4 Eliciação de Probabilidades

Para que se inicie a tarefa de eliciar as probabilidades subjetivas dos avaliadores deve-se ter em vista alguns questionamentos:

a) Que distribuição melhor representa a característica geológica do prospecto a ser avaliado?
b) Quantos percentis ou outras estatísticas são necessários para descrever adequadamente a distribuição?
c) Como realizar o cálculo da distribuição de probabilidades do volume de hidrocarbonetos através do produto das distribuições dos parâmetros de petróleo?
d) Que vieses poderão ocorrer neste procedimento?
e) Que técnica utilizar para obter estes percentis ou outras estatísticas, de modo a reduzir o surgimento destes vieses?
f) Qual a probabilidade deste volume estimado, de fato, conter hidrocarbonetos? Ou seja, como eliciar o chamado risco geológico?

A escolha da distribuição que melhor representa a característica desejada do prospecto avaliado geralmente é um problema resolvido sem muita dificuldade pelos geólogos. Entretanto, algumas tendências vêm surgindo no âmbito de exploração de petróleo. Alguns autores defendem fortemente a utilização da distribuição lognormal (Capen, 1996, 2001) para descrever as características geológicas, outros afirmam que a utilização da distribuição triangular é mais acessível aos geólogos, outra corrente garante que o uso da distribuição beta representa melhor a distribuição desejada. Em qualquer caso parece haver concordância quanto à existência de assimetria nestas distribuições.

O uso das distribuições beta e lognormal é atualmente muito difundido no processo de avaliação de recursos não descobertos de petróleo e tem destaque especial na literatura de eliciação de probabilidades (Rose, 2004; Schuenemeyer, 2002; Schneiderman, 1997). Distribuições triangulares são defendidas com argumentos de simplicidade, sendo, por muitos, consideradas aproximações aceitáveis. É importante perceber que quando o especialista não identifica a similaridade de seu parecer com nenhuma das distribuições disponíveis, forçá-lo a se expressar por meio de uma delas pode introduzir vieses e, por isso, deve-se permitir a geração de uma distribuição empírica (Schuenemeyer, 2002).

É realmente difícil avaliar, em curto prazo, a adequação destas distribuições à realidade, pois o custo e o tempo associados a isto seriam impraticáveis para a indústria do petróleo. A solução aparentemente mais conveniente apresentada na literatura é deixar ao avaliador a escolha da forma da distribuição mais semelhante àquela que ele imagina e quando isto não se faz possível a melhor distribuição é a distribuição empírica criada diretamente das estimativas do avaliador (Schuenemeyer, 2002).

4.1 Distribuição Empírica

Como foi brevemente citado anteriormente, a distribuição empírica deve ser utilizada quando o avaliador não acreditar que as distribuições padronizadas capturem a forma essencial da natureza da característica avaliada.

Para utilizar este tipo de distribuição, o avaliador deve estar treinado para fornecer 6 ou 7 percentis para que a distribuição tenha uma forma apropriada para realizar as operações que vêm a seguir.

Schuenemeyer (2002), em um caso de avaliação de parâmetros de recursos não descobertos de petróleo, utiliza os percentis 25%, 50%, 75%, de fácil percepção e pede aos avaliadores também que definam dois pontos de truncamento que se chamam de ponto máximo e ponto mínimo (aproximadamente P_{99} e P_1) e por último os percentis 95% e 5%. O autor não sugere definições qualitativas para diferenciar o ponto máximo do percentil 95%, por exemplo, ou o ponto mínimo do percentil 5%.

O quadro 4.1 e a figura 4.1 a ele correspondente mostram as distribuições empíricas de algumas características do sistema geológico eliciados no exemplo de Schuenemeyer (2002):

Caracterítica/Percentis	**Mín**	**5%**	**25%**	**50%**	**75%**	**95%**	**Máx**
Espessura do reservatório	50	63	100	150	250	425	500
Área de enclausuramento	0,5	1	1,5	2	5	10	20

Quadro 4.1 – Características do Sistema Geológico e Percentis Eliciados
Reproduzido de Schuenemeyer (2002)

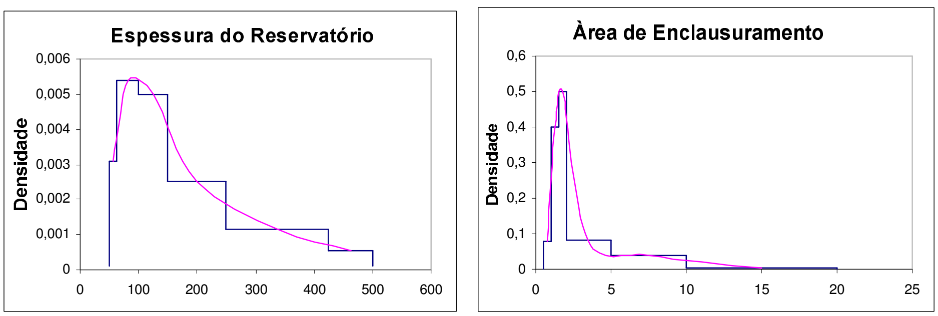

Figura 4.1 – Gráficos das Distribuições Empíricas dos Parâmetros Geológicos
Reproduzido de Schuenemeyer (2002)

Como se observam nos gráficos da figura 4.1 acima, cada intervalo entre percentis P_α gera uma classe, ou seja, *n* percentis geram *n-1* classes, que nos permitem calcular suas densidades, da seguinte forma:

$$f_\alpha = (P_{\alpha+1} - P_\alpha)/(x_{\alpha+1} - x_\alpha) \qquad (4.1.1)$$

Onde f_α, P_α, x_α são, respectivamente, a densidade da distribuição, a probabilidade cumulativa e o valor do percentil α.

