

Distribuições Discretas: Geométrica

Prof. Alejandra, Cristian, Paula e Rafael

Introdução à Probabilidade e Estatística



Estes slides tem por objetivo apenas complementar e esclarecer os textos principais indicados para a disciplina, e não devem ser usados como único texto de referência.

Eles tratam apenas dos principais aspectos da teoria abordada, deixando para os textos e vídeos os detalhes teóricos e exemplos mais elaborados.

Junto à essas notas o aluno deve ler os textos e ver os vídeos indicados no site do curso no Moodle, além da referência principal do curso.

Começemos com um exemplo...

Um colecionador de álbuns de figurinhas está tentando completar um certo álbum, e lhe falta apenas uma figura. Ele sabe que a probabilidade de encontrar essa figura em um pacote comprado na banca é de 0,02. Seja N o total de pacotes comprados até conseguir encontrar a figura desejada. Supondo que cada pacote pode ou não ter a figura de maneira independente dos demais, encontre a função probabilidade de N .

Solução: Defina os eventos

$$A_k = \{\text{a figura desejada veio no } k\text{-ésimo pacote comprado}\},$$

para $k = 1, 2, 3, \dots$. O problema nos diz que para qualquer k , $P(A_k) = 0,02$. Além disso, os eventos $A_1, A_2, \dots$ são independentes.

Observe agora que

$$\{N = 1\} = A_1, \quad \{N = 2\} = A_1^c \cap A_2, \quad \{N = 3\} = A_1^c \cap A_2^c \cap A_3,$$

e em geral

$$\{N = k\} = A_1^c \cap \dots \cap A_{k-1}^c \cap A_k.$$

Segue da independência dos eventos $A_1, A_2, A_3, \dots$, que

$$P(N = 1) = 0,02, \quad P(N = 2) = 0,98 \cdot 0,02, \quad P(N = 3) = (0,98)^2 \cdot 0,02$$

e assim, para $k = 1, 2, 3, \dots$,

$$p(k) = P(N = k) = (0,98)^{k-1} 0,02.$$

Observação 1

Neste problema o colecionador está realizando ensaios independentes de Bernoulli, cada um deles com probabilidade 0,02 de sucesso, até observar o evento desejado pela primeira vez.

Esta é uma situação bastante comum em modelagem estatística, e a variável aleatória que conta o total de repetições é o tema destas notas.

Definição 1

Diremos que X é uma variável aleatória **Geométrica** com parâmetro p , quando esta representa o número tentativas até a ocorrência do primeiro sucesso em sucessivos ensaios de Bernoulli independentes, cada um com probabilidade p de sucesso.

Neste caso X tem função de probabilidade dada por

$$p(k) = P(X = k) = (1 - p)^{k-1}p, \quad k \in \{1, 2, 3, \dots\}$$

Alternativamente, podemos dizer que X tem **distribuição Geométrica** com parâmetro p .

Notação $X \sim \text{Geom}(p)$ ou $X \sim \text{Geometrica}(p)$.

A seguir seremos confrontados pela necessidade de calcular uma soma de infinitos termos, e por isso vamos dar uma pausa na descrição das variáveis geométricas para entender um pouco as somas que enfrentaremos.

Estamos interessados nas somas do tipo

$$\sum_{k=1}^{\infty} r^{k-1} = 1 + r + r^2 + r^3 + \dots,$$

para $0 < r < 1$, conhecidas como *séries geométricas*.

Antes de calcular esta soma, vamos entender como somas infinitas (conhecidas como *séries*) são definidas.

Dada uma sequência $(a_1, a_2, a_3, \dots)$, definimos as *somas parciais* S_n , $n \geq 1$, como a soma

$$S_n := a_1 + a_2 + \dots + a_n = \sum_{k=1}^n a_k.$$

Definimos então a *série*

$$\sum_{k=1}^{\infty} a_k = a_1 + a_2 + \dots = \lim_{n \rightarrow \infty} S_n.$$

Não vamos entrar em mais detalhes sobre isso, e os alunos interessados nesse assunto podem buscar uma disciplina chamada *Sequências e Séries*, onde entramos mais a fundo deste assunto.

