# 1. TERCEIRA AULA

## Variáveis Aleatórias Discretas

Os fenômenos aleatórios quantitativos admitem ferramentas analíticas que permitem um estudo abrangente destes experimentos. Mesmo os experimentos qualitativos são passíveis de estudos analíticos com os devidos cuidados de interpretação e utilização dos resultados. Imagine que entrevistamos milhares de pessoas em uma pesquisa de opinião em que o entrevistado responde a uma questão tipo **certo x errado**. Ao registrarmos a resposta seria conveniente codificarmos o **certo** pelo número 1 e o **errado** pelo número 0. Assim, se estamos interessados no número de certos somaríamos as respostas. Ao interpretarmos os dados concluiríamos que a proporção do certo é a soma dos 1s sobre o número de entrevistados, mas seria inapropriado dizer que a média da variável em questão é a resposta certo.

A formalização de tais procedimentos analíticos é através das variáveis aleatórias que podem ser discretas, contínuas ou mistas. A variável aleatória é discreta se assume valores em um conjunto enumerável, finito ou infinito, Em nosso contexto, a variável aleatória é contínua quando assume valores em um subconjunto dos números reais, $\Re$.

1.1. **Variáveis Aleatórias Discretas.** Introduzimos este conceito através de um exemplo: Lançamos, independentemente, uma moeda equilibrada três vezes obtendo o espaço amostral

$$\Omega = \{(c,c,c),(c,c,r),(c,r,c),(r,c,c),(c,r,r),(r,c,r),(r,r,c),(r,r,r)\}.$$

Como na Definição 1.3, consideramos $\Im$, o conjunto das partes de $\Omega$, e em $\Im$ definimos a medida de probabilidade que leva a cada elemento amostral a probabilidade $\frac{1}{8}$. Argumentamos que a probabilidade de cara é $\frac{1}{2}$, pois a moeda é equilibrada e os lançamentos são independentes. Assim

$$P(\{(c,c,r)\}) = P(\{c\}\cap\{c\}\cap\{r\}) = P(\{c\}).P(\{c\}).P(\{r\}) = \frac{1}{2}.\frac{1}{2}.\frac{1}{2} = \frac{1}{8}.$$

À terna $(\Omega, \Im, P)$ chamamos espaço de probabilidades. Nosso objetivo é definir uma função $X$ do espaço amostral $\Omega$ no conjunto dos números reais $\Re$. Na realidade construimos um outro espaço de probabilidades $(\Re, \Im_X, P_X)$, onde $\Im_X$ é o conjunto das partes da imagem da aplicação $X, X(\Omega)$. Para prosseguirmos definimos a imagem inversa através de $X$ de um conjunto em $\Im_X$.

**Definição 1.1.** A imagem inversa de um conjunto $B$ de $\Im_X$, através de $X$, é o conjunto de elementos em $w \in \Omega$ tais que $X(w) \in B$, isto é

$$X^{-1}(B) = \{w \in \Omega : X(w) \in B\}.$$

**Lema 1.2.** *Se $A$ e $B$ são dois conjuntos em $\Im_X$, temos $X^{-1}(A \cup B) = X^{-1}(A) \cup X^{-1}(B)$ e $X^{-1}(A \cap B) = X^{-1}(A) \cap X^{-1}(B)$.*

*Prova A prova da primeira afirmação segue das equivalências:*

$$w \in X^{-1}(A \cup B) \leftrightarrow X(w) \in A \cup B \leftrightarrow$$

$$X(w) \in A \quad ou \quad X(w) \in B \leftrightarrow w \in X^{-1}(A) \quad ou \quad w \in X^{-1}(B) \leftrightarrow$$

$$w \in X^{-1}(A) \cup X^{-1}(B).$$

*Semelhantemente*

$$w \in X^{-1}(A \cap B) \leftrightarrow X(w) \in A \cap B \leftrightarrow$$

$$X(w) \in A \quad e \quad X(w) \in B \leftrightarrow w \in X^{-1}(A) \quad e \quad w \in X^{-1}(B) \leftrightarrow$$

$$w \in X^{-1}(A) \cap X^{-1}(B).$$

Tais operações são válidas para uma sequência enumerável de eventos.

Se assumimos que a imagem inversa de qualquer conjunto $A$ em $\Im_X$ pertence a $\Im$, isto é, $X^{-1}(A) \in \Im, \quad \forall A \in \Im_X$, podemos definir

$$P_X(A) = P(X^{-1}(A)), \quad \forall A \in \Im_X.$$

Observe que:

$P_X(\Re) = P(X^{-1}(\Re)) = P(\Omega) = 1$, e que se $(A_n)_{n \geq 1}$ é uma sequência de eventos em $\Im_X$, dois a dois disjuntos

$$P_X(\cup_{n=1}^{\infty} A_n) = P(X^{-1}(\cup_{n=1}^{\infty} A_n)) = P(\cup_{n=1}^{\infty} X^{-1}(A_n)) =$$

$$\sum_{n=1}^{\infty} P(X^{-1}(A_n)) = \sum_{n=1}^{\infty} P_X(A_n).$$

Portanto os axiomas de Kolmogorov estão satisfeitos e o espaço de probabilidades $(\Re, \Im_X, P_X)$ esta bem definido e procedemos na formalização do conceito:

**Definição 1.3.** Seja $(\Omega, \Im, P)$ um espaço de probabilidade e $X$ uma aplicação de $\Omega$ em $\Re$. $X$ é uma variável aleatória se $X^{-1}(A) \in \Im, \ \forall A \in \Im_X$, where $\Im_X$ é o conjunto das partes de $X(\Omega)$.

Denominamos $(\Re, \Im_X, P_X)$ como o espaço de probabilidade induzido por $X$.

Observemos que a configuração de um conjunto $A$ de $\Im_X$ é $A = \{x_{i_1}, ..., x_{i_k}\}$, onde $x_{i_j} = X(w_{i_j})$ para algum $w_{i_j} \in \Omega$. Portanto $A$ pode ser escrito como a união disjunta dos conjuntos unitários $\{x_{i_j}\}$, $A = \cup_{j=1}^{k}\{x_{i_j}\}$ e

$$X^{-1}(A) = X^{-1}(\cup_{j=1}^{k}\{x_{i_j}\}) = \cup_{j=1}^{k} X^{-1}(\{x_{i_j}\}))$$

com

$$P_X(A) = P(X^{-1}(A)) = P(\cup_{j=1}^{k} X^{-1}(\{x_{i_j}\})) =$$
$$\Sigma_{j=1}^{k} P(X^{-1}(\{x_{i_j}\})) = \Sigma_{j=1}^{k} P_X(\{x_{i_j}\}).$$

Consequentemente o espaço de probabilidade induzido por $X$ fica completamente caracterizado pelos valores $x_i$ que $X$ assume e suas probabilidades $P(X = x_i)$ com $0 \leq P(X = x_i) \leq 1$ e $\Sigma P(X = x_i) = 1$.

**Definição 1.4.** A função que associa o valor da variável aleatória à sua probabilidade é denominada função de probabilidade.

| $x$ | $x_1$ | $x_2$ | ... | $x_n$ |
|---|---|---|---|---|
| $P(X = x)$ | $P(X = x_1)$ | $P(X = x_2)$ | ... | $P(X = x_n)$ |

**Exemplo 1.5.** Lançamos, independentemente, uma moeda equilibrada três vezes obtendo o espaço amostral

$$\Omega = \{(c, c, c), (c, c, r), (c, r, c), (r, c, c), (c, r, r), (r, c, r), (r, r, c), (r, r, r)\}.$$

Seja $X$ a variável aleatória definida como o número de caras nos três lançamentos. Portanto $X$ assume os valores $0, 1, 2$ ou $3$ com probabilidades $\frac{1}{8}, \frac{3}{8}, \frac{3}{8}$ e $\frac{1}{8}$, respectivamente, correspondendo às imagens inversas de cada valor da variável.

Podemos ter interesse na probabilidade de obtermos no máximo duas caras e calculamos:

$$P(X \leq 2) = P(X \in \{0, 1, 2\}) = P(X = 0) + P(X = 1) + P(X = 2) =$$
$$\frac{1}{8} + \frac{3}{8} + \frac{3}{8} = \frac{7}{8},$$

que, tambem, pode ser obtida pelo complementar

$$P(X \leq 2) = 1 - P(X > 2) = 1 - P(X = 3) = 1 - \frac{1}{8} = \frac{7}{8}.$$

A probabilidade de obtermos mais do que uma cara conhecendo que o lançamento resultou em no máximo duas caras é

$$P(X > 1 | X \leq 2) = \frac{P(X > 1, X \leq 2)}{P(X \leq 2)} = \frac{P(X = 2)}{P(\leq 2)} = \frac{\frac{3}{8}}{\frac{7}{8}} = \frac{3}{7}.$$

**Exemplo 1.6.** Seja N um número natural positivo e $Y$ uma variável aleatória assumindo os valores $1, 2, ..., N$ com probabilidades $P(Y = n) = c2^n, \quad 1 \leq n \leq N$ onde c é uma constante. Qual o valor de $c$?

