M (3)	A (4)	M (5)	J (6)	J (7)	A (8)	S (9)	O (10)	N (11)	D (12)
NecMed	8	8	8	9	10	13	15	15	15	13	9	8
NecDp	1,0	0,8	0,7	0,8	0,9	1,1	1,2	1,3	1,1	0,7	0,8	0,9
PDist	1	1	1	1	2	2	2	2	2	1	1	1
P1	30	40	40	30	20	5	0	0	6	18	29	30
P2	2,5	3,0	3,5	2,0	30	15	10	5	18	2,0	2,5	3,0

Para elaborarmos a **rotina 'Pára'** devemos recordar que a amplitude do intervalo de confiança a 95 % para o valor médio do Lucro é igual a 2 . 1,96 . s' / √n , sendo s' uma estimativa pontual do desvio padrão do Lucro mensal (que se obtém facilmente a partir de SL e SL2). Assim, o fluxograma desta rotina resume-se a:

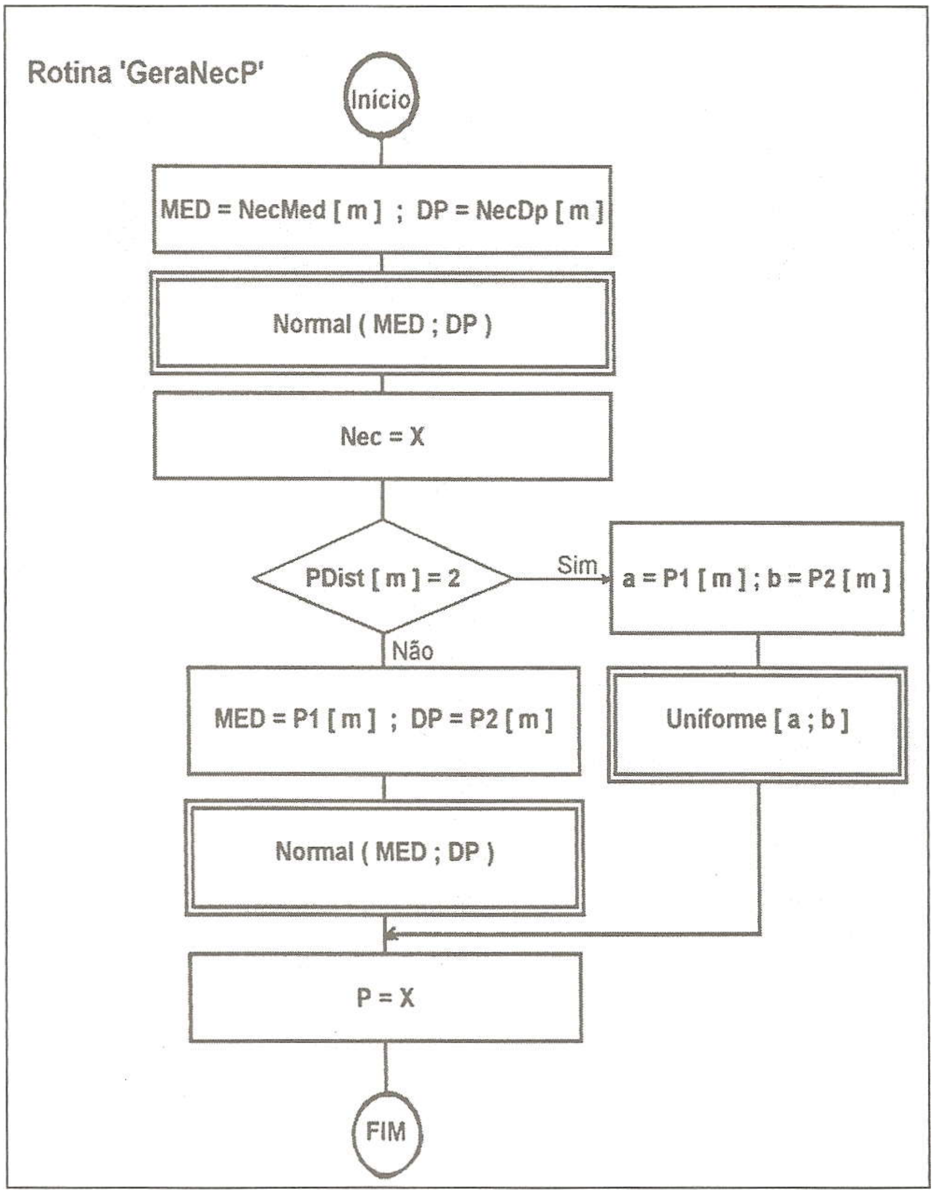


Antes de avançarmos para o fluxograma da **rotina 'Output'** , chama-se a atenção para o facto de, quando a rotina "Pára" (ou a rotina "Output") é invocada, o número de anos simulados é igual a k, sendo o número de 'observações' de lucro mensal correspondentes igual a 12 . k .



Refira-se que os resultados apurados na rotina 'Output' relativos a lucros são expressos em u.m. e que os volumes são expressos em u.vol. .