Os percentis da distribuição da espessura do reservatório geram as seguintes classes, com suas respectivas médias e freqüências, com observado no quadro 4.2:

Classe	Densidade	Média da Classe ($\bar{x}$)
P1%-P5%	0,0031	56,50
P5%-P25%	0,0054	81,50
P25%-P50%	0,0050	125,00
P50%-P75%	0,0025	200,00
P75%-P95%	0,0011	337,50
P95%-P99%	0,0005	462,50
Soma	**0,0177**	

Quadro 4.2 – Classes da Distribuição da Espessura do Reservatório

A média e o desvio padrão são calculados de acordo com as seguintes expressões:

$$\mu = \frac{\sum_{\alpha=1}^{n-1} f_\alpha \cdot \bar{x}_\alpha}{\sum_{\alpha=1}^{n-1} f_\alpha} \qquad (4.1.2)$$

$$\sigma = \frac{\sum_{\alpha=1}^{n-1} (\bar{x}_\alpha - \mu)^2}{n-1} \qquad (4.1.3)$$

Onde $\bar{x}_\alpha$ é o ponto médio da classe α.

O volume de óleo no reservatório é calculado através do produto de diversos parâmetros do prospecto, entre eles as características geológicas. Portanto quando estes parâmetros são representados por variáveis aleatórias, a distribuição probabilística deste produto pode ser obtida numericamente pelo método de Monte Carlo ou por outros métodos numéricos aproximados. Em casos particulares, como o caso de todos as características geológicas serem representadas por distribuições lognormais; a distribuição do volume pode ser obtida por métodos analíticos. Na seção seguinte, discute-se esse caso de distribuições lognormais.

4.2 Distribuição Lognormal

Muitos autores, entre eles o enfático Capen (2001), afirmam que a distribuição lognormal descreve bem as características geológicas de um prospecto. Rose (2004) expõe que este tipo de distribuição começou a ter forte aceitação na ciência do petróleo a partir da década passada. Afirma também que os geocientistas que realizam suas estimativas conscientes da lognormalidade das características geológicas tendem a realizar melhores predições.

Ainda segundo Rose (2004), algumas poucas características do petróleo têm distribuições exponenciais e menos características ainda, distribuições normais.

Para realizar o ajuste da distribuição eliciada a uma curva lognormal são necessários apenas 2 percentis, pois 2 desses parâmetros são estatísticas suficientes para uma distribuição lognormal; entretanto tais percentis devem corresponder a probabilidades complementares para facilitar o cálculo da média da distribuição normal associada, além de parecer mais simples para o avaliador.

A estimativa de parâmetros de petróleo, quando utilizada a distribuição lognormal, os geólogos costumam estimar os percentis 10% e 90% e, portanto, essa simetria facilitadora é bem conveniente.

Dados esses percentis, é possível calcular a média e o desvio padrão da distribuição. Supondo o fator geológico uma variável aleatória X de distribuição lognormal, sabe-se que existe uma variável aleatória $Y = \ln X$ que possui uma distribuição normal de média μ e variância σ^2.

A função densidade da *distribuição lognormal* do fator geológico, então, é dada por:

$$f(x,\mu,\sigma) = \frac{1}{x.\sqrt{2\pi\sigma}} \cdot \exp\left[-\frac{(\ln x - \mu)^2}{2 \cdot \sigma^2}\right], \text{ onde } -\infty < \mu < \infty \text{ e } \sigma > 0 \quad (4.2.1)$$

Sua variância é

$$Var[X] = \exp\left[2 \cdot \mu + 2 \cdot \sigma^2\right] - \exp\left[2 \cdot \mu + \sigma^2\right] \quad (4.2.2)$$

Sua média é

$$E[X] = \exp\left[\mu + \frac{1}{2}\cdot\sigma^2\right] . \tag{4.2.3}$$

É fácil verificar que com os percentis eliciados, pode-se encontrar essa média e desvio padrão da seguinte forma:

- Fazer a transformação dos percentis da distribuição lognormal ($x_{\alpha i}$) para percentis de uma distribuição normal ($y_{\alpha i}$) através da expressão $y_{\alpha i}$ = ln($x_{\alpha i}$).

- Tendo os percentis eliciados simétricos em relação à média, pode-se calcular μ através da expressão

$$\mu = (y_{\alpha i} + y_{1-ai})/2 \ . \tag{4.2.4}$$

- Em uma distribuição normal, a média μ é igual ao percentil y_{50}.

- Em seguida, para o cálculo do desvio-padrão σ utiliza-se a seguinte expressão

$$\sigma = \frac{y_{\alpha_i} - \mu}{\Phi^{-1}(\alpha_i)} \tag{4.2.5},$$

onde Φ^{-1} é a função cumulativa inversa da distribuição normal.

- Com μ e σ determinados à distribuição lognormal está perfeitamente descrita e então se pode encontrar quaisquer outros percentis calculando a inversa cumulativa da função lognormal.