O número $c$ deve ser positivo pois as probabilidades são valores positivos. Como a soma das probabilidades é igual a um temos

$$1 = \sum_{n=1}^{N} c.2^n = c.\sum_{n=1}^{N} 2^n = c.\{2+2^2+...+2^N\} = c.2.\frac{1-2^N}{1-2} = c.2.(2^N-1)$$

de maneira que $c = \frac{1}{2^{N+1}-2}$.

Uma variável aleatória pode assumir um número infinito de valores como no exemplo:

**Exemplo 1.7.** Uma váriavel aleatória $Z$, discreta, assumindo valores nos números naturais positivos com probabilidades $P(Z = n) = \frac{1}{2^n}$ é bem definida pois $P(Z = n) \geq 0$ e

$$\sum_{n=1}^{\infty} P(Z = n) = \sum_{n=1}^{\infty} \frac{1}{2^n} = \frac{\frac{1}{2}}{1 - \frac{1}{2}} = 1.$$

Alem da função de probabilidade a função de distribuição da variável aleatória $X$ caracteriza completamente o modelo de probabilidade

**Definição 1.8.** Se $X$ é uma variável aleatória, a função de distribuição de $X$, do conjunto dos números reais, $\Re$ em $[0, 1]$ é definida por

$$F_X : \Re \to [0, 1]$$
$$F_X(x) \mapsto P(X \leq x).$$

**Exemplo 1.9.** Seja $X$ o número de caras em três lançamentos independentes de uma moeda equilibrada, a função de probabilidade de $X$ é

| $x$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $P(X = x)$ | 1/8 | 3/8 | 3/8 | 1/8 |

e a função de distribuição $F_X(x) = F(x)$.

$$F(x) = \begin{cases} 0 & : \quad x < 0 \\ \frac{1}{8} & : \quad 0 \leq x < 1 \\ \frac{4}{8} & : \quad 1 \leq x < 2 \\ \frac{7}{8} & : \quad 2 \leq x < 3 \\ 1 & : \quad x \geq 3 \end{cases} \}.$$

Observe que $P(X = 2) = F(2) - F(2^-) = \frac{7}{8} - \frac{4}{8} = \frac{3}{8}$.

**Teorema 1.10. Propriedades da função de distribuição**

*Seja $X$ uma variável aleatória com função de distribuição $F(x) = P_X((-\infty, x]) = P(X \leq x)$. Então*

*A) A função de distribuição é uma função não decrescente em $\Re$.*

*B) $\lim_{x\to a^+} F(x) = F(a)$ e $\lim_{x\to a^-} F(x) = F(a^-)$,$\forall a \in \Re$, isto é, $F(x)$ é contínua à direita, mas pode não ser contínua à esquerda.*

*C)$\lim_{x\to -\infty} F(x) = 0$ e $\lim_{x\to\infty} F(x) = 1$.*

*D) A função de distribuição determina completamente a função de probabilidade, isto é,*

$$P(X = x) = F(x) - F(x^-), \forall x \in \Re.$$

*Prova*

*A) Se $x$ e $y$ são dois números reais com $x \leq y$, temos $(-\infty, x] \subseteq (-\infty, y]$ e $F(x) = P_X((-\infty, x]) \leq P_X((-\infty, y]) = F(y)$.*

*B) Para todo número real $a$, $((-\infty, a + \frac{1}{n}])_{n\geq 1}$ é uma sequência decrescente s $(-\infty, a] = \cap_{n=1}^{\infty}{}^{(} - \infty, a + \frac{1}{n}]$. Portanto*

$$F(a) = P_X((-\infty, a]) = P_X(\cap_{n=1}^{\infty}{}^{(}-\infty, a+\frac{1}{n}]) = P_X(\lim_{n\to\infty}(-\infty, a+\frac{1}{n}]) =$$

$$\lim_{n\to\infty} P_X((-\infty, a + \frac{1}{n}]) = \lim_{n\to\infty} F(a + \frac{1}{n}].$$

*Equivalentemente, a sequência $((-\infty, a+\frac{1}{n}])_{n\geq 1}$ pode ser substituida por qualquer outra sequência decrescente $((-\infty, a_n])_{n\geq 1}$, com $(a_n)_{n\geq 1}$ decrescente, $\lim_{n\to\infty} a_n = a$ e $\lim_{n\to\infty} F(a_n) = F(a)$, isto é, $F$ é contínua pela direita.*

*Considerando o limite pela esquerda, observamos que a sequência de intervalos $((-\infty, a - \frac{1}{n}])_{n\geq 1}$ cresce para $(-\infty, a)$, isto é $(-\infty, a) = \cup_{n=1}^{\infty}(-\infty, a - \frac{1}{n}]$. Portanto*

$$F(a-) = P_X((-\infty, a)) = P_X(\cup_{n=1}^{\infty}(-\infty, a-\frac{1}{n}]) = P_X(\lim_{n\to\infty}(-\infty, a-\frac{1}{n}]) =$$

$$\lim_{n\to\infty} P_X((-\infty, a - \frac{1}{n}] = \lim_{n\to\infty} F(a - \frac{1}{n}).$$

*Da mesma maneira, a sequência $((-\infty, a - \frac{1}{n}])_{n\geq 1}$ pode ser substituida por qualquer outra sequência crescente $((-\infty, a_n])_{n\geq 1}$, com $(a_n)_{n\geq 1}$, $\lim_{n\to\infty} a_n = a$ e $\lim_{n\to\infty} F(a_n) = F(a-)$, isto é, $F$ pode não ser contínua pela esquerda.*

*C) Observe que*

$$\lim_{n\to\infty} F(-n) = \lim_{n\to\infty} P_X((-\infty,-n]) =$$

$$P_X(\lim_{n\to\infty}(-\infty,-n]) = P_X(\cap_{n=1}^{\infty}(-\infty,-n]) = P_X(\emptyset) = 0$$

*Da mesma forma a sequência* $(-n)_{n\geq 1}$ *pode ser substituida por qualquer outra sequência* $(a_n)_{n\geq 1}$ *que converge para* $-\infty$ *e concluimos que* $\lim_{x\to-\infty} F(x) = 0$.

*De maneira semelhante*

$$\lim_{n\to\infty} F(n) = \lim_{n\to\infty} P_X((-\infty,n]) =$$

$$P_X(\lim_{n\to\infty}(-\infty,n]) = P_X(\cup_{n=1}^{\infty}(-\infty,n]) = P_X(\Re) = 1$$

*e podemos substituir a sequência* $(n)_{n\geq 1}$ *por qualquer outra sequência* $(a_n)_{n\geq 1}$ *que converge para* $\infty$ *e concluimos que* $\lim_{x\to\infty} F(x) = 1$.

*D) Como* $\{x\} = (-\infty,x] - (-\infty,x)$ *e* $(-\infty,x) \subset (-\infty,x]$, *temos*

$$P(X=x) = P_X(\{x\}) = P_X((-\infty,x] - (-\infty,x)) =$$

$$P_X((-\infty,x]) - P_X((-\infty,x))) = F(x) - F(x-).$$

Uma medida centralizadora dos valores da variável aleatória é o seu Valor Esperado ou Média da variável.

**Definição 1.11.** Se $X$ é uma variável aleatória com função de probabilidade

| $x$ | $x_1$ | $x_2$ | ... | $x_n$ |
|---|---|---|---|---|
| $P(X=x)$ | $P(X=x_1)$ | $P(X=x_2)$ | ... | $P(X=x_n)$ |

definimos a média de $X$, que denotamos por $E[X]$ ou $\mu$, à soma

$$\mu = \sum_{i=1}^{n} x_i P(X=x_i).$$

Se $X$ assume valores em um conjunto de cardinalidade infinita, a série $\mu = \sum_{i=1}^{\infty} x_i P(X=x_i)$ deve ser absolutamente convergente, isto é, $\sum_{i=1}^{\infty} |x_i| P(X=x_i) < \infty$, caso contrário dizemos que a média não existe.