Finalizaremos a concepção do modelo de simulação d' "A Barragem de Pós-Boa" com a apresentação do fluxograma da **rotina 'GeraNecP'**. Esta rotina invoca as **rotinas 'Normal (μ ; σ)' e rotina 'Uniforme [a ; b]'** tal como foram definidas na secção anterior destes apontamentos.



Recordemos que quer a rotina 'Normal', quer a rotina 'Uniforme' (tal como foram definidas na secção anterior) agora inV0cadas pela rotina 'GeraNecP' afectam o número pseudo-aleatório gerado à variável X. Assim, teve-se o cuidado de afectar as variáveis **Nec** e **P** os correspondentes valores pseudo-aleatórios de Necessidades e Precipitação mensal gerados e correspondentes ao mês genérico **m** (de recordar que o ciclo 'Mês' é controlado pela variável m !).

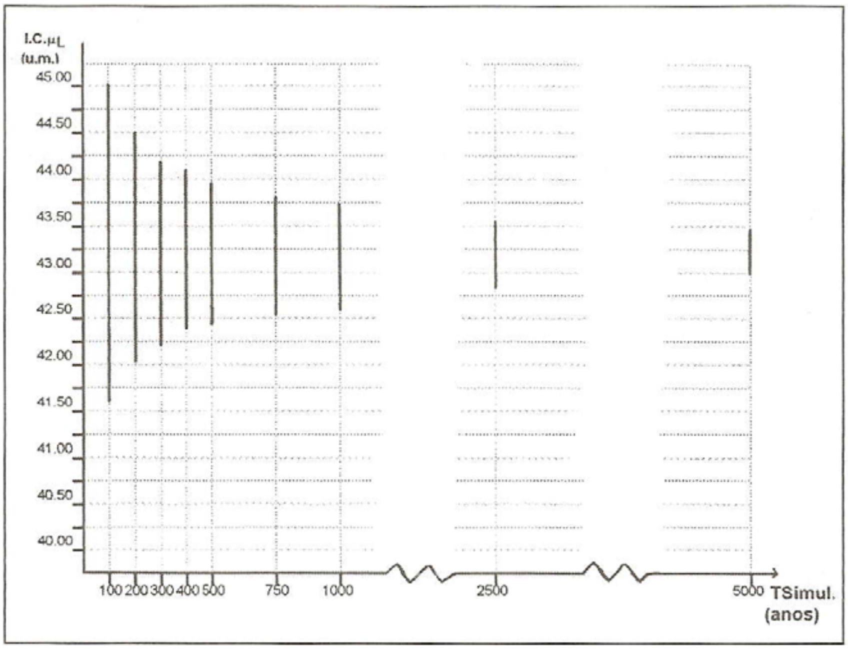
E assim fica completamente especificado o nosso **modelo de simulação d' "A Barragem de Pós-Boa", que, para alem do programa principal, inclui as rotinas 'Inic', 'GeraNecP', 'Mês', 'Pára' e 'Output'**.

☞ E se sabe programar computadores, meta maos a' obra e converta Os fluxogramas apresentados num programa e... divirta-se !