O exemplo abaixo resume o método supracitado para a definição da distribuição lognormal eliciada. O método é facilmente programável em planilhas eletrônicas (Veja Apêndice 2). Supondo que os percentis utilizados foram P_{10} e P_{90} e o parâmetro de petróleo utilizado foi a área medida em acres:

x_{10} = 60 ac e x_{90} = 600 ac,

logo,

$y_{10} = \ln(60) = 4{,}094$ e $y_{90} = \ln(600) = 6{,}397$

Calcula-se a média normalizada:

$$\mu = (4{,}094 + 6{,}397)/2 = 5{,}245$$

Escolhendo y_{90}, temos;

$$\sigma = \frac{6{,}397 - 5{,}245}{1{,}282} = 0{,}898$$

Outros percentis da distribuição lognormal, com os valores μ e σ acima, são exibidos no quadro 4.3 abaixo, em seguida, observa-se o gráfico da distribuição eliciada:

Percentil	Área (ac)
P1	23,470
P5	43,292
P10	60,000
P25	103,513
P50	189,736
P75	347,780
P90	600,000
P95	831,560
P99	1533,840

Quadro 4.3 – Percentis da Distribuição Lognormal

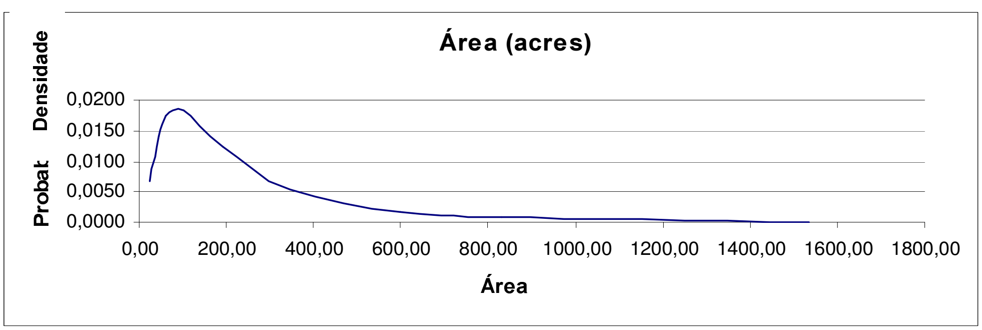

Figura 4.2 – Distribuição Lognormal da Área do Play

Supondo que as distribuições das características geológicas são lognormais, e independentes (que em geral não é uma hipótese válida), existe uma forma simplificada de para calcular a distribuição do volume de hidrocarbonetos que é dada a partir da multiplicação das distribuições destas características.

Se desejarmos encontrar a distribuição de *Xv* que é a variável aleatória que representa o volume de hidrocarbonetos, podemos utilizar a seguinte expressão.

$$Xv = X_1 \times X_2 \times \cdots \times X_n \qquad (4.2.6)$$

Onde *Xn* é a variável aleatória que representa n-ésima característica geológica.

Tomando os logaritmos da expressão, temos:

$$\ln Xv = \ln\left(X_1 \times X_2 \times \cdots \times X_n\right)$$

$$\ln Xv = \ln X_1 + \ln X_2 + \cdots + \ln X_n$$

Reescrevendo a expressão com $\ln X_i = Y_i$:

$$Yv = Y_1 + Y_2 + \cdots + Y_n \qquad (4.2.7)$$

Onde *Yi* são variáveis aleatórias de distribuição normal.

Logo,

$$\mu_v = \mu_1 + \mu_2 + \cdots + \mu_n \qquad (4.2.8)$$

$$\text{e } \sigma_v = \sqrt{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2}\ , \qquad (4.2.9)$$

que são os parâmetros para a distribuição lognormal que representa o volume.

4.3 Distribuição Triangular

O uso da distribuição triangular nas atividades de avaliação de recursos de petróleo ainda é muito comum, pela sua aparente simplicidade e facilidade de uso.

Não se observam na literatura autores que possuam bons argumentos para a defesa da utilização desta distribuição para descrever os parâmetros de petróleo, entretanto, na prática ela é freqüentemente utilizada com aproximação válida.

Pode-se definir esse tipo de distribuição por meio dos parâmetros ponto máximo (*a*), ponto mínimo (*b*) e moda (*c*) (veja figura 4.3). Os pontos máximo e mínimo geralmente são eliciados através da idéia de visão mais otimista ou mais pessimista, ou seja, na melhor das hipóteses ou na pior das hipóteses, e a moda é associada ao valor mais provável. Os métodos de eliciação destes parâmetros devem ajudar ao avaliador evitar ancorar em uma estimativa inicial que é muito comum quando se pretende eliciar pontos centrais.

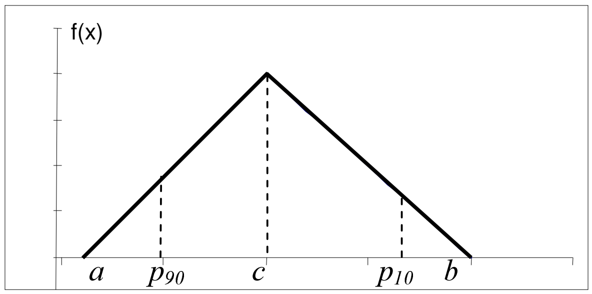

Figura 4.3 – Distribuição Triangular

A distribuição triangular, simétrica ou não, possui a função densidade de probabilidade ilustrada na figura 4.3 (para o caso simétrico) e definida como segue:

$$f(x;a,b,c)=\begin{cases}\dfrac{2\cdot(x-a)}{(b-a)\cdot(c-a)};a\le x\le c\\[2ex]\dfrac{2\cdot(b-x)}{(b-c)\cdot(b-a)};c\le x\le b\end{cases}\qquad(4.3.1)$$

Onde:

- a = min
- b = máx
- c = moda.

Com os parâmetros *a*, *b* e *c* eliciados, a média e o desvio padrão da distribuição podem ser facilmente obtido pelas suas definições.