**Exemplo 1.12.** Se $X$ é uma variável aleatória com função de probabilidade

| $x$ | $-3$ | $-1$ | $0$ | $1$ | $2$ | $3$ | $5$ | $8$ |
|---|---|---|---|---|---|---|---|---|
| $P(X=x)$ | $0,1$ | $0,2$ | $0,15$ | $0,2$ | $0,1$ | $0,15$ | $0,05$ | 0,05 |

Então

$\mu = -3.0,1 + -1.0,2 + 0.0,15 + 1.0,2 + 2.0,1 + 3.0,15 + 5.0,05 + 8.0,05 = 1.$

**Exemplo 1.13.** Se $X$ é uma variável aleatória com função de probabilidade

| $x$ | 1 | 2 | ... | $n$ | ... |
|---|---|---|---|---|---|
| $P(X=x)$ | $\frac{1}{2}$ | $\frac{1}{2^2}$ | ... | $\frac{1}{2^n}$ | ... |

Então

$$\mu = \Sigma_{n=1}^{\infty} \frac{n}{2^n} = \Sigma_{n=0}^{\infty} \frac{(n+1)}{2^{n+1}} = \Sigma_{n=0}^{\infty} \frac{1}{2}(\frac{n+1}{2^n}) = \frac{1}{2}[\Sigma_{n=0}^{\infty}(\frac{n}{2^n}) + \Sigma_{n=0}^{\infty}(\frac{1}{2^n})].$$

Portanto

$$\Sigma_{n=1}^{\infty} \frac{n}{2^n} - \frac{1}{2}\Sigma_{n=1}^{\infty}(\frac{n}{2^n}) = \frac{1}{2}(\frac{1}{1-\frac{1}{2}})$$

e $\mu = \Sigma_{n=1}^{\infty} \frac{n}{2^n} = 2.$

**Exemplo 1.14.** Se $X$ é uma variável aleatória com função de probabilidade

| $x$ | 2 | $2^2$ | ... | $2^n$ | ... |
|---|---|---|---|---|---|
| $P(X=x)$ | $\frac{1}{2}$ | $\frac{1}{2^2}$ | ... | $\frac{1}{2^n}$ | ... |

Então $\Sigma_{n=1}^{\infty} \frac{2^n}{2^n} = \infty$ e a média não existe.

Por vezes estamos interessados em variáveis aleatórias que resultam de composição de outra variável. Por exemplo, seja $X$ uma variável aleatória e $Y = g(X)$, onde $g$ é uma função real. Pense no esquema

$$(\Omega, \Im, P) \Rightarrow (\Re, \Im_X, P_X) \Rightarrow (\Re, \Im_{g(X)}, P_{g(X)}).$$

$Y$ é completamente caracterizada assumindo valores $y_i = g(x_i)$ com probabilidades $P(Y = y_i) = \sum_{g(x_i)=y_i} P(X = x_i)$. Portanto

$$E[g(X)] = E[Y] = \sum_i y_i P(Y = y_i) =$$

$$\sum_i y_i \sum_{g(x_i)=y_i} P(X = x_i) = \sum_i g(x_i) P(X = x_i).$$

**Exemplo 1.15.** Considere a variável aleatória $X$ no Exemplo 1.12 e a transformação $Y = g(X) = X^2$, A variável aleatória $Y$ tem função de probabilidade

| $y$ | 0 | 1 | 4 | 9 | 25 | 64 |
|---|---|---|---|---|---|---|
| $P(Y=y)$ | $0,15$ | $0,4$ | $0,1$ | $0,25$ | $0,05$ | $0,05$ |

Portanto

$$E[Y] = 0.0,15 + 1.0,4 + 4.0,1 + 9.0,25 + 25.0,05 + 64.0,05 = 7,5.$$

Por outro lado

$$E[X^2] = (-3)^2.0,1 + (-1)^2.0,2 + 0^2.0,15 + 1^2.0,2+$$
$$2^2.0,1 + 3^2.0,15 + 5^2.0,05 + 8^2.0,05 = 7,5.$$

Em particular a esperança da $n$-ésima potência de variável aleatória, se existir, é denominada momento.

**Definição 1.16.** Seja $X$ uma variável aleatória. O $n$-ésimo momento de $X$, denotado por $\mu_n$ é definida como:

$$\mu_n = E[X^n], \quad se \quad existir.$$

Observe que $\mu_1 = \mu = E[X]$. Em adição a esperança de potências da forma $E[(X-\mu)^n]$ são denominadas de momentos centrais.

**Definição 1.17.** Seja $X$ uma variável aleatória com média $\mu$. O $n$-ésimo momento central de $X$, denotado por $\sigma_n$ é definida como:

$$\sigma_n = E[(X-\mu)^n)], \quad se \quad existir.$$

Em particular o segundo momento central em torno da média é denominado variância, interpretado como uma medida de dispersão, em torno da média.

**Definição 1.18.** Se $X$ é uma variável aleatória com média $\mu$ e função de probabilidade

| $x$ | $x_1$ | $x_2$ | ... | $x_n$ |
|---|---|---|---|---|
| $P(X=x)$ | $P(X=x_1)$ | $P(X=x_2)$ | ... | $P(X=x_n)$ |

definimos a variância de $X$, que denotamos por $E[(X-\mu)^2] = \sigma^2$ , à soma

$$\sigma_2 = \sigma^2 = \sum_{i=1}^{n}(x_i-\mu)^2 P(X=x_i).$$

Se $X$ assume valores em um conjunto de cardinalidade infinita, a série $\sigma^2 = \sum_{i=1}^{\infty}(x_i-\mu)^2 P(X=x_i)$ deve ser absolutamente convergente, caso contrário dizemos que a variância não existe.

A definição da variância resulta em unidades de medida ao quadrado que dificulta sua comparação com outras quantidades mensuráveis em unidades. Este vicio é eliminado definindo-se o **Desvio Padrão** como a raiz quadrada da variância, denotado por $\sigma = DP(X) = \sqrt{\sigma^2}$.

**Exemplo 1.19.** Calculemos o desvio padrão da variável aleatória $X$ do exemplo 1.12 em que $\mu = E[X] = 1$.

$$Var(X) = \sigma^2 = \sum_{i=1}^{8}(x_i - 1)^2 P(X = x_i) =$$

$$(-3-1)^2 + (-1-1)^2 + (0-1)^2 + (1-1)^2 +$$

$$(2-1)^2 + (3-1)^2 + (5-1)^2 + (8-1)^2 = 6,5.$$

Portanto $\sigma = 2,55$.

*Observação* 1.20. O cálculo da variância de uma variável aleatória $X$ com média $\mu = E[X]$ pode ser conduzido da seguinte maneira:

$$\sigma^2 = \sum_i (x_i - \mu)^2 . P(X = x_i) = \sum_i (x_i^2 - 2\mu x_i + \mu^2) . P(X = x_i) =$$

$$\sum_i x_i^2 . P(X = x_i) - 2\mu \sum_i x_i P(X = x_i) + \mu^2 \sum_i . P(X = x_i) = E[X^2] - \mu^2.$$

Portanto, utilizando os resultados dos exemplos 1.12 e 1.15, a resposta do exemplo 1.19 é $\sigma^2 = 7,5 - 1 = 6,5$.

*Observação* 1.21. **Desigualdade de Markov**

Se $X$ é uma variável aleatória e $g(.)$ é uma função real positiva, com domínio em $\Re$, então

$$P(g(X) \geq k) \leq \frac{E[g(X)]}{k}.$$

Prova:

$$E[g(X)] = \sum_i g(x_i) P(X = x_i) = \sum_{\{i : g(x_i) < k\}} g(x_i) P(X = x_i) +$$

$\sum_{\{i:g(x_i)\geq k\}} g(x_i) P(X = x_i) \geq \sum_{\{i:g(x_i)\geq k\}} g(x_i) P(X = x_i) \geq \sum_{\{i:g(x_i)\geq k\}} k P(X = x_i) = k P(g(X) \geq k)$.

**Desigualdade de Chebyshev** Se $X$ é uma variável aleatória com média $\mu$ e variância $\sigma^2$, finitas, e se consideramos a função $g(x) = |x - \mu|$ na desigualdade de Markov temos:

$$P(|x - \mu| \geq k\sigma) = P((x - \mu)^2 \geq k^2 \sigma^2) \leq \frac{1}{k^2}, \quad \forall k > 0.$$

Portanto, se $X$ é uma variável aleatória com variância finita

$$P(\mu - k\sigma < X < \mu + k\sigma) = P(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2},$$

isto é, a probabilidade de que $X$ está entre $k\sigma$ unidades de sua média é maior do que $1 - \frac{1}{k^2}$.

A Desigualdade de Chebyshev permite o cálculo de limites inferiores e superiores de eventos que dependem da variável aleatória, de sua média e de sua variância independente de sua função de distribuição.

**Exemplo 1.22.** Existe uma variável aleatória, com média $\mu$ e variância $\sigma^2$ com

$$P(\mu - 2\sigma < X < \mu + 2\sigma) = 0,6?$$

Pela desigualdade de Chebyshev temos

$$P(|X - \mu| < 2\sigma) \geq 1 - \frac{1}{4} = 0,75$$

e não podemos ter tal variável.

### Desigualdade de Jensen

Observemos que uma função real contínua, $g(.)$, é chamada convexa se para todo $x \in \Re$ existe uma reta que passa por $(x, g(x))$ que está sobre e inteiramente abaixo do gráfico de $g(x)$.