Antes de prosseguirmos, gostaríamos de chamar a atenção para um **aspecto muito importante: na implementação informática da rotina "INIC" do modelo apresentado não nos deveríamos esquecer de <u>inicializar o processo de geração de números pseudo-aleatórios</u>, indicando um valor para 'semente' desse processo !** No entanto, a nível de fluxograma, entendemos não ser imperativa a indicação expressa desta inicialização.

Aproveitaremos Os resultados da simulação com o modelo desenV0lvido para "A Barragem de Pós-Boa" (II) para comentarmos algumas quest5es referidas anteriormente.

No segundo tópico deste capítulo, já havíamos referido um aspecto muito importante a considerar na simulação: **à medida que se aumenta a duração da simulação, a diminuição da amplitude dos intervalos de confiança para os valores médios das medidas de desempenho do sistema é cada vez menor** (embora continue a ocorrer), isto é, **verifica-se um acréscimo de precisão dos resultados cada vez menor.** Na figura seguinte esboça-se a variação da amplitude do intervalo de confiança para o valor médio do Lucro Médio Mensal, L, com a duração da simulação, para a política V1 = 30 ; T1 = 10 (com V0 = 50):



Como se pode observar, a amplitude do intervalo para T = 100 anos é aproximadamente dupla da correspondente a T = 500 anos ! Com efeito, simular 100 anos corresponde a obter um intervalo com 3,44 u.m. de amplitude, que passa a 1,54 para 500 anos. Se duplicarmos a 'duração' da simulação para 1000 anos, a amplitude do intervalo baixará para 1,09 u.m. . Provavelmente, para a maior parte das aplicações, já não se justificaria este esforço de simulação/computação adicional ! Assim, seria natural fixar a 'duração' desta simulação nos 500 anos, já que a correspondente precisão dos resultados parece aceitável para a maior parte das aplicações.

É interessante notar que para reduzir em aproximadamente 50 % a amplitude do intervalo correspondente a T = 100 anos, tivemos que aumentar a duração da simulação em <u>mais 400 anos</u> (3,44 → 1,54; 55,2 %); uma nova redução em aproximadamente 50 % implica um acréscimo da duração da simulação de aproximadamente <u>2000 anos</u> (1,54 → 0,69; 55,2 %) e, se aumentarmos a duração da simulação em mais <u>2500 anos</u>, a redução da amplitude do intervalo já nem atingirá Os 30 % (0,69 ~ 0,49; 29,0 %) !

Também no segundo tópico deste capítulo, já havíamos referido que é importante avaliar a **influência das condições iniciais nos resultados obtidos** com um modelo de simulação. No modelo desenV0lvido, apenas se tem que fixar o valor de V0 (V0lume de água inicial na albufeira).

No Quadro seguinte apresentamos os resultados obtidos nas simulações das duas políticas que destacamos na secção **I** (simulação com dados históricos): (V1 = 30 ; T1 = 10) e (V1 = 60 ; TI = 30). Consideraremos três valores para a condição inicial V0 = 25, 50 e 75 (u.V0l.) e nove valores de duração da simulação T = 100, 200, 300, 400, 500, 750,1000, 2500 e 5000.

