Tópico 2: Variável Aleatória

Ben Dêivide

6 de outubro de 2021

Descrição sobre o ponto - Função de distribuição acumulada; Principais distribuições de probabilidade: propriedades e exemplos de aplicações; Esperança; Variância; Função característica.

1 Revisão

Devemos relembrar alguns conceitos na Teoria de probabilidade.

Definição 1 (Espaço de probabilidade)**.** *Seja Ω o espaço amostral e $\mathcal{F}$ uma coleção de subconjuntos de Ω que pode ser atribuído probabilidade. Se P é uma função que mede a probabilidade dos eventos em $\mathcal{F}$, então a tripla $(\Omega, \mathcal{F}, P)$ é chamada de espaço de probabilidade.* $\square$

Com o Espaço de probabilidade definido poderemos calcular a probabilidade de um evento $A \in \mathcal{F}$, do qual $\mathcal{F}$ é uma coleção de subconjuntos de Ω. Entretanto, quando $\Omega = \mathbb{R}$, poderemos ter problemas quanto a medida de probabilidade quando o conjunto for não-contável, isto é, dado que A é um evento não-contável tal que $A \in \mathbb{R}$, como medir a probabilidade A, se o espaço de probabilidade não pode ser facilmente obtido? Para solucionar esse problema, vamos definir a σ-álgebra de Borel.

Definição 2 (σ-álgebra de Borel)**.** *A σ-álgebra de Borel, denotada por $\mathcal{B}(\mathbb{R})$ ou $\mathcal{B}$, é a menor σ-álgebra que contém todos os intervalos abertos dos reais (a, b), isto é, $(-\infty \leq a < b \leq \infty)$. Os conjuntos $B \subseteq \mathbb{R}$ tais que $B \in \mathcal{B}$ são chamados conjuntos de Borel.* $\square$

Agora, como os elementos da σ-álgebra de Borel inclui os conjuntos abertos, conjuntos fechados (os complementares dos conjuntos abertos), interseções enumeráveis de conjuntos abertos (lembrando que uniões enumeráveis e conjuntos abertos já são abertos), uniões enumeráveis de conjuntos fechados (lembrando que interseções enumeráveis de conjuntos fechados já são fechados), é possível obter uma medida de probabilidade. Isso nos leva ao espaço de probabilidade $(\mathbb{R}, \mathcal{B}, P)$.

2 Variável aleatória

Em estatística, avaliamos um experimento não pelos eventos em si, mas por uma função definida no espaço amostral, que associa o evento a um número real. Essa função chamamos de variável aleatória, denotada por uma letra maiúscula, X ou $X(.)$. Alguns autores criticam o termo "variável aleatória", já que a mesma é uma função. Como essa definição ficou conhecida com esse nome, seria um equívoco tentar renomeá-la.

Definição 3 (Variável Aleatória)**.** *Uma variável aleatória X é uma função mensurável, $X : \Omega \to \mathbb{R}$, que mapeia $(\Omega, \mathcal{F})$ em $(\mathbb{R}, \mathcal{B})$, tal que*

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

sendo B o conjunto de Borel e $\mathcal{B}$ é a σ-álgebra dos conjuntos de Borel, X^{-1} é a imagem inversa de conjuntos $B \in \mathcal{B}$ que pertencem a σ-álgebra $\mathcal{F}$. $\square$

Em outras palavras, se considerarmos $B = (-\infty, x]$

$$X^{-1}((-\infty, x)) = \{w \in \Omega : X(w) \leq x\} \in \mathcal{F} \ \forall x \in \mathbb{R},$$

isto quer dizer, que X é uma variável aleatória (v.a.) de Ω em $\mathbb{R}$, se e somente se, para todo x que a função assumir, a imagem inversa do conjunto dos X menores ou iguais a x, pertencer a σ-álgebra de eventos $\mathcal{F}$, para todo x pertencente aos reais. Veja a ilustração de uma variável aleatória na Figura 1.

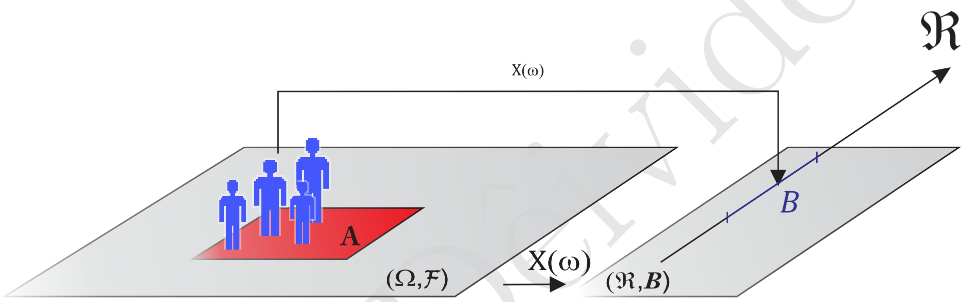


Figura 1: Variável aleatória X definida de Ω em $\mathbb{R}$.

Ao definir uma variável aleatória X, saímos de um espaço mensurável $(\Omega, \mathcal{F})$ com uma medida de probabilidade P, para um novo espaço mensurável $(\mathbb{R}, \mathcal{B})$ nos reais com medida de probabilidade P_X induzida por X. Para que este espaço induzido por X também seja mensurável, definimos os eventos numa σ-álgebra de Borel $\mathcal{B}$, de tal forma que a variável aleatória será mensurável se $X^{-1}(B) \in \mathcal{F}$, sendo B um conjunto tal que $B \subset \mathcal{B}$.

O material do Ross (2007, p.13-15) elucida bem o assunto de espaço de probabilidade e σ-álgebra de Borel

Definição 4 (Espaço de probabilidade induzido por X)**.** *Seja X uma variável aleatória em $(\Omega, \mathcal{F}, P)$, tal que $X : \Omega \to \mathbb{R}$, então definimos o espaço de probabilidade induzido por X como $(\mathbb{R}, \mathcal{B}, P_X)$, em que*

$$P_X(X(\omega) \in B) = P(\{w \in \Omega : X(\omega) \in B\}), \ \forall B \in \mathcal{B}, \tag{1}$$

em que $\mathcal{B}$ é a σ-álgebra de Borel e P_X é a medida de probabilidade induzida por X. $\square$

A partir de agora, entendamos que o cálculo das probabilidades estará relacionado aos eventos Borelianos na σ-álgebra de Borel $\mathcal{B}$.

Sabemos que uma variável aleatória X, é função dos possíveis resultados $\omega \in \Omega$, e que estes possíveis resultados assumem diferentes valores com diferentes probabilidades. Assim, w é aleatório, e por consequência X também. Por isso, o termo variável aleatória, em que a probabilidade de seu resultado assumir um determinado valor x, passa a ser importante. Assim, conceituamos essa relação de distribuição de X.

Definição 5 (Distribuição de X)**.** *O conjunto de probabilidades*

$$P_X(X(\omega) \in B) = P(\{w \in \Omega : X(\omega) \in B\}),$$

para todos os subconjuntos de $B \in \mathcal{B}$, em que B é um subconjunto da σ-álgebra de borel $\mathcal{B}$, é a distribuição de uma variável aleatória X. □

Para entendermos o comportamento estatístico de uma variável aleatória, temos que definir sua função de distribuição, em que são caracterizados por eventos da forma $B = (-\infty, x]$.

Definição 6 (Função de distribuição de X)**.** *A função de distribuição ou função de distribuição acumulada de uma variável aleatória X, no espaço de probabilidade $(\Omega, \mathcal{F}, P)$, é definida por*

$$F_X(x) = P(X \in (-\infty, x]) = P(X \leq x), \ x \in \mathbb{R}.$$

□

Para que não haja confusão sobre função de distribuição a qual está relacionada com a variável aleatória X, usamos o subscrito F_X.

Para entendermos melhor sobre a função de distribuição, mostraremos suas propriedades.

Teorema 1 (Propriedades da função de distribuição)**.** *Uma função de distribuição de uma variável aleatória X em $(\Omega, \mathcal{F}, P)$ obedece as seguintes propriedades:*

i) $\lim_{x\to\infty} F_X(x) = 1$ *e* $\lim_{x\to-\infty} F_X(x) = 0$*;*

ii) $F_X(x)$ *é uma função não decrescente, isto é,* $F_X(x) \leq F_X(y)$ *sempre que* $x \leq y$*,* $\forall x, y \in \mathbb{R}$*;*

iii) $F_X(x)$ *é contínua à direita, ou seja, para um número* x*,* $\lim_{x_n \downarrow x} F_X(x_n) = F_X(x)$*.* □

Demonstração. i) Aplicando a continuidade da probabilidade. Observe que para $x_n \downarrow -\infty$, os eventos $[X \leq x_n] = \{w \in \Omega : X(w) \leq x_n\}$ têm como o limite o conjunto vazio. Logo $F_X(x_n) = P(X \leq x_n) \downarrow 0$. Tomando $x_n \uparrow \infty$, os eventos $[X \leq x_n] \uparrow \Omega$ e, portanto $F_X(x_n) = P(X \leq x_n) \uparrow 1$;

ii) Note que $[X \leq x] \subset [X \leq y]$ sempre que $x \leq y$. Logo as probabilidades satisfazem à desigualdade:

$$F_X(x) = P(X \leq x) \leq P(X \leq y) = F_X(y).$$

Como x e y são arbitrários, concluímos que F é não decrescente.

iii) Seja $x \in \mathbb{R}$ e considere uma sequência $\{x_n, \ n \geq 1, \ n \in \mathbb{N}\}$ tal que $x_n \downarrow x$, ou seja, os x_n's se aproximam de x pela direita ou por valores superiores a x. Então, $[X \leq x_n] \downarrow [X \leq x]$, e assim, $F_X(x_n) \downarrow F_X(x)$. Como o resultado vale para qualquer x, a propriedade está verificada.

□

Se X é uma variável aleatória no espaço de probabilidade $(\Omega, \mathcal{F}, P)$, estamos interessados em calcular probabilidade de eventos que envolvem X, isto é, $P(w \in \Omega : X(w) \in B)$ com $B \in \mathcal{B}$, $\mathcal{B}$ uma σ-álgebra de Borel. A forma que essas probabilidades são calculadas depende da natureza particular de X.

Definição 7 (Variável aleatória discreta). *Uma variável aleatória discreta X em $(\Omega, \mathcal{F}, P)$, é uma função mensurável $X : \Omega \to \mathbb{R}$ tal que a imagem $Im(X(w))$ é um subconjunto contável finito ou infinito dos reais e $X^{-1}(x) = \{\omega \in \Omega : X(\omega) = x\} \in \mathcal{F}$ para todo $x \in \mathbb{R}$.* □

Na realidade, dizer que os valores que a variável aleatória discreta assume em $\mathbb{R}$ é dizer que $Im(X(w)) \in B$ tal que $B \in \mathcal{B}$. A probabilidade dos valores de X que pertencem a B é dada por

$$\begin{aligned} P(X \in B) &= P(X = x_1 \text{ ou } X = x_2 \text{ ou } \ldots) \\ &= P(X = x_1) + P(X = x_2) + \ldots = \sum_{x \in B} P_X(x), \end{aligned}$$

em que $P_X(x)$, $x \in \mathbb{R}$, é a função de probabilidade de X, definida por $P_X(x) = P(X = x)$. Assim, a probabilidade de um evento envolvendo X é encontrado pela sumarização das funções de probabilidades dos conjuntos de pontos favoráveis ao evento. Em particular, a função de probabilidade determina as probabilidades de todos os eventos envolvendo X, para o caso discreto.

Dessa forma, podemos definir a função de probabilidade da variável aleatória discreta X.