Se $X$ é uma variável aleatória com média $E[X]$ e $g(.)$ é uma função convexa, então

$$E[g(X)] \geq g(E[X]).$$

Óbviamente

$$g(E[X]) = a + bE[X] = E[a + bX] \leq E[g(X)].$$

Observemos que em geral $E[g(X)] \neq g(E[X])$. Note que $g(x) = x^2$ é convexa e $E[X^2] \geq (E[X])^2$, e $Var(X) \geq 0$.

Quando existem, os momentos de uma variável aleatória determinam completamente sua distribuição. A demonstração do teorema seguinte pode ser encontrada em literatura mais específica.

**Teorema 1.23.** *Sejam $X$ e $Y$ duas variáveis aleatórias com momentos finitos $\mu_n^X$ e $\mu_n^Y$, respectivamente. Então, se $\mu_n^X = \mu_n^Y, \forall n$, as funções de distribuições $F_X(.)$ e $F_Y(.)$ são iguais.*

Como a função de distribuição caracteriza completamente a medida de probabilidade induzida pela variável aleatória o Teorema acima é de grande importância e, por sua vez, assim o é, a função geradora de momentos definida por

**Definição 1.24.** Se $X$ é uma variável aleatória. A função geradora de momentos de $X$, $M_X(t)$, é o valor esperado $E[\exp[tX]]$, se existir, em um intervalo simétrico, $(-s, s)$, de números reais.

$$M_X(t) = E[\exp[tX]] = \sum_i \exp[tx_i]P(X = x_i).$$

*Observação* 1.25. O desenvolvimento de Taylor, em torno do ponto 0, de uma função infinitamente diferenciavel permite escrever

$$M_X(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} M_X^{(n)}(0)$$

onde $M_X^{(n)}(t)$ é a $n$-ésima derivada no ponto $t$.

A série de Taylor para a função $\exp[tX]$ é $\sum_{n=0}^{\infty} \frac{(tX)^n}{n!}$ e

$$M_X(t) = E[\exp[tX]] = \sum_{n=0}^{\infty} \frac{t^n E[X^n]}{n!}.$$

Considerando o Teorema 1.21 e comparando as duas expressões para $M_X(t)$ temos $M_X^{(n)}(0) = E[X]^n$, para todo $n$ e, por sua vez, a função geradora de momentos, quando existe, caracteriza completamente a distribuição da variável aleatória.

**Exemplo 1.26.** No Exemplo 5.10, a função de probabilidade é

| $x$ | $-3$ | $-1$ | $0$ | $1$ | $2$ | $3$ | $5$ | $8$ |
|---|---|---|---|---|---|---|---|---|
| $P(X = x)$ | $0,1$ | $0,2$ | $0,15$ | $0,2$ | $0,1$ | $0,15$ | $0,05$ | 0,05 |

e a função geradora de momentos é

$$M_X(t) = 0,1\exp[-3t] + 0,2\exp[-t] + 0,15\exp[0t] +$$

$$0,2\exp[t] + 0,1\exp[2t] + 0,15\exp[3t] + 0,05\exp[5t] + 0,05\exp[8t].$$

A primeira derivada é

$$M_X^{(1)}(t) = -0,3\exp[-3t] - 0,2\exp[-t] + 0,2\exp[t] +$$

$$0,2\exp[2t] + 0,45\exp[3t] + 0,25\exp[5t] + 0,4\exp[8t],$$

que no valor zero vale

$$\mu = M_X^{(1)}(0) = -0,3 - 0,2 + 0,2 + 0,2 + 0,45 + 0,25 + 0,4 = 1.$$

A segunda derivada é

$$M_X^{(2)}(t) = 0,9\exp[-3t] + 0,2\exp[-t] + 0,2\exp[t] +$$

$$0,4\exp[2t] + 1,35\exp[3t] + 1,25\exp[5t] + 3,2\exp[8t],$$

que no valor zero vale

$$M_X^{(2)}(0) = 0,9 + 0,2 + 0,2 + 0,4 + 1,35 + 1,25 + 3,2 = 7,5.$$

Tais valores coincidem com os calculados anteriormente e tal argumento pode sempre ser utilizado.

*Observação* 1.27. **Função geradora de probabilidade**

**Definição 1.28.** Seja $X$ uma variável aleatória a valores inteiros não negativos. Definimos a função geradora de probabilidades de $X$ como a função

$$\varphi_X(s) = E[s^X] = \Sigma_{i=0}^{\infty} P(X = i)s^i,$$

quando a esperança existe.

**Exemplo 1.29.** Uma variável aleatória com distribuição binomial de parâmetros $p$, $0 < p < 1$ e $n$ assume valores em $\{0, 1, 2, ..., n\}$ com probabilidades

$$P(X = i) = \binom{n}{i} p^i (1-p)^{n-i}$$

A sua função geradora de probabilidades é

$$\varphi_X(s) = E[s^X] = \Sigma_{i=0}^{n} \binom{n}{i} (sp)^i (1-p)^{n-i} = (sp + (1-p))^n.$$

*Observação* 1.30. A) Observe que se $|s| < 1$, temos

$$|\varphi_X(s)| = |\Sigma_{i=0}^{\infty} P(X = i)s^i| \leq \Sigma_{i=0}^{\infty} P(X = i)|s^i| \leq \Sigma_{i=0}^{\infty} P(X = i) = 1$$

e $\varphi_X(s)$ existe para $|s| < 1$.

B) Se $\varphi_X(s)$ existe e a derivamos continuamente temos

$$\varphi_X(s) = \Sigma_{i=0}^{\infty} P(X = i)s^i \Rightarrow \varphi_X(0) = P(X = 0);$$

$$\varphi_X^{'}(s) = \Sigma_{i=0}^{\infty} iP(X = i)s^{i-1} \Rightarrow \varphi_X^{'}(0) = P(X = 1);$$

$$\varphi_X)^{"}(s) = \Sigma_{i=0}^{\infty} i(i-1)P(X = i)s^{i-2} \Rightarrow \frac{\varphi_X)^{"}(0)}{2!} = P(X = 2);$$

.....

$$\varphi_X^{(r)}(s) = \Sigma_{i=0}^{\infty} i(i-1)...(i-r+1)P(X = i)s^{i-r} \Rightarrow \frac{\varphi_X^{(r)}(0)}{r!} = P(X = r).$$

Concluimos que a função geradora de probabilidades determina completamente a função de probabilidade.

C) A função geradora de probabilidades é uma série de potência particular. Se duas séries de potências são iguais, isto é

$$\Sigma_{i=0}^{\infty} a_i s^i = \Sigma_{i=0}^{\infty} b_i s^i, \quad -1 < s < 1,$$

podemos concluir que $a_i = b_i, \quad i = 0, 1, 2, ....$

Concluimos que se duas variáveis aleatórias inteiras não negativas tem mesma função geradora de probabilidade, então tem a mesma distribuição de probabilidade.

D)

$$\varphi_X^{'}(s) = \Sigma_{i=0}^{\infty} iP(X = i)s^{i-1} \Rightarrow \varphi_X^{'}(1) = \Sigma_{i=0}^{\infty} iP(X = i) = E[X];$$

$$\varphi_X)^{"}(s) = \Sigma_{i=0}^{\infty} i(i-1)P(X = i)s^{i-2} \Rightarrow$$

$$\varphi_X)^{"}(1) = \Sigma_{i=0}^{\infty} i(i-1)P(X = i) = E[X(X-1)];$$

.....

$$\varphi_X^{(r)}(s) = \Sigma_{i=0}^{\infty} i(i-1)...(i-r+1)P(X = i)s^{i-r} \Rightarrow$$

$$\varphi_X^{(r)}(1) = \Sigma_{i=0}^{\infty} i(i-1)...(i-r+1) = E[X(X-1)...((X-r+1)].$$

Concluimos que a função geradora de probabilidades determina completamente a função de probabilidade.

**Exemplo 1.31.** Uma variável aleatória com distribuição de Poisson de parâmetros $\lambda, \quad \lambda > 0$ e $n$ assume valores em $\{0, 1, 2, ...\}$ com probabilidades $P(X = n) = \frac{e^{-\lambda}\lambda^n}{n!}$ tem função geradora de probabilidades $\varphi_X(s) = e^{\lambda(s-1)}$.

Portanto

$$\varphi_X(s) = e^{-\lambda}e^{\lambda s} = e^{-\lambda}\Sigma_{n=0}^{\infty}\frac{(\lambda s)^n}{n!} = \Sigma_{n=0}^{\infty}\frac{e^{-\lambda}\lambda^n}{n!}s^n.$$

**Exemplo 1.32.** Os salários de uma pequena empresa são distribuidos de acordo com a variável aleatória $Z$, em unidades monetárias, que tem função de probabilidade.

| $z$ | 2 | 4 | 6 | 8 |
|---|---|---|---|---|
| $P(Z = z)$ | $0,4$ | $0,3$ | $0,2$ | $0,1$ |

Perguntamos qual a média salarial? Qual o desvio padrão dos salários?