Para o cálculo dos percentis utiliza-se a expressão da função cumulativa da distribuição triangular:

$$F(x;a,b,c)=\begin{cases}0;\ x<a\\ \dfrac{(x-a)^2}{(b-a)\cdot(c-a)};\ a\le x\le c\\ 1-\dfrac{(b-x)^2}{(b-a)\cdot(b-c)};\ a\le x\le c\\ 1;\ x<a\end{cases} \tag{4.3.4}$$

Os pontos máximo e mínimo da distribuição triangular não são recomendados para eliciação, pois são pontos de difícil estimação. Isso porque, conforme mencionado, a distribuição triangular geralmente não é reconhecida como uma boa representação da realidade porque a faixa de valores muito próxima ao máximo concentra uma probabilidade muito grande, ou seja, ela não permite que os avaliadores expressem adequadamente sua melhor estimativa da mediana, nem o fato de acreditarem que embora com uma probabilidade muito pequena valores muitos grandes podem ocorrer. Então de forma alternativa, deve-se utilizar percentis mais familiares aos especialistas como, por exemplo, os percentis p_{10} e p_{90}. Por geometria, pode-se chegar às seguintes expressões.

$$(p_{90} - a)^2 = 0{,}1\ (b - a)\ (c - a)$$
$$(b - p_{10})^2 = 0{,}1\ (b - a)\ (\mathrm{b} - c)$$

Estas expressões são utilizadas para calcular os parâmetros *a* e *b* por métodos numéricos. Todas as expressões apresentadas acima são facilmente programáveis em planilhas eletrônicas. Para calcular o volume deve-se utilizar simulação de Monte Carlo para realizar o produto das distribuições (Keefer, D. L. & Bodily, S. E., 1983).

4.4 Distribuição Beta

A distribuição beta, como a triangular, também não possui caudas, logo a sua vantagem em relação à distribuição triangular é uma melhor adaptação a distribuição empírica na região central e tem fórmulas simples para o cálculo da média e desvio padrão utilizadas em PERT. (Schuenemeyer, 2002; Jenkinson, 2005).

A função densidade de probabilidade da distribuição beta é dada pela seguinte expressão:

$$f(x) = \frac{1}{B(\alpha,\beta)} \cdot x^{(\alpha-1)} \cdot (1-x)^{(\beta-1)} \tag{4.4.1}$$

Onde α e β são parâmetros da distribuição e $B(\alpha,\beta)$ é a função beta. A função beta é constante que normaliza a função densidade de probabilidade para assegurar que a integral de *f(x)* entre x = 0 e x = 1 seja 1. A função beta se relaciona com a função gama a partir da seguinte expressão:

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \tag{4.4.2}$$

Onde,

$$\Gamma(X) = \int_0^\infty t^{x-1} \cdot e^{-t}\,dt \tag{4.4.3}$$

é a função gama.

Jenkinson (2005) realiza uma abrangente revisão sobre dez métodos utilizados para eliciar os parâmetros da distribuição beta. Destes apenas um é mais relevante para a área de petróleo, já que os demais se preocupam em eliciar média ou desvio-padrão da distribuição, conceitos que apresentam grande complexidade para a compreensão da maioria dos avaliadores desta área.

O método de Fox para a estimação dos parâmetros da distribuição beta foi desenvolvido em 1966. Neste método, primeiramente, pede-se ao avaliador que estime a moda ($\hat{m}$) que este imagina ter a distribuição. Em seguida, pede-se ao avaliador que diga qual a probabilidade (p) de que a moda esteja contida no intervalo $(\hat{m} - K\hat{m}; \hat{m} + K\hat{m})$, onde 0<K<1. Tendo estes valores sido eliciados, resolvemos o seguinte sistema de equações por métodos numéricos para descobrir os parâmetros da distribuição beta:

$$\frac{\hat{\alpha}-1}{\hat{\alpha}+\hat{\beta}-2} = \hat{m} \tag{4.4.4}$$

$$\int_{\hat{m}-K\hat{m}}^{\hat{m}+K\hat{m}} \frac{x^{(\hat{\alpha}-1)} \cdot x^{(\hat{\beta}-1)}}{B(\alpha,\beta)} = p \tag{4.4.5}$$

Pleguezuelo e Rodriguez (1991) mencionam que os criadores da metodologia PERT sugerem estimar a média e o desvio-padrão de uma distribuição beta através da opinião mais otimista (*a*), o p_1, a mais pessimista (*b*), o p_{99} e da mais provável (*c*), a moda, utilizando as seguintes aproximações:

$$\hat{\mu} = \frac{a + 4m + b}{6} \tag{4.4.6}$$

$$\hat{\sigma}^2 = \frac{(b-a)^2}{36} \tag{4.4.7}$$

Podem-se obter as estimativas dos parâmetros α e β da distribuição beta resolvendo-se o sistema de suas equações para a média e a variância:

$$\mu = \frac{\alpha}{\alpha + \beta} \tag{4.4.8}$$

$$\sigma^2 = \frac{\alpha \cdot \beta}{(\alpha + \beta + 1) \cdot (\alpha + \beta)^2} \tag{4.4.9}$$

A utilização dessas aproximações pode trazer consigo vieses difíceis de controlar quando os especialistas não estão bem calibrados, já que é difícil ter uma visão dos percentis p_1 e p_{99}, principalmente quando se sofre pressão por descobertas ou redução de custos de exploração. Uma visão mais abrangente sobre os casos que levam ao surgimento de vieses será dada na subseção 4.6.

4.5 Eliciação de Distribuições de Probabilidades Subjetivas

Os métodos acima pressupõem uma distribuição conhecida, exigindo, portanto, apenas a determinação dos parâmetros. Entretanto, conforme já dito, há quem defenda a eliciação da distribuição sem a indicação prévia de qualquer distribuição.

Os seguintes métodos mostram algumas formas de eliciar distribuições de probabilidades subjetivas sem indicação de uma distribuição conhecida. Jenkinson (2005) apresenta quatro métodos que são divididos em duas categorias. A primeira é composta pelo método conhecido como *EPS* (*equivalent prior sample*) ou *Amostra Prévia Equivalente* e o método *HFS* (*hypothetical future sample*) ou *Amostra Futura Hipotética*, estes métodos são chamados métodos indiretos, pois o especialista não reconhece previamente a forma da distribuição. Estes métodos são utilizados para eliciar proporções, ou seja variáveis aleatórias que somente podem se apresentar na faixa de 0 a 1, o que não pertence ao escopo desta dissertação.

