3.3.6. Modelo de Poisson

3.3.6.1. Introdução

Os modelos geométrico e binomial (seções 3.3.4 e 3.3.5), e também o hipergeométrico e o binomial negativo (seção 3.3.7), são usados nos experimentos em que várias *tentativas* são feitas, e em cada tentativa um *sucesso* pode ocorrer uma vez, ou pode não ocorrer. Veremos agora um modelo de um tipo bem diferente, usado para modelar variáveis em problemas como estes:

> (1) O número médio de raios alfa (emitidos por rádio) que atingem por segundo uma pequena área e são registrados por um contador Geiger é de λ=1,4. Qual é a probabilidade de, num dado intervalo de um segundo, nenhum raio alfa ser registrado pelo contador?
>
> (2) O número de carros que param numa certa esquina quando fecha o sinal de trânsito é uma variável cuja média foi estimada em $\lambda = 2{,}3$. Se neste sinal pararem cinco ou mais carros, haverá um congestionamento causado pelo bloqueio de uma das ruas transversais. Qual é a probabilidade de isto ocorrer?
>
> (3) Uma tubulação de 10 km de extensão tem um custo anual de manutenção que varia da seguinte forma: se não houver nenhum vazamento ao longo dela, o custo será de \$500; se houver um vazamento, será de \$700; se houver mais de um, de \$800. Se este tipo de tubulação tem uma média anual de 0,08 vazamentos por quilômetro, qual será o custo anual médio de manutenção?

Estes problemas têm em comum estas características:

- Em vez de considerarmos *tentativas*, consideramos *intervalos* de alguma variável contínua no espaço (distância, área, volume), ou no tempo;
- Em cada intervalo, o evento de interesse pode não ocorrer, ou pode ocorrer uma ou mais vezes;
- A variável de interesse X é o número de vezes em que o evento ocorre num dado intervalo;
- Não conhecemos a probabilidade de ocorrência do evento, mas conhecemos o número médio de ocorrências dentro de um intervalo.

Outra característica, que talvez não seja tão evidente nos exemplos, é que número de ocorrências é sempre relativamente pequeno (geralmente da ordem de algumas unidades). O modelo usado é por isso também chamado por alguns autores de modelo para *eventos raros*.

3.3.6.2. Função de probabilidades, valor esperado e variância

Um modelo adequado para problemas deste tipo foi publicado pela primeira vez em 1837 por Siméon–Denis Poisson (1781-1840). Partindo do modelo binomial, Poisson deduziu um modelo com a seguinte função de probabilidades:

$$p(x) = P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad x = 0,1,2,...$$

cujo único parâmetro é λ (letra grega *lambda*), o número médio de ocorrências por intervalo. Para indicar que uma variável X segue este modelo, usamos a forma:

$$X \sim Pois(\lambda)$$

Uma curiosidade deste modelo é que seu valor esperado e sua variância são ambas iguais ao parâmetro λ:

$$E(X) = \lambda \qquad Var(X) = \lambda$$

Para deduzir a expressão, Poisson partiu dos seguintes pressupostos:

(i) O número de ocorrências do evento em cada intervalo é independente do número de ocorrências em qualquer outro intervalo (se estes dois intervalos não se sobrepõe);

(ii) A distribuição de probabilidades do número de ocorrências em qualquer intervalo depende apenas da amplitude deste intervalo, e não de seus valores extremos; o número médio de ocorrências por intervalo portanto é constante, qualquer que seja a localização do intervalo;

(iii) A probabilidade de *uma* ocorrência é diretamente proporcional a amplitude do intervalo, para intervalos suficientemente pequenos;

(iv) A probabilidade de *mais de uma* ocorrência, para intervalos suficientemente pequenos, pode ser considerada igual a 0.

