1 Distribuição Geométrica

Material didático preparado pelo professor Maurício Mota para a disciplina CC0282- Probabilidade I ministrada no semestre 2021.1.

1.1 Introdução

Fato 1. Considere que um experimento de Bernoulli com probabilidade de sucesso em cada repetição constante e igual a p, $0 < p < 1$. é ensaiado independentemente até que o primeiro sucesso aconteça. Seja X a variável aleatória correspondente ao número total de repetições até a obtenção do primeiro sucesso. A f.p. de X é dada por

$$f(x) = p\, q^{x-1} I_{\{1,\ldots,\infty\}}(x) \tag{1}$$

Prova: O suporte da distribuição é o conjunto $A = \{1, \ldots, \infty\}$ e seja $x \in A$. Vai-se calcular a $f(x) = P(X = x)$, isto é, a probabilidade de que sejam necessários x repetições. A probabilidade de que as $x - 1$ primeiras repetições sejam fracassos e a última sucesso é dada por:

$$P(FF \ldots FS) = q^{x-1} p,$$

pois as repetições são independentes.

Assim,

$$f(x) = P(X = x) = p\, q^{x-1} I_{\{1,2,\ldots,\infty\}}(x).$$

Esta distribuição é chamada na literatura de distribuição Geométrica e é uma das distribuições mais importantes da Estatística.

1.2 Definição

Uma variável aleatória discreta X é dita possuir distribuição Geométrica de parâmetro e p, onde $q = 1-p$, $0 \leq p \leq 1$, se sua função de probabilidade ($f.p.$) é da forma:

$$f(x) = p\, q^{x-1} I_{\{1,2,\ldots,\infty\}}(x). \tag{2}$$

Notação: $X \sim Geo(p)$.

Observação: Lê-se a notação acima do seguinte modo: X segue distribuição Geométrica de parâmetro p.

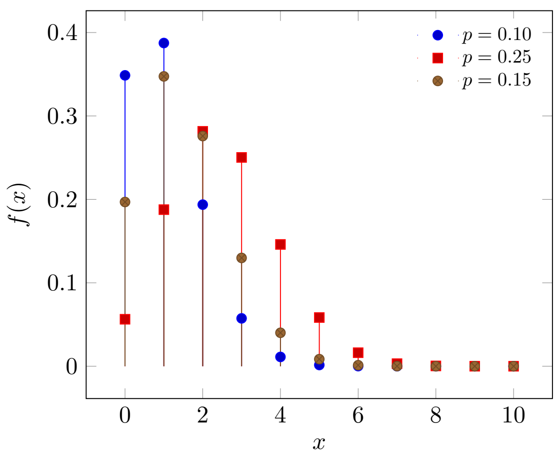

Figura 1: Gráfico da Função de Probabilidade Geométrica

A Figura 1 apresenta o gráfico da distribuição Geométrica para certos valores do parâmetro p.

1.3 Propriedades da função de probabilidade

Fato 2. A expressão (2) é realmente uma função de probabilidade.

<u>*Prova*</u>: deve-se verificar que

i. $f(x) > 0, \; x \in A.$; e

ii. $\sum_A f(x) = 1.$

sendo $A = \{x \in R | f(x) > 0\}$ o suporte da distribuição de X. Como $A = \{1, 2, \ldots, \}$ e para $0 < p < 1$ tem-se $f(x) = pq^{x-1} > 0$ para qualquer ponto do suporte . A segunda propriedade nos diz que a soma dos valores das probabilidades para os pontos do suporte é 1. Assim

$$\sum_{x=1}^{\infty} p\, q^{x-1} = p \sum_{x=1}^{\infty} q^{x-1} = p \frac{1}{1-q} = 1, \qquad \blacksquare$$

pois temos um progressão geométrica de razão $0 < q < 1$.

1.4 Função geradora de probabilidades

Fato 3. Se $X \sim Geo(p)$, então

$$G(t) = \frac{pt}{1-qt}, \quad |t| < \frac{1}{q}.$$

<u>*Prova*</u>:

$$G(t) = E(t^X) = \sum_{x=1}^{\infty} t^x \, p \, q^{x-1},$$

Colocando $\dfrac{p}{q}$, em evidência temos:

$$\varphi(t) = \frac{p}{q}\sum_{x=1}^{\infty}(qt)^x,$$

que é uma progressão geométrica infinita de razão $a = qt$.

Sabemos que para $|a| < 1$ a soma da progressão geométrica infinita

$$\sum_{i=1}^{\infty} a^i = \frac{a}{1-a}.$$

Finalmente,

$$G(t) = \frac{p}{q}\frac{qt}{1-qt} = \frac{pt}{1-qt}, \quad |t| < \frac{1}{q},$$

A condição existência aparece

$$|qt| = q|t| < 1,$$

$$|t| < \frac{1}{q}.$$

1.5 Função geradora de momentos

Fato 4. Se $X \sim Geo(p)$, então

$$M(t) = \frac{pe^t}{1-qe^t}, \quad |t| < -ln(q).$$

Prova:

$$M(t) = E(e^{tX}) = \sum_{x=1}^{\infty} e^{tx}\, p\, q^{x-1} = \frac{p}{q}\sum_{x=1}^{\infty}(qe^t)^x \;= \frac{pe^t}{1-qe^t}, \quad t < -ln(q) \qquad \blacksquare,$$

pois temos uma progressão geométrica infinita de razão $a = qe^t$ e cuja condição de existência é dada por:

$$|qe^t| = q\, e^t < 1,$$

assim

$$ln(qe^t) = lnq + t < ln(1) = 0,$$

o que acarreta

$$t < -ln(q).$$

1.6 Momentos fatoriais

Fato 5. Se $X \sim Geo(p)$, então os quatro primeiros momentos fatoriais são dados por:

$$E\left(X_{[r]}\right) = \begin{cases} \dfrac{1}{p}, & \text{se } r = 1 \\ \\ \dfrac{2q}{p^2}, & \text{se } r = 2 \\ \\ \dfrac{6q^2}{p^3}, & \text{se } r = 3 \\ \\ \dfrac{24q^3}{p^4}, & \text{se } r = 4 \end{cases}$$