A segunda categoria, a dos métodos diretos, contém *o método da função da distribuição cumulativa* (CDF – cumulative distribution function) e o *método da função densidade de probabilidade* (PDF – probability density function).

No método CDF é dado ao especialista uma probabilidade e é perguntado a ele qual o valor da variável cuja probabilidade de não ser excedido é a dada, ou seja, pede-se ao avaliador que estabeleça diversos percentis. Por exemplo, se a probabilidade dada for 0,5, então o especialista daria o valor o qual ele sente ser igualmente provável que o valor real esteja abaixo e acima deste valor. Este valor é chamado de mediana do especialista ou de percentil 50º. Este procedimento é repetido várias vezes para vários percentis e os valores dos demais quantís podem ser estimados livremente e confirmados posteriormente com o avaliador.

O método PDF consiste em representar graficamente a função densidade. Do especialista são eliciados primeiramente os valores correspondentes à média e à moda, outros valores de percentis devem ser eliciados para aperfeiçoar a forma da função. Em ambos os métodos os dados eliciados serão conectados através do ajuste em uma curva, sendo que no método PDF a curva pode ser estimada livremente.

Enquanto os métodos EPS e HFS são utilizados de forma restrita para a estimação de proporções, os demais métodos apresentados podem ser utilizados para a eliciação da distribuição de diversas variáveis aleatórias contínuas.

Os quantís muitas vezes são eliciados pedindo-se ao avaliador que divida a extensão em regiões de iguais probabilidades.

Entretanto, o método mais simples é o *método da bisecção*, no qual especialista determina o valor que divide um dado intervalo em dois intervalos equiprováveis como mostram as figuras 4.4 e 4.5. Por exemplo pede-se inicialmente ao avaliador que divida o espaço amostra em dois intervalos equiprováveis.

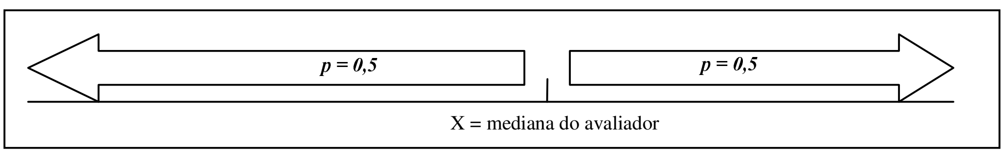

Figura 4.4– Método da Bisecção (Passo 1)

Em seguida, é solicitado ao avaliador que considere o valor X eliciado anteriormente como a sua *mediana* e que agora considere que o valor real **está acima** dela. Pede-se então para que o avaliador defina qual o valor que dividiria a região acima da mediana em duas regiões de iguais probabilidades. Este valor seria o *quartil superior* (Qs), mostrado na figura 4.5.

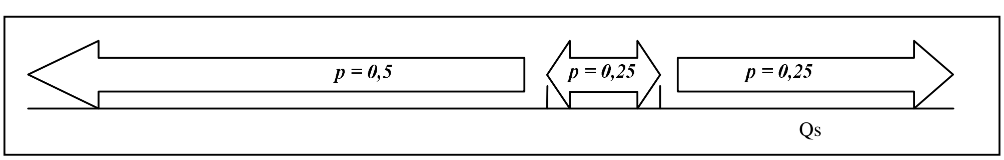

Figura 4.5 - Método da Bisecção (Passo 2)

Segue-se com o mesmo procedimento para a determinação do *quartil inferior* (Qi) e dos quantís 12,5%, 27,5%, 62,5% e 87,5%. O número de quantís eliciados depende do nível de detalhe desejado para a forma da distribuição.

Loveridge (2004), entretanto, sugere eliciar os quantís altos, baixos e limites da seguinte forma:

Solicita ao avaliador que defina dois valores (um inferior e um superior) para o valor da variável para o qual este ficaria “surpreso” se fosse o valor real. Estes valores seriam considerados subjetivamente como os quantís 10% e 90%. Como visto anteriormente, estes percentis são suficientes para definir uma distribuição lognormal.

Em seguida, solicita-se ao avaliador que defina outros dois valores (um inferior e um superior) para a variável para o qual este ficaria “perplexo” se fosse o valor real. Estes valores seriam considerados subjetivamente como os quantís 1% e 99%, equivalentes aos limites mínimo e máximo de Schuenemeyer.

O autor acrescenta que não existem definições qualitativas boas o suficiente para serem utilizadas para a o processo de eliciação dos quantís intermediários.

Outros quantís que têm sido eliciados são os tercís (33,3% e 66,6%) e os percentis 17% e 83%. Os tercís são eliciados similarmente ao método da bisecção apenas pedindo ao especialista que divida primeira a faixa de valores possíveis em três faixas igualmente prováveis. Jenkinson (2005) comenta que o uso de tercís é preferível ao uso dos quartís em relação a calibração das distribuições eliciadas. Os tercís também evitam um viés comum na estimativa da mediana, a ancoragem.

Segundo Daneshkah (2005) outros o objetivo de se eliciar mais percentis do que o mínimo necessário para definir uma distribuição é possibilitar a identificação de vieses. O que será tratado na subseção a seguir.

4.6 Identificação de Vieses

Consideramos que os especialistas utilizados em uma avaliação têm um grande conhecimento sobre dado evento; entretanto, quando este especialista é influenciado por fatores externos, há discrepâncias entre o seu conhecimento e as probabilidades eliciadas.