É claro que, quando um modelo matemático é usado em aplicações no mundo real, os pressupostos são atendidos apenas de forma aproximada. No problema (2), por exemplo, o número médio de carros que param no sinal deve variar ao longo do dia (é provavelmente maior na hora do *rush* e menor durante a madrugada); a média mencionada no problema deve ser válida apenas em algumas horas específicas do dia, não para o dia inteiro. No problema (3), estamos supondo que a probabilidade de ocorrer um defeito seja constante, de um segmento da tubulação para outro; só um engenheiro bem familiarizado com o assunto pode decidir se este pressuposto descreve, pelo menos aproximadamente, o que acontece em uma determinada tubulação.

3.3.6.3. Exemplos

Usando o modelo de Poisson, podemos resolver os três problemas enunciados acima.

Problema (1)
X: número de partículas registradas pelo contador Geiger; λ=1,4
P(X = 0) = p(0) = ?

$$p(0) = \frac{e^{-1,4}1,4^0}{0!} = e^{-1,4} \cong 0,2466$$

resposta: $p \approx 0,2466$

Problema (2)
X: número de carros parados no sinal; λ = 2,3
P(X ≥ 5) = ?
P(X ≥ 5) = 1–P(0 ≤ X ≤ 4)

$$p(0) = \frac{e^{-2,3}2,3^0}{0!} = e^{-2,3} \cong 0,1003$$

$$p(1) = \frac{e^{-2,3}2,3^1}{1!} = 2,3 \times e^{-2,3} \cong 0,2306 \text{ , etc.}$$

Estas probabilidades são mostradas na Tabela 1.

Tabela 1. Distribuição de Poisson, λ=2,3

X	p(x)
0	0,1003
1	0,2306
2	0,2652
3	0,2033
4	0,1169
5+	0,0838

P(X ≥ 5) = 1–P(0≤X≤4) ≈ 1-(0,1003+0,2306+0,2652+0,2033+0,1169) ≈ 0,0838
resposta: $p \approx 0,0838$

Problema (3)
X : número de defeitos; λ = 0,08 por km → 0,80 por 10 km

$$p(0) = \frac{e^{-0.8}0.8^0}{0!} = e^{-0.8} \cong 0,4493$$

$$p(1) = \frac{e^{-0,8}0,8^1}{1!} \cong 0,3595$$

Tabela 2. Número de defeitos (X) e custo anual de manutenção

X	p(x)	custo ($)
0	0,4493	500
1	0,3595	700
2+	0,1912	800
Σ	1,0000	-

A Tab. 2 mostra os valores assumidos pelo número de defeitos e pelo custo, e as probabilidades associadas a estes valores. O custo esperado pode então ser calculado a partir dos valores na tabela:

custo esperado ≈ 500×0,4493+700×0,3595+800×0,1912 ≈ \$ 629,26

resposta: \$ 629,26

3.3.6.4. O modelo de Poisson como aproximação do modelo binomial

Poisson propôs originalmente o modelo como uma aproximação da distribuição binomial, embora mais tarde ele tenha sido empregado para resolver vários outros tipos de problemas.

Quando o número n de repetições é muito grande, não é possível usar o modelo binomial, pois o fatorial de números grandes não pode ser calculado (veja nota na Seção 3.3.5.4). O modelo de Poisson fornece uma boa aproximação quando $n \to \infty$ e $p \to 0$. Por exemplo, a Tab. 3 compara as probabilidades calculadas por um modelo binomial $B(n,p)$ e por um de Poisson $Pois(\lambda)$, para diversos valores de n e de p, mantendo constante a média $\lambda = np = 2$. É possível observar que o modelo de Poisson dá uma boa aproximação, que melhora à medida que aumenta n e diminui p.

Tabela 3. Exemplos da aproximação da distribuição binomial pela de Poisson

		Probabilidades calculadas por modelo binomial		
n	*p*	P(X = 0)	P(X = 1)	P(X = 4)
10	0,2	0,1074	0,2684	0,0881
20	0,1	0,1216	0,2702	0,0898
50	0,04	0,1299	0,2706	0,0902
100	0,02	0,1326	0,2707	0,0902
200	0,01	0,1340	0,2707	0,0902
Prob. calculadas pelo modelo de Poisson →		0,1353	0,2707	0,0902