Definição 8 (Função de Probabilidade). *Seja X uma variável aleatória discreta, então sua função de probabilidade, $P_X : \mathbb{R} \to [0, 1]$, é definida por:*

$$P_X(x) = P_X(X = x) = P(\{w \in \Omega : X(w) = x\}),$$

sendo $\sum_x P_X(x) = 1$. □

Essa definição nos permite observar que para valores distintos de X, se $\mathcal{X} = \{x_1, x_2, \ldots\} \in B$, com $B \in \mathcal{B}$, então $\Omega = \bigcup_n \{w : X(w) = x_n\} = \bigcup_n \{X = x_n\}$ e $\{X = x_i\} \cap \{X = x_j\} = \emptyset$ para $i \neq j$. Logo, $1 = P(\Omega) = \sum_n P(X = x_n)$.

Definição 9 (Função de distribuição de uma v.a. discreta). *A função de distribuição de uma variável aleatória discreta X é a função $F_X : \mathbb{R} \to [0, 1]$, definida por*

$$F_X(x) = P(X \leq x) = \sum_{\{i : x_i \leq x\}} P_X(X = x_i),$$

para todo $x \in \mathbb{R}$. □

Essa função de distribuição tem a forma de escada sendo descontínua nos valores assumidos pela variável aleatória X.

Uma característica interessante é que da distribuição de probabilidade obtemos a função de distribuição e vice-versa, do qual apresentamos a seguir.

Teorema 2 (Relação entre F_X e P_X). *Seja X uma variável aleatória discreta em $(\Omega, \mathcal{F}, P)$. Então, $F_X(.)$ pode ser obtida de $P_X(.)$, e vice-versa.* □

Demonstração. Seja os valores assumidos pela variável aleatória X, $x_1, x_2, \ldots$,

i) suponha que $P_X(.)$ seja conhecido. Então $F_X(x) = P_X(X \leq x) = \sum_{x_i \leq x} P_X(X = x_i)$;

ii) suponha que $F_X(.)$ seja conhecido. Então $P_X(X = x_j) = F_X(x_j) - \lim_{x \to x_j^-} F_X(x)$; portanto $P_X(X = x_j)$ pode ser encontrado para cada massa de ponto x_j; entretanto, $P_X(X = x_j) = 0$ para $x \neq x_j$, $j = 1, 2, \ldots$, então $P_X(X = x_j)$ é determinado por todos os números reais.

$\square$

Uma outra classe de variáveis aleatórias é a v.a. contínua. Inicialmente, podemos dizer que uma variável aleatória X contínua é uma função que assume em uma sequência não contável de números reais distintos, pertencentes a $B \in \mathcal{B}$, sendo $\mathcal{B}$ a σ-álgebra de Borel.

Definição 10 (Variável Aleatória Contínua). *Uma variável aleatória X é contínua se $P_X(X = x) = 0$.*

Isto é, como $\{X = x\} \subset \{x - \varepsilon \leq X \leq x\}$, temos

$$\begin{aligned} P_X(X = x) &\leq P_X(x - \varepsilon \leq X \leq x) \\ &\leq F_X(x) - F_X(x - \varepsilon), \end{aligned}$$

para qualquer $\varepsilon > 0$. Assim,

$$0 \leq P_X(X = x) \leq \lim_{\varepsilon \downarrow 0} [F_X(x) - F_X(x - \varepsilon)] = 0,$$

pela continuidade de F_X.

Assim, como poderíamos atribuir probabilidade às variáveis aleatórias? Com o intuito de resolver esse problema, definimos a função densidade de probabilidade, que servirá como uma "probabilidade pontual" da variável aleatória contínua. Assim, fazemos a seguinte definição,

Definição 11 (Função densidade de probabilidade). *Seja $f : \mathbb{R} \to \mathbb{R}$ uma função. Então f é uma função densidade de X se $f_X(x) \geq 0$ para todo $x \in \mathbb{R}$, e $\int_{-\infty}^{\infty} f_X(x)dx = 1$.*

Observe que a função densidade de probabilidade foi definida sem fazer referência a uma variável aleatória contínua. Apenas que satisfaça as duas condições. Com essa definição, podemos redefinir formalmente uma variável aleatória contínua.

Definição 12 (Variável aleatória absolutamente contínua). *Uma variável aleatória X é absolutamente contínua se existe uma função densidade $f_X(x)$, tal que*

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt, \tag{2}$$

para todo $x \in \mathbb{R}$.

Nessas duas definições, percebemos que $f_X(x)$ não é unicamente definida. O que é requerido é que a integral de $f_X(x)$ exista para todo x. Observe como exemplo, supondo que $F_X(x) = xI_{[0,1)}(u) + I_{[1,\infty)}(u)$, $f_X(u) = I_{(0,1)}(u)$ satisfaz $F_X(x) = \int_{-\infty}^{x} f_X(u)du$. Agora, se considerarmos $f_X(u) = I_{(0,1/2)}(u) + 35I_{[1/2]}(u) + I_{(1/2,1)}(u)$, também satistaz $F_X(x) = \int_{-\infty}^{x} f_X(u)du$. Ou seja, a função densidade é mudada em alguns pontos, e a função distribuição não se altera. Portanto, o correto ao invés de dizermos "a" função densidade de probabilidade, seria "uma" função densidade de probabilidade.

Há casos patológicos, em que mesmo tendo uma variável aleatória contínua, $F_X(x)$ não é diferenciável devido a $f_X(x)$. Nesses casos, a variável aleatória não é absolutamente contínua. Por isso, uma variável é absolutamente contínua se (2) se mantiver.

Assim, iremos omitir a palavra "absolutamente" ao mencionarmos uma variável aleatória absolutamente contínua, uma vez que os casos em que isso não ocorre são considerados patológicos e não merecedores de atenção, exceto em casos muito particulares e avançados.

A partir da Definição 6, poderemos definir a função de distribuição de uma variável aleatória contínua a seguir.

Definição 13 (Função de distribuição de uma variável aleatória contínua)**.** *Se X é uma variável aleatória contínua em $(\Omega, \mathcal{F}, P)$, a função de distribuição F_X, se existir uma função densidade f_X, é definida por*

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt, \; x \in \mathbb{R}.$$

□

Assim, como no caso discreto, para a variável aleatória contínua, podemos a partir de uma função distribuição podemos obter uma função densidade, como também o contrário é válido. Portanto, fazemos a seguinte definição,

Teorema 3. *Relação entre F_X e f_X Seja X uma variável aleatória contínua. Então, a partir de uma função densidade $f_X(x)$ pode ser obtido a função de distribuição $F_X(x)$, e vice-versa.*

Demonstração. Se X é uma v. a. contínua e $f_X(x)$ é dado, então $F_X(x)$ é obtida pela integração $\int_{-\infty}^{x} f_X(u)du$. Se $F_X(x)$ é conhecida, então a função densidade pode ser obtida por $f_X(x) = dF_X(x)/dx$, nos pontos em que x seja diferenciável. □

Definição 14 (Função quantil)**.** *A função quantil, considerando uma variável aleatória com função de distribuição $F_X(x)$ é a função $F^{-1} : [0, 1] \to \mathbb{R}$ definida por*

$$F^{-1}(p) = x,$$

para todo $p \in [0, 1]$, em que x é o menor valor real tal que $F_X(x) \geq p$. Assim, $F_X^{-1}(p)$ é o $p-$ésimo quantil de X ou $100p\%$ percentil de X. □

3 Esperança e Momentos

Um conceito extremamente importante envolvendo as variáveis aleatórias é a esperança.

Definição 15 (Esperança matemática)**.** *Seja X uma variável aleatória em $(\Omega, \mathcal{F}, P)$. A esperança matemática (ou média) de X, denotada por μ_X ou $E[X]$, é definida:*

i) se X for discreta,

$$E[X] = \sum_i x_i P_X(x_i), \quad (3)$$

para um conjunto de pontos x_1, x_2, ..., x_i, ...;

ii) se X for contínua,

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx, \tag{4}$$

sendo $f_X(x)$ uma função densidade de probabilidade.

A esperança $E[X]$ é conhecida como uma medida de locação. Uma outra medida que envolve a esperança é a variância, que mede a dispersão da função densidade ou função de probabilidade de X, por isso uma medida de dispersão ou de forma.

Definição 16 (Variância)**.** *Seja X uma variável aleatória em $(\Omega, \mathcal{F}, P)$ e sua esperança matemática dada por $E[X] = \mu$. A variância de X, denotada por σ_X^2 ou $Var[X]$, é definida:*

i) se X for discreta, por:

$$Var[X] = \sum_i (x_i - \mu)^2 P_X(x_i), \tag{5}$$

para um conjunto de pontos x_1, x_2, ..., x_i, ...;

ii) se X for contínua, por:

$$Var[X] = \int_{-\infty}^{\infty} (x_i - \mu)^2 f_X(x) dx, \tag{6}$$

sendo $f_X(x)$ uma função densidade de probabilidade.

Definição 17 (Desvio padrão)**.** *Seja X uma variável aleatória em $(\Omega, \mathcal{F}, P)$. O desvio padrão de X, denotado por σ_X, é definido por $\sigma_X = +\sqrt{Var[X]}$.* □

Definição 18 (Esperança de $g(X)$)**.** *Seja X uma variável aleatória em $(\Omega, \mathcal{F}, P)$ e $g(.)$ uma função com domínio e contradomínio nos reais. A esperança da função $g(.)$ de X que também é uma variável aleatória no mesmo espaço de probabilidade, denotada por $E[g(X)]$, é definida por:*

i) se X for discreta,

$$E[g(X)] = \sum_i g(x_i) P(x_i), \tag{7}$$

para um conjunto de pontos x_1, x_2, ..., x_i, ...;

ii) se X for contínua,

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx, \tag{8}$$

sendo $f_X(x)$ uma função densidade de probabilidade.

Existe uma prova em Magalhães p. 219 para $Y = g(X)$ e em Daniel Estatística Matemática (versão 20.01.2016, p.244); Mood p.153.

Teorema 4. *Se $g(X) = X$, então $E[g(X)] = E[X] = \mu_X$ é a média. Se $g(X) = (X - \mu_X)^2$, então $E[g(X)] = E[(X - \mu_X)^2] = Var[X]$.* □

Demonstração. Se $g(x) = x$ segue a Definição 18. Se $g(x) = (x - \mu_x)^2$ segue a Definição 16. □

Teorema 5. *Se X um variável aleatória em $(\Omega, \mathcal{F}, P)$, então*

$$Var[X] = E[X^2] - (E[X])^2, \tag{9}$$

dado que $E[X^2]$ existe. □

Demonstração. Observe inicialmente que se $E[X^2]$ existe, então $E[X]$ também existe. Logo,

$$\begin{aligned} Var[X] &= E[(X - E[X])^2] \\ &= E[X^2 - 2XE[X] + (E[X])^2] \\ &= E[X^2] - 2(E[X])^2 + (E[X])^2 \\ &= E[X^2] - (E[X])^2. \end{aligned}$$

□

Uma forma alternativa para a expressão (9) pode ser dado da seguinte forma:

$$\begin{aligned} Var[X] &= E[X^2] + (E[X] - E[X]) - (E[X])^2 \\ &= E[X(X-1)] + E[X] - (E[X])^2. \end{aligned} \tag{10}$$

Esse resultado é importante quando formos demonstrar a variância de algumas distribuições, como por exemplo a distribuição geométrica.

Teorema 6 (Propriedades da Esperança)**.** *Seja X um variável aleatória em $(\Omega, \mathcal{F}, P)$. Então,*

i) Unilateralidade: $E[c] = c$ para uma constante c;

ii) Homogeneidade: $E[cX] = cE[X]$ para uma constante c;

iii) Linearidade: $E[c_1X + c_2] = c_1E[X] + c_2$;

iv) Monotonicidade: $E[X_1] \leq E[X_2]$ se $x_1 \leq x_2$ $\forall x$.