A média dos salários é $\mu = E[Z] = 2.0,4 + 4.0,3 + 6.0,2 + 8.0,1 = 4$.

O segundo momento é $\mu_2 = E[Z^2] = 4.0,4 + 16.0,3 + 36.0,2 + 64.0,1 = 20$, de forma que $\sigma^2 = Var(Z) = E[Z^2] - \mu^2 = 20 - 16 = 4$ e o desvio padrão de $Z$ é $\sigma = 2$.

Se o empresário decide dobrar o salário e conceder um abono de três unidades monetárias, como fica a nova distribuição dos salários, qual a nova média salarial e qual o desvio padrão?

A nova representação salarial é através da variável aleatória $Y = 2.Z + 3$ que tem função de probabilidade

| $z$ | 7 | 11 | 15 | 19 |
|---|---|---|---|---|
| $P(Z = z)$ | $0,4$ | $0,3$ | $0,2$ | $0,1$ |

com média $E[Y] = 2,8 + 3,3 + 3 + 1,9 = 11 = 2.4 + 3 = 2.E[Z] + 3$.

Em adição $E[Y^2] = 19,6 + 36,3 + 45 + 36,1 = 137$ e portanto $Var(Y) = E[Y^2] - E[Y]^2 = 16 = 4.4 = 4.Var(Z)$.

O propósito do exemplo anterior é aplicar o Teorema seguinte que prova a linearidade da esperança de uma variável aleatória.

**Teorema 1.33.** *Seja $X$ uma variável aleatória com $\mu = E[X] < \infty$. Então, se $a$ e $b$ são números reais, E[a.X + b] = a E[X] + b. Em adição, se $E[X^2] < \infty$, temos $Var(a.X + b) = a^2 Var(X)$.*

*Prova*

*Se $\{x_1, x_2, ..., x_n, ...\}$ são os valores que a variável $X$ assume, a variável $Y$ assume os valores $\{y_1, y_2, ..., y_n, ...\}$, onde $y_i = ax_i + b$, com probabilidades $P(Y = y_i) = P(X = x_i)$.*

*Portanto*

$$E[a.X + b] = E[Y] = \sum_i y_i P(Y = y_i) = \sum_i (a.x_i + b)P(X = x_i) =$$

$$a \sum_i x_i P(X = x_i) + b \sum_i P(X = x_i) = aE[X] + b.$$

*E ainda*

$$Var(a.X+b) = Var(Y) = E[(Y-E[Y])^2] = E[(a.X+b-a.E[X]-b)^2] =$$

$$E[(a.X - a.E[X])^2] = a^2 E[(X - E[X])^2] = a^2 Var(X).$$

1.2. **Modelos Probabilísticos Discretos.** *Frequentemente o Estatístico tem um conjunto de dados a ser adaptado a um modelo probabilístico conveniente. No que segue reproduziremos modelos clássicos que aparecem frequentemente em tais situações, como as distribuições, uniforme, Bernoulli, binomial, hipergeométrica, geométrica, binomial negativa e Poisson. Em cada caso procuramos interpretar as situações em que tal modelo é conveniente.*

**Distribuição uniforme.**

*Uma variável aleatória $X$ tem distribuição uniforme em um espaço amostral finito e enumerável, $\Omega = \{x_1, ..., x_n\}$, de cardinalidade $n$, se*

*a cada elemento amostral associamos a probabilidade* $\frac{1}{n}$. *A função de probabilidade é dada por*

| $x$ | $x_1$ | $x_2$ | ... | $x_n$ |
|---|---|---|---|---|
| $P(X=x)$ | $\frac{1}{n}$ | $\frac{1}{n}$ | ... | $\frac{1}{n}$ |

*A função geradora de momentos é definida por*

$$M_X(t) = \sum_{i=1}^{n} \frac{1}{n} \exp[tx_i].$$

*No caso particular em que* $x_i = i$ *temos* $M_X(t) = \frac{1}{n}\sum_{i=1}^{n} \exp[ti]$, *com primeira derivada* $M_X^{(1)}(t) = \frac{1}{n}\sum_{i=1}^{n} i \exp[ti]$ *e segunda derivada* $M_X^{(2)}(t) = \frac{1}{n}\sum_{i=1}^{n} i^2 \exp[ti]$. *Portanto, avaliadas em zero, temos*

$$M_X^{(1)}(0) = \frac{1}{n}\sum_{i=1}^{n} i = \frac{n+1}{2} = E[X] = \mu$$

$$M_X^{(2)}(0) = \frac{1}{n}\sum_{i=1}^{n} i^2 = \frac{(n+1)(2n+1)}{6} = E[X^2]$$

*de maneira que* $\sigma^2 = E[X^2] - \mu^2 = \frac{n^2-1}{12}$.

**Distribuição Hipergeométrica.**

*Uma variável aleatória* $Y$ *tem distribuição hipergeométrica se a sua função de probabilidade é*

$$P(Y=k) = \frac{\binom{K}{k}.\binom{M-K}{n-k}}{\binom{M}{n}}; \quad k = 0, 1, ..., n,$$

*onde* $M$, $K$ *e* $n$ *são inteiros positivoscom* $K \leq M$ *e* $n \leq M$.

*Considerando que*

$$\sum_{i=0}^{m} \binom{a}{i}\binom{b}{n-i} = \binom{a+b}{m}$$

*observamos que*

$$\sum_{k=0}^{n} P(Y=k) = \sum_{k=0}^{n} \frac{\binom{K}{k}.\binom{M-K}{n-k}}{\binom{M}{n}} = \frac{\binom{M}{n}}{\binom{M}{n}} = 1.$$

$$E[Y] = \sum_{k=0}^{n} k \frac{\binom{K}{k}.\binom{M-K}{n-k}}{\binom{M}{n}} =$$

$$n.\frac{K}{M}\sum_{k=1}^{n} \frac{\binom{K-1}{k-1}.\binom{M-K}{n-k}}{\binom{M-1}{n-1}} =$$

$$n.\frac{K}{M}\sum_{k=0}^{n-1}\frac{\binom{K-1}{k}.\binom{M-1-K+1}{n-1-k}}{\binom{M-1}{n-1}} = n.\frac{K}{M}$$

$$E[Y.(Y-1)] = \sum_{k=0}^{n} k.(k-1)\frac{\binom{K}{k}.\binom{M-K}{n-k}}{\binom{M}{n}} =$$

$$n.(n-1)\frac{K.(K-1)}{M.(M-1)}\sum_{k=2}^{n}\frac{\binom{K-2}{k-2}.\binom{M-K}{n-k}}{\binom{M-2}{n-2}} =$$

$$n.(n-1)\frac{K.(K-1)}{M.(M-1)}\sum_{k=0}^{n-2}\frac{\binom{K-2}{k}.\binom{M-2-K+2}{n-2-k}}{\binom{M-2}{n-2}} = n.(n-1)\frac{K.(K-1)}{M.(M-1)}.$$

*Portanto*

$$Var(Y) = E[Y^2] - (E[Y])^2 = EY.(Y-1)] + E[Y] - (E[Y])^2 =$$

$$n.(n-1)\frac{K.(K-1)}{M.(M-1)} + n.\frac{K}{M} - (n.\frac{K}{M})^2 =$$

$$n.\frac{K}{M}[\frac{(M-K)(M-n)}{M.(M-1)}].$$

**Exemplo 1.34.** Dentre os $M$ componentes do depósito de certa fábrica retira-se uma amostra casual de $n$ componentes. Sabe-se que $K, K \leq M$ componentes são defeituosos. Qual a probabilidade de que a amostra contenha no máximo 2 componentes defeituosos?

**Distribuição de Bernoulli.**

*O modelo é adequado quando um fenômeno aleatório apresenta dois possíveis resultados interpretados como sucesso e fracasso.*

**Exemplo 1.35.** As apólices direcionadas a indivíduos com alto potencial de risco devem ser analisadas separadamente. Uma Cia de seguros de automóveis, baseada nos registros do Departamento de Trânsito, avalia que a probabilidade de que um indivíduo escolhido aleatóriamente deste registro tenha alto potencial de risco é $0,001$. Aqui podemos interpretar como sucesso o fato de que o indivíduo tenha alto potencial de risco.