Prova: como $E\left(X_{[1]}\right) = E(X)$ tem-se:

$$E(X) = \sum_{x=1}^{\infty} xp\ q^{x-1} = p\sum_{x=1}^{\infty} x\ q^{x-1} = p\sum_{x=1}^{\infty} (q^x)' = p\left(\sum_{x=1}^{\infty} q^x\right)',$$

onde a derivada é tomada em relação a q.

mas

$$S = \sum_{x=1}^{\infty} q^x = \frac{q}{1-q},$$

e a derivada de S em relação a q fica:

$$S' = \frac{q'(1-q) - q(-1)}{(1-q)^2} = \frac{1}{(1-q)^2} = \frac{1}{p^2}.$$

Assim,

$$E(X) = p\frac{1}{p^2} = \frac{1}{p}.$$

Vai-se calcular agora o segundo momento fatorial, assim

$$E\left(X_{[2]}\right) = E\left[X(X-1)\right] = \sum_{x=2}^{\infty} x(x-1)p\ q^{x-1} = pq\sum_{x=2}^{\infty} x(x-1)\ q^{x-2} = pqS_1,$$

mas $x(x-1)\ q^{x-2} = (q^x)''$ é a derivada segunda de q^x em relação a q. logo,

$$S_1 = \sum_{x=2}^{\infty} x(x-1)\ q^{x-2} = \sum_{x=2}^{\infty} (q^x)^{''} = \left(\sum_{x=2}^{\infty} q^x \right)^{''}.$$

Seja

$$S_2 = \sum_{x=2}^{\infty} q^x = \frac{q^2}{1-q},$$

e a derivada primeira de S_2 em relação a q fica:

$$S_2' = \frac{(2q)(1-q) - q^2(-1))}{(1-q)^2} = \frac{2q - 2q^2 + q^2}{(1-q)^2} = \frac{2q - q^2}{(1-q)^2} = \frac{2q - q^2)}{(1-q)^2}.$$

A derivada segunda de S_2 em relação a q fica:

$$S_2^{''} = \frac{(2-2q)(1-q)^2 - (2q-q^2)(-2(1-q))}{(1-q)^4} = \frac{2(1-q)(1-q)^2 + 2(2q-q^2)(1-q)}{(1-q)^4} = 2\ \frac{(1-q)^2 + 2q - q^2}{(1-q)^3},$$

Assim,

$$S_2^{''} = 2\ \frac{1 - 2q + q^2 + 2q - q^2)}{(1-q)^3} = \frac{2}{p^3}.$$

Finalmente,

$$E\left(X_{[2]}\right) = pq\ \ \frac{2}{p^3} = \frac{2q}{p^3}.$$

Vai-se calcular agora o terceiro momento fatorial, assim

$$E\left(X_{[3]}\right) = E\left[X(X-1)(X-2)\right] = \sum_{x=3}^{n} x(x-1)(x-2)\ \ pq^{x-1},$$

mas,

$$\begin{aligned} x(x-1)(x-2)q^{x-1} &= q^2\ x(x-1)(x-2)q^{x-3} \\ &= (q^x)^{(3)}, \end{aligned}$$

a derivada de terceira ordem de q^x.

Assim,

$$E\left(X_{[3]}\right) = \sum_{x=3}^{n} x(x-1)(x-2)\ \ pq^{x-1} = pq^2 \left(\sum_{x=3}^{n}\ q^x \right)^{(3)} = pq^2\ S_3,$$

onde

Seja

$$S_3 = \sum_{x=3}^{\infty} q^x = \frac{q^3}{1-q},$$

e a derivada primeira de S_3 em relação a q fica:

$$S_3' = \frac{(3q^2)(1-q) - q^3(-1))}{(1-q)^2} = \frac{3q^2 - 3q^3 + q^3}{(1-q)^2} = \frac{3q^2 - 2q^3}{(1-q)^2}.$$

A derivada segunda de S_3 em relação a q fica:

$$S_3^{''} = \frac{(6q-6q^2)(1-q)^2 - (3q^2-2q^3)(-2(1-q))}{(1-q)^4} = \frac{6q(1-q)(1-q)^2 + 2(3q^2-2q^3)(1-q)}{(1-q)^4},$$

vamos simplificar um pouco mais

$$S_3^{''} = 2\ \frac{3q(1-q)^3 + (3q^2-2q^3)(1-q)}{(1-q)^4} = 2\ \frac{3q(1-q)^2 + (3q^2-2q^3)}{(1-q)^3},$$

Continuando a simplificação

$$S_3^{''} = 2\ \frac{3q - 6q^2 + 3q^3 + 3q^2 - 2q^3)}{(1-q)^3} = 2\ \ \frac{3q - 3q^2 + q^3}{(1-q)^3}.$$

A derivada terceira de S_3 em relação a q fica:

Assim,

$$S_3^3 = 2\ \frac{3(1-2q+q^2)(1-q)^3 - (3q-3q^2+q^3)(-3(1-q)^2)}{(1-q)^6},$$

Colocando $3(1-q)^2$ em evidência no numerador e e cancelando com o denominador teremos:

$$S_3^3 = 6\ \frac{(1-2q+q^2)(1-q) + (3q-3q^2+q^3)}{(1-q)^4},$$

lembrando que $(1-q)^2 = 1 - 2q + q^2$

$$S_3^3 = 6\ \frac{(1-q)^3 + 3q - 3q^2 + q^3}{(1-q)^4},$$

Como $(1-q)^3 = 1 - 3q + 3q^2$ temos então:

