3.4. Variáveis aleatórias contínuas

3.4.1. Introdução

No capítulo passado, vimos alguns modelos para variáveis aleatórias *discretas*. Estas variáveis geralmente resultam de contagens de ocorrências de algum evento: por exemplo, do número de sucessos em *n* repetições de um experimento, de mulheres numa turma de estudantes, de ovos nos ninhos de uma espécie de ave, de vezes que um interruptor elétrico pode ligar e desligar antes de quebrar, de carros parados num sinal de trânsito. Estes valores podem ser *enumerados* – isto é, podem ser contados e organizados numa lista. Não é necessário que esta lista seja *finita*. Por exemplo, a variável "número de tentativas necessárias até que seja obtida uma cara, numa seqüência de lançamentos de uma moeda" é uma VA discreta que não tem limite superior, pois não um há número máximo de tentativas possíveis; o que é importa é que estes valores sejam enumeráveis, e que possamos calcular a probabilidade de ocorrência de cada um dos valores possíveis por meio de um modelo probabilístico (no exemplo, o modelo *geométrico*, seção 3.3.4).

Veremos agora as variáveis aleatórias *contínuas*. Estas variáveis geralmente provém de *medições*, cujos resultados são representados por números reais: por exemplo, o peso ou a altura de uma mulher sorteada de uma população conhecida, a distância que um carro pode percorrer com um litro de gasolina, o tempo de funcionamento de um aparelho antes que sua bateria se esgote. Peso, altura, distância, tempo, são variáveis contínuas, cujos valores possíveis não podem ser enumerados; não podemos, por exemplo, fazer uma lista dos pesos ou alturas possíveis de uma pessoa, pois entre dois valores quaisquer sempre há infinitos outros, já que qualquer intervalo contínuo pode ser subdividido infinitamente.

É claro que a distinção entre variáveis discretas e contínuas é teórica; na prática, medições não podem ser feitas com precisão infinita, e os resultados são sempre arredondados e se tornam valores discretos. Primeiro, porque nenhum aparelho de medição tem precisão infinita. Segundo, porque medições extremamente precisas, mesmo que possíveis, em geral não são necessárias, ou não são úteis. O peso de uma pessoa adulta, por exemplo, geralmente é dado em quilogramas (às vezes em centenas de gramas), e a altura em centímetros. Não faz sentido medir o peso em gramas, ou a altura em milímetros, já que quaisquer medidas de uma pessoa – mesmo o peso e a altura −, variam ao longo do dia. Por outro lado, há variáveis que são fundamentalmente discretas, mas que é mais prático tratar como contínuas; veremos isto com mais detalhes abaixo, na seção 3.4.1.5.

3.4.1.1. Exemplo de um modelo para VAC

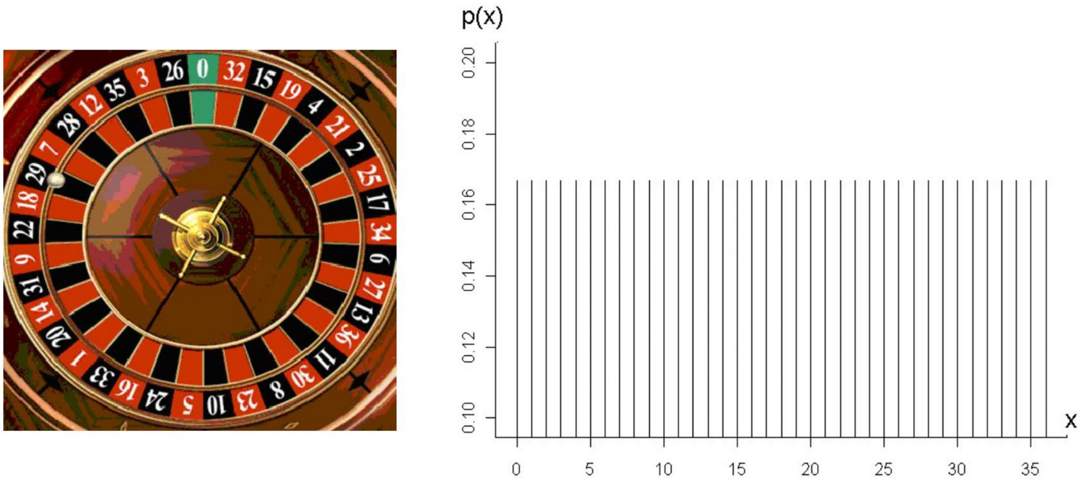
Nos modelos para VADs, usamos a *função de probabilidades p(x)* para calcular a probabilidade de ocorrência de um valor determinado de X:

$$p(x) = P(X=x)$$

Como exemplo disto, usamos na seção 3.3.1.3 uma roleta comum (Fig. **1A**). Os resultados possíveis de cada rodada são os número inteiros de 0 a 36, e a função de probabilidades é

$$p(x) = P(X=x) = 1/37 \qquad \text{para qualquer } x$$

Esta função pode ser representada por um gráfico de barras, como na Fig. **1B**; a variável X é representada no eixo horizontal, a função de probabilidades no eixo vertical. Este é um exemplo de modelo de distribuição *uniforme* para uma VAD.