□

Demonstração. Assuma X uma variável aleatória contínua. Para provar (i), temos

$$E[c] = \int_{-\infty}^{\infty} cf_X(x)dx = c\int_{-\infty}^{\infty} f_X(x)dx = c.$$

Para (ii), temos

$$E[cX] = \int_{-\infty}^{\infty} cXf_X(x)dx = c\int_{-\infty}^{\infty} Xf_X(x)dx = cE[X].$$

Para (iii), temos

$$\begin{aligned} E[c_1X + c_2] &= \int_{-\infty}^{\infty} [c_1X + c_2]f_X(x)dx \\ &= \int_{-\infty}^{\infty} c_1Xf_X(x)dx + \int_{-\infty}^{\infty} c_2f_X(x)dx \\ &= c_1\int_{-\infty}^{\infty} Xf_X(x)dx + c_2\int_{-\infty}^{\infty} f_X(x)dx \\ &= c_1E[X] + c_2. \end{aligned}$$

Finalmente,

$$0 \leq E[X_2 - X_1] = E[X_2] - E[X_1],$$

que implica em (iv). □

Outras propriedades de esperança podem ser encontradas na apostila de Rolla p. 77-78/84; Casella port. p. 52.

Teorema 7 (Propriedade da variância)**.** *Seja X um variável aleatória em $(\Omega, \mathcal{F}, P)$, então para quaisquer constantes a e b, as propriedades da variância são:*

i) $Var[b] = 0$*;*

ii) $Var[aX] = a^2 Var[X]$*;*

iii) $Var[aX + b] = a^2 Var[X]$

Demonstração. Para (i), pela Definição 16 temos

$$\begin{aligned} Var[b] &= E[(b - E[b])^2] \\ &= E[(b - b)^2] \text{ (Teorema 6 - Unilarelidade)} \\ &= 0. \end{aligned}$$

Para (ii), temos

$$\begin{aligned} Var[aX] &= E[(aX - E[aX])^2] \\ &= E[(aX - aE[X])^2] \text{ (Teorema 6 - Homogeneidade)} \\ &= a^2 E[(X - E[X])^2] \text{ (Teorema 6 - Homogeneidade)} \\ &= a^2 Var[X] \text{ (Definição 16)}. \end{aligned}$$

Para (iii), usando o Teorema 5, temos

$$Var[aX + b] = E[(aX + b)^2] - (E[aX + b])^2,$$

e sabendo da linearidade da esperança, Teorema 6, temos que $E[aX + b] = aE[X] + b = a\mu_X + b$. Logo

$$\begin{aligned} Var[aX + b] &= E[(aX + b)^2] - (a\mu_X + b)^2 \\ &= E[(aX + b - a\mu_X - b)^2] \\ &= E[(aX - a\mu_X)^2] \\ &= E[a^2(X - \mu_X)^2] \\ &= a^2 E[(X - \mu_X)^2] \\ &= a^2 Var[X]. \end{aligned}$$

□

Outras propriedades de variância e desvio padrão podem ser encontradas na apostila de Rolla p. 79-80.

3.1 Algumas desigualdades interessantes

Para não precisarmos está repetindo sempre as notações quando a variável aleatória X é contínua ou discreta, denotemos $f_X(x)$ uma função densidade de probabilidade (X contínua) ou função de probabilidade (X discreta).

Teorema 8 (Desigualdade de Chebychev)**.** *Seja X uma variável aleatória definida no espaço de probabilidade $(\Omega, \mathcal{F}, P)$, com $f_X(x;.)$ sendo uma função de probabilidade ou uma função densidade de probabilidade. Considere ainda $g(x)$ uma função não negativa de X. Então para qualquer $k > 0$,*

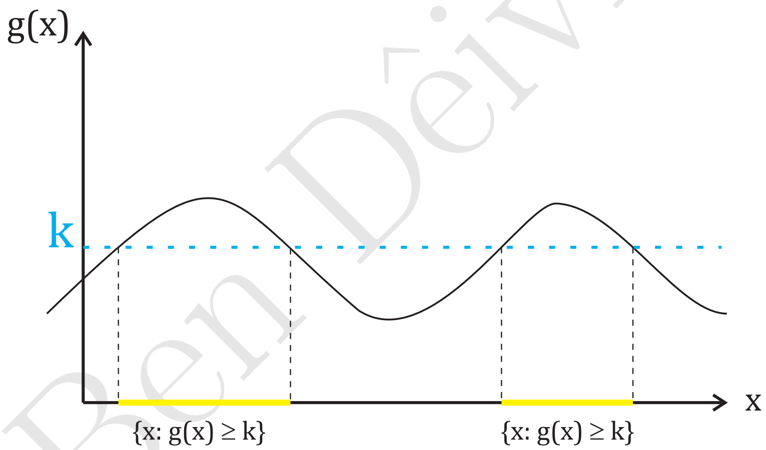
$$P(g(X) \geq k) \leq \frac{E[g(X)]}{k}. \tag{11}$$

□

Demonstração. A prova será realizada considerando X uma variável aleatória absolutamente contínua. Então poderemos expressar esperança de $g(x)$ da seguinte forma:

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x;.) dx. \tag{12}$$

Observe pelo gráfico:



que a $E[g(X)]$ pode ser particionada em

$$\begin{aligned} E[g(X)] &= \int_{\{x:g(X) \leq k\}} g(x) f_X(x;.) dx + \int_{\{x:g(X) \geq k\}} g(x) f_X(x;.) dx \\ &\geq \int_{\{x:g(X) \geq k\}} g(x) f_X(x;.) dx \\ &\geq \int_{\{x:g(X) \geq k\}} k f_X(x;.) dx = k \int_{\{x:g(X) \geq k\}} f_X(x;.) dx = kP(g(x) \geq k). \end{aligned} \tag{13}$$

Logo,

$$P(g(X) \geq k) \leq \frac{E[g(X)]}{k}, \tag{14}$$

como queríamos provar. □

Se a distribuição de X for discreta, basta substituir as integrais por somatórios. Esse Teorema tem uma grande utilidade nas provas dos Teoremas limites, tais como convergência em probabilidade e Lei Fraca dos Grandes Números.

Teorema 9 (Desigualdade de Markov)**.** *Seja X uma variável aleatória definida no espaço de probabilidade $(\Omega, \mathcal{F}, P)$, com $f_X(x;.)$ sendo uma função de probabilidade ou uma função densidade de probabilidade. Considere ainda $g(x)$ uma função não negativa de X. Então para quaisquer $t > 0$ e $k > 0$,*

$$P(|g(X)| \geq k) \leq \frac{E[|g(X)|^t]}{k^t}. \tag{15}$$

□

Demonstração. Temos

$$P(|g(X)| \geq k) = P(|g(X)|^t \geq k^t) \leq \frac{E[|g(X)|^t]}{k^t},$$

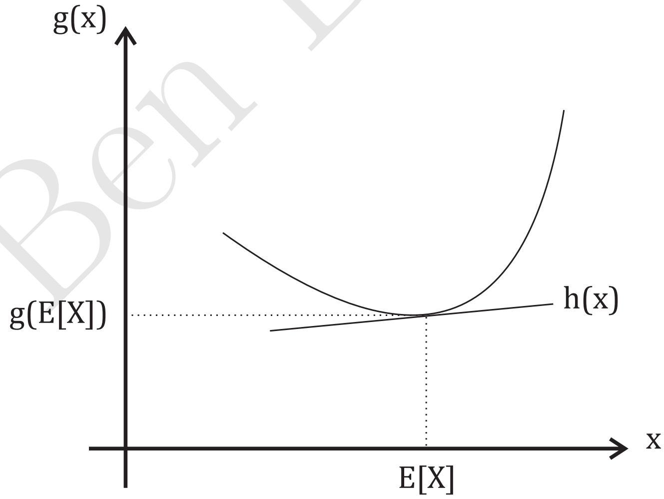
pela Desigualdade de Chebychev. □

Teorema 10 (Desigualdade de Jensen)**.** *Seja X uma variável aleatória definida no espaço de probabilidade $(\Omega, \mathcal{F}, P)$ com $E[X] < \infty$ e $g(.)$ uma função convexa, então*

$$E[g(X)] \geq g(E[X]). \tag{16}$$

□

Demonstração. Sendo $g(.)$ uma função convexa, existe uma reta $h(.)$ que passa no ponto $(x_0, g(x_0))$ que fica abaixo do gráfico da função $g(.)$, ver Figura .



Considerando o ponto $(x_0, g(x_0)) = (E[X], g(E[X]))$ e $h(x) = ax + b$, temos que $h = g$ no ponto $(E[X], g(E[X]))$, e nos demais casos $h(x) \leq g(x)$. Logo,

$$g(E[X]) = h(E[X]) = aE[X] + b = E[aX + b] = E[h(x)] \leq E[g(X)],$$

em que a suposição da esperança finita, permitiu a aplicação das funções g e h no valor esperado de X. □

Para detalhes sobre uma função convexa aplicado na Desigualdade de Jensen, ver Casella, port. p. 171.

4 Momentos

Definição 19 (Momentos amostrais)**.** *Seja uma amostra aleatória* $X_1, X_2, \ldots, X_n$ *com fdp ou fp* $f_X(x;\theta)$*, com* $\theta = (\theta_1, \theta_2, \ldots, \theta_k) \in \Theta$ *em que* Θ *é o espaço paramétrico. O k-ésimo momento amostral, denotado por* M_k*, é definido por*

$$M_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k, \tag{17}$$

e o k-ésimo momento amostral em torno da média amostral $\bar{X}$*, denotado por* M'_k*, é definido por*

$$M'_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k. \tag{18}$$

□

Definição 20 (Momentos populacionais)**.** *Seja uma amostra aleatória* $X_1, X_2, \ldots, X_n$ *com fdp ou fp* $f_X(x;\theta)$*, com* $\theta = (\theta_1, \theta_2, \ldots, \theta_k) \in \Theta$ *em que* Θ *é o espaço paramétrico. O k-ésimo momento populacional, denotado por* μ_k*, é definido por*

$$\mu_k = E[X^k], \tag{19}$$

e o k-ésimo momento populacional em torno da média populacional $\mu = E[X]$*, denotado por* μ'_k*, é definido por*

$$\mu'_k = E[(X - \mu)^k]. \tag{20}$$

□

Geralmente μ_k ou μ'_k é uma função conhecida e função dos k parâmetros $\theta_1, \theta_2, \ldots, \theta_k$. Portanto, poderíamos reescrever $\mu_k = \mu_k(\theta_1, \theta_2, \ldots, \theta_k)$ e $\mu'_k = \mu_k(\theta_1, \theta_2, \ldots, \theta_k)$.

Sabemos que para $k = 1$ em (19) $\mu_1 = E[X^1] = E[X] = \mu$ é a esperança de X. E para $k = 2$ em (20) $\mu'_2 = E[(X_i - \mu)^2] = Var[X]$ é a variância de X.

Entretanto, poderemos gerar momentos populacionais usando a função geradora de momentos (fgm).

Definição 21 (Função Geradora de Momentos (fgm))**.** *Seja* X *uma variável aleatória definida no espaço de probabilidade* $(\Omega, \mathcal{F}, P)$*, com função de distribuição* $F_X(x)$*. A função geradora de momentos (fgm), denotada por* $m_X(t)$ *ou* $m(t)$*, é definida por*

$$m_X(t) = E[e^{tX}], \tag{21}$$

se o valor esperado de e^{tX} *existe para todo* t *em* $-h < t < h$*;* $h > 0$*.* □

Contudo, seu principal uso não é gerar momentos, mas ajudar a caracterizar uma distribuição, pois se a fgm existe ela é única.

Para obter os momentos pela fgm, basta diferenciar (21) em relação a t, isto é,

$$\frac{d^k}{dt^k} m_X(t) = \int x^k e^{xt} f_X(x) dx.$$

Tomando $t \to 0$, obtemos

$$\frac{d^k}{dt^k} m_X(t) = E[X^k] = \mu_k.$$

Em (Casella, port. p. 57 Teorema 2.3.7) e (Magalhães, p. 268) podem mostrar como obter os momentos populacionais.

Entretanto, nem sempre é garantido que uma fmg exista, pois para o caso discreto

$$m_X(t) = E[e^{tX}] = \sum_x e^{tx} P(X = x), \tag{22}$$

e para o caso contínuo

$$m_X(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f_X(x) dx, \tag{23}$$

nem sempre a soma em (22) e a integral em (23) são finitas numa vizinhança de t em torno do zero, e portanto, a fgm não existirá.