*A distribuição de Bernoulli tem um parâmetro, $p$, $0 < p < 1$, a probabilidade de obtermos sucesso. A variável aleatória assume os valores reais 1 e 0 correspondendo aos resultados sucesso e fracasso, respectivamente. A função de probabilidade é dada por*

| $x$ | 0 | 1 |
|---|---|---|
| $P(X=x)$ | $1-p$ | $p$ |

*A função geradora de momentos de $X$ é*

$$M_X(t) = p\exp[t] + (1-p)$$

*e suas primeira e segunda derivadas são iguais a $M_X^{(1)}(t) = M_X^{(2)}(t) = p\exp[t]$ que, avaliadas em 0 resulta em $M_X^{(0)}(0) = M_X^{(2)}(0) = p$ e consequentemente*

$$\mu = p \quad e \quad \sigma^2 = p.(1-p).$$

*Observamos que definimos uma família de distribuições de Bernoulli, isto é, para cada valor de $p$, $0 < p < 1$ temos uma distribuição.*

**Distribuição Binomial.**

**Exemplo 1.36.** Suponha que no exemplo anterior escolhemos ao acaso 100 indivíduos do registro do DETRAN e que estamos interessados em calcular a probabilidade de que, dentre os escolhidos, no máximo 1 tenha alto potencial de risco, isto é, procuramos o número de sucessos em 100 ensaios de Bernoulli independentes e identicamente distribuidos.

*O exemplo acima nos da idéia das situações modeladas pela distribução binomial. Utilizamos o modelo binomial quando observamos o número de sucessos em $n$ ensaios de Bernolli independentes e identicamente distribuidos com probabilidade de sucesso igual a $p$, $0 < p < 1$. Se $Y$ é a variável aleatória binomial, denotamos, $Y \sim B(n,p)$ onde $n$ e $p$ são os parâmetros da distribuição. Claramente $Y$ assume os valores $0, 1, 2, ..., n$.*

*Analisemos a probabilidade do evento de obtermos $k$ sucessos, isto é, $\{Y = k\}$. Observamos que qualquer realização desse evento tem probabilidade $p^k(1-p)^{(n-k)}$. Em adição o número de realizações que satisfazem esse evento é $\binom{n}{k}$ e concluimos que*

$$P(\{Y = k\}) = \binom{n}{k} p^k (1-p)^{(n-k)}, 0 \leq k \leq n.$$

*Observe que $0 \leq P(\{Y = k\}) \leq 1$ e que, pelo Binômio de Newton temos*

$$\sum_{k=0}^{n} P(\{Y = k\}) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{(n-k)} = (p + 1 - p)^n = 1.$$

*Utilizando o mesmo argumento verificamos que a função geradora de momentos de $Y$ é*

$$M_X(t) = \sum_{k=0}^{n} \exp[kt] \binom{n}{k} p^k (1-p)^{(n-k)} = [p\exp[t] + (1-p)]^n.$$

*Portanto* $M_Y^{(1)}(t) = n[p\exp[t] + (1-p)]^{n-1}p\exp[t]$, *com* $\mu = E[Y] = M_Y^{(1)}(0) = np$ *e*

$$M_Y^{(2)}(t) = n(n-1)[p\exp[t] + (1-p)]^{(n-2)}p^2\exp[2t] +$$

$$n[p\exp[t] + (1-p)]^{(n-1)}p\exp[t].$$

*No valor* 0 *temos*

$$E[Y^2] = M_Y^{(2)}(0) = n.(n-1).p^2 + n.p$$

*e*

$$\sigma^2 = n.(n-1).p^2 + n.p - n^2.p^2 = n.p.(1-p).$$

**Exemplo 1.37.** No exemplo acima a variável aleatória $Y$ é o número de indivíduos que tenham alto potencial de risco. $Y \sim B(100, 0,001)$. Esperamos que entre os 100 escolhidos $\mu = np = 100.0,001 = 0,1$ indivíduos tenham alto potencial de risco. A probabilidade de que no máximo 1 tenha tenha alto potencial de risco é

$$P(Y \leq 1) = \binom{100}{0} 0,999^{100} + \binom{100}{1} 0,001.0,999^{99}.$$

**Distribuição Binomial Negativa.**

*Como a distribuição binomial a distribuição binomial negativa, denotada por* $Z$, *considera ensaios de Bernoulli independentes e identicamente distribuidos com probabilidade de sucesso igual a* $p$, $0 < p < 1$, *mas em um número infinito. A variável aleatória* $Z$ *é definida como o número de fracassos até o* $r$*-ésimo sucesso na série de ensaios. Claramente* $Z$ *assume os valores no conjunto* $\mathbf{N}$, *dos números naturais. O evento* $\{Z = k\}$ *ocorre quando em* $r + k$ *ensaios de Bernoulli, o último ensaio é sucesso e existirem* $k$ *fracassos nos* $r+k-1$ *ensaios anteriores. O número de maneiras de tais realizações é* $\binom{r+k-1}{k}$ *e a probabilidade de cada realização é* $p^r.(1-p)^k$, *de forma que*

$$P(Z = k) = \binom{r+k-1}{k} p^r.(1-p)^k, k \geq 0.$$

*Cumpre observar que o número combinatório* $\binom{x}{k} = \frac{x.(x-1)...(x-k+1)}{k!}$ *esta bem definido quando* $x$ *é qualquer número real desde que* $k$ *seja um inteiro positivo. Para valores inteiros negativos de* $k$ *definimos* $\binom{x}{k} = 0$. *Pode-se provar que*

$$\binom{r+k-1}{k} = (-1)^k \binom{-r}{k}$$

*e que*

$$\sum_{k=0}^{\infty}(-1)^k\binom{-r}{k}(1-p)^k = p^{-r}.$$

*Portanto*

$$\sum_{k=0}^{\infty}P(Z=k) = \sum_{k=0}^{\infty}\binom{r+k-1}{k}p^r.(1-p)^k =$$

$$p^r\sum_{k=0}^{\infty}\binom{-r}{k}(-1)^k(1-p)^k = p^rp^{-r} = 1.$$

*A função geradora de momentos da distribuição binomial negativa é calculada através de*

$$M_Z(t) = E[\exp[tZ]] = \sum_{k=0}^{\infty}\exp[tk].\binom{r+k-1}{k}p^r.(1-p)^k =$$

$$p^r\sum_{k=0}^{\infty}\binom{-r}{k}(-1)^k((1-p)\exp[t])^k =$$

$$p^r(1-(1-p)\exp[t])^{-r} = [\frac{1-((1-p)\exp[t])}{p}]^{-r}.$$

*A primeira derivada de $M_Z(t)$ é*

$$M_Z^{(1)}(t) = r(\frac{p}{1-(1-p)\exp[t]})^{r-1}\frac{p.(1-p)\exp[t]}{(1-(1-p)\exp[t])^2}$$

*e* $\mu = E[Z] = M_Z^{(1)}(0) = r\frac{(1-p)}{p}$.

*O leitor pode verificar que a variância de $Z$ é dada por*

$$\sigma^2 = Var(Z) = M_Z^{(2)}(0) - M_Z^{(1)}(0)^2 = r\frac{(1-p)}{p^2}.$$

**Exemplo 1.38.** O número anual de sinistros contra uma apólice de seguro de automóvel é modelado por uma distribuição binomial negativa com parâmetros $r = 5$ e $p = 0,9$.

O número esperado de sinistros é $\mu = \frac{5.0,1}{0,9} = 0,56$ e a variância do número de sinistros, $\sigma^2 = \frac{5.0,1}{0,81} = 0,62$. Observe que a variância é maior do que o valor esperado, constratando com a variância da distribuição binomial que é menor do que seu valor esperado. Na prática atuarial este fato é relevante para modelar o número de sinistros contra uma apólice, ou um grupo de apólices, por certo período,como uma binomial negativa.

A probabilidade de que nenhum sinistro ocorra no próximo ano é $P(Z=0)=\binom{4}{0}0,9^5=0,59$.

**Exemplo 1.39.** O número anual de perdas contra uma apólice de seguros é modelado por uma distribuição binomial negativa com média 4 e variância 8. Com tal informação os valores dos parâmetros $p$ e $r$ são calculados a partir das equações

$$\frac{r(1-p)}{p}=4; \frac{r(1-p)}{p^2}=8$$

com soluções $p=0,5$ e $r=4$

A probabilidade de que ocorra ao menos 3 perdas é

$$P(Z\geq 3)=1-P(Z<3)=1-[P(Z=0)+P(Z=1)+P(Z=2)]=$$

$$1-[\binom{3}{0}(0,5)^4+\binom{4}{1}(0,5)^5+\binom{5}{2}(0,5)^6=0,34.$$

**Exemplo 1.40.** Ao modelar o número de sinistros , com respeito a uma apólice de seguro de automóvel, durante um período de três anos, o Atuário assumiu simplesmente que, para todo inteiro $n, n\geq 0$, a probabilidade, $p_n$, de que o existam $n$ sinistros satisfaça a equação $p_{n+1}=\frac{1}{5}p_n$.