Para calcularmos então o valor de uma série geométrica precisamos primeiro calcular o valor de

$$S_n = \sum_{k=1}^n r^{k-1} = 1 + r + r^2 + \dots + r^{n-1}.$$

Para isso note que

$$r \cdot S_n = r(1 + r + r^2 + \dots + r^{n-1}) = r + r^2 + r^3 + \dots + r^n.$$

Segue daí que

$$S_n - r \cdot S_n = (1 + r + r^2 + \dots + r^{n-1}) - (r + r^2 + r^3 \dots + r^n) = 1 - r^n,$$

de onde concluímos que

$$(1 - r) \cdot S_n = 1 - r^n$$

e

$$S_n = \frac{1 - r^n}{1 - r}.$$

Observe agora que, como $0 < r < 1$, então $\lim_{n \rightarrow \infty} r^n = 0$, e assim

$$\lim_{n \rightarrow \infty} S_n = \frac{1}{1 - r},$$

revelando que a série geométrica, para $0 < r < 1$, é dada por

$$\sum_{k=1}^{\infty} r^{k-1} = 1 + r + r^2 + \dots = \frac{1}{1 - r}.$$

Acabamos de ver que se $X \sim \text{Geométrica}(p)$ então a função de probabilidade de X é dada por

$$p(k) = P(X = k) = (1 - p)^{k-1} p, \quad k = 1, 2, 3, \dots$$

Para confirmar que $p(\cdot)$ está realmente bem definida, precisamos verificar que

$$\sum_{k=1}^{\infty} p(k) = p(1) + p(2) + p(3) + \dots = 1.$$

Para isso usemos a série geométrica e notemos que

$$\begin{aligned} \sum_{k=1}^{\infty} p(k) &= \sum_{k=1}^{\infty} (1 - p)^{k-1} p \\ &= p \cdot \left[\sum_{k=1}^{\infty} (1 - p)^{k-1} \right] \\ &= p \cdot \left[\frac{1}{1 - (1 - p)} \right] \\ &= 1. \end{aligned}$$

Exemplo

Rolamos um dado não-viciado repetidamente, até observarmos o número 6 pela primeira vez. Chame de X o total de lançamentos necessários e encontre $P(X > 10)$, $P(X \leq 12)$ e a função distribuição acumulada de X .

Solução: Sabemos que $X \sim \text{Geometrica}(1/6)$ e portanto

$$P(X = k) = \left(\frac{5}{6}\right)^{k-1} \frac{1}{6},$$

para $k = 1, 2, 3, \dots$.

Assim

$$\begin{aligned} P(X > 10) &= P(X = 11) + P(X = 12) + P(X = 13) + \dots \\ &= \left(\frac{5}{6}\right)^{10} \frac{1}{6} + \left(\frac{5}{6}\right)^{11} \frac{1}{6} + \left(\frac{5}{6}\right)^{12} \frac{1}{6} + \dots \\ &= \left(\frac{5}{6}\right)^{10} \left[\frac{1}{6} + \frac{5}{6} \cdot \frac{1}{6} + \left(\frac{5}{6}\right)^2 \frac{1}{6} + \dots \right] \end{aligned}$$

Note agora que

$$\frac{1}{6} + \frac{5}{6} \cdot \frac{1}{6} + \left(\frac{5}{6}\right)^2 \frac{1}{6} + \dots = P(X=1) + P(X=2) + P(X=3) + \dots = 1,$$

e portanto

$$P(X > 10) = \left(\frac{5}{6}\right)^{10}.$$

Para calcular $P(X \leq 12)$, primeiro calculamos $P(X > 12)$ usando as mesmas ideias acima ([deixamos os detalhes como exercício](#)), e encontramos

$$P(X > 12) = \left(\frac{5}{6}\right)^{12},$$

de onde segue que

$$P(X \leq 12) = 1 - \left(\frac{5}{6}\right)^{12}.$$

O queremos agora é seguir os cálculos do exemplo anterior, e assim tentar encontrar a função distribuição acumulada de uma v.a. $X \sim \text{Geometrica}(p)$.

Para começar, sabemos que, como X é discreta, sua função distribuição é uma função escada, com saltos nos valores $k \in \mathbb{R}$ tal que $P(X = k) > 0$. Ou seja, em valores de $\{1, 2, 3, \dots\}$.

Para começar tomemos $a < 1$. Como X não assume valores abaixo de a , segue que $P(X \leq a) = 0$.

Para $a \geq 1$, sabemos que os saltos acontecerão nos valores inteiros, de modo que para cada $k \in \{1, 2, 3, \dots\}$ devemos analisar a função para $a \in [k, k+1)$.

Seguindo os passos usado nos exemplo anterior, vamos começar analisando $P(X > a)$.