$(1-q)^3 + 3q - 3q^2 = 1,$

finalmente,

$$S_3^3 = 6\ \frac{1}{(1-q)^4} = \frac{6}{p^4}.$$

e por fim,

$$E\left(X_{[3]}\right) = pq^2 \ \frac{6}{p^4} = \frac{6q^2}{p^3}.$$

Vai-se calcular agora o quarto momento fatorial, assim

$$E\left(X_{[4]}\right) = E\left[X(X-1)(X-2)(X-3)\right] = \sum_{x=4}^{\infty} x(x-1)(x-2)(x-3)pq^{x-1} = pq^3 \sum_{x=4}^{\infty} x(x-1)(x-2)(x-3)pq^{x-4}$$

mas,

$$E\left(X_{[4]}\right) = pq^3 \sum_{x=4}^{\infty} (q^x)^{(4)},$$

a derivada de quarta ordem de q^x.

Assim,

$$E\left(X_{[4]}\right) = \sum_{x=4}^{\infty} x(x-1)(x-2)(x-3) \ \ pq^{x-1} = pq^3 \left(\sum_{x=4}^{\infty} \ q^x \right)^{(4)} \ = \ pq^3 \ S_4,$$

onde

Seja

$$S_4 = \sum_{x=4}^{\infty} \ q^x = \frac{q^4}{1-q},$$

e a derivada primeira de S_4 em relação a q fica:

$$S_4' = \frac{(4q^3)(1-q) - q^4(-1))}{(1-q)^2} = \frac{4q^3 - 4q^4 + q^4}{(1-q)^2} = \frac{4q^3 - 3q^4}{(1-q)^2}.$$

A derivada segunda de S_4 em relação a q fica:

$$S_4^{''} = \frac{(12q^2 - 12q^3)(1-q)^2 - (4q^3 - 3q^4)(-2(1-q))}{(1-q)^4} = \frac{12q^2(1-q)(1-q)^2 + 2(4q^3 - 3q^4)(1-q)}{(1-q)^4},$$

vamos simplificar um pouco mais

$$S_4^{''} = 2 \ \frac{6q^2(1-q)^2 + 4q^3 - 3q^4}{(1-q)^3} = 2 \ \frac{6q^2(1-2q+q^2) + 4q^3 - 3q^4}{(1-q)^3},$$

Continuando a simplificação

$$S_4^{''} = 2 \ \frac{6q^2 - 12q^3 + 6q^4 + 4q^3 - 3q^4}{(1-q)^3} = 2 \ \ \frac{6q^2 - 8q^3 + 3q^4}{(1-q)^3}.$$

A derivada terceira de S_4 em relação a q fica:

Assim,

$$S_4^3 = 2\ \frac{12q - 24q^2 + 12q^3)(1-q)^3 - (6q^2 - 8q^3 + 3q^4)(-3(1-q)^2)}{(1-q)^6},$$

Colocando $3(1-q)^2$ em evidência no numerador e e cancelando com o denominador teremos:

$$S_4^3 = 6\ \frac{4(q - 2q^2 + q^3)(1-q)^3 + 3q - 3q^2 + q^3}{(1-q)^4},$$

$$S_4^3 = 6\ \frac{4q - 8q^2 + 4q^3 - 4q^2 + 8q^3 - 4q^4 + 6q^2 - 8q^3 + 3q^4}{(1-q)^3},$$

finalmente,

$$S_4^3 = 6\ \frac{4q - 6q^2 + 4q^3 - q^4}{(1-q)^4}.$$

A derivada quarta de S_4 em relação a q fica:

$$S_4^4 = 6\ \frac{(4 - 12q + 12q^2 - 4q^3)(1-q)^4 - (4q - 6q^2 + 4q^3 - q^4)(-4(1-q)^3)}{(1-q)^8},$$

Colocando $4(1-q)^3$ em evidência no numerador e e cancelando com o denominador teremos:

$$S_4^4 = 24\ \frac{(1 - 3q + 3q^2 - q^3)(1-q) + 4q - 6q^2 + 4q^3 - q^4}{(1-q)^3},$$

simplificando um pouco mais

$$S_4^4 = 24\ \frac{(1 - 3q + 3q^2 - q^3 - q + 3q^2 - 3q^3 + q^4 + 4q - 6q^2 + 4q^3 - q^4}{(1-q)^5} = 24\ \frac{1}{(1-q)^5},$$

$$S_4^4 = 24\ \frac{1}{p^5}.$$

e por fim,

$$E\left(X_{[4]}\right) = pq^3\ \ \frac{24}{p^5} = \frac{24q^3}{p^4}.$$

1.7 Momentos em relação à origem

Fato 6. Se $X \sim Geo(p)$, então os quatro primeiros momentos em relação à origem são dados por

$$E\left(X^r\right) = \begin{cases} \frac{1}{p}, & \text{se } r = 1 \\ \\ \frac{1+q}{p^2}, & \text{se } r = 2 \\ \\ \frac{1+4q+q^2}{p^3}, & \text{se } r = 3 \\ \\ \frac{1+11q+11q^2+q^3}{p^3}, & \text{se } r = 4 \end{cases}$$

<u>*Prova*</u>: O primeiro momento em relação à origem é igual ao primeiro momento fatorial e portanto

$$E(X) = \mu = \frac{1}{p}.$$

O segundo momento em relação à origem, $E(X^2)$, é calculado por:

$$E(X^2) = E(X(X-1)) + E(X) = \frac{2q}{p^2} + \frac{1}{p} = \frac{2q+p}{p^2} = \frac{q+q+p}{p^2} = \frac{1+q}{p^2}.$$