Entretanto, uma outra função que gera momentos e que sempre existe desde que os momentos existam, é a função característica.

Definição 22 (Função Característica)**.** *Seja X uma variável aleatória definida no espaço de probabilidade $(\Omega, \mathcal{F}, P)$. Então a função característica, denotada por φ_X, é definida por*

$$\varphi_X(t) = E[e^{itX}] = E[\cos(tX)] + iE[\sin(tX)], \tag{24}$$

em que $i = \sqrt{-1}$. □

A grande vantagem da função característica é que além de sempre existir, pois $\varphi_X(0) = 1$ e $|\varphi_X| \leq 1$, ela pode gerar os momentos de uma variável aleatória X, como também determina a função de distribuição da variável aleatória X. Por exemplo, se X tem esperança $E[X]$, os momentos de X podem ser expressos por $\varphi^d(0) = i^d E[X^d]$. □

5 Distribuições especiais de Probabilidade

5.1 Distribuições especiais para variáveis aleatórias discretas

Definição 23 (Distribuição Bernoulli)**.** *Uma variável aleatória discreta X tem distribuição Bernoulli de parâmetro p, se a distribuição de probabilidade de X, é definida por*

$$P(X = x) = \begin{cases} p^x (1-p)^{1-x}, & \text{para } x = 0 \text{ ou } 1, \\ 0, & \text{caso contrário} \end{cases} \tag{25}$$

em que p satisfaz $0 < p < 1$.

Em outras palavras, define-se uma v.a. X que assume dois possíveis resultados: o valor 1, se ocorrer o evento desejado (sucesso); e o valor 0, se ocorrer o valor não desejado (fracasso). Definindo-se ainda p como sendo a probabilidade de ocorrer o sucesso, e $(1 - p)$, a probabilidade de ocorrer o fracasso, então X tem distribuição de probabilidade Bernoulli.

Para validade se a distribuição de Bernoulli é uma função de probabilidade, temos que demonstrar:

i) Mostrar que $P(X = x) \geq 0$:
   Prova: Como os valores de $x = 0$ ou 1 e $0 < p < 1$, então $p^x(1-p)^{1-x} \geq 0$.

ii) Mostrar que $\sum_{x=0}^{1} P(X = x) = 1$:
   Prova: $\sum_{x=0}^{1} P(X = x) = P(X = 0) + P(X = 1) = (1 - p) + p = 1$.

Teorema 11 (Esperança, Variância e Desvio padrão da distribuição de Bernoulli)**.** *Se a variável aleatória X possui distribuição Bernoulli, cuja função de probabilidade é expressa em* (25)*, a esperança de X é dada por*

$$E[X] = p, \tag{26}$$

a variância

$$Var[X] = p(1-p), \tag{27}$$

e o desvio padrão

$$\sigma_X = \sqrt{p(1-p)}. \tag{28}$$

□

Demonstração. Sendo a esperança definida por:

$$E[X] = \sum_{x=0}^{n} xP(X = x),$$

e considerando o experimento da v.a. X com distribuição Bernoulli, em que $x = 1$, sucesso para o experimento com probabilidade p e $x = 0$, o fracasso, com probabilidade $(1 - p)$, então a esperança é dada por

$$E[X] = 1 \times p + 0 \times (1-p) = p,$$

e a variância sendo $Var(X) = E[X^2] - E^2[X]$, então,

$$E[X^2] = 1^2 \times p + 0^2 \times (1-p) = p$$

e

$$E^2[X] = p^2,$$

portanto,

$$Var[X] = p - p^2 = p(1-p).$$

O desvio padrão é dado por:

$$\sigma_x = \sqrt{p(1-p)}.$$

□

Exemplo 1. *Uma urna contém 30 bolas brancas e 20 verdes. Retira-se uma bola dessa urna. Seja X: número de bolas verdes, determine $P(x)$.*

$$X = \begin{cases} 0 \Rightarrow q = 30/50 = 3/5, \\ 1 \Rightarrow p = 20/50 = 2/5, \end{cases} \therefore P(X = x) = (2/5)^x(3/5)^{1-x}.$$

Assim, como $P(X = 1) = (2/5)^1(3/5)^0 = 2/5$.

Considere a repetição de n ensaios de Bernoulli independentes e todos com a mesma probabilidade de sucesso p. A v.a. que conta o número total de sucessos tem distribuição Binomial, definida a seguir.

Definição 24 (Distribuição Binomial)**.** *Uma variável aleatória X discreta, tem distribuição Binomial, se sua função de probabilidade é dada por*

$$P(X = x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & \text{para } x = 0, 1, 2, \ldots, n, \\ 0, & \text{caso contrário,} \end{cases} \tag{29}$$

em que $\binom{n}{x} = n!/[x!(n-x)!]$, e p é o parâmetro que indica o número de sucessos do evento ocorrer, sendo $0 < p < 1$.

Para mostrarmos que a distribuição Binomial é uma função de probabilidade, devemos mostrar as seguintes propriedades:

i) Mostrar que $P(X = x) \geq 0$;

ii) Mostrar que $\sum_{x=0}^{n} P(X = x) = 1$.

Demonstração. i) Como $n > 0$, $x \geq 0$, $n > x$ e $0 < p < 1$, então $\binom{n}{x} \geq 0$, $p^x \geq 0$ e $(1-p)^{n-x} \geq 0$.

Portanto,

$$P[X = x] = \binom{n}{x} p^x (1-p)^{n-x} \geq 0.$$

ii) Antes de chegarmos ao resultado, vamos entender uma propriedade do binômio de Newton. Para $n = 3$, temos

$$\begin{aligned} x = 0 \quad &\Rightarrow \quad \tbinom{3}{0} p^0 (1-p)^{3-0} = (1-p)^3 = q^3, \\ x = 1 \quad &\Rightarrow \quad \tbinom{3}{1} p^1 (1-p)^{3-1} = 3p(1-p)^2 = 3pq^2, \\ x = 2 \quad &\Rightarrow \quad \tbinom{3}{2} p^2 (1-p)^{3-2} = 3p^2(1-p)^1 = 3p^2 q, \\ x = 3 \quad &\Rightarrow \quad \tbinom{3}{3} p^3 (1-p)^{3-3} = p^3 (1-p)^0 = p^3. \end{aligned}$$

Agora observe,

$$(p + q)^3 = p^3 + 3p^2 q + 3pq^2 + q^3,$$

então, podemos concluir que

$$\begin{aligned} \sum_{x=0}^{3} P(X = x) &= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \\ &= (p + q)^3. \end{aligned}$$

Expandindo esse resultado para n, podemos observar que

$$\sum_{x=0}^{n} P(X = x) = (p + q)^n. \tag{30}$$

Sabendo que $q = 1 - p$, então

$$\begin{aligned} \sum_{x=0}^{n} P(X = x) &= [p + (1-p)]^n, \\ &= 1^n = 1. \end{aligned} \quad (31)$$

Portanto, $P(X = x)$ é uma função de probabilidade.

$\square$

Teorema 12 (Esperança, Variância e Desvio padrão da distribuição Binomial)**.** *Se a variável aleatória X possui distribuição Bernoulli, cuja função de probabilidade é expressa em* (25)*, a esperança de X é dada por*

$$E[X] = np, \quad (32)$$

a variância

$$Var[X] = np(1-p). \quad (33)$$

e o desvio padrão

$$\sigma_X = \sqrt{np(1-p)}. \quad (34)$$

$\square$

Demonstração. Sabemos que se $X_1, X_2, \ldots, X_n$ são variáveis aleatórias iid com distribuição de Bernoulli de parâmetro p, então $Y = \sum_{i=1}^{n} X_i \sim Binomial(n, p)$. Logo,

$$E[Y] = E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i] = nE[X] = np,$$

considerando $E[X]$ em (26). A variância pode ser obtida pelo mesmo raciocínio feito para obter a esperança, isto é,

$$Var[Y] = Var[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} Var[X_i] = nVar[X] = np(1-p),$$

considerando $Var[X]$ em (27). O desvio é a raiz da variância, logo

$$\sigma_Y = \sqrt{np(1-p)}.$$

$\square$

Seja uma sequência de experimentos independentes de uma variável aleatória com distribuição Bernoulli. Se desejássemos calcular a probabilidade de verificar o número de fracassos até o primeiro sucesso, essa medida seria analisada pela distribuição geométrica, isto é, X é uma variável aleatória que representa o tempo de espera (número de fracassos) até o primeiro sucesso de um evento. Dizemos que X tem distribuição geométrica, como segue a Definição.

Definição 25 (Distribuição Geométrica)**.** *Uma variável aleatória X discreta tem distribuição Geométrica, se sua função de probabilidade é definida por*

$$P(X = x) = \begin{cases} p(1-p)^x, & \textit{para } x = 0, 1, 2, \ldots, \\ 0, & \textit{caso contrário}, \end{cases} \quad (35)$$

sendo $0 < p < 1$.

Poderíamos redefinir a variável aleatória X com distribuição Geométrica, dizendo que X é o número de experimentos até o primeiro sucesso. O que acontece agora é que o primeiro sucesso entra na contagem. Desse modo, a distribuição Geométrica se modifica, como pode ser visto na seguinte definição.

Definição 26 (Distribuição Geométrica)**.** *Uma variável aleatória X discreta, tem distribuição Geométrica, se sua função de probabilidade é dada por*

$$P(X = x) = \begin{cases} p(1-p)^{x-1}, & \text{para } x = 1, 2, \ldots, \\ 0, & \text{caso contrário,} \end{cases} \tag{36}$$

sendo $0 < p < 1$.

Assim, como X contou o primeiro sucesso, a variável aleatória passará a ser definida $X \in \mathbb{N}^*+$, e o expoente da probabilidade do fracasso se altera para $(1-p)^{x-1}$.

Para verificar que a distribuição Geométrica é uma função de probabilidade, devemos mostrar as seguintes propriedades:

i) Mostrar que $P(X = x) \geq 0$;

ii) Mostrar que $\sum_x P(X = x) = 1$,

sabendo que para a definição 25 é algo similar.

Demonstração. Para o item (i), como $x \geq 0$ e $0 < p \leq 1$, então $p(1-p)^{x-1} \geq 0$, logo $P(X = x) \geq 0$. Para o item (ii), pela Definição (25) temos

$$\sum_{x=0}^{\infty} P(X = x) = \sum_{x=0}^{\infty} p(1-p)^x = p \sum_{x=0}^{\infty} (1-p)^x = \frac{p}{p} = 1.$$

Pela Definição (26) temos

$$\begin{aligned} \sum_{x=1}^{\infty} P(X = x) &= \sum_{x=1}^{\infty} p(1-p)^{x-1} \\ &= p \sum_{x=1}^{\infty} (1-p)^{x-1} \\ &= \frac{p}{1-p} \sum_{x=1}^{\infty} (1-p)^x \\ &= \frac{p}{1-p} \sum_{y=0}^{\infty} (1-p)^{y+1} \qquad (y = x - 1) \\ &= \frac{p(1-p)}{1-p} \sum_{y=0}^{\infty} (1-p)^y \\ &= p \sum_{y=0}^{\infty} (1-p)^y = \frac{p}{p} = 1. \end{aligned}$$

□

Teorema 13 (Esperança, Variância e Desvio padrão da distribuição Binomial)**.** *Se a variável aleatória X possui distribuição Geométrica, cuja função de probabilidade é expressa em* (35) *ou* (36)*, a esperança de X é dada por*

$$E[X] = \frac{1-p}{p} \qquad \textit{Definição (25)} \tag{37}$$

$$= \frac{1}{p}, \qquad \textit{Definição (26)} \tag{38}$$

a variância

$$Var[X] = \frac{1-p}{p^2}, \qquad \textit{(Para as duas situações)} \tag{39}$$

e o desvio padrão

$$\sigma_X = \sqrt{\frac{1-p}{p^2}}. \qquad \textit{(Para as duas situações)} \tag{40}$$