Observe que $\sum_{n=0}^{\infty}p_n=1$, e as equivalências

$$p_0+\sum_{n=1}^{\infty}p_n=p_0+\sum_{n=0}^{\infty}p_{n+1}=p_0+\sum_{n=0}^{\infty}\frac{1}{5}p_n=1\leftrightarrow p_0+\frac{1}{5}=1\leftrightarrow p_0=\frac{4}{5}$$

sâo verdadeiras e se $X$ é o número de sinistros reclamados

$$P(X=k)=\frac{4}{5}\frac{1}{5}^k$$

um caso particular da distribuição binomial negativa com parâmetros $r=1$ e $p=\frac{4}{5}$.

Esta distribuição tem importância por sí e é denominada distribuição geométrica.

**Distribuição Geométrica**

*A distribuição geométrica tambem é relacionada com uma série infinita de ensaios de Bernoulli, independentes e identicamente distribuidos, com probabilidade de sucesso $p, 0<p<1$.*

*Seja $X$ o número de fracassos até o primeiro sucesso. Obviamente $X$ assume valores em $\mathbf{N}$, o conjunto dos números naturais. A realização do evento $\{X=k\}$ acontece com fracassos nos $k-1$ ensaios iniciais*

*seguidos de um sucesso. Assim, a distribuição de probabilidade de $X$ é dada por*

$$P(X = k) = (1-p)^k p, \quad k = 0, 1, 2, ...$$

*Como caso particular da distribuição binomial negativa temos*

$$M_X(t) = \frac{p}{1 - ((1-p)\exp[t])},$$

$E[X] = \frac{(1-p)}{p}$ *e* $Var(X) = \frac{(1-p)}{p^2}$.

*A distribuição geométrica pode modelar tambem o número de ensaios até o primeiro sucesso, $X^*$, que tem função de distribuição*

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

*Note que $X* = X + 1$ e assim $E[X^*] = E[X] + 1 = \frac{1}{p}$ e $Var(X^*) = Var(X)$.*

**Exemplo 1.41.** Suponha que um engenheiro inspeciona determinada máquina em períodos discretos de tempo e observa se a máquina quebrou durante o período anterior. Independente do período, a probabilidade da máquina quebrar em determinado período é $p, 0 < p < 1$. Qual a probabilidade da máquina não quebrar até o $n$-ésimo período.

Se $X^*$ representa o número de períodos até a primeira falha, a probabilidade da máquina não quebrar até o $n$-ésimo período é

$$P(X^* > n) = \sum_{k=n+1}^{\infty} (1-p)^{k-1}.p = p \sum_{k=n+1}^{\infty} (1-p)^{k-1} = (1-p)^n.$$

Se a máquina funcionou por $n$ períodos consecutivos, a probabilidade de funcionar por $m$ períodos adicionais é

$$P(X^* > m+n | X^* > n) = \frac{P(X^* > m+n, X^* > n)}{P(X^* > n)} =$$

$$\frac{P(X^* > m+n}{P(X^* > n)} = \frac{(1-p)^{m+n}}{(1-p)^n} = (1-p)^m = P(X^* > m).$$

Esta é a propriedade de falta de memória de uma distribuição. Ela caracteriza a distribuição geométrica como a única variável aleatória discreta com tal propriedade. Concluimos que a máquina em uso é equivalente à máquina nova.

### Distribuição de Poisson

*O principal uso da distribuição de Poisson é na modelagem da ocorrência de eventos aleatórios no tempo, $(\Re^+)$, no plano cartesiano, $(\Re^+ x \Re^+)$ou no espaço, $(\Re^+ x \Re^+ x \Re^+)$. Como função aleatória e do tempo é interpretada como um processo estocástico. Uma outra propriedade é que a*

*distribuição de Poisson aproxima assintóticamente a distribuição binomial, $B(n,p)$, quando os parâmetros $n$ é grande e $p$ pequeno.*

*Em termos analíticos, trabalhar com a distribuição de Poisson é extremamente simples: $N$ é uma variável aleatória de Poisson se assume valores no conjunto dos números naturais, $\mathbf{N}$, com probabilidades*

$$P(N=k)=\frac{\exp[-\lambda]\lambda^k}{k!},\quad k=0,1,2,...,$$

*onde $\lambda$, o parâmetro de distribuição é um número real positivo. Na realidade estamos definindo uma família de distribuições pois a cada $\lambda$ temos, em correspondência, uma distribuição.*

*Note que*

$$\Sigma_{k=0}^{\infty}P(N=k)=\Sigma_{k=1}^{\infty}\frac{\exp[-\lambda]\lambda^k}{k!}=\exp[-\lambda].\Sigma_{k=1}^{\infty}\frac{\lambda^k}{k!}=\exp[-\lambda].\exp[\lambda]=1.$$

*A esperança de $N$ é*

$$\mu=E[N]=\Sigma_{k=0}^{\infty}kP(N=k)=\Sigma_{k=1}^{\infty}k\frac{\exp[-\lambda]\lambda^k}{k!}=$$

$$\lambda.\exp[-\lambda].\Sigma_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}=\lambda.$$

*Utilizando do mesmo argumento temos*

$$E[N(N-1)]=\Sigma_{k=0}^{\infty}k(k-1)P(N=k)=\Sigma_{k=2}^{\infty}k.(k-1)\frac{\exp[-\lambda]\lambda^k}{k!}=$$

$$\lambda^2.\Sigma_{k=2}^{\infty}\frac{\exp[-\lambda]\lambda^{k-2}}{(k-2)!}=\lambda^2.$$

*Portanto*

$$\sigma^2=Var(N)=E[N^2]-\mu^2=E[N(N-1)]+E[N]-\lambda^2=\lambda^2+\lambda-\lambda^2=\lambda.$$

*A função geradora de momentos é*

$$M_N(t)=E[\exp[tN]]=\exp[\lambda.(\exp[t]-1)].$$

*Observação* 1.42. A distribuição binomial, $Y\sim B(n,p)$, pode ser aproximada por uma distribuição de Poisson de parâmetro $\lambda=np$ quando $n$ é grande e $p$ pequeno. Observe que

$$P(Y=k)=\binom{n}{k}p^k(1-p)^{n-k}=$$

$$\frac{1}{k!}\frac{n.(n-1).(n-2)....(n-k+1)}{n^k}\lambda^k(1-\frac{\lambda}{n})^n(1-\frac{\lambda}{n})^{-k}\to\frac{\exp[-\lambda]\lambda^k}{k!},$$

quando $n\to\infty$.

**Exemplo 1.43.** Uma Cia de Seguros vende 10.000 apólices residênciais em determinado período. A probabilidade de que uma residência seja destruida por um incêndio é $0,001$ e neste caso, o benefício a ser pago pela perda total é de 200.000 unidades monetárias.

Se $Y$ é o número de segurados que tem perda total no período, $Y$ tem distribuição binomial de parâmetros $n = 10.000$ e $p = 0,001$, o número esperado de segurados que solicitam o benefício no período é $E[Y] = 10.000.0,001 = 10$.

A a probabilidade de que, no período, 15 segurados solicitem o benefício é

$$P(Y = 15) = \binom{10.000}{15} 0,001^{15} 0,999^{9985},$$

uma expressão com cálculos complicados. Como o parâmetro $n$ é grande e $p$, pequeno podemos aproximar $Y$ pela variável aleatória de Poisson $N$, com parâmetro $\lambda = np = 10$.

$$P(Y = 15) \approx P(N = 15) = \frac{\exp[-10]10^{15}}{15!} = 0,0347.$$

Como a cada solicitação a Cia. de Seguros paga 200.000 unidades monetária, o custo esperado para o pagamento é

$$E[200.000Y] = 200.000.E[Y] = 200.000.10 = 2.000.000.$$

quantia que pode ser interpretada como uma reserva mínima para pagamento das solicitações, pois existem outros custos, como os administrativos e de pessoal, que devem ser contabilizados.

**Exemplo 1.44.** A probabilidade de sofrer os efeitos colaterais de determinada droga é de $0,005$. Se 1000 pessoas são inoculadas, qual a probabilidade de que 20 pessoas sofram estes efeitos?

Observe que cada pessoa inoculada pode sofrer, ou não, os efeitos colaterais e portanto é considerada como uma Bernoulli. O número de pessoas que sofrem dos efeitos é uma variável binomial de parâmetros $n = 1000$ e $p = 0,005$. Portanto

$$P(Y = 20) = \binom{1000}{20} (0,005)^{20} (0,995)^{980},$$

que pode ser aproximado por

$$P(N = 20) = \frac{\exp[-5]5^{20}}{20!}$$

onde $N$ tem distribuição com média $np = 5$.