O terceiro momento em relação à origem, $E(X^3)$, é calculado por:

$$E(X^3) = E\left[X(X-1)(X-2)\right] + 3E(X^2) - 2E(X).$$

Logo

$$E(X^3) = \frac{6q^2}{p^3} + 3\,\frac{(1+q)}{p^2} - \frac{2}{p} = \frac{6q^2+3p+3pq-2p^2}{p^3},$$

Mas,

$$num = 6q^2+3p+3pq-2p^2 = 6q^2+3(1-q)+3(1-q)q-2(1-q)^2 = 6q^2+3-3q+3q-3q^2-2+4q-2q^2,$$

assim,

$$num = 1+4q+q^2.$$

e

$$E(X^3) = \frac{1+4q+q^2}{p^3}.$$

O quarto momento em relação à origem, $E(X^4)$, é calculado por:

$$E(X^4) = E\left[X(X-1)(X-2)(X-3)\right] + 6E(X^3) - 11E(X^2) + 6E(X).$$

Logo,

$$E(X^4) = \frac{24q^3}{p^4} + 6\,\frac{1+4q+q^2}{p^3} - 11\,\frac{1+q}{p^2} + 6\,\frac{1}{p}$$

Tirando o mínimo múltiplo comum, temos:

$$E(X^4) = \frac{24q^3 + 6p(1+4q+q^2) - 11p^2(1+q) + 6p^3}{p^4} = \frac{num}{p^4}.$$

Mas

$$p(1+4q+q^2) = (1-q)(1+4q+q^2) = (1+4q+q^2-q-4q^2-q^3) = (1+3q-3q^2-q^3,)$$

$$p^2(1+q) = (1-q)^2(1+q) = (1-2q+q^2)(1+q) = 1-2q+q^2+q-2q^2+q^3 = 1-q-q^2+q^3,$$

$$p^3 = (1-q)^3 = 1-3q+3q^2-q^3.$$

O numerador fica:

$$num = 24q^3 + 6 + 18q - 18q^2 - 6q^3 - 11 + 11q + 11q^2 - 11q^3 + 6 - 18q + 18q^2 - 6q^3,$$

que simplificando fica:

$$num = 1 + 11q + 11q^2 + q^3.$$

$$E(X^4) = \frac{1+11q+11q^2+q^3}{p^4}.$$

1.8 Momentos centrais

Fato 7. Se $X \sim Geo(p)$, então$\mu_2 = E\left[(X-\mu)^2\right] = E(X^2) - E^2(X) = \dfrac{q}{p^2}$.

Prova:

$$var(X) = \frac{1+q}{p^2} - \frac{q^2}{p^2} = \frac{q}{p^2}. \qquad \blacksquare$$

Assim, a variância de $X \sim G(p)$, é dada por

$$Var(X) = \frac{q}{p^2}. \tag{3}$$

O desvio padrão é dado por:

$$\sigma = \frac{\sqrt{q}}{p}.$$

O terceiro momento central de $X \sim Geo(p)$

$$\mu_3 = \frac{q(2-p)}{p^3} \tag{4}$$

Como

$$\begin{aligned}
\mu_3 &= E(X^3) - 3E(X^2) + 2E(X)^3 \\
&= \frac{1+4q+q^2}{p^3} - 3\,\frac{1+q}{p^2} - 2\,\frac{1}{p^3} \\
&= \frac{1+4q+q^2-3-3q+2}{p^3} \\
&= \frac{q+q^2}{p^3} \\
&= \frac{q(1+q)}{p^3} \\
&= \frac{q(2-p)}{p^3}.
\end{aligned}$$

O quarto momento central de $X \sim Geo(p)$

$$\mu_4 = \frac{q(1+7q+q^2}{p^4}. \tag{5}$$

Como

$$\begin{aligned}
\mu_4 &= E(X^4) - \ 4E(X^3)E(X) + 6E(X^2)E(X)^2 - 3E(X)^4 \\
&= \frac{1+11q+11q^2+q^3}{p^4} - 4\frac{1+4q+q^2}{p^3}\frac{1}{p} + 6\,\frac{1+q}{p^2}\frac{1}{p^2} - 3\,\frac{1}{p^4} \\
&= \frac{1+11q+11q^2+q^3-4-16q-4q^2+6+6q-3}{p^4} \\
&= \frac{q+7q^2+q^3}{p^4} \\
&= \frac{q(1+7q+q^2)}{p^4}.
\end{aligned}$$

Vamos calcular $E(X)$ de outra maneira:

$$
\begin{aligned}
E(X) &= \sum_{x=1}^{\infty} x \; p \; q^{x-1} \\
&= \sum_{x=1}^{\infty} (x-1+1) \; p \; q^{x-1} \\
&= \sum_{x=1}^{\infty} (x-1) \; p \; q^{x-1} + \sum_{x=1}^{\infty} \; p \; q^{x-1} \\
&= \sum_{y=0}^{\infty} y \; q^{y} p + \sum_{x=1}^{\infty} \; p \; q^{x-1} \\
&= q \sum_{y=1}^{\infty} y \; q^{y-1} p + 1 \\
E(X) &= qE(X) + 1,
\end{aligned}
$$

Resolvendo equação temos:

$$(1-q)E(X) = 1,$$

O que acarreta $E(X) = \dfrac{1}{p}$.