T Simul	V0	V1 = 30 ; T1 = 10			V1 = 60 ; T1 = 30		
		I.C.95%(μL)	**%Cheia**	**VCheia**	**I.C.95%(μL)**	**%Cheia**	**VCheia**
100	25	[41,53; 44,97]	12,3 %	3,3	[197,02; 222,90]	8,5 %	3,3
	50	[41,58; 45,02]	12,3 %	3,3	[197,41; 223,30]	8,6 %	3,2
	75	[41,65; 45,08]	12,3 %	3,3	[197,77; 223,66]	8,6 %	3,2
200	25	[42,03; 44,47]	11,6 %	3,4	[201,83; 220,14]	8,0 %	3,5
	50	[42,06; 44,49]	11,6 %	3,4	[202,03; 220,34]	8,1 %	3,4
	75	[42,09; 44,52]	11,6 %	3,4	[202,21; 220,52]	8,1 %	3,4
300	25	[42,19; 44,18]	11,5 %	3,4	[203,46; 220,14]	8,0 %	3,3
	50	[42,21; 44,20]	11,5 %	3,4	[203,59; 220,34]	8,0 %	3,3
	75	[42,23; 44,22]	11,5 %	3,4	[203,71; 218,69]	8,0 %	3,3
400	25	[42,36; 44,09]	11,3 %	3,4	[204,90; 217,89]	7,9 %	3,3
	50	[42,38; 44,10]	11,3 %	3,4	[205,00; 217,99]	7,9 %	3,3
	75	[42,39; 44,11]	11,3 %	3,4	[205,09; 218,08]	7,9 %	3,3
500	25	[42,39; 43,93]	11,3 %	3,4	[205,15; 216,74]	7,9 %	3,4
	50	[42,40; 43,94]	11,3 %	3,4	[205,23; 216,82]	7,9 %	3,4
	75	[42,41; 43,95]	11,3 %	3,4	[205,30; 216,90]	7,9 %	3,4
750	25	[42,53; 43,79]	11,3 %	3,4	[206,28; 215,75]	8,0 %	3,3
	50	[42,54; 43,80]	11,3 %	3,4	[206,33; 215,80]	8,0 %	3,3
	75	[42,54; 43,80]	11,4 %	3,4	[206,38; 215,85]	8,0 %	3,3
1000	25	[42,65; 43,74]	11,5 %	3,4	[206,82; 215,02]	8,0 %	3,3
	50	[42,65; 43,74]	11,5 %	3,4	[206,86; 215,06]	8,0 %	3,3
	75	[42,66; 43,75]	11,5 %	3,4	[206,90; 215,10]	8,0 %	3,3
2500	25	[42,84; 43,53]	11,6 %	3,5	[208,25; 213,44]	8,2 %	3,3
	50	[42,84; 43,53]	11,6 %	3,5	[208,27; 213,45]	8,2 %	3,3
	75	[42,85; 43,54]	11,6 %	3,5	[208,28; 213,47]	8,2 %	3,3
5000	25	[42,96; 43,45]	11,6 %	3,5	[208,97; 212,64]	8,2 %	3,4
	50	[42,96; 43,45]	11,6 %	3,5	[208,98; 212,65]	8,2 %	3,4
	75	[42,96; 43,45]	11,6 %	3,5	[208,99; 212,65]	8,2 %	3,4

Notas: 1) Para estas duas políticas, nunca foram observadas situações de 'restrições'.
2) T Simul (anos) ; V0 (u.V0l.) ; I.C.95%(μL) (u.m.) ; VCheia (V0lume médio de cheia nos meses com cheias; em u.V0l.).

De notar que, como se esperava, com o aumento da duração da simulação, diminui a influência da condição inicial. Com efeito, para uma simulação de 100 anos observa-se uma 'clara' influência de V0, no tocante ao Lucro – à medida que V0 aumenta, também aumenta o Lucro Médio Mensal. Refira-se que, no entanto, as diferenças observadas são <u>mínimas</u>, não atingindo 1 u.m. ! Quando se simulam 1000 anos essa influência praticamente desapareceu ... De notar ainda que para as medidas de desempenho '%Cheias' e Vol.Med.Cheia' praticamente não se detectou qualquer influência de V0.

O Quadro anterior permite ainda observar o aumento de precisão que está associado ao aumento da duração da simulação (que já se referira). Aparentemente, esse aumento parece ser maior quando se estuda a política "Vi = 30 ; T1 = 10" do que quando se considera "V1 = 60 ; T1 = 30". Tal, no entanto, é apenas *aparentemente...* Com efeito, as diferenças observadas entre as duas séries de resultados devem-se ao facto de, no primeiro caso, os valores do lucro médio mensal se situarem cerca das 40 u.m., enquanto que, no segundo caso, rondam as 200 u.m. ... Assim, a amplitude observada de 0,49 u.m. para a primeira política (e T = 5000 anos) corresponde a aproximadamente 1,1 % do valor médio do intervalo; para a segunda política (e mesma duração da simulação) observamos uma amplitude de 3,67 u.m. que corresponde a 1,7 % do valor médio do intervalo. Como se vê, não há uma grande diferença...