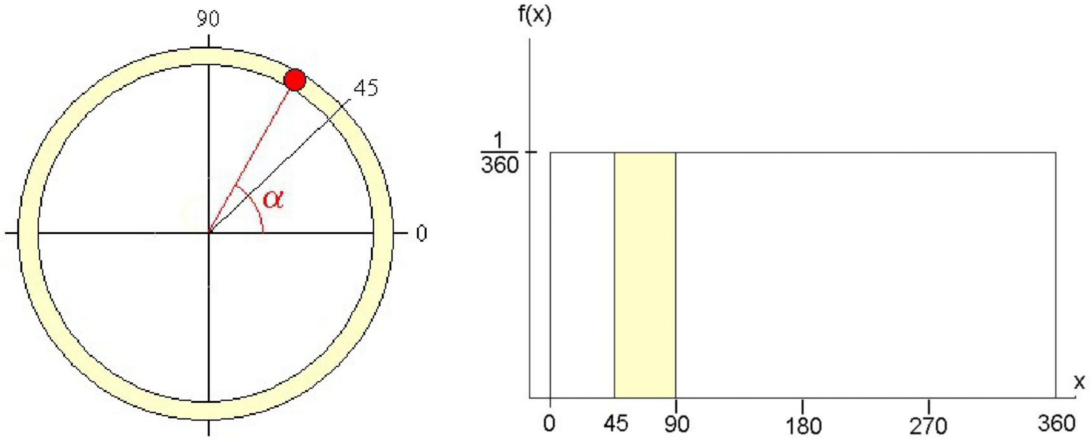
A B

Figura 1. Distribuição de probabilidades em uma roleta

Nos modelos de VAC, como não é possível enumerar todos os valores possíveis de uma variável contínua, não tem sentido calcular a probabilidade de cada valor. Em vez disto, iremos calcular as probabilidades de que a variável X assuma valores dentro de *intervalos* determinados.

Por exemplo, suponhamos uma roleta como na Fig. **2A**, da qual tenhamos retirado as barrinhas de metal que separam as casas numeradas onde a bola pode cair (estas barrinhas são como os trastes no braço de um violão). A bola (em vermelho, na figura) pode agora parar sobre qualquer dos infinitos pontos da circunferência da roleta. A posição da bola será definida pelo angulo α entre o ponto sobre o qual ela parou e o eixo horizontal. Não faz sentido perguntar qual é a probabilidade de ela parar exatamente em cima de um ponto determinado; em vez disto, iremos calcular a probabilidade de a bola cair dentre de um intervalo desejado; por exemplo, no setor entre os ângulos de 45º e 90º.

Para representar as probabilidades dos intervalos, usaremos um conceito que já foi aplicado aos histogramas (Seção 2.1.5): nos histogramas, a freqüência de cada classe é dada pela *área* do retângulo que representa a classe, e não pela altura do retângulo. Nos modelos de VACs, usaremos também gráficos nos quais a probabilidade de cada intervalo será representada pela *área* sob uma curva limitada por aquele intervalo, e não pelo eixo vertical. Para uma o modelo uniforme de distribuição, o gráfico é simplesmente um retângulo como o da Fig. 2B. Veremos mais detalhes destes gráficos na próxima seção.



(A) (B)

Figura 2. Exemplo de variável aleatória contínua

3.4.1.2. Função de densidade de probabilidade

O modelo para uma VAC será portanto uma função a partir da qual poderemos calcular a probabilidade de qualquer intervalo da variável; esta probabilidade será dada pela área do gráfico delimitada pelo intervalo.

Nos histogramas, a variável do eixo vertical é chamada *densidade de freqüência*. Da mesma forma, no gráfico da distribuição de probabilidades na Fig. 2B, a variável do eixo vertical será chamada de *densidade de probabilidade*; não dá diretamente as probabilidades, mas mostra como a probabilidade total, igual a 1, está distribuída ao longo do eixo *X*. Neste exemplo, a densidade é constante, o que quer dizer que a probabilidade se distribui uniformemente, e intervalos com a mesma base terão a mesma probabilidade, qualquer que seja a sua posição ao longo do eixo *X* (este é o modelo mais simples possível para uma VAC; veja seção 3.4.2).