Para calcular $E(X^2)$ vamos usar um argumento similar:

$$
\begin{aligned}
E(X^2) &= \sum_{x=1}^{\infty} x^2 \; p \; q^{x-1} \\
&= \sum_{x=1}^{\infty} (x-1+1)^2 \; p \; q^{x-1} \\
&= \sum_{x=1}^{\infty} (x-1)^2 \; p \; q^{x-1} + 2\sum_{x=1}^{\infty} (x-1) \; p \; q^{x-1} + \sum_{x=1}^{\infty} \; p \; q^{x-1} \\
&= \sum_{y=0}^{\infty} y^2 \; q^{y} p + 2\sum_{y=1}^{\infty} y \; p \; q^{y-1} + 1 \\
&= q\sum_{y=1}^{\infty} y^2 \; q^{y-1} p + 2\sum_{y=1}^{\infty} y \; q^{y-1} p1 \\
E(X^2) &= qE(X^2) + 2qE(X) + 1,
\end{aligned}
$$

Resolvendo equação temos:

$$(1-q)E(X^2) = 2qE(X) + 1 = \frac{2q}{p} + 1 = \frac{2q+1}{p},$$

O que acarreta $E(X^2) = \dfrac{2q+1}{p^2}$.

De maneira similar pode-se calcular $E(X^3)$ e $E(X^4)$.

Vamos usar um argumento válido para uma variável aleatória inteira não negativa X, isto é, o suporte dela é $A = \{0, 1, 2, \dots, \}$. A esperança de X pode ser calculada através de sua função de sobrevivência.

$$\begin{aligned} E(X) &= \sum_{x=0}^{\infty} x \, P(X = x) \\ &= \sum_{x=1}^{\infty} x \, P(X = x) \end{aligned}$$

Note que $x = \sum_{y=1}^{x} 1$, assim

$$\begin{aligned} E(X) &= \sum_{x=1}^{\infty} \sum_{y=1}^{x} 1 \, P(X = x) \\ &= \sum_{x=1}^{\infty} \sum_{y=1}^{x} \, P(X = x) \\ &= \sum_{y=1}^{\infty} \sum_{x=y}^{\infty} \, P(X = x) \\ &= \sum_{y=1}^{\infty} \, P(X \geq y) \\ &= \sum_{y=0}^{\infty} \, P(X > y) \\ &= \sum_{y=0}^{\infty} \, S(y). \end{aligned}$$

Vamos calcular a esperança de X usando esta nova técnica . Sabemos que o suporte de X é o conjunto $A = \{1, 2, \dots, \}$. Devemos fazer $W = X - 1$ para a variável começar no ponto zero.

$$E(X) = E(W) + 1$$

Para y no suporte

$$S(y) = p(X > y) = \sum_{x=y+1}^{\infty} pq^{x-1} = p \, \frac{q^y}{1-q} = q^y.$$

Assim,

$$EX) = \sum_{y=0}^{\infty} \, S(y) = \sum_{y=0}^{\infty} \, q^y = \frac{1}{p}.$$

1.9 Coeficiente de Assimetria

Fato 8. O coeficiente de assimetria de $X \sim Geo(p)$

$$\alpha_3 = \frac{2-p}{\sqrt{q}}.$$

Prova:

$$\alpha_3 = \frac{E\left(X-\mu\right)^3}{\sigma^3} = \frac{\mu_3}{\mu_2^{3/2}} = \frac{\frac{q(2-p)}{p^3}}{\frac{q\sqrt{q}}{p^3}} = \frac{q(2-p)}{\sqrt{q}}. \qquad \blacksquare$$

Assim pode-se classificar a distribuição geométrica quanto à assimetria como assimétrica positiva.

1.10 Coeficiente de Curtose

Fato 9. O coeficiente de curtose de $X \sim Geo(p)$

$$\alpha_4 = \frac{1+7q+7q^2}{q}.$$

Prova:

$$\alpha_4 = \frac{\mu_4}{\mu_2^2} = \frac{\frac{q(q^2+7q+1)}{p^4}}{\frac{q^2}{p^4}} = \frac{q^2+7q+1}{q} = 7+q+\frac{1}{q} > 3.$$

Assim a distribuição geométrica é sempre leptocúrtica.

1.11 Coeficiente de Variação

Fato 11. O coeficiente de variação de $X \sim Geo(p)$

$$CV = \sqrt{q}.$$

Prova:

$$CV = \frac{\sigma}{\mu} = \frac{\sqrt{q}}{p} = \sqrt{\frac{1}{p}} = \sqrt{q} \qquad \blacksquare$$

1.12 Moda

Fato 11. A moda da distribuição de $X \sim Geo(p)$ é $M_o = 1$.

Como

$$f(x) = pq^{x-1} < p = P(X=1),$$

logo a moda é o ponto $x = 1$. Observe que $f(x)$ é sempre uma função decrescente.

1.13 Função de distribuição

Fato 12. A função de distribuição de $X \sim Geo(p)$

$$F(x) = \sum_{i=0}^{[x]} p\, q^{i-1} I_{[\ 1,\ \infty)}(x),$$

em que $[x]$ é o maior inteiro que não ultrapassa x.

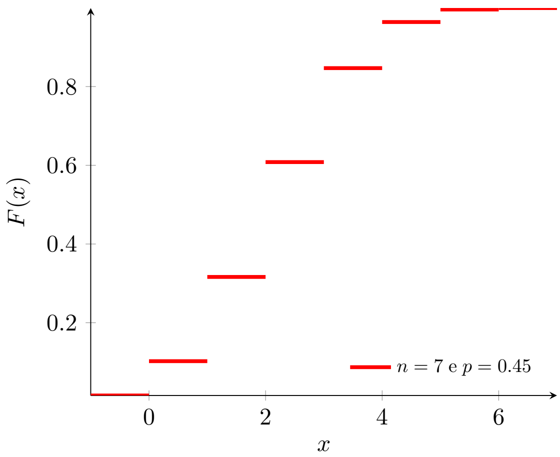

Figura 2: Gráfico da Função de Distribuição Geométrica

A Figura 2 mostra a função de distribuição geométrica com parâmetro $p = 0,5$.