Assim, para $a \in [k, k+1)$, note que os valores que X assume que são maiores que a são $k+1, k+2, \dots$, e portanto

$$\begin{aligned}P(X > a) &= P(X = k+1) + P(X = k+2) + P(X = k+2) + \dots \\&= (1-p)^k p + (1-p)^{k+1} p + (1-p)^{k+2} p + \dots \\&= (1-p)^k (p + (1-p)p + (1-p)^2 p + \dots) \\&= (1-p)^k (P(X=1) + P(X=2) + P(X=3) + \dots) \\&= (1-p)^k.\end{aligned}$$

Segue assim que $P(X \leq a) = 1 - (1-p)^k$ para $a \in [k, k+1)$, e juntando tudo temos

$$P(X \leq a) = \begin{cases} 0 & , \text{ se } a < 1; \\ 1 - (1-p)^k & , \text{ se } k \leq a < k+1, k = 1, 2, 3, \dots \end{cases}$$

Mais um exemplo...

Lançamos um dado honesto até observar o número 6 pela primeira vez. Qual a média de lançamentos que esperamos realizar?

Solução: Se chamamos de N o total de lançamentos realizados, sabemos que $N \sim \text{Geométrica}\left(\frac{1}{6}\right)$. Ou seja

$$p(k) = P(N = k) = \left(\frac{5}{6}\right)^{k-1} \frac{1}{6}.$$

Para calcular $E[N]$ vamos fazer um pequeno *truque*... Esse truque que fica mais fácil de ser percebido com a notação de somatório, então lembremos que

$$E[N] = 1 \cdot p(1) + 2 \cdot p(2) + 3 \cdot p(3) + \dots = \sum_{k=1}^{\infty} k \cdot p(k),$$

e portanto

$$E[N] = \sum_{k=1}^{\infty} (k-1+1) \cdot p(k) = \sum_{k=1}^{\infty} (k-1) \cdot p(k) + p(k) \quad (1)$$

Somando as partes em azul e em vermelho separadamente, temos

$$\sum_{k=1}^{\infty} p(k) = \sum_{k=1}^{\infty} P(N=k) = 1,$$

e

$$\begin{aligned}\sum_{k=1}^{\infty} (k-1) \cdot p(k) &= 0 \cdot \frac{1}{6} + 1 \cdot \frac{5}{6} \cdot \frac{1}{6} + 2 \cdot \left(\frac{5}{6}\right)^2 \frac{1}{6} + 3 \cdot \left(\frac{5}{6}\right)^3 \frac{1}{6} + \dots \\ &= \frac{5}{6} \cdot \left(1 \cdot \frac{1}{6} + 2 \cdot \frac{5}{6} \cdot \frac{1}{6} + 3 \cdot \left(\frac{5}{6}\right)^2 \frac{1}{6} + \dots\right) \\ &= \frac{5}{6} \cdot (1 \cdot p(1) + 2 \cdot p(2) + 3 \cdot p(3) + \dots) \\ &= \frac{5}{6} \cdot E[N].\end{aligned}$$

Mais um exemplo...

Voltando à equação (1)

$$E[N] = \sum_{k=1}^{\infty} (k-1) \cdot p(k) + p(k) = \sum_{k=1}^{\infty} (k-1) \cdot p(k) + \sum_{k=1}^{\infty} p(k),$$

e assim

$$E[N] = 1 + \frac{5}{6} \cdot E[N].$$

E resolvendo a equação acima encontramos $E[N] = 6$.

Observação 2

O resultado acima é consistente com a interpretação da probabilidade como frequência média de observação de um evento em várias repetições do experimento.

De fato, pensando assim, a probabilidade $1/6$ nos diz que em sucessivas rolagens de um dado esperamos encontrar o número 6, em média, em 1 a cada 6 rolagens.

Assim, entre duas observações sucessivas do número 6 (em particular entre o início e a primeira observação) devem haver em média 6 rolagens.

Lançamos um dado honesto até observar o número 6 pela primeira vez. Denote por N o total de lançamentos necessários.

Você consegue adaptar a ideia usada acima no cálculo de $E[N]$ para calcular $E[N^2]$, e assim encontrar a variância $Var[N]$?

Resposta: $Var[N] = 30$.

As ideias acima podem ser usadas para encontrar o valor esperado de uma variável aleatória $N \sim \text{Geométrica}(p)$.

Para isso lembre que

$$p(k) = P(N = k) = (1 - p)^{k-1} p.$$

Assim, temos

$$E[N] = 1 \cdot p(1) + 2 \cdot p(2) + 3 \cdot p(3) + \dots = \sum_{k=1}^{\infty} k \cdot p(k),$$

e portanto, como anteriormente,

$$E[N] = \sum_{k=1}^{\infty} (k - 1 + 1) \cdot p(k) = \sum_{k=1}^{\infty} (k - 1) \cdot p(k) + p(k) \quad (2)$$