A diferença entre a resposta do avaliador e a descrição acurada de seu conhecimento é chamada de viés (Sptzler et al., 1975). Alguns vieses ocorrem freqüentemente em avaliações. Ampla literatura expõe uma grande variedade destas distorções, explica suas causas e mostra como reduzir suas incidências. Algumas destas distorções são apresentadas a seguir.

4.6.1 Ancoragem

A ancoragem é uma distorção que se dá em decorrência da estratégia utilizada para a eliciação de uma dada quantidade, pois os avaliadores que geralmente partem de uma quantidade inicial, geralmente a mediana, ou a moda, ao determinar uma faixa de probabilidade de ocorrência, terminam por estabelecer valores demasiadamente próximos ao valor inicialmente determinado (Spetzler et al., 1975; Schuenemeyer, 2002; Bratvold et al. 2002).

A resposta inicial em uma entrevista geralmente conduz as respostas posteriores, e quando a primeira pergunta é relacionada com a mediana da quantidade eliciada, a ancoragem se apresenta como um viés centralizador. A ancoragem ocorre devido a uma falha do processo da informação sobre os outros pontos da distribuição independente do ponto sob consideração.

É menos comum ocorrer este viés quando as extremidades da distribuição eliciada, P_{95} e P_5, por exemplo, são requeridas antes dos pontos centrais. Porém quando se deseja certificar que o avaliador está realmente de acordo com a distribuição escolhida para descrever a característica avaliada é importante avaliar outros parâmetros da distribuição eliciada que servirão como certificadores, por exemplo, se desejamos eliciar uma distribuição lognormal, além de eliciar suas extremidades que é suficiente para definir a distribuição podemos solicitar ao avaliador a mediana P_{50} que servirá para comparação com a mediana calculada a partir das extremidades, com isso o avaliador poderá realizar os ajustes que lhe convier com maior clareza.

4.6.2 Representatividade

Representatividade significa que a amostra é muito pequena para permitir que eventos de baixa probabilidade estejam nela representados com um número de ocorrências suficiente para se fazer inferências. O tamanho da amostra afeta a probabilidade de obter certos resultados. As pessoas geralmente tendem a acreditar em uma pequena fatia de informação ignorando a informação global (Speltzer at al. 1975; Daneshkah, 2004).

Deve-se salientar que quando se deseja observar um grande número de variáveis que possam gerar um evento a amostra deve ser tanto maior quanto maior for o número de variáveis para avaliar.

Por exemplo, se desejamos verificar a probabilidade de que $\boldsymbol{n}$ características que podem se apresentar de $\boldsymbol{m}$ formas gerem um evento θ, $\boldsymbol{n}^{\boldsymbol{m}}$ classes deveriam ser definidas. E, para cada uma delas deveria haver um número significante de ocorrências para uma avaliação representativa.

Em um estudo sobre avaliação de prospectos apresentados por Schneiderman e Otis (1997), os autores definem quatro fatores que influenciam na probabilidade de existir ou não hidrocarbonetos em uma dada área, e cada uma destas quatro características podem ser apresentadas como desfavorável, questionável, neutra, encorajadora, e favorável e associam a estas denominações valores de probabilidade. É fácil perceber que seria mais interessante manter um histórico que associasse estas combinações de fatores com a ocorrência ou não do evento, entretanto, para uma amostra representativa haveria a necessidade de registrar ocorrências suficientes em 4^5 classes. E para isto, precisaríamos de alguns milhões de observações que é claramente impossível de se obter.

4.6.3 Dependências

Dependências probabilísticas entre distribuições de parâmetros podem ocorrer em muitos estágios de um processo de avaliação. Geralmente, em exploração de petróleo algumas características avaliadas como porosidade e espessura do reservatório podem ser dependentes entre si, mas geralmente não são consideradas desta forma pela complexidade da formulação final. Em exploração, dependências também podem ocorrer entre plays ou outras unidades de avaliação e deveriam ser consideradas na fase de agregação de resultados (Schuenemeyer, 2002).

Outro tipo de ocorrência da dependência é entre as avaliações feitas pelos especialistas que podem ter tido treinamentos e experiências passadas em comum quando avaliam a mesma situação com base nas mesmas informações. É de se esperar que, se todos são bons avaliadores, todos façam avaliações similares, isto é, que haja concordância face às mesmas evidências. Entretanto, se a opinião de um determinado especialista, por motivos não estritamente racionais, influenciar as opiniões dos demais, informação relevante pode ser perdida e uma dependência não desejável (viés) pode ser introduzida.

Uma boa prática para evitar este viés é montar uma boa estrutura para o processo de eliciação e regras rígidas para evitar conversas paralelas. A interação entre avaliadores é sob alguns aspectos desejável, mas deve ser controlada para evitar dependências que signifiquem perda de informação relevante. Isso será discutido mais adiante no contexto de combinação de estimativas.

4.6.4 Coerência

As pessoas costumam construir, para seus resultados, cenários compatíveis para que eles lhes pareçam verossímeis, ou seja, a probabilidade a posteriori em alguns casos está mais relacionada com a **coerência dos cenários** encontrados pelos especialistas do que pelas evidências. O que quer dizer que, especialistas com um conhecimento substantivo sobre um determinado assunto pode facilmente “provar” que sua previsão está coerente e torná-la aceita. As metodologias de combinação bayesiana reduzem este tipo de viés, pois não consideram o conhecimento do especialista, mas sim a probabilidade de sua opinião ser acertada (Schuenemeyer, 2002).

Outro tipo de incoerência pode ocorrer quando os resultados apresentam impossibilidades matemáticas ou físicas, como por exemplo, probabilidade resultante maior que 1 ou, também, volume incompatível com a área estudada.