1.14 Função de sobrevivência

Fato 13. A função de distribuição de $X \sim Geo(p)$

$$S(x) = I_{(-\ \infty,\ 1)}(x) + \sum_{i=[x]+1}^{\infty} pi\, q^{i-1} I_{[\ 1,\ \infty)}(x).$$

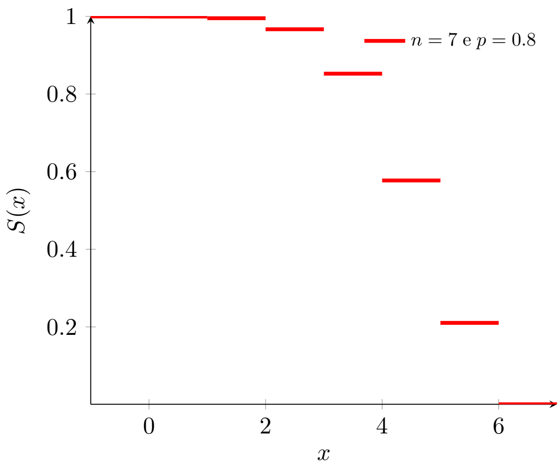

Figura 3: Gráfico da Função de Sobrevivência da Geométrica

A Figura 3 mostra a função de sobrevivência da Geométrica com parâmetro $p = 0,5$.

1.15 Propriedade da Falta de Memória

A propriedade da falta de memória pode ser definida como:

Suponha que $X \sim Geo(p)$. Então, para quaisquer inteiros positivos a e b tem-se:

$$P(X > a + b \mid X > a) = P(X > b).$$

Prova: sabemos que a função de sobrevivência de $X \sim Geo(p)$ é dada por:

$$S(x) = q^x, \quad x = 1, 2, \ldots,$$

1. Logo,

$$P(X > a + b \mid X > a) = \frac{P(X > a + b, X > a)}{P(X > a)} = \frac{P(X > a + b}{P(X > a)} = \frac{S(a+b)}{P(S(a)} = \frac{q^{a+b}}{q^a} = q^b = P(X > b).$$

Vamos comentar esta propriedade :

Comentário 2: A distribuição Geométrica é a única distribuição discreta que satisfaz a esta propriedade. Entre as contínuas a única distribuição a satisfazer a propriedade da falta de memória é a distribuição exponencial.

Comentário 1. A propriedade afirma que a distribuição não tem memória ou não tem desgaste no seguinte sentido: Considere que um componente tenha distribuição Geométrica de parâmetro p como modelo para seu tempo de vida que é medido em unidades inteiras. Sabendo que o componente já sobreviveu por a unidades de tempo de operação, então a probabilidade que ele opere por mais b unidades de tempo, isto é, ainda esteja funcionando após o tempo $(a+b)$ é a mesma que a probabilidade de que um componente novo funcione por mais de b unidades de tempo de operação.

Comentário 2. A recíproca da propriedade da falta de memória também é verdadeira. Assim se houver uma variável aleatória inteira não negativa distribuição para a qual

$$P(X > a+b \mid X > a) = P(X > b),$$

é válida. Assim obrigatoriamente X tem distribuição Geométrica.

Comentário 3.

1.16 Quantis

Fato 10. Vamos obter os quantis da $Geo(p)$.

Seja x_a o quantil de ordem $0 < a < 1$. Vamos supor x_a inteiro. Então

$$S(x_a) = 1 - a.$$

$$q^{x_a} = 1 - a.$$

assim

$$x_a = \frac{ln(1-a)}{q} = \frac{ln(1-a)}{q}.$$

Como a distribuição é discreta este não exatamente o valor do quantil pois nem sempre será um número inteiro.

Vamos usar o R para calcular os quantis:

```
>
> ####Vamos obter usando o R os quantis da Geométrica(p)
>
> y=0:10
>
> p=0.5
> fy=dgeom(y,p)
> Fy=pgeom(y,p)
> tab1=cbind(y,fy,Fy);tab1
y          fy        Fy
```

```
[1,]  0 0.5000000000 0.5000000
[2,]  1 0.2500000000 0.7500000
[3,]  2 0.1250000000 0.8750000
[4,]  3 0.0625000000 0.9375000
[5,]  4 0.0312500000 0.9687500
[6,]  5 0.0156250000 0.9843750
[7,]  6 0.0078125000 0.9921875
[8,]  7 0.0039062500 0.9960938
[9,]  8 0.0019531250 0.9980469
[10,]  9 0.0009765625 0.9990234
[11,] 10 0.0004882812 0.9995117
>
> a=c(0.10,0.25,0.4,0.5,0.75,0.80,0.95);a
[1] 0.10 0.25 0.40 0.50 0.75 0.80 0.95
>
> quantis=qgeom(alfa,p);quantis
[1] 0 0 0 0 1 2 4
>
>
>
> ####Para a Geométrica começando no 1 soma-se um ao valor obtido na Geométrica começando
> #### no 1.
>
```

1.17 Transformações Importantes

Fato K. Se $X \sim Geo(p)$ e $Y \sim Geo(p)$ são variáveis aleatórias independentes, então

$$S = X + Y \sim Pascal(2, p).$$

<u>*Prova:*</u> A função geradora de S é dada por

$$\begin{aligned}
\varphi_S(t) &= \varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t) \\
&= \frac{pt}{1-qt}\ \frac{pt}{1-qt} \\
&= = \left(\frac{pt}{1-qt}\right)^2, \ \ t < \frac{1}{q},
\end{aligned}$$

que é a f.g.p. de uma Pascal de parâmetros 2 e p e cuja f.p. é dada por

$$f(s) = \binom{s-1}{1} p^2 q^{s-2} I_{\{2,3,\ldots,\infty\}}(s) = (s-1); p^2 q^{s-2} I_{\{2,3,\ldots,\infty\}}(s) \qquad \blacksquare.$$

Fato Q. Se $X \sim Geo(p)$ e $Y \sim Geo(p)$,independentes. Então e $X|S = s \sim Unif(A = \{1, 2, s-1\})$.