Somando as partes em azul e em vermelho separadamente, temos

$$\sum_{k=1}^{\infty} p(k) = \sum_{k=1}^{\infty} P(N=k) = 1,$$

e

$$\begin{aligned}\sum_{k=1}^{\infty} (k-1) \cdot p(k) &= 0 \cdot p + 1 \cdot (1-p) \cdot p + 2 \cdot (1-p)^2 p + 3 \cdot (1-p)^3 p + \dots \\ &= (1-p) \cdot (1 \cdot p + 2 \cdot (1-p) \cdot p + 3 \cdot (1-p)^2 p + \dots) \\ &= (1-p) \cdot (1 \cdot p(1) + 2 \cdot p(2) + 3 \cdot p(3) + \dots) \\ &= (1-p) \cdot E[N].\end{aligned}$$

Voltando à equação (2)

$$E[N] = \sum_{k=1}^{\infty} (k-1) \cdot p(k) + p(k) = \sum_{k=1}^{\infty} (k-1) \cdot p(k) + \sum_{k=1}^{\infty} p(k),$$

e assim

$$E[N] = 1 + (1-p) \cdot E[N].$$

E resolvendo a equação acima encontramos

$$E[N] = \frac{1}{p}.$$

Seguindo uma ideia análoga encontramos que

$$\text{Var}[N] = \frac{1-p}{p^2}.$$

Proposição

Se X uma variável aleatória com distribuição geométrica de parâmetro $p \in (0,1)$, então

$$P(X = k) = (1 - p)^{k-1} p, \quad k \in \{1, 2, 3, \dots\},$$

$$E[X] = \frac{1}{p}$$

e

$$\text{Var}[X] = \frac{1-p}{p^2}.$$

Um apostador obstinado joga na mega-sena semanalmente. Esperando ganhar acertar ao menos uma quina (acertar 5 das 6 dezenas sorteadas), ele joga em 8 dezenas a cada semana.

Com um jogo de 8 dezenas, a probabilidade de acertar as seis dezenas sorteadas é de $1/1.787.995$, enquanto a probabilidade de acertar a quina é de $1/17.192$.

Qual o valor esperado e variância e do total de jogos que este apostador deve jogar até ganhar ao menos a quina pela primeira vez?

Solução: A probabilidade do jogador ganhar ao menos a quina é de

$$p = \frac{1}{1.787.995} + \frac{1}{17.192} \approx 0,00005872.$$

Assim, se N é o total de jogos necessários até ganhar ao menos a quina pela primeira vez, então $N \sim \text{Geométrica}(0,00005872)$.

Exemplo

Com isso temos

$$E[N] = \frac{1}{0,00005872} \approx 17.029,97 \text{ jogos,}$$

e

$$Var[N] = \frac{1 - 0,00005872}{(0,00005872)^2} \approx 297.247.919,14.$$

Assim, jogando 8 dezenas semanalmente, o apostador deve esperar em média 17.029 semanas (aproximadamente 326 anos) para ganhar ao menos a quina pela primeira vez!!

Observação 3

Se $X \sim \text{Geom}(p)$, então

- Se $p = 1$, então $P(X = 1) = 1$;
- Para $0 < p < 1$, $p(k) = P(X = k) = (1 - p)^{k-1}p$ é decrescente em k ;
- A esperança $E[X] = \frac{1}{p}$ cresce quando p decresce. Mais que isso

$$\lim_{p \rightarrow 0^+} E[X] = \infty;$$

- Se $p = 1$, então $\text{Var}[X] = 0$;
- Assim como a esperança, $\text{Var}[X]$ cresce quando p decresce e

$$\lim_{p \rightarrow 0^+} \text{Var}[X] = \infty;$$

Atenção: Para visualizar o gráfico e verificar as propriedades acima acesse o applet Geogebra no site do do curso no Moodle na atividade sobre a distribuição Geométrica.

Após olhar nossos cálculos, o apostador do exemplo anterior (que ainda não se libertou da sua obsessão com jogos de azar) resolveu tentar a sorte em um jogo com maiores probabilidades de vitória. Sabendo que em um jogo de roleta, ao apostar no vermelho, a probabilidade de vitória é de $9/19$ ele resolve ir à um cassino e fazer sucessivas apostas, sempre apostando no vermelho.

Após 5 apostas, e sabendo que só tem dinheiro para mais 4 apostas, ele pergunta a um amigo matemático qual a probabilidade dele perder ao menos outras 4 apostas antes de vencer pela primeira vez, dado que ele já perdeu 5 apostas.

Qual foi a resposta do amigo? Calcule também a probabilidade dele ter perdido ao menos 4 apostas antes de ganhar a primeira, e compare os dois resultados.

(**Dica:** Se X é o total de apostas até ganhar a primeira, o evento “perder ao menos k apostas antes de ganhar a primeira pode ser escrito como $\{X \geq k + 1\}$.)