Finalmente, é legitimo pretender-se obter uma perspectiva dos resultados da simulação com este modelo para as diferentes políticas que se pode adoptar. No Quadro seguinte apresentamos Os resultados obtidos nas simulações correspondentes às políticas traduzidas por V1 = 20, 30, …, 90 / T1 = 10, 20, …, 90 (T = 500 anos / V0 = 50 u.vol.):

V1→ T1↓	20		30		40		50		60		70		80		90	
10	(45,87)		(43,17)		(40,29)		(36,82)		(32,22)		(25,72)		(15,73)		(1,50)	
	10,5 % **3,4**	0 % –	**11,3 %** **3,4**	0 % –	**11,8 %** **3,4**	0 % –	**11,8 %** **3,4**	0 % –	**11,8 %** **3,4**	0 % –	**11,8 %** **3,4**	0 % –	**11,8 %** **3,4**	0 % –	**11,8 %** **3,4**	0 % –
20	142,76		138,24		132,68		(125,83)		(114,95)		(96,54)		(60,86)		(6,00)	
	4,8% 3,6	0 % –	5,9% 3,6	0 % –	7,3% 3,4	0 % –	**10,0 %** **3,3**	0 % –	**11,6 %** **3,4**	0 % –	**11,8 %** **3,4**	0 % –	**11,8 %** **3,4**	0 % –	**11,8 %** **3,4**	0 % –
30	224,72		226,17		225,40		219,86		211,03		(192,12)		(130,90)		(13,51)	
	0,1% 1,6	5,5% 2,6	1,1% 2,2	0,8% 1,7	3,9% 2,9	0,0% 0,3	5,8% 3,6	0 % –	7,9% 3,4	0 % –	**10,9 %** **3,3**	0 % –	**11,8 %** **3,4**	0 % –	**11,8 %** **3,4**	0 % –
40	(262,52)		260,36		267,76		276,89		(277,77)		261,80		203,61		(24,01)	
	1,5% 0,9	**12,8 %** **5,2**	3,0% 1,3	6,8 % 3,1	5,4% 1,7	0,9% 1,7	9,4% 2,0	0,0% 0,5	**11,6 %** **2,2**	0 % –	8,3% 3,0	0 % –	9,9% 3,3	0 % –	**11,8 %** **3,4**	0 % –
50	(288,05)		(281,01)		283,77		(291,39)		(303,18)		(307,29)		(251,00)		(37,50)	
	4,8% 1,5	**15,7 %** **6,7**	6,9% 1,8	**10,4 %** **4,1**	9,4% 2,3	3,2% 2,6	**11,0 %** **3,2**	0,2% 1,4	**12,8 %** **4,3**	0 –	**15,6 %** **5,0**	0 % –	**12,3 %** **2,9**	0 % –	**11,8 %** **3,4**	0 % –
60	(305,95)		(295,80)		(295,57)		(301,50)		(320,62)		(353,49)		(301,71)		(53,48)	
	8,0% 2,0	**17,8 %** **7,3**	**10,0 %** **2,4**	**12,3 %** **5,2**	**11,0 %** **3,2**	5,5% 3,0	**12,4 %** **4,1**	0,6% 1,6	**13,9 %** **6,4**	0 –	**15,6 %** **9,9**	0 % –	**18,3 %** **5,1**	0 % –	**11,7 %** **3,4**	0 % –
70	(319,08)		(306,57)		(306,01)		(313,42)		(356,00)		(430,86)		(361,73)		(71,47)	
	10,2 % **2,5**	**19,3 %** **7,7**	**11,0 %** **3,2**	**13,9 %** **5,8**	**12,2 %** **4,0**	6,8% 3,4	**13,7 %** **5,3**	1,0% 1,8	**14,5 %** **10,0**	0,0% 0,6	**19,3 %** **13,3**	0,0% 2,1	**20,6 %** **9,3**	0 % –	**11,6 %** **3,4**	0 % –
80	(329,43)		(315,24)		(317,59)		(332,28)		(426,93)		(506,53)		(433,04)		(90,01)	
	11,0 % **3,2**	**20,6 %** **7,9**	**11,9 %** **3,9**	**14,6 %** **6,3**	**13,7 %** **4,9**	7,7% 3,7	**14,0 %** **7,7**	1,4% 2,1	**17,8 %** **13,1**	0,1% 1,7	**19,8 %** **17,6**	0,0% 0,8	**20,9 %** **14,0**	0 % –	**11,5 %** **3,4**	0 % –
90	—		—		—		—		—		—		—		—	
	11,7 % **3,8**	**21,9 %** **7,9**	**12,4 %** **4,7**	**15,2 %** **6,5**	**14,8 %** **6,6**	8,4% 3,9	**15,1 %** **10,9**	1,6% 2,3	**18,8 %** **16,2**	0,5% 2,5	**18,5 %** **21,1**	0,3% 2,6	**19,9 %** **18,5**	0 % –	**11,6 %** **3,4**	0 % –