<u>*Prova:*</u>

$$\begin{aligned}
P(X|S = s) &= \frac{P(X = x,\, X + Y = s)}{P(S = s)} = \frac{P(X = x,\, Y = s - x)}{P(S = s)} \quad \text{(por independência)}\\
&= \frac{P(X = x)P(Y = s - x)}{P(S = s)} = \frac{pq^{x-1}pq^{s-x-1}}{(s-1)p^2q^{s-2}} = \frac{1}{s-1} I_{\{1,\ldots,s-1\}},
\end{aligned}$$

que é a função de probabilidade da Uniforme discreta de parâmetro $N = s - 1$ ■

Fato J. Sejam $X_1, X_2, \ldots, X_r \stackrel{iid}{\sim} Geo(p)$, então $S = \sum_{i=1}^{r} X_i \sim Pascal(r, p)$.

<u>*Prova:*</u> Sabemos que $\varphi_{X_i}(t) = \dfrac{pt}{1-qt}$ $t < \dfrac{1}{q}$ e que $\varphi_S(t) = \prod_{i=1}^{n} \varphi_{X_i}(t) = \left(\dfrac{pt}{1-qt}\right)^r$, $t < \dfrac{1}{q}$. Assim $S \sim Pascal(r, p)$ ■

1.18 Exercícios Resolvidos

1. Seja $X \sim G(0,5)$. Calcule usando o R:

   a $P(X = 2)$ e $P(X > 4)$

   b. Qual a média e a variância de X?

```
> x=0:20
> fx=dbinom(x,20,0.3)
> EX=sum(x*fx);EX
[1] 6
> VX=sum((x-6)^2*fx);VX
[1] 4.2
```

2. Vamos reproduzir o exemplo 8.17 do Meyer páginas 261 e 262. É um exemplo que servirá de base para vários problemas práticos pois envolve o cálculo do lucro esperado.

   Suponha que o custo de realização de realização de um experimento seja 1000 u.m. Se o experimento falhar, ocorrerá um custo adicional de 300 u.m. em virtude de serem necessárias algumas alterações antes que a próxima tentativa seja executada. Se a probabilidade de sucesso em uma tentativa qualquer for 0,2, se as provas forem independentes, e se os experimentos continuarem até que o primeiro resultado frutuoso seja encontrado, qual será o custo esperado do procedimento completo?

Solução Sejam as variáveis aleatórias: $C=$ custo do procedimento completo e $X=$ número de provas necessárias para alcançar o primeiro sucesso.

Sabemos que pagaremos 1000 u.m para fazer cada prova (há no total X) das quais $(X-1)$ são fracassos acrescentando $300(X-1)$ u.m. no custo total.

Logo,

$$C = 1000X + 300(X-1) = 1300X - 300.$$

Mas $X \sim Geo(p = 0.2)$ e cuja $E(X) = \dfrac{1}{0.2} = 5$ tentativas

Dessa maneira

$$E(C) = E(1300X - 300 = 1300E(X) - 300 = 6500 - 300 = 6200 u.m..$$

Vamos resolver agora com a variável aleatória

$Y=$ número de tentativas fracassadas antes do primeiro sucesso

Sabemos que $Y + 1 = X$ e $Y = X - 1$ .. A esperança de Y é dada por $E(Y) = \dfrac{q}{p} = 4$. Cada tentativa malsucedida tem um cisto de 1300 u.m..

O custo da experiência é dado por:

$$C = 1300Y + 1000,$$

adicionando o custo de 1000 u.m. da última tentativa que é o sucesso.

assim.

$$E(C) = E(1300Y + 1000) = 1300E(Y) + 1000 = 1300 \times 4 + 1000 = 5200 + 1000 = 6200 u.m..$$

Vamos fazer o exemplo 8.8 da página 202 do Meyer.

3. Em determinada localidade, a probabilidade da ocorrência de uma tormenta em algum dia durante o verão (nos meses de dezembro e janeiro)é igual a 0,1. Admitindo independência de um dia para outro, qual a probabilidade de ocorrência da primeira tormenta da estação de verão no dia 3 de janeiro?

   Solução: Seja X o número de dias começando no primeiro de dezembro até a primeira tormenta. Assim

   $$X \sim Geo(p = 0,1),$$

   e

$$P(X=x)=pq^{x-1}\,I_A(x)=0,1(0,9)^{x-1};I_A(x),$$

em o suporte é dado por $A=\{1,2,\ldots\}$

Até o dia 3 de janeiro temos (31+3)=34 dias, logo

$$P(X=34)=0,1(0,9)^{33}=0,003.$$

```
>
>
> p=0.1;q=1-p;p;q
[1] 0.1
[1] 0.9
>
> x=34
>
> p34=p*q^(x-1);p34;round(p34,3)
[1] 0.003090315
[1] 0.003
>
> ##Fazendo direto no R
>
> ###Y=número de dias sem tormenta até a primeira tormenta
>
> ###Y +1=X-----------------Y=X-1.
>
> ###P(X=34)=P(Y+1=34)=P(Y=33)
>
> dgeom(33,0.1);round(dgeom(33,0.1),3)
[1] 0.003090315
[1] 0.003
>
```

Vamos fazer agora o exemplo 5.9 da Meyer página 202.