A função que gera este gráfico é chamada de *função de densidade de probabilidades* (*f.d.p.*), e é normalmente representada por *f(x)*. Esta função tem como características :

(i) sempre assume valores positivos (já que não existem probabilidades negativas)

$f(x) \geq 0$ para todo x

(ii) a área total sob a curva que representa *f(x)* é igual a 1

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

(iii) a probabilidade de *X* estar num intervalo entre os valores *a* e *b* será dada pela integração da função no intervalo (*a*,*b*); ou, em termos geométricos, pela área sob o gráfico da função *f(x)* limitada pelos valores *X*=*a* e *X*=*b*.

$$P(a < X < b) = \int_a^b f(x)dx$$

Duas observações importantes que devem ser feitas aqui. Primeiro: Quando trabalhamos com VAs discretas, as probabilidade P(X>x) e P(X≥x) têm valores diferentes. Por exemplo, se *X* é o número obtido no lançamento de um dado

$P(X>4) = p(5) + p(6) \quad = 2/6$
$P(X\geq 4) = p(4) + p(5) + p(6) \quad = 3/6$

Quando trabalhamos com VAs contínuas, porém, esta distinção perde o sentido. Uma vez que uma VAC pode assumir infinitos valores, a probabilidade de qualquer valor x em particular é nula; por isso, tanto faz incluir x no intervalo, $P(X\geq x)$, ou não incluir, $P(X>x)$, que o resultado é o mesmo. A expressão em (iii) por isso às vezes se encontra em livros como

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

o que não faz diferença.

Segundo: Se você não sabe Cálculo, não se preocupe. Na maioria dos modelos de VACs (com exceção dos modelo Uniformes e Exponencial, que serão vistos à seguir), as *f.d.p.s* têm expressões muito complicadas, que não podem ser integradas analiticamente, e as áreas têm que ser calculadas por meio de técnicas de integração numérica (ou seja, num computador). Tradicionalmente, os estatísticos usavam tabelas com valores aproximados (estas tabelas são sempre encontradas como apêndices no final dos livros de Estatística); atualmente, também é possível usar programas estatísticos para fazer estes cálculos. Neste site, mostraremos como usar estes dois métodos.

3.4.1.3. Função de distribuição

Outra função importante, derivada da função de densidade, é a *função de distribuição* ou *função acumulada de probabilidade*, definida como:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

Esta função é muito usada para calcular probabilidades a partir de um modelo de VAC (por exemplo, o modelo normal na seção 3.4.4.2); além disso, ela é fundamental nos estudos teóricos, pois muitas das demonstrações das propriedades dos modelos, e derivações de novos modelos, são feitas usando esta função, ao invés da função de densidade (estes estudos porém estão fora do escopo deste site).

3.4.1.4. Valor esperado e variância de uma VAC

(i) Valor esperado

O valor esperado de uma VAC é definido por:

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

Esta expressão é equivalente àquela que dá o valor esperado de uma VAD (seção 3.3.1.3):

$$E(X) = \sum_{i=1}^{n} x_i p(x_i)$$

A diferença é que aqui o somatório discreto foi substituído por seu equivalente contínuo, que é a integral.

(ii) Variância

O variância de uma VAC é definido por:

$$V(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x)dx$$

Esta expressão também é equivalente à usada em VADs (seção 3.3.1.4):

$$V(X) = \sum_{i=1}^{n} \left(x_i - E(X)\right)^2 p(x_i)$$

Aqui também o somatório discreto foi substituído pela integral.

(iii) Variância como medida de risco

Na análise financeira, o *retorno* (o valor que o investidor ganha, por fazer o investimento) pode ser tratado como uma variável aleatória contínua, e sua variância é freqüentemente usada como base nas medidas de risco do investimento. Se os retornos de dois investimentos têm médias iguais mas variâncias diferentes, o investimento de maior variância é considerado o mais arriscado. Por exemplo, suponha que dois investimentos tenham tido retornos anuais médios de 5% nos últimos anos. Num deles a variância é baixa, e o retorno tem variado geralmente entre 4% e 6%; no outro, a variância é alta, e o retorno tem variado entre -10% a +20%. Este segundo investimento é considerado mais arriscado (o investidor pode até perder dinheiro em alguns anos), embora em média ambos tenham o mesmo rendimento.