□

Demonstração. Uma prova interessante e rápida. Seja $U \sim Bernoulli(1-p)$ e $X \sim Geo(p)$. Considerando que U e X são variáveis aleatórias independentes, então $X^* = U(1+X)$ é uma variável aleatória que tem a mesma distribuição de X, distribuição geométrica de parâmetro p, isto é,

$$\begin{array}{lrl}
P(X^* = 0) \Rightarrow & P(U=0) = & p \\
P(X^* = 1) \Rightarrow & P(U=1)P(X=0) = & (1-p)p \\
P(X^* = 2) \Rightarrow & P(U=1)P(X=1) = & (1-p)^2 p \\
\vdots & \vdots & \vdots \\
P(X^* = x) \Rightarrow & P(U=1)P(X=x-1) = & (1-p)^x p.
\end{array}$$

Dessa forma podemos calcular a esperança e a variância facilmente.

$$\begin{aligned}
E[X^*] &= E[U(1+X)] \\
&= E[U] + E[U]E[X] \\
&= (1-p) + (1-p)E[X^*], \qquad (E[X] = E[X^*]) \\
E[X^*] - [(1-p)E[X^*]] &= (1-p) \\
E[X^*] &= \frac{1-p}{p}.
\end{aligned}$$

Para a variância temos

$$Var[X^*] = E[X^{*2}] - (E[X^*])^2,$$

em que $E[X^*]$ já foi obtido. A segunda parte é dada por

$$\begin{aligned}
E[X^{*2}] &= E[U^2(1+X)^2] = E[U^2]E[(1+X)^2] \\
&= (Var[U] + (E[U])^2)(Var[1+X] + (E[1+X])^2) \\
&= (p(1-p) + (1-p)^2)\left[Var[X] + \left(1 + \frac{q}{p}\right)^2\right] \\
Var[X^*] + (E[X^*])^2 &= (p(1-p) + (1-p)^2)\left[Var[X] + \frac{1}{p^2}\right], \qquad (q + p = 1) \\
Var[X^*] + \left(\frac{1-p}{p}\right)^2 &= (1-p)[p + (1-p)]\left[Var[X^*] + \frac{1}{p^2}\right], \qquad Var[X^*] = Var[X] \\
Var[X^*] + \left(\frac{1-p}{p}\right)^2 &= (1-p)Var[X^*] + \frac{1-p}{p^2} \\
Var[X^*]p &= \frac{1-p}{p^2} - \frac{(1-p)^2}{p^2} \\
Var[X^*] &= \frac{1-p}{p^2}.
\end{aligned}$$

O desvio padrão é imediato. □

Demonstração. Para a Definição (25) temos

$$\begin{aligned}
E[X] &= \sum_{x=0}^{\infty} xp(1-p)^x \\
&= p\sum_{x=0}^{\infty} x(1-p)^x \\
&= p(1-p)\sum_{x=0}^{\infty} x(1-p)^{x-1} \\
&= p(1-p)\left[-\sum_{x=0}^{\infty}\frac{d}{dp}(1-p)^x\right] \\
&= p(1-p)\left[-\frac{d}{dp}\sum_{x=0}^{\infty}(1-p)^x\right] \qquad \text{(Sob condições de regularidade)} \\
&= p(1-p)\left[-\frac{d}{dp}\frac{1}{p}\right] = \frac{p(1-p)}{p^2} = \frac{(1-p)}{p}.
\end{aligned}$$

Para a Definição (26) temos

$$\begin{aligned}
E[X] &= \sum_{x=1}^{\infty} xp(1-p)^{x-1} \\
&= p\sum_{x=1}^{\infty} x(1-p)^{x-1} \\
&= p\sum_{y=0}^{\infty}(y+1)(1-p)^y \qquad (y = x - 1)
\end{aligned}$$

$$
\begin{aligned}
&= p\left[-\sum_{y=0}^{\infty}\frac{d}{dp}(1-p)^{y+1}\right] \\
&= p\left[-\frac{d}{dp}(1-p)\sum_{y=0}^{\infty}(1-p)^{y}\right] \\
&= p\left[\frac{1+(p-1)}{p^2}\right] = \frac{1}{p}.
\end{aligned}
$$

A variância pode ser expressa por

$$
\begin{aligned}
Var[X] &= E[X^2] - (E[X])^2 \\
&= E[X^2] - E[X] + E[X] - (E[X])^2 \\
&= E[X(X-1)] + E[X] - (E[X])^2.
\end{aligned}
$$

Para $E[X(X-1)]$, temos

$$
\begin{aligned}
E[X(X-1)] &= \sum_{x=1}^{\infty} x(x-1)p(1-p)^{x-1} \\
&= \sum_{y=0}^{\infty}(y+1)yp(1-p)^{y} \qquad (y = x-1) \\
&= p(1-p)\sum_{y=0}^{\infty}(y+1)y(1-p)^{y-1},
\end{aligned}
$$

sabendo que $\frac{d^2}{dp^2}(1-p)^{y+1} = y(y+1)(1-p)^{y-1}$, então

$$
\begin{aligned}
E[X(X-1)] &= p(1-p)\sum_{y=0}^{\infty}\frac{d^2}{dp^2}(1-p)^{y+1} \\
&= p(1-p)\frac{d^2}{dp^2}\sum_{y=0}^{\infty}(1-p)^{y+1} \\
&= p(1-p)\frac{d^2}{dp^2}(1-p)\sum_{y=0}^{\infty}(1-p)^{y} \\
&= p(1-p)\frac{d^2}{dp^2}\frac{(1-p)}{p} \\
&= p(1-p)\left[\frac{2}{p^3}\right] \\
&= \frac{2(1-p)}{p^2}.
\end{aligned}
$$

Assim,

$$
\begin{aligned}
Var[X] &= \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} \\
&= \frac{2(1-p)+p-1}{p^2}
\end{aligned}
$$

$$= \frac{1-p}{p^2}.$$

Finalmente, o desvio padrão segue a expressão (40). $\square$

Considere agora, um experimento em que a variável aleatória X, representa o número de fracassos até o r-ésimo sucesso. Dizemos que X tem distribuição binomial negativa.

Definição 27 (Distribuição Binomial Negativa)**.** *Uma variável aleatória X discreta, tem distribuição Binomial negativa, se sua função de probabilidade é dada por*

$$P(X=x) = \begin{cases} \binom{x+r-1}{x} p^r (1-p)^x, & \text{para } x = 0, 1, 2, \ldots, \\ 0, & \text{caso contrário,} \end{cases} \tag{41}$$

sendo $0 < p < 1$. $\square$

Um exemplo para elucidar essa definição.

Exemplo 2 (Binomial negativa, expressão (41))**.** *Um experimento com uma moeda honesta, em que cada lançamento pode resultar em cara ou coroa. O objetivo do experimento é verificar a probabilidade de obter 3 vezes a face coroa até sair pela segunda vez a face cara. Observamos que a variável aleatória X quantifica o número de fracassos (número de coroas). Dessa forma, $X = 3$ fracassos até a ocorrência do segundo sucesso ($r = 2$). Assim,*

$$P(X=3) = \binom{3+2-1}{3} (1/2)^2 (1-1/2)^3 = 0,125.$$

$\square$

A expressão nos mostra que para as $x + r - 1$ primeiras realizações, precisamos ter $r - 1$ sucessos. Poderíamos expressar essa distribuição de outro modo,

$$P(X=x) = \binom{x+r-1}{r-1} p^r (1-p)^x, \tag{42}$$

ou

$$P(X=x) = (-1)^x \binom{-r}{x} p^r (1-p)^x. \tag{43}$$

Este último, é o que justifica o nome da distribuição.

Agora, seja $Y = X + r$, sendo que e a variável aleatória Y o número de ensaios até r sucessos. Y também tem distribuição binomial negativa, com a seguinte função de probabilidade

$$P(Y=y) = \begin{cases} \binom{y-1}{r-1} p^r (1-p)^{y-r}, & \text{para } y = r, r+1, \ldots, \\ 0, & \text{caso contrário,} \end{cases} \tag{44}$$

A distribuição Binomial Negativa está relacionado com a distribuição Geométrica. Isto é, se $X \sim Geo(p)$, então $Y = \sum_{i=1}^{r} X_i$, sendo r o número de sucessos. Veja o esquema abaixo, supondo um sequência de experimentos Bernoulli sendo F - fracasso e S - Sucesso:

$$\overbrace{\underbrace{F\quad F\quad F\quad S}_{X_1}\quad \underbrace{F\quad S}_{X_2}\quad \underbrace{F\quad F\quad S}_{X_3}\quad \ldots\quad \underbrace{F\quad F\quad S}_{X_r}}^{Y}$$

Por isso $Y = X + r$, sendo X o número de fracassos e r o número de sucessos. Se $r = 1$, então $Y \sim BN(p, 1)$ que é $Y \sim Geo(p)$.

Exemplo 3 (Binomial negativa, expressão (44))**.** *Suponha que num experimento com 1 dado honesto, desejamos obter 3 vezes na face superior, o valor dois. Para isso, iremos dispor de 6 tentativas. Observe que nessa situação, estamos querendo em 6 tentativas obter 3 vezes o valor dois na face superior. Na expressão* (44)*, temos que* $Y = 6$ *e* $r = 3$*. Isso implica dizer, qual probabilidade de em 6 tentativas,* $P(Y = 3)$*, obtermos 3 sucessos (face superior igual a dois). Observe na combinatória que temos* $\binom{y-1}{r-1}$*, apenas* $y - 1$ *tentativas e* $r - 1$ *sucessos, porque na* x*-ésima tentativa já está definido um sucesso. Em outras palavras, na sexta tentativa, o modelo fixa um sucesso, restando apenas, as outras* $y - 1 = 5$ *tentativas combinar os outros* $r - 1 = 2$ *sucessos. Assim,*

$$P(Y = 6) = \binom{6-1}{3-1}\left(\frac{1}{6}\right)^3\left(1 - \frac{1}{6}\right)^3 \approx 0,0267918.$$

□

Usando a definição 43, iremos mostrar que a distribuição Binomial negativa é uma função de probabilidade, devemos mostrar as seguintes propriedades:

i) Mostrar que $P(X = x) \geq 0$;

ii) Mostrar que $\sum_{x=0}^{\infty} P(X = x) = 1$,

Demonstração. i) Como $x \geq 0$ e $0 < p \leq 1$, logo, $P(X = x) \geq 0$;

ii) Antes de iniciarmos a prova, vamos apresentar do resultado do teorema binomial, que afirma que para duas variáveis x e y, tem-se:

$$\begin{aligned}(x+y)^n &= \binom{n}{0}x^n y^0 + \binom{n}{1}x^{n-1}y^1 + \binom{n}{2}x^{n-2}y^2 + \cdots \\ &\quad \cdots + \binom{n}{n-1}x^1 y^{n-1} + \binom{n}{n}x^0 y^n, \\ &= \sum_{k=0}^{n}\binom{n}{k}x^{n-k}y^k \\ &= \sum_{k=0}^{n}\binom{n}{k}x^k y^{n-k}.\end{aligned}$$

Considerando apenas uma variável, substituindo y por 1, temos

$$\begin{aligned}(1+x)^n &= \binom{n}{0}x^0 + \binom{n}{1}x^1 + \binom{n}{2}x^2 + \cdots \\ &\quad \cdots + \binom{n}{n-1}x^{n-1} + \binom{n}{n}x^n,\end{aligned}$$

ou equivalentemente,

$$(1+x)^n = \sum_{k=0}^{n}\binom{n}{k}x^k.$$

Se considerarmos uma série binomial, que é a mesma que uma série taylor para qualquer constante α e $|x| < 1$, temos

$$\begin{aligned}(1+x)^\alpha &= \sum_{k=0}^{\infty} \binom{\alpha}{k} x^k \\ &= 1 + \alpha x + \frac{\alpha(\alpha-1)}{2!} x^2 + \cdots .\end{aligned}$$