4.6.5 Disponibilidade

Segundo Daneshkah (2005) a disponibilidade está relacionada com a facilidade com a que os indivíduos se recordam de eventos similares quando avaliam um dado evento. Em outras palavras, é possível afirmar que as pessoas, inclusive os especialistas, consistentemente superestimam a probabilidade de eventos que observaram mais recentemente em relação as que atribuem a eventos menos recentes.

Na prática, para eventos independentes, a probabilidade de ocorrência não é alterada, por exemplo, se um avião cai num dado dia, a probabilidade de um avião cair continua a mesma e a probabilidade de dois aviões caírem em um mesmo dia é ainda menor, mas se observa que as pessoas tendem se preocupar mais com uma nova queda.

Quando este viés é observado é interessante pedir aos avaliadores que estimem a probabilidade condicional do evento ocorrer dado que ocorreu recentemente. Esta probabilidade depende da correlação entre os dois eventos (Bratvold et al., 2002).

4.7 Especificação da Incerteza Geológica

Dentro do processo de avaliação de prospectos, consideramos o risco denominado geológico, isto é, o risco (ou a probabilidade) que um volume de hidrocarbonetos viável para produção exista.

Para avaliar este risco é preciso levar em consideração alguns fatores determinantes para a existência de um play. Alguns autores divergem sobre quantos e quais são estes fatores. Schneiderman (1997, p.1090) afirma que quatro são os fatores que devem ser considerados: presença de uma rocha geradora, presença de uma rocha reservatório, presença de uma trapa e a dinâmica do *play*, ou seja, o tempo apropriado para a formação de uma trapa, para a migração, para a formação dos caminhos para a migração de hidrocarbonetos etc.

Rose (2004, p. 34) afirma que são cinco os fatores determinantes: a presença de uma rocha reservatório, a presença de rochas formadoras de hidrocarbonetos, a existência de tempo e caminhos necessários à migração, avaliação da área de enclausuramento, e por último a existência de rochas próximas à rocha reservatório que estejam promovendo a contenção dos hidrocarbonetos.

De qualquer forma não estamos interessados em definir que riscos geológicos deverão ser avaliados e sim como devem ser tratados para definir o risco geológico global.Uma prática freqüente utilizada é a utilização de listas de verificação (*checklists*).

Schneiderman (1997, p.1090) apresenta uma lista de verificação que pode ser utilizada para analisar todos os subfatores pertinentes a cada uma das características que se deseja avaliar. No quadro 4.4 é apresentado um exemplo:

A. Rocha Reservatório
1. Presença
    a. Litologia
    b. Distribuição
    c. Modelo Deposicional
2. Qualidade
    a. Continuidade lateral e extensão
    b. Espessura e ciclicidade vertical
    ⋮

Quadro 4.4 – Lista de verificação para análise de fatores geológicos

Fonte: Adaptado de Schneiderman (1997)

Este tipo de lista de verificação é importante para que o avaliador tenha em vista, tudo o que deve ser avaliado para não ser tendencioso em seus resultados.

Depois de avaliar cada um dos elementos apresentados na lista o geólogo deverá codificar os conceitos positivos e negativos em probabilidades, ou seja, a partir da avaliação dos subfatores este poderá qualificá-los como: desfavorável, questionável, neutro e encorajador, identificando a faixa de probabilidades em que se encontra. Um exemplo é mostrado no quadro 4.5.

Rose (2004, p. 37) e Schneiderman (1997, p.1092) indicam diferentes conceitos qualitativos sobre a certeza do avaliador e escalas que o permitem associar suas idéias a uma probabilidade. Um importante ponto de convergência é que a probabilidade do fator será a igual menor probabilidade dentre seus subfatores.

Na prática, o formulário de avaliação da probabilidade dos fatores que geram o risco geológico toma a seguinte forma:

A. Reservatório	**Desfavorável**	**Questionável**	**Neutro**	**Encorajador**	**Favorável**
1. Presença					***X***
2. Qualidade			***X***		
3. Outros				***X***	
Faixa de Probabilidade	**0,1 a 0,3**	**0,3 a 0,4**	**0,5**	**0,6 a 0,7**	**0,7 a 0,9**

Quadro 4.5 – Formulário de avaliação de probabilidades dos fatores geológicos

Fonte: Adaptado de Schneiderman (1997)

No caso apresentado acima a probabilidade do fator reservatório está associada à probabilidade do subfator qualidade, ou seja, 0,5.

A probabilidade de descoberta de hidrocarbonetos é obtida a partir do produto das probabilidades dos n fatores geológicos, ou seja:

$$P_{HC} = \prod_{i=1}^{n} Pi \qquad (4.7.1)$$

Onde:

P_{HC} é a probabilidade de descoberta de hidrocarbonetos;

Pi é a probabilidade do i-ésimo fator geológico.

A expressão acima admite que os fatores são independentes o que em grande parte dos casos é apenas uma aproximação. A abordagem bayesiana de combinação que será explicada em detalhes no capítulo 5 desta dissertação pode ser utilizada para aumentar a precisão da probabilidade resultante.

A expressão bayesiana abaixo pretende definir a probabilidade da ocorrência de hidrocarbonetos (evento O) dadas as avaliações qualitativas dos fatores geológicas.

$$P(\theta|F_1 = Quali_k;\cdots;F_n = Quali_k) = \frac{\prod_{i=1}^{n} P(F_i = Quali_k|\theta)}{\prod_{i=1}^{n} P(F_i = Quali_k|\theta) + \prod_{i=1}^{n} P(F_i = Quali_k|\overline{\theta})} \quad (4.7.2)$$

Onde:

F_i = i-ésimo fator relacionado;

$Quali_k$ = expressão que qualifica o fator; ex.: (favorável, neutro, desfavorável, etc.)

n = número de fatores relacionados.