**Exemplo 1.45.** Em uma Central Telefônica, o fluxo de chamadas em certos períodos é intenso e o número de tais chamadas pode ser modelado com uma distribuição de Poisson. Suponha que o parâmetro $\lambda$ seja 8 chamadas por segundo. A probabilidade de que em um segundo tenhamos ao menos 5 chamadas é

$$P(N \geq 5) = 1 - P(N < 5) = 1 - \{P(N = 0) + P(N = 1) +$$

$$P(N = 2) + P(N = 3) + P(N = 4)\} =$$

$$1 - \{0,0003 + 0,0027 + 0,0107 + 0,0286 + 0,0573\} = 0,9004.$$

Se a média da distribuição é 8 chamadas por segundo, em 3 segundos teremos uma média de $\lambda\hat{*} = 3.8 = 24$. Portanto, a probabilidade de que tenhamos 2 chamadas em 3 segundos é

$$P(N^* = 2) = \frac{\exp[-24]24^2}{2!} = 1,0872 \exp[-8].$$

*Observação* 1.46. No exemplo anterior poderíamos perguntar qual a probabilidade de que ter $k$ chamadas em $t$ segundos. De maneira lógica escreveríamos

$$P(N(t) = N = N^* = k) = \frac{\exp[-8.t]8t^k}{k!}, \quad k = 0, 1, 2, ....$$

Na expressão em que adotamos $N(t) = N^* = N$, descrevemos uma distribuição (processo) de Poisson com média $8t$ e de parâmetro $\lambda = 8$ que denominamos intensidade do processo.

O processo de Poisson $(N(t))_{t\geq 0}$ com intensidade $\lambda$, tem distribuição de Poisson

$$P(N(t) = k) = \frac{\exp[-\lambda.t](\lambda.t)^k}{k!}, \quad k = 0, 1, 2, ....$$

que é uma função do tempo $t \in \Re^+$ e aleatória $w \in \Omega$, isto é, para cada $t > 0$, $N(t)$ é uma variável aleatória. Portanto $(N(t))_{t\geq 0}$ é uma familia de variáveis aleatórias denominada de processo estocástico (homogêneo)de Poisson. O processo de Poisson caracteriza a ocorrência de eventos aleatórios no tempo.

**Exemplo 1.47.** O Sinistros ocorrem contra uma apólice de acordo com um processo de Poisson com uma intensidade de $0,3$ a cada ano. Qual a probabilidade de que, nos próximos 5 anos ocorram 3 sinistros?

Claramente,

$$P(N(5) = 3) = \frac{\exp[-0,3.5](0,3.5)^3}{3!} = 0,08.$$

**Exemplo 1.48.** Devido aos fortes nevoeiros no inverno o aeroporto de Congonhas, S.P., fecha para pousos e decolagens de acordo com uma distribuição de Poisson com média de 2 vezes ao ano. Determine a probabilidade de que o aeroporto paralize suas atividades pela segunda vez, conhecendo que ja paralizou uma vez no corrente ano.

Seja $N$ o número de paralizações anuais. Queremos calcular

$$P(N=2|N\geq 1)=\frac{P(N=2,N\geq 1)}{P(N\geq 1)}=\frac{P(N=2)}{P(N\geq 1)}=\frac{\frac{\exp[-2]2^2}{2!}}{1-\exp[-2]}.$$

**Ocorrência de eventos aleatórios no tempo**

*A escolha aleatória, ou casual, de um elemento da população finita* $\{x_1,x_2,...,x_n\}$*, é através da distribuição uniforme que associa a cada elemento a probabilidade* $\frac{1}{n}$*. Como veremos no próximo capítulo a ocorrência aleatória de eventos em* $(0,t]$ *se dá de acordo com a distribuição uniforme em* $(0,t]$*, através da função densidade de probabilidade* $f(s)=\frac{1}{t}$ *se* $s\in(0,t]$ *e* $0$ *caso contrário, que atribui probabilidades iguais a cada intervalo de igual comprimento. O que podemos dizer quando* $t\to\infty$*, isto é, da ocorrência de eventos em* $\Re^+$ *?*

*Seja* $(N(t))_{t\geq 0}$ *o processo estocástico que, para cada* $t$*, conta o número de eventos que ocorrem em* $(0,t]$*. O número de eventos no intervalo* $(s,t]$ *é denotado por* $N(t)-N(s)$*. A ocorrência aleatória de eventos no tempo se caracteriza pelas seguintes suposições:*

*I -* $(N(t))_{t\geq 0}$ *tem incrementos independentes, isto é, a distribuição das ocorrências em intervalos de tempos disjuntos são independentes: se* $(s_1,t_1]\cap(s_2,t_2]=\emptyset$*, então*

$$P(N(t_1)-N(s_1)=i,N(t_2)-N(s_2)=j)=$$

$$P(N(t_1)-N(s_1)=i).P(N(t_2)-N(s_2)=j).$$

*II -* $(N(t))_{t\geq 0}$ *tem incrementos estacionários, isto é, as ocorrências em intervalos de mesmo comprimento tem mesma distribuição:*

$$P(N(t)-N(s)=k)=P(N(t-s)=k).$$

*III - A probabilidade de uma ocorrência em um intervalo infinitesimal de tempo,* $(t,t+\triangle t]$ *é*

$$P(N(\triangle t)=1)=\lambda\triangle t+o(\triangle t),$$

*onde* $o(\triangle t)$ *é tal que* $\lim_{\triangle t\downarrow 0}\frac{o(\triangle t)}{t}=0$.

*IV - A probabilidade de mais de uma ocorrência em um intervalo infinitesimal de tempo,* $(t,t+\triangle t]$ *é*

$$P(N(\triangle t)>1)=o(\triangle t).$$

*Observação* 1.49. $o(t)$ representa uma função real que, quando $t \to 0$, $o(t) \to 0$ mais rapidamente, isto é $\frac{o(t)}{t} \to 0$. Por exemplo $f(t) = t^2$ é $o(t)$ pois $\lim_{t\downarrow 0} \frac{t^2}{t} = 0$, contudo, $f(t) = \sqrt{t}$ não é $o(t)$ pois $\lim_{t\downarrow 0} \frac{\sqrt{t}}{t} = \infty$.

*Se o processo $(N(t))_{t\geq 0}$ satisfaz as suposições acima, então é um processo de Poisson. Delinearemos uma prova heurística:*

*Dividimos o intervalo $(0, t]$ em intervalos infinitesimais da forma $(\frac{kt}{n}, \frac{(k+1)t}{n}], 0 \leq k \leq n-1$, de comprimento $\frac{t}{n}$.*

*Observe que, das suposições III e IV, o número de ocorrências em cada intervalo $(\frac{kt}{n}, \frac{(k+1)t}{n}]$ é uma variável aleatória de Bernoulli com probabilidade de sucesso $\frac{\lambda.t}{n} + o(\frac{t}{n})$ . Das suposições I e II podemos afirmar que tais Bernoullis são independentes e identicamente distribuidas. Portanto, $N(t)$, o número de ocorrências em $(0, t]$ tem distribuição binomial de parâmetros $n$ e $p = \frac{\lambda.t}{n} + o(\frac{t}{n})$.*

*Como $n$ é grande, podemos aproximar a distribuição binomial pela distribuição de Poisson com parâmetro $\lambda.t$ pois*

$$np = n(\frac{\lambda.t}{n} + o(\frac{t}{n})) = \lambda.t + t.\frac{o(\frac{t}{n})}{\frac{t}{n}} \to \lambda.t$$

*quando $n \to \infty$.*

**Exemplo 1.50.** Em uma indústria, para verificar se a ocorrência de acidentes com operários ocorriam ao acaso, observou-se o número de acidentes por hora durante um certo número de dias (24 horas por dia)obtendo os dados:

| No. de acidentes por hora | No. de horas |
|---|---|
| 0 | 200 |
| 1 | 152 |
| 2 | 60 |
| 3 | 30 |
| 4 | 13 |
| 5 | 9 |
| 6 | 7 |
| 7 | 5 |
| 8 | 4 |

Sobre as 480 horas observadas, o número médio de acidentes por hora é

$$\frac{1}{480}\{200.0 + 152.1 + 60.2 + 30.3 + 13.4 + 9.5 + 7.6 + 5.7 + 4.8\} = 1,2.$$

Se os acidentes ocorrecem aleatóriamente, $N$, o número de acidentes por hora teria, aproximadamente, uma distribuição de Poisson com

média $1,2$. Aceitando tal distribuição como verdadeira, o número esperado de horas com k acidentes é $480.P(N=k) = 480.\frac{\exp[-1,2](1,2)^k}{k!}$ e o o número esperado de horas com $0, 1, 2, ...$ acidentes são

| No. de acidentes por hora | No. esperado de horas |
|---|---|
| 0 | 144,6 |
| 1 | 173,5 |
| 2 | 104,1 |
| 3 | 41,6 |
| 4 | 12,9 |
| 5 | 3 |
| 6 | 0,6 |
| 7 | 0,1 |
| 8 | 0 |

que podem ser considerados diferentes dos valores observados e concluimos que os acidentes não estão ocorrendo aleatóriamente.

*E-mail address*: `bueno@ime.usp.br`

Departamento de Estatística, Instituto de Matemática e Estatística, Universidade de São Paulo, Caixa Postal 66281, CEP 05311-970, São Paulo, Brazil