E considerando ainda uma série binomial com números inteiros negativos, temos

$$(1-q)^{-r} = \sum_{k=0}^{\infty} \binom{-r}{k} (-q)^k. \tag{45}$$

Assim, de acordo com a expressão (43), e fazendo $q = 1 - p$, temos

$$\begin{aligned}\sum_{x=0}^{\infty} P(X=x) &= \sum_{x=0}^{\infty} (-1)^x \binom{-r}{x} p^r (q)^x \\ &= p^r \sum_{x=0}^{\infty} \binom{-r}{x} (-1)^x q^x \\ &= p^r \sum_{x=0}^{\infty} \binom{-r}{x} (-q)^x,\end{aligned}$$

usando o resultado da expressão (45), segue

$$\sum_{x=0}^{\infty} P(X=x) = p^r (1-q)^{-r},$$

como $p = 1 - q$, logo

$$\sum_{x=0}^{\infty} P(X=x) = p^r \times p^{-r} = 1.$$

□

Teorema 14 (Esperança, Variância e Desvio padrão da distribuição Binomial Negativa)**.** *Se a variável aleatória Y possui distribuição Binomial Negativa expressa em* (44)*, então a esperança de X é dada por*

$$E[X] = \frac{k}{p}, \tag{46}$$

a variância

$$Var[X] = \frac{r(1-p)}{p^2}, \tag{47}$$

e o desvio padrão

$$\sigma_X = \sqrt{\frac{r(1-p)}{p^2}}. \qquad \textit{(Para as duas situações)} \tag{48}$$

□

Demonstração. Sabemos que se $X_1, X_2, \ldots, X_r$ são variáveis aleatórias iid com distribuição Geométrica de parâmetro p, então $Y = \sum_{i=1}^{r} X_i \sim BN(p, r)$. Logo,

$$E[Y] = E[\sum_{i=1}^{r} X_i] = \sum_{i=1}^{r} E[X_i] = rE[X] = \frac{r}{p},$$

considerando $E[X]$ em (38). A variância pode ser obtida pelo mesmo raciocínio feito para obter a esperança, isto é,

$$Var[Y] = Var[\sum_{i=1}^{r} X_i] = \sum_{i=1}^{r} Var[X_i] = rVar[X] = \frac{r(1-p)}{p^2},$$

considerando $Var[X]$ em (39). O desvio é a raiz da variância, logo

$$\sigma_Y = \sqrt{\frac{r(1-p)}{p^2}}.$$

$\square$

Um outro tipo de distribuição pode ser visto, num experimento clássico com urnas. Suponha que temos uma urna grande com N bolas. Essas bolas podem ter M bolas vermelhas e $N - M$ bolas verdes. Assim, retiramos uma amostra de tamanho K dessa urna, e pretendemos saber, qual a probabilidade de se ter x bolas vermelhas nessa amostra K?. Note que as K bolas retiradas da urna são todas de uma vez (sem reposição). Estamos diante de uma experimento, em que a variável aleatória X, tem distribuição hipergeométrica.

O número total de amostras possíveis de retirar uma amostra K dentre as N bolas, é $\binom{N}{K}$. O número possível de retirar x num total de M bolas vermelhas, é $\binom{M}{x}$. E o número possível de retirar $K - x$ num total de $N - M$ bolas verdes, é $\binom{N-M}{K-x}$. Assim, definimos

Definição 28 (Distribuição hipergeométrica)**.** *Uma variável aleatória X discreta, tem distribuição hipergeométrica, se sua função de probabilidade é dada por*

$$P(X = x) = \begin{cases} \frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}}, & \textit{para } x = 0, 1, 2, \ldots, K \\ 0, & \textit{caso contrário,} \end{cases} \tag{49}$$

em que $M \geq x$, $N - M \geq K - x$, e $M - (N - K) \leq x \leq M$. $\square$

Para mostrar que a distribuição Hipergeométrica é uma função de probabilidade, devemos provas as seguintes propriedades:

i) Mostrar que $P(X = x) \geq 0$;

ii) Mostrar que $\sum_x P(X = x) = 1$.

Demonstração. Para o item (i) verificamos claramente que $P[X = x] > 0$ para $x = 0, 1, \ldots, K$. Para o item (ii), pela identidade de Vondermonde:

$$\sum_{x=0}^{K} \binom{M}{x}\binom{N-M}{K-x} = \binom{M}{K},$$

logo,

$$\sum_{x=0}^{K} P(X=x) = \sum_{x=0}^{K} \frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}} = \frac{1}{\binom{N}{K}} \sum_{x=0}^{K} \binom{M}{x}\binom{N-M}{K-x} = \frac{\binom{N}{K}}{\binom{N}{K}} = 1.$$

□

Teorema 15 (Esperança, Variância e Desvio padrão da distribuição Binomial Negativa)**.** *Se a variável aleatória X possui distribuição Hipergeométrica expressa em* (49)*, então a esperança de X é dada por*

$$E[X] = K\frac{M}{N}, \tag{50}$$

a variância

$$Var[X] = K\frac{M}{N}\frac{N-M}{N}\frac{N-K}{N-1}, \tag{51}$$

e o desvio padrão

$$\sigma_X = \sqrt{K\frac{M}{N}\frac{N-M}{N}\frac{N-K}{N-1}}. \tag{52}$$

□

Demonstração.

$$\begin{aligned}
E[X] &= \sum_{x=0}^{K} x\frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}}\\
&= 0 \times \frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}} + \sum_{x=1}^{K} x\frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}}\\
&= \sum_{x=1}^{K} x\frac{\frac{M}{x}\binom{M-1}{x-1}\binom{N-M}{K-x}}{\frac{N}{K}\binom{N-1}{K-1}}\\
&= M\frac{K}{N}\sum_{x=1}^{K} \frac{\binom{M-1}{x-1}\binom{N-M}{K-x}}{\binom{N-1}{K-1}}\\
&= M\frac{K}{N}\sum_{y=0}^{K-1} \frac{\binom{M-1}{y}\binom{N-M}{K-1-y}}{\binom{N-1}{K-1}} \qquad (y = x-1)\\
&= M\frac{K}{N}\frac{\binom{N-1}{K-1}}{\binom{N-1}{K-1}} \qquad \text{(Identidade de Vandermonde)}\\
&= M\frac{K}{N}.
\end{aligned}$$

A variância pode ser expressa por

$$\begin{aligned}
Var[X] &= E[X^2] - (E[X])^2\\
&= E[X^2] - E[X] + E[X] - (E[X])^2\\
&= E[X(X-1)] + E[X] - (E[X])^2.
\end{aligned}$$

Para $E[X(X-1)]$, temos

$$
\begin{aligned}
E[X(X-1)] &= \sum_{x=0}^{K} x(x-1)\frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}}\\
&= 0(0-1)\times\frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}} + 1(1-1)\times\frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}} + \sum_{x=2}^{K} x\frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}}\\
&= \sum_{x=2}^{K} x(x-1)\frac{\frac{M(M-1)}{x(x-1)}\binom{M-2}{x-2}\binom{N-M}{K-x}}{\frac{N(N-1)}{K(K-1)}\binom{N-1}{K-1}}\\
&= M(M-1)\frac{K(K-1)}{N(N-1)}\sum_{x=2}^{K}\frac{\binom{M-2}{x-2}\binom{N-M}{K-x}}{\binom{N-2}{K-2}}\\
&= M(M-1)\frac{K(K-1)}{N(N-1)}\sum_{y=0}^{K-2}\frac{\binom{M-2}{y}\binom{N-M}{K-2-y}}{\binom{N-2}{K-2}} \qquad (y = x-2)\\
&= M(M-1)\frac{K(K-1)}{N(N-1)}\frac{\binom{N-2}{K-2}}{\binom{N-2}{K-2}} \qquad \text{(Identidade de Vandermonde)}\\
&= M(M-1)\frac{K(K-1)}{N(N-1)}.
\end{aligned}
$$

Assim,

$$
\begin{aligned}
Var[X] &= M(M-1)\frac{K(K-1)}{N(N-1)} + M\frac{K}{N} - M^2\frac{K^2}{N^2}\\
&= M\frac{K}{N}\left[(M-1)\frac{(K-1)}{(N-1)} + 1 - M\frac{K}{N}\right]\\
&= M\frac{K}{N}\left[\frac{(N-M)(N-K)}{N(N-1)}\right].
\end{aligned}
$$

Portanto, o desvio padrão é dado por

$$
\sigma_X = \sqrt{M\frac{K}{N}\left[\frac{(N-M)(N-K)}{N(N-1)}\right]}.
$$

□

Se considerássemos $p = \frac{M}{N}$ e $n = K$, reescreveríamos a esperança e a variância de X por:

$$
E[X] = np, \tag{53}
$$

a variância

$$
Var[X] = np(1-p)\frac{N-n}{N-1}, \tag{54}
$$

em que p representa o número de sucessos numa amostra de tamanho N. Isso significa, que a esperança da distribuição hipergeométrica é idêntica da distribuição binomial. Entretanto, sua variância é igual a da distribuição binomial acrescido um fator de correção para populações finitas dada por $\frac{N-n}{N-1}$. Entretanto, se a amostra fosse com reposição, X deve ter uma distribuição binomial e sua variância deve ser $np(1-p)$.

Uma distribuição discreta amplamente aplicada em diversos tipos de experimentos, é a distribuição de Poisson. Eventos do tipo, fenômenos que esperamos dado um tempo ou um espaço, como o número de telefonemas de uma atendente em uma hora, ou o número de acidentes por 15km em uma rodovia. Uma das ideias por trás dessa distribuição é que para pequenos intervalos de tempo, a probabilidade de uma chegada é proporcional ao tempo de espera (Ver portal action, tem uma introdução legal!!!). Isso faz com que a distribuição consiga modelar os eventos citados anteriormente. Assim, definimos

Definição 29 (Distribuição de Poisson)**.** *Uma variável aleatória X discreta, tem distribuição de Poisson, se sua função de probabilidade é dada por*

$$P(X=x)=\begin{cases}\frac{e^{-\lambda}\lambda^x}{x!}, & \textit{para } x=0,1,2,\ldots,\\ 0, & \textit{caso contrário},\end{cases} \tag{55}$$

em que $\lambda > 0$. □

Para provar que a distribuição de Poisson é uma função de probabilidade, temos que demostrar os itens:

i) Mostrar que $P(X=x)\geq 0$;

ii) Mostrar que $\sum_{x=0}^{\infty} P(X=x)=1$,

Demonstração. i) Sabe-se que $\lambda > 0$, $x=0,1,2,\ldots$, e a constante $e = 2,7188281...$, então

- $e^{-\lambda} > 0$,
- $\lambda^x > 0$,
- $x! > 0$,

portanto, $P(X=x)\geq 0$.

ii) Tendo,

$$\begin{aligned}\sum_{x=0}^{\infty} P(X=x) &= \sum_{x=0}^{\infty}\frac{e^{-\lambda}\lambda^x}{x!},\\ &= e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^x}{x!},\\ &= e^{-\lambda}\left[\frac{\lambda^0}{0!}+\frac{\lambda^1}{1!}+\frac{\lambda^2}{2!}+\ldots+\frac{\lambda^x}{x!}+\ldots\right],\\ &= e^{-\lambda}\left[\underbrace{1+\lambda+\frac{\lambda^2}{2!}+\ldots+\frac{\lambda^x}{x!}+\ldots}_{\text{série de Taylor}}\right],\end{aligned}$$

e sabendo que $e^x = 1 + x + \frac{x^2}{2!} + \ldots$, é a expressão da série de Taylor, portanto

$$\sum_{x=0}^{\infty} P(X=x) = e^{-\lambda}e^{\lambda} = e^0 = 1.$$

Assim, provamos que a distribuição de Poisson é uma distribuição de probabilidade. □