☞ **Ver a Legenda e a Nota relativas a este Quadro apresentadas na página seguinte !**

Legenda do Quadro anterior

	V1 = 30		
	260,36		← Lucro Médio Mensal (u.m.)
T1 = 40	3,0%	6,8 %	← % de tempo com 'restrições'
	1,3	3,1	← vol. médio de 'restrição' no tempo com 'restrições'(u.vol.)

↖ % de tempo com 'cheias'
vol. médio de cheia no tempo com 'cheias'

Nota relativa ao Quadro anterior: Sempre que tenha ocorrido uma percentagem de 'tempo com restrições', ou de 'tempo com cheias', superior a 10 % do tempo total simulado, considerou-se a correspondente política é 'desaconselhável', pelo que o correspondente valor médios mensais de lucro foi indicado entre parêntesis.

Relativamente aos resultados obtidos, pode-se referir em primeiro lugar que dos 72 cenários estudados, 58 são de imediato classificados como 'desaconselháveis' ! Se considerarmos que é desejável que a percentagem de ocorrência de situações de cheia (ou de situações de 'restrições') não exceda, por exemplo, 5 %, dos 14 cenários restantes somos reduzidos a três:

V1 = 20 u.vol.; T1 = 20 u.vol. $\Rightarrow$ L = 142,76 u.m.; C:(4,8%; 3,4 u.vol.); R:(0%; –)

V1 = 30 u.vol.; T1 = 30 u.vol. $\Rightarrow$ L = 226,17 u.m.; C:(1,1%; 2,2 u.vol.); R:(0,8%; 1,7 u.vol.)

V1 = 40 u.vol.; T1 = 30 u.vol. $\Rightarrow$ L = 225,40 u.m.; C:(3,9%; 2,9 u.vol.); R (0,0%; 0,3 u.vol.)

0 primeiro dos três cenários destacados não apresenta situações de 'restrições', mas, infelizmente, é o que apresenta a maior percentagem de situações de cheia (4,8 %) e o maior volume médio de cheia (3,4 u.vol.), apresentando ainda, o menor valor de lucro médio mensal (142,76 u.m.).

0 terceiro cenário praticamente não apresenta situações de 'restrições', apresenta 3,9 % de situações de cheia (com uma média de 2,9 u.vol.) e um lucro médio mensal de 225,40 u.m., pelo que não temos dúvida em considerá-lo melhor do que o primeiro cenário.

Globalmente, no entanto, parece-nos ser de **destacar a segunda política (V1 = 30 ; T1 = 30)**, já que as situações de 'restrições' são ainda negligenciáveis e a ocorrência de situações de cheia já é também muito minorada (apenas 1,1 % com 2,2 u.vol.), correspondendo esta política ao valor mais elevado de lucro médio mensal (226,17 u.m.)

[Atenção: Não se comparem os resultados agora apresentados com os obtidos aquando da simulação com dados históricos (secção I destes apontamentos)! Qualquer semelhança é pura coincidência, já que os 'dados históricos' utilizados não estão directamente relacionados com as distribuições estatísticas apresentadas nesta secção !]