4. Se a probabilidade de que um certo ensaio dê reação positiva for igual a 0,4, qual será a probabilidade de que menos de 5 reações negativas ocorram antes da primeira positiva?

   solução: Seja Y=número de reações negativas ocorram antes da primeira positiva. O suporte de Y é o conjunto:

$$A = \{0, 1, 2, \ldots\}.$$

Assim $Y \sim Geo(p = 0,4)$ começando no zero cuja função de probabilidade é dada por:

$$P(Y = y) = 0,4(0,6)^y \; I_A(y),$$

e sua função de sobrevivência é dada por:

$$S(y) = q^{y+1} = 0,6^{y+1}.$$

A probabilidade pedida é dada por:

$$P(Y < 5) = P(Y \geq 4) = 1 - P(Y > 4) = 1 - S(4) - 1 - (0,6)^5 = 1 - 0,07 = 0,92.$$

```
>
>
> ####Como já é a geométrica começando no zero. A P(Y<=4) é dada por:
>
>
> pgeom(4,0.4);round(pgeom(4,0.4),3)
[1] 0.92224
[1] 0.922
>
```

1.19 Exercícios propostos

1. (Marcos Magalhães & Antonio Carlos Lima-Exercício 1-pg 83)

   Seja $X \sim Geo(0,4)$, calcule:

   a. $P(X = 3)$.
   b. $P(2 \leq X < 4)$.
   c. $P(X > 1|X \geq 2)$.
   d. a curtose de X e classifique a distribuição.
   e. a moda de X.
   f. a mediana de X
   g. Faça tudo no R.

2. (Marcos Magalhães & Antonio Carlos Lima-Exercício 2-pg 83)

   Uma moeda equilibrada é lançada sucessivamente, de modo independente, até que a primeira cara ocorra. Seja X a variável aleatória que conta o número de lançamentos anteriores à ocorrência de cara. Determine:

   a. $P(X \leq 2)$.

   b. $P(X > 1)$.

   c. $P(3 < X \leq 5)$.

   d. Quantas vezes deve, no mínimo, ser lançada a moeda para garantir a ocorrência de cara com pelo menos 0,8 de probabilidade.

   e. a mediana e a moda de X

   g. Faça tudo no R.

3. (Marcos Magalhães & Antonio Carlos Lima-Exercício 24-pg 91)

   Considere a variável aleatória $X \sin GEO(0,8)$. Construa uma nova variável Y tal que $Y = X$ para os valores $0, 1, 2, 3, 4, 5$ e $Y = 6$ para $X \geq 6$. Obtenha a função de probabilidade de Y e calcule:

   a. $P(Y = 2)$.

   b. O valor da função de distribuição acumulada no ponto 2,5.

   c. $P(Y = 3|Y <\leq 5)$.

   d. $P(Y \geq 3 \;\; e \;\; X < 8)$.

   e. Mostre que $Y = min(X, 6)$.

4. (Marcos Magalhães & Antonio Carlos Lima-Exercício 25-pg 91)

   A duração (em centenas de horas) de uma lâmpada especial segue o modelo geométrico com parâmetro $p = 0,7$. Determine a probabilidade da lâmpada:

   a. Durar menos de 500 horas.

   b. Durar mais de 200 menos de 400 horas.

   c. Sabendo-se que vai durar mais de 300 horas, durar mais de 880 horas.

5. (Airton e Teresinha Xavier-pg 183) Suponhamos que um envelope sobre 100, de figurinhas para o Álbum Universal, traz a figura de número 137, que um menino tem necessidade para completar seu álbum. Seja Y o número de envelopes precisaria comprar até encontrar a figura desejada

   a. Que valores pode assumir a variável Y?

   b. Qual a distribuição de Y?

   c. Calcule $E(Y)$ e $Var(Y)$.

   d. Calcule $P(Y > 5)$.

6. João deve a Antônio 1300 reais. Cada viagem de Antônio à casa de João custa 50 reais e a probabilidade de João ser encontrado em casa é $\frac{1}{3}$. Se Antônio encontrar João, conseguirá cobrar a dívida.

   a. Mostre que a probabilidade de Antônio ter de ir mais de 3 vezes à casa de João para conseguir cobrar a dívida é $\frac{8}{27}$
   b. Se na segunda vez em que Antônio foi à casa de João ainda não o encontrou, mostre que a probabilidade de conseguir cobrar a dívida na terceira vez é $\frac{1}{3}$.
   c. O resultado do item b lembra alguma propriedade? Justifique.

7. (Costa Neto& Cymbalista-pg 90) Bolas são retiradas sucessivamente de uma urna que contem milhares de bolas, sendo 30% das bolas vermelhas, 65% pretas e 5% das brancas.

   a. Qual a probabilidade de sair a primeira bola branca na sexta retirada?
   b. Qual o número médio de retiradas até sair a primeira bola vermelha?

8. Identifique a variável aleatória que possui a seguinte função geradora de momentos:

$$M_X(t) = \frac{e^t}{4 - 3e^t}, \quad t < ln(4) - ln(3).$$

9. (George Roussas-Introduction to Probability-pg 117) Um dado imparcial é jogado seguidamente e independentemente até que a face 6 apareça pela primeira vez. Ache a probabilidade que:

   a. isto aconteça na terceira vez.
   b. pelo menos 5 lançamentos sejam necessários.

10. De um baralho comum de 52 cartas retiramos cartas uma a uma , com reposição, até que um ás seja encontrado? Qual a probabilidade que sejam necessárias, no mínimo, 10 retiradas?

11. (Meyer-pg 180) As 5 primeiras repetições de um experimento custam 10 u.m.. Todas as repetições subsequentes custam 5 u.m. cada uma. Suponha que o experimento seja repetido até o primeiro resultado bem sucedido ocorra. Se a probabilidade de um resultado bem sucedido for sempre igual a 0,9, e se as repetições forem independentes , qual será o custo esperado da operação completa?