Teorema 16 (Esperança, Variância e Desvio padrão da distribuição Poisson)**.** *Se a variável aleatória Y possui distribuição Poisson expressa em* (55)*, então a esperança de X é dada por*

$$E[X] = \lambda, \tag{56}$$

a variância

$$Var[X] = \lambda, \tag{57}$$

e o desvio padrão

$$\sigma_X = \sqrt{\lambda}. \tag{58}$$

$\square$

Demonstração. A dedução da esperança matemática pode ser deduzida da seguinte forma:

$$\begin{aligned} E[X] &= \sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!}, \\ &= 0 \times \frac{e^{-\lambda}\lambda^0}{0!} + \sum_{x=1}^{\infty} x\frac{e^{-\lambda}\lambda^x}{x!}, \end{aligned}$$

fazendo $y = x - 1$, então,

$$\begin{aligned} E[X] &= \sum_{y=0}^{\infty} (y+1)\frac{e^{-\lambda}\lambda^{y+1}}{(y+1)!}, \\ &= \sum_{y=0}^{\infty} (y+1)\frac{e^{-\lambda}\lambda^{y+1}}{(y+1)y!}, \\ E[X] &= \sum_{y=0}^{\infty} \frac{e^{-\lambda}\lambda^{y}\lambda}{y!}, \\ &= \lambda \underbrace{\sum_{y=0}^{\infty} \frac{e^{-\lambda}\lambda^{y}}{y!}}_{=1}, \\ &= \lambda. \end{aligned}$$

A variância pode ser desenvolvida da forma a seguir.

$$\begin{aligned} E[X^2] &= \sum_{x=0}^{\infty} x^2\frac{e^{-\lambda}\lambda^x}{x!}, \\ &= 0^2 \times \frac{e^{-\lambda}\lambda^0}{0!} + \sum_{x=1}^{\infty} x^2\frac{e^{-\lambda}\lambda^x}{x!}, \\ &= \sum_{x=1}^{\infty} x^2\frac{e^{-\lambda}\lambda^x}{x!} \end{aligned}$$

Como $Var(X) = E[X^2] - E^2[X]$, então,

$$Var(X) = \sum_{x=1}^{\infty} x^2\frac{e^{-\lambda}\lambda^x}{x!} - \lambda^2,$$

fazendo $y = x - 1$, então,

$$
\begin{aligned}
Var(X) &= \sum_{y=0}^{\infty}(y+1)^2\frac{e^{-\lambda}\lambda^{(y+1)}}{(y+1)!} - \lambda^2,\\
&= \sum_{y=0}^{\infty}(y+1)^2\frac{e^{-\lambda}\lambda^{(y+1)}}{(y+1)y!} - \lambda^2,\\
&= \sum_{y=0}^{\infty}(y+1)\frac{e^{-\lambda}\lambda^{(y+1)}}{y!} - \lambda^2,\\
&= \sum_{y=0}^{\infty}y\frac{e^{-\lambda}\lambda^{y}\lambda}{y!} + \sum_{y=0}^{\infty}\frac{e^{-\lambda}\lambda^{y}\lambda}{y!} - \lambda^2,\\
&= \lambda\left[\underbrace{\sum_{y=0}^{\infty}y\frac{e^{-\lambda}\lambda^{y}}{y!}}_{=E[X]=\lambda} + \underbrace{\sum_{y=0}^{\infty}\frac{e^{-\lambda}\lambda^{y}}{y!}}_{=1} - \lambda\right],\\
&= \lambda(\cancel{\lambda} + 1 - \cancel{\lambda}) = \lambda. \qquad (59)
\end{aligned}
$$

o desvio padrão é dado por

$$\sigma_X = \sqrt{\lambda}. \qquad (60)$$

□

5.1.1 Relação entre variáveis aleatórias discretas

Teorema 17 (Distribuição Binomial e Distribuição Poisson). *Mood p. 96-97; Magalhaes p. 91-92; Ross p.29-30. Prova como teorema em DeGroot p. 291; wikipedia/Poisson_limit_theorem (ótimo).*

Exemplo 4 (DeGroot(2014) p. 292). *Suponha que uma grande população tenha uma proporção de 0,01 que uma certa doença aconteça. Desejamos determinar a probabilidade de que numa amostra aleatória de 200 pessoas possa encontrar ao menos 4 pessoas com essa doença.* □

Ver Magalhães p.90-92.

5.2 Distribuições especiais para variáveis aleatórias contínuas

Definição 30 (Distribuição Normal). *Uma variável aleatória X contínua, tem distribuição normal se sua função densidade de probabilidade é dada por*

$$
f(x) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} & \text{, } para \quad -\infty < x < \infty, \\ 0 & \text{, } caso\ contrário, \end{cases} \qquad (61)
$$

em que os parâmetros μ e σ^2 satisfazem $\mu \in \mathbb{R}$ e $\sigma^2 > 0$. □

Algumas propriedade da densidade da normal podem ser facilmente observadas de seu gráfico:

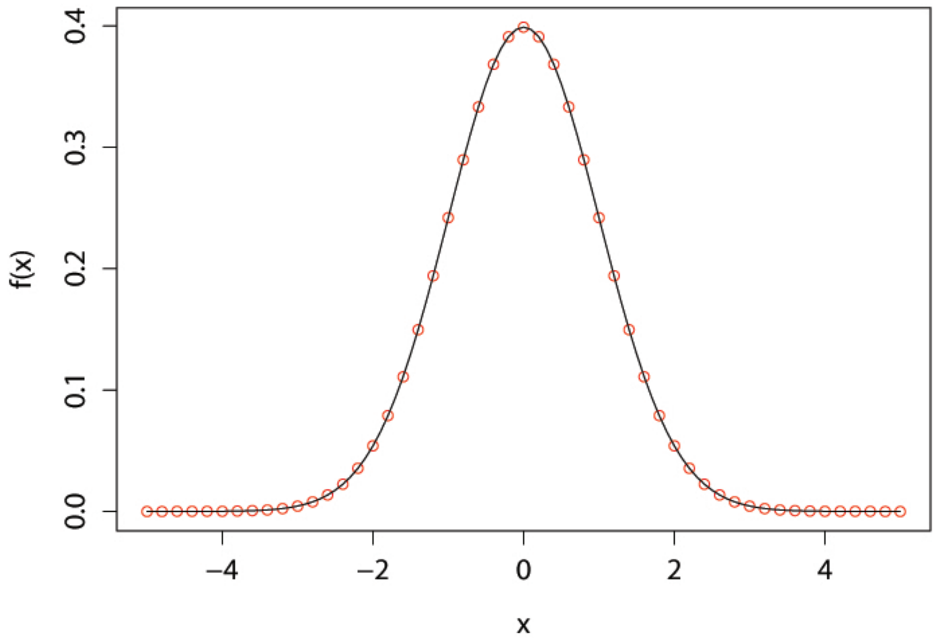


Figura 2: Função densidade da distribuição normal.

i) $f(x)$ é simétrica em relação a μ;

ii) $f(x) \to 0$ quando $x \to \pm\infty$;

iii) o valor máximo de x se dar para $x = \mu$.

Para verificar se $f(x)$ é uma função densidade de probabilidade, temos que mostrar que:

i) $f(x) \geq 0$, para todo $x \in (-\infty, \infty)$;

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

Prova:

i) Vamos fazer as seguintes observações:

   – a constante $\frac{1}{\sqrt{2\pi\sigma^2}} \geq 0$;
   – todo valor e elevado a um valor negativo, sempre retorna um valor positivo.

   Dessa forma, $f(x) \geq 0$.

• Para mostrar que $A = \int_{-\infty}^{\infty} f(x)dx = 1$, vamos partir do seguinte raciocínio. Se $A = 1$, então $A^2 = 1$. Assim,

$$\begin{aligned} A^2 &= A \times A, \\ &= \int_{-\infty}^{\infty} f(x)dx \times \int_{-\infty}^{\infty} f(x)dx, \\ &= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx \times \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx. \end{aligned}$$

Fazendo uma mudança de variável, $y = \frac{(x-\mu)}{\sigma}$ e $dy = \frac{1}{\sigma}dx$, então

$$A^2 = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy \times \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy.$$

Mudando uma variável y de uma das densidades, simplesmente para diferenciá-las, chamando-a de w, temos

$$\begin{aligned} A^2 &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy \times \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}w^2} dw, \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sqrt{2\pi}} e^{-\frac{1}{2}(y^2+w^2)} dydw. \end{aligned}$$

Usando a técnica do cálculo da integral de coordenadas polares, temos

$$\begin{aligned} y &= r \sin\theta \\ w &= w \cos\theta. \end{aligned}$$

Fazendo,

$$\begin{aligned} y^2 + w^2 &= r^2 \sin^2\theta + r^2 \cos^2\theta, \\ &= r^2(\sin^2\theta + \cos^2\theta) = r^2, \end{aligned}$$

considerando $dA = rdrd\theta$, e que o intervalos das variáveis r sendo $[0, \infty]$ e θ sendo $[0, 2\pi]$, então A^2 pode ser expresso da seguinte forma

$$\begin{aligned} A^2 &= \int_{0}^{\infty} \int_{0}^{2\pi} \frac{1}{\sqrt{2\pi}\sqrt{2\pi}} e^{-\frac{1}{2}r^2} d\theta r dr, \\ &= \int_{0}^{\infty} \int_{0}^{2\pi} \frac{1}{2\pi} e^{-\frac{1}{2}r^2} d\theta r dr, \\ &= \frac{1}{2\pi} \int_{0}^{\infty} e^{-\frac{1}{2}r^2} r \left[\theta\right]_0^{2\pi} dr, \\ &= \frac{1}{2\pi} \int_{0}^{\infty} e^{-\frac{1}{2}r^2} r(2\pi) dr - \underbrace{\frac{1}{2\pi} \int_{0}^{\infty} e^{-\frac{1}{2}r^2} r(0) dr,}_{=0} \\ &= \int_{0}^{\infty} e^{-\frac{1}{2}r^2} r dr. \end{aligned}$$

Faremos um artifício de multiplicar $1 = (-1) \times (-1)$ na expressão, que não alterará o resultado, ou seja,

$$\begin{aligned} A^2 &= 1 \times \int_{0}^{\infty} e^{-\frac{1}{2}r^2} r dr, \\ &= (-1) \times (-1) \times \int_{0}^{\infty} e^{-\frac{1}{2}r^2} r dr, \\ &= (-1) \times \int_{0}^{\infty} (-1) e^{-\frac{1}{2}r^2} r dr. \end{aligned}$$

Usando mais uma mudança de variável $u = -(1/2)r^2$ e $du = -rdr$, podemos concluir

$$\begin{aligned} A^2 &= (-1) \times \int_0^{\infty} e^u du, \\ &= -\left[e^u\right]_0^{\infty}, \\ &= -[0-1] = 1. \end{aligned} \tag{62}$$

Se $A^2 = 1$, então $A = 1$. Assim, fica concluído que a distribuição normal é uma função densidade de probabilidade.