Uma última referência a um outro tópico que ainda não tínhamos abordado: a **influência da 'semente' utilizada no processo de geração de números pseudo-aleatórios nos resultados.** Nesta aplicação para simular, por exemplo, 1000 anos, precisamos de gerar mais de 233 000 (!) valores pseudo-aleatórios U[0;1], pelo que temos que utilizar 4 sementes distintas (ver secção **III**), 0 que reduzira, à partida, qualquer eventual influência deste factor !

Em aplicações que justifiquem esse cuidado e em que não esteja envolvida a geração de um número tão elevado de números pseudo-aleatórios, poderíamos obter resultados com sementes distintas e 'combinar' os resultados (ou, pelo menos, avaliar a sua variabilidade...).

E já chega de comentários a propósito dos resultados obtidos com o modelo d' **"A Barragem de Pós-Boa (II)"** ! No entanto, esta barragem ainda nos pode ser útil...

Não a abandonemos já ! Consideremos, agora, o **problema "A Barragem de Pós--Boa (III)"** :

☞ Admitamos que o enunciado apresentado anteriormente se mantém, mas que se pretende **considerar que as necessidades e as precipitações dependem do grau de secura / humidade do ano**.

Recordemos os valores indicados anteriormente para as Precipitações Mensais e para as Necessidades Mensais de Água a Jusante:

Mês	**J**	**F**	**M**	**A**	**M**	**J**	**J**	**A**	**S**	**O**	**N**	**D**
Nec μ	8	8	8	9	10	13	15	15	15	13	9	8
Nec σ	1,0	0,8	0,7	0,8	0,9	1,1	1,2	1,3	1,1	0,7	0,8	0,9
Distr. P	N	N	N	N	U	U	U	U	U	N	N	N
μ ou a	30	40	40	30	20	5	0	0	6	18	29	30
σ ou b	2,5	3,0	3,5	2,0	30	15	10	5	18	2,0	2,5	3,0

Nota: Se a distribuição de P for **Normal (μ ; σ)**, a penúltima linha do Quadro anterior diz respeito a μ e a última linha a σ ; se a distribuição for **Uniforme [a ; b]**, a penúltima linha do Quadro anterior diz respeito a **a** e a última a **a**.

Admita que, por simplicidade, os valores médios anteriormente indicados das distribuições referidas devem ser encarados como 'valores médios de referência'. Um 'valor médio real' obtém-se por multiplicação do correspondente 'valor médio de referência' pelo respectivo **'factor de humidade do ano' (θ (para a precipitação) e δ (para as necessidades))**. Os factores de humidade referidos têm as distribuições indicadas no Quadro seguinte:

	θ	δ
Ano seco	Uniforme [0,5 ; 0,7]	Normal (μ = 1,2 ; σ = 0,10)
Ano normal	Normal (μ = 1,0 ; σ = 0,10)	Normal (μ = 1,0 ; σ = 0,05)
Ano húmido	Uniforme [1,3 ; 1,5]	Normal (μ = 0,9 ; σ = 0,10)

Admita que se mantém inalterada a variância das distribuições, independentemente do grau de humidade do ano e que a geração de valores pseudo-aleatórios se faz a partir das distribuições indicadas com os valores médios 'reais' e as variâncias indicadas.

Admita ainda que a 'probabilidade de transição' do grau de secura / humidade de um ano para o seguinte é dada por:

ano$_i$ ↓ ano$_{i+1}$ →	**seco**	**normal**	**húmido**
seco	0,50	0,30	0,20
normal	0,10	0,75	0,15
húmido	0,05	0,45	0,50

Que alterações deveriam ser introduzidas no modelo anteriormente concebido para aconselhar o gestor da barragem relativamente à política de gestão de águas a adoptar, tendo em conta que se pretende considerar que as necessidades e as precipitações dependem do grau de secura / humidade do ano ?

As alterações introduzidas nada têm de especial ... Poderemos começar por, **na rotina 'Inic', solicitar ao utilizador a indicação da nova condição inicial G0 (grau de humidade do ano anterior ao ano inicial).** Pode-se codificar G0 adoptando, por exemplo, G0 = 1 (seco) ; 2 (normal) e 3 (húmido).

☞ Deixamos ao leitor a tarefa de proceder à adaptação sugerida. Mãos à obra !