(iv) Propriedades do valor esperado e da variância

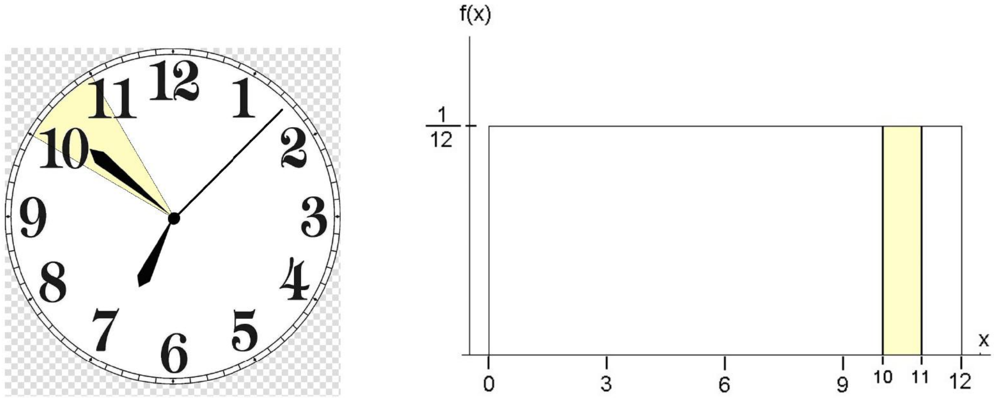
O valor esperado e da variância de uma VAC têm as mesmas propriedades daqueles das VADs (seção 3.3.1.5). Veremos exemplos da utilização do valor esperado, da variância e de suas propriedades nas seções seguintes, que estudam os modelos principais de VAC.

3.4.1.5. De VAC para VAD, e vice-versa

Na teoria, a distinção entre VACs e VADs é bem clara. Nas aplicações práticas, porém, freqüentemente acontece de usarmos modelos de VACs para variáveis que na verdade são discretas, ou modelos de VAD para variáveis contínuas.

Para ilustrar isto, suponhamos um experimento simples, relacionado às roletas das Figs. **1** e **2**: deixamos o ponteiro de um relógio de quartzo girar até que acabe a bateria, e anotamos a posição onde parou o ponteiro dos minutos (Fig. **3A**).

Nestes relógios há um circuito eletrônico, controlado pelas oscilações de um cristal de quartzo, que emite um pulso elétrico a cada segundo; este pulso move simultaneamente os três ponteiros (*). O ponteiro dos segundos dá uma volta completa em torno do mostrador a cada minuto, e pára portanto em 60 posições diferentes. O ponteiro dos minutos dá uma volta a cada 60 min = 3600 segundos, e pára em 3600 posições diferentes. Se queremos fazer um modelo probabilístico para a posição deste ponteiro, portanto, deveríamos a rigor usar um diagrama de barras como o da Fig. **1B**, mas com 3600 barras. Numa aplicação real, porém, provavelmente optaríamos por usar um modelo de VAC uniforme, como o da Fig. **3B**, e trabalharíamos não com a probabilidade de o ponteiro parar sobre um ponto dado, mas sim, a probabilidade de ele parar dentro de um intervalo.



(A) (B)

Figura 3. Exemplo de variável aleatória: posição do ponteiro de um relógio

A probabilidade de ele parar entre as marcas de 10 e 11 horas, por exemplo, será igual à área do retângulo em amarelo na figura, dada pelo produto de sua base por sua altura :

$$P(10 \leq X \leq 11) = (11 - 10) \times 1/12 = 1/12$$

Este problema é um intermediário entre os das Figs. **1** e **2**: entre uma roleta de movimento discreto, na qual a bola somente pode parar num número limitado de posições definidas, e uma roleta de movimento contínuo, onde a bola pode parar em infinitas posições. Um exemplo mais prático pode ser encontrado em pesquisas sobre salários de trabalhadores. Salários são medidos em *reais*, cuja subdivisão mínima é o *centavo*; não existem salários que usem meios centavos. A variável *salário*, portanto, é discreta. Num levantamento dos salários dos trabalhadores de uma cidade, poderíamos fazer uma lista de todos os salários possíveis, por exemplo

1.634,77 1.634,78 1.634,79 1.634,80 etc.

Isto porém seria trabalhoso, e provavelmente inútil. Economistas que fazem este tipo de estudo devem estar mais interessados em conhecer a proporção dos trabalhadores que têm salários entre R$ 1.500,00 e R$ 2.000,00, por exemplo, do que a proporção dos que têm salários de exatamente R$ 1.634,78. É mais prático portanto tratar a variável como contínua, e calcular as probabilidades de intervalos, e não as de valores pontuais. (Outros exemplos serão vistos na seção 3.4.4.5, onde o modelo *normal* de VAC será usado como aproximação para os modelos *binomial* e de *Poisson*, que são de VAD)

(*) Num relógio mecânico o ponteiro de segundos se move várias vezes por segundo (geralmente de 4 a 8 vezes, dependendo do modelo). Um relógio de quartzo também poderia fazer isto, se fossem acrescentadas engrenagens multiplicadoras no mecanismos. Isto não é feito, porém, porque aumentaria o consumo de energia do relógio − ao invés de o mecanismo se mover uma vez por segundo, ele teria que se mover várias vezes.