Para que a abordagem bayesiana tenha um resultado positivo é necessário que haja um histórico razoável de eventos observados. O exemplo apresentado no quadro 4.6 supõe um histórico para cada fator geológico, em três classes qualitativas.

Fator A	θ	$\overline{\theta}$	**Fator B**	θ	$\overline{\theta}$	**Fator C**	θ	$\overline{\theta}$	**Fator D**	θ	$\overline{\theta}$
Desfavorável	5	25	Desfavorável	6	23	Desfavorável	7	25	Desfavorável	6	24
Neutro	20	20	Neutro	22	18	Neutro	22	20	Neutro	24	18
Favorável	30	18	Favorável	31	20	Favorável	27	17	Favorável	29	18
Totais	**55**	**63**		**59**	**61**		**56**	**62**		**59**	**60**

Quadro 4.6 – Histórico de Ocorrência dos Fatores Geológicos

Em termos de probabilidade temos as apresentadas no quadro 4.7:

Fator A	θ	$\overline{\theta}$	**Fator B**	θ	$\overline{\theta}$	**Fator C**	θ	$\overline{\theta}$	**Fator D**	θ	$\overline{\theta}$
Desfavorável	0,091	0,397	Desfavorável	0,107	0,377	Desfavorável	0,125	0,403	Desfavorável	0,102	0,4
Neutro	0,364	0,317	Neutro	0,373	0,295	Neutro	0,393	0,323	Neutro	0,407	0,3
Favorável	0,545	0,285	Favorável	0,525	0,328	Favorável	0,482	0,274	Favorável	0,492	0,3
Totais	**1**	**1**		**1**	**1**		**1**	**1**		**1**	**1**

Quadro 4.7 – Probabilidade de Ocorrência dos Fatores Geológicos

Utilizando o histórico acima, se calcularmos a probabilidade da descoberta de hidrocarbonetos através da expressão 4.7.2, dado que os fatores geológicos foram avaliados como A = Favorável, B=Neutro, C = Favorável e D= Neutro, expressamos por:

$$P(\theta | A = F; B = N; C = F; D = N) = \frac{0{,}545 \cdot 0{,}373 \cdot 0{,}482 \cdot 0{,}407}{0{,}545 \cdot 0{,}373 \cdot 0{,}482 \cdot 0{,}407 + 0{,}285 \cdot 0{,}295 \cdot 0{,}274 \cdot 0{,}300} = 0{,}85$$

A abordagem bayesiana tem como principais virtudes o fim da necessidade de se considerar fatores independentes quando não são, e a possibilidade de ajustar a probabilidade final para uma solução coerente, mesmo que os especialistas venham sistematicamente sendo tendenciosos em suas avaliações sobre os fatores geológicos.

O principal problema da utilização desta abordagem na atividade de exploração de petróleo é a dificuldade de se gerar históricos confiáveis, já que, a formação de um *play* não necessariamente tem relação com outros avaliados pela equipe.

4.8 Interação Permitida entre Especialistas

Segundo Schuenemeyer (2002, p.105), a quantidade e o tempo de interação deve ser estabelecido e documentado antes do início do processo de avaliação. À luz da estatística é desejável que haja independência entre as opiniões dos avaliadores. Entretanto, em avaliação de recursos de petróleo isto geralmente não é possível pela escassez de especialistas com um conhecimento geral sobre as diferentes formações geológicas. Acreditar na plena independência é ilusão e pode prejudicar no reconhecimento de vieses.

Então, quando com 2 ou 3 especialistas avaliam plays é comum que tenham conversas freqüentes sobre o objeto de avaliação, desde que haja concordância em apenas um resultado.

Quando os especialistas trabalham individualmente a comunicação deverá ser feita por um intermediário, um tipo de mediador, geralmente o chefe da equipe de avaliação, para evitar que ocorra dependência excessiva entre os resultados e que possam gerar um resultado final pouco enviesado após a combinação.

4.9 Quantidade e Tempo entre Feedbacks

Segundo Schuenemeyer (2002, p.105), os resultados serão mais acurados e terão vieses reduzidos quando se oferece aos avaliadores as distribuições para as variáveis já encontradas durante o processo. Elas proporcionam aos avaliadores uma visão mais completa sobre a formação geológica.

Por exemplo, a distribuição de um dos parâmetros geológicos pode dar à distribuição do volume de hidrocarbonetos uma forma que se julgue absurda, logo, uma revisão se fará necessária nas avaliações dos parâmetros até que o avaliador se sinta confortável com a forma final da distribuição.

Entretanto, o avaliador não deverá fazer disto uma justificativa para sua ancoragem, modificando os parâmetros à sua conveniência para gerar um resultado volumétrico predeterminado.

4.10 Documentação

Toda a informação utilizada para gerar os resultados deverá ser anexada como documentos de avaliação. Por exemplo, listas de verificação, testes sísmicos, etc (Schuenemeyer, 2002; Bratvold et al. 2002).

Além disto, as avaliações devem conter o nome do avaliador, data, objeto avaliado e outras informações que possam identificar o processo de avaliação e as ferramentas que permitiram aos avaliadores a chegar ao resultado.

Todas as discussões no processo de detalhamento do problema, dos eventos realizados e as quantidades eliciadas devem ser cuidadosamente documentadas. Pelo menos uma das sessões de treinamento deve ser registrada através de vídeo, embora registros escritos sobre os materiais utilizados e as tarefas executadas sejam suficientes. Deve-se criar uma hierarquia da documentação de modo que se possam observar os diferentes níveis de agregação da informação (Keeney e Von Winterfeldt, 1991).

Uma boa documentação é indispensável para uma posterior avaliação dos resultados, verificar que erros, acertos e omissões foram relevantes para o resultado final.