Teorema 18 (Esperança, Variância e Desvio padrão da distribuição Normal)**.** *Se a variável aleatória Y possui distribuição Normal expressa em* (61)*, então a esperança de X é dada por*

$$E[X] = \mu, \tag{63}$$

a variância

$$Var[X] = \sigma^2, \tag{64}$$

e o desvio padrão

$$\sigma_X = \sigma. \tag{65}$$

□

Demonstração. A esperança pode ser deduzida da seguinte forma:

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x f(x) dx, \\ &= \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx, \end{aligned}$$

fazendo uma mudança de variável, $y = (x-\mu)/\mu \Rightarrow x = y\sigma + \mu$ e $du = (1/\sigma)dx \Rightarrow dx = \sigma dy$, então

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} \left(\frac{y\sigma + \mu}{\cancel{\sigma}\sqrt{2\pi}}\right) e^{-\frac{1}{2}y^2} \cancel{\sigma} dy, \\ &= \int_{-\infty}^{\infty} \left(\frac{y\sigma + \mu}{\sqrt{2\pi}}\right) e^{-\frac{1}{2}y^2} dy, \\ &= \int_{-\infty}^{\infty} \left(\frac{y\sigma}{\sqrt{2\pi}}\right) e^{-\frac{1}{2}y^2} dy + \int_{-\infty}^{\infty} \left(\frac{\mu}{\sqrt{2\pi}}\right) e^{-\frac{1}{2}y^2} dy, \\ &= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y e^{-\frac{1}{2}y^2} dy + \mu \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy}_{=1}, \\ &= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y e^{-\frac{1}{2}y^2} dy + \mu. \end{aligned}$$

Algo interessante acontece com a integral $\int_{-\infty}^{\infty} y e^{-\frac{1}{2}y^2} dy$. Observe graficamente na Figura 3. Como $A = B$, a integral se anula por simetria. Assim,

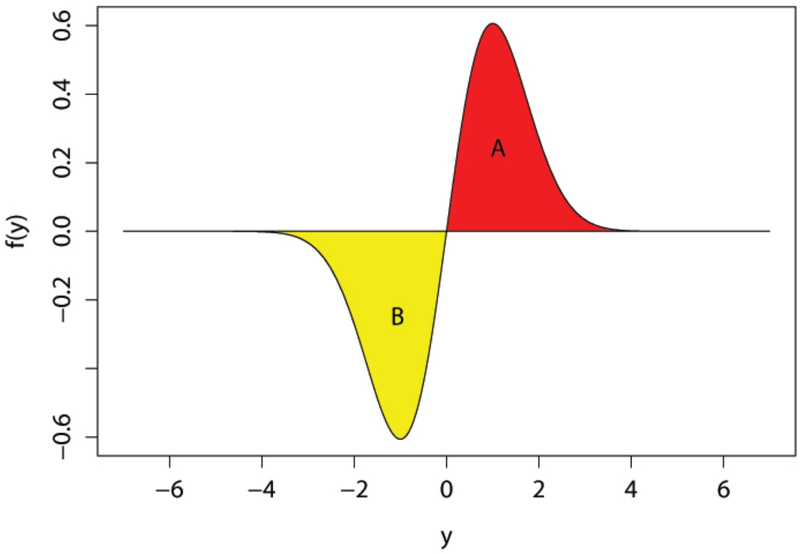


Figura 3: Gráfico da função $\int_{-\infty}^{\infty} y e^{-\frac{1}{2}y^2} dy$.

$$
\begin{aligned}
E[X] &= \frac{\sigma}{\sqrt{2\pi}} \underbrace{\int_{-\infty}^{\infty} y e^{-\frac{1}{2}y^2} dy}_{=0} + \mu, \\
&= \mu. \qquad (66)
\end{aligned}
$$

A variância será deduzida a seguir.

$$
E[X^2] = \int_{-\infty}^{\infty} x^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx,
$$

usando a mudança de variável $y = (x-\mu)/\sigma \Rightarrow x = \mu + \sigma y$ e $dy = (1/\sigma)dx \Rightarrow dx = \sigma dy$, então

$$
\begin{aligned}
E[X^2] &= \int_{-\infty}^{\infty} \frac{(\mu+\sigma y)^2}{\cancel{\sigma}\sqrt{2\pi}} e^{-\frac{1}{2}(y)^2} \cancel{\sigma} dy, \\
&= \int_{-\infty}^{\infty} \frac{\mu^2 + 2\mu\sigma y + \sigma^2 y^2}{\sqrt{2\pi}} e^{-\frac{1}{2}(y)^2} dy, \\
&= \int_{-\infty}^{\infty} \frac{\mu^2}{\sqrt{2\pi}} e^{-\frac{1}{2}(y)^2} dy + \int_{-\infty}^{\infty} \frac{2\mu\sigma y}{\sqrt{2\pi}} e^{-\frac{1}{2}(y)^2} dy + \int_{-\infty}^{\infty} \frac{\sigma^2 y^2}{\sqrt{2\pi}} e^{-\frac{1}{2}(y)^2} dy, \\
&= \mu^2 \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y)^2} dy}_{=1} + 2\mu\sigma \underbrace{\int_{-\infty}^{\infty} \frac{y}{\sqrt{2\pi}} e^{-\frac{1}{2}(y)^2} dy}_{=0} + \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 e^{-\frac{1}{2}(y)^2} dy.
\end{aligned}
$$

A primeira integral da expressão de $E[X^2]$ é igual a 1 como provado pela expressão (62), uma distribuição normal com parâmetros $\mu = 0$ e $\sigma = 1$. A segunda integral foi mostrado pela Figura 3 que a integral se anula por simetria. Agora a terceira integral pode ser observado o gráfico na Figura 4 mostrando que é uma função par (simétrica no eixo y), $A = B$, ou seja,

$$\text{Área} = A + B = A + A = 2A,$$

e portanto,

$$\int_{-\infty}^{\infty} y^2 e^{-\frac{1}{2}y^2} dy = 2\int_{0}^{\infty} y^2 e^{-\frac{1}{2}y^2} dy.$$

Assim,

$$E[X^2] \;=\; \mu^2 + 2\int_0^\infty \frac{\sigma^2}{\sqrt{2\pi}} y^2 e^{-\frac{1}{2}(y)^2} dy,$$

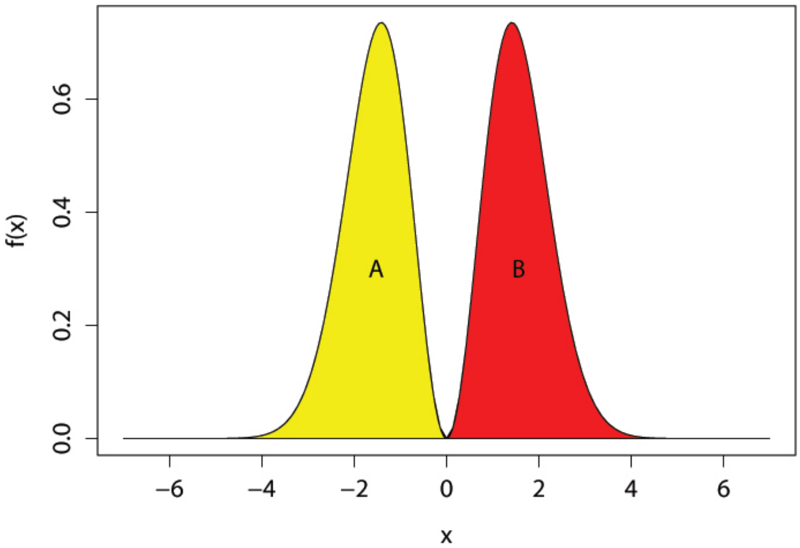


Figura 4: Gráfico da função $\displaystyle\int_{-\infty}^{\infty} y^2 e^{-\frac{1}{2}y^2} dy.$

fazendo uma mudança de variável $z = y^2/2 \Rightarrow y = \sqrt{z}$ e $dz = ydy \Rightarrow dy = dz/y$, temos

$$\begin{aligned}
E[X^2] &= \mu^2 + 2\sigma^2 \int_0^\infty \frac{(\sqrt{2z})^2}{\sqrt{2\pi}} e^{-\frac{1}{2}(\sqrt{2z})^2} \frac{dz}{\sqrt{dz}}, \\
E[X^2] &= \mu^2 + 2\sigma^2 \int_0^\infty \frac{2z}{\sqrt{2\pi}} e^{-z} \frac{dz}{\sqrt{2z}},
\end{aligned}$$

sabendo que $2 = \sqrt{2} \times \sqrt{2}$ e $z = \sqrt{z} \times \sqrt{z}$, então

$$\begin{aligned}
E[X^2] &= \mu^2 + 2\sigma^2 \int_0^\infty \frac{\sqrt{2} \times \sqrt{2} \times \sqrt{z} \times \sqrt{z}}{\sqrt{2} \times \sqrt{\pi}} e^{-z} \frac{dz}{\sqrt{2} \times \sqrt{z}}, \\
&= \mu^2 + 2\sigma^2 \int_0^\infty \frac{\sqrt{z}}{\sqrt{\pi}} e^{-z} dz, \\
&= \mu^2 + \frac{2\sigma^2}{\sqrt{\pi}} \int_0^\infty z^{1/2} e^{-z} dz,
\end{aligned}$$

A função gama é definida por:

$$\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}dx, \quad t > 0.$$

Reescrevendo $1/2 = 3/2 - 1$, então

$$\begin{aligned} E[X^2] &= \mu^2 + \frac{2\sigma^2}{\sqrt{\pi}} \underbrace{\int_0^\infty z^{3/2-1}e^{-z}dz}_{=\Gamma(3/2)}, \\ &= \mu^2 + \frac{2\sigma^2}{\sqrt{\pi}}\Gamma(3/2). \end{aligned}$$

Uma outra propriedade da função gama é $\Gamma(t+1) = t\Gamma(t)$, e um caso particular dessa propriedade é que $\Gamma(1/2) = 2\Gamma(3/2) = \sqrt{\pi}$. Assim,

$$\begin{aligned} E[X^2] &= \mu^2 + \frac{2\sigma^2}{\sqrt{\pi}}\left(\frac{\sqrt{\pi}}{2}\right), \\ &= \mu^2 + \sigma^2. \end{aligned}$$

Como a $Var(X) = E[X^2] - E^2[X]$, então

$$Var(X) = [\mu^2 + \sigma^2] - \mu^2 = \sigma^2.$$

O desvio padrão é dado por:

$$\sigma_X = \sqrt{\sigma^2} = \sigma. \tag{67}$$

□

Definição 31 (Distribuição Gama). *Uma variável aleatória X contínua, tem distribuição normal se sua função densidade de probabilidade é dada por*

$$f(x) = \begin{cases} \frac{\lambda}{\Gamma(r)}(\lambda x)^{r-1}e^{-\lambda x} & \text{, para} \quad 0 < x < \infty, \\ 0 & \text{, caso contrário,} \end{cases} \tag{68}$$

em que os parâmetros r e λ satisfazem $r > 0$ e $\lambda > 0$, e $\Gamma(.)$ é a função gama definida por:

$$\Gamma(r) = \int_0^\infty t^{r-1}e^{-t}dt, \qquad r > 0. \tag{69}$$

□

A função gama em (69), apresenta seguintes propriedades:

- $\Gamma(r+1) = r\Gamma(r)$ (Integral por partes);
  Observe que $\Gamma(r+1) = \int_0^\infty t^r e^{-t}dt$. Fazendo $u = t^r \Rightarrow du = rt^{r-1}dt$ e $dv = e^{-t}dx \Rightarrow v = -e^{-t}$, logo

$$\begin{aligned} \Gamma(r+1) &= (udv)|_0^\infty - \int_0^\infty vdu \\ &= \left(t^r e^{-t}\right)|_0^\infty + \int_0^\infty e^{-t}rt^{r-1}dt \\ &= 0 + r\Gamma(r) \\ &= r\Gamma(r). \end{aligned}$$

- $\Gamma(n) = (n-1)!$, para $n \in \mathbb{N}^+$;
- $\Gamma(1/2) = \sqrt{\pi}$.

-1 Para provar que a distribuição Gama é uma função de probabilidade, temos que demostrar os itens:

i) Mostrar que $f_X(x) \geq 0$;

ii) Mostrar que $\int_{-\infty}^{\infty} f_X(x)dx = 1$.

Demonstração. Para o item (i), como $X > 0$, os parâmetros λ e r são maiores que 0, logo $f_X(x) \geq 0$. Para o item (ii) temos que

$$\begin{aligned}\int_{-\infty}^{\infty} f_X(x)dx &= \int_0^{\infty} \frac{\lambda}{\Gamma(\lambda)}(\lambda x)^{r-1}e^{-\lambda x}dx \\ &= \int_0^{\infty} \frac{\lambda}{\Gamma(\lambda)} t^{r-1}e^{-t}\frac{1}{\lambda}dt, \qquad (t = \lambda x) \\ &= \frac{1}{\Gamma(r)}\int_0^{\infty} t^{r-1}e^{-t}dt \\ &= \frac{\Gamma(r)}{\Gamma(r)} = 1. \qquad \text{(Função gama, expressão (69))}\end{aligned}$$

□

Teorema 19 (Esperança, Variância e Desvio padrão da distribuição Gama)**.** *Se a variável aleatória X possui distribuição Gama expressa em* (68)*, então a esperança de X é dada por*

$$E[X] = \frac{r}{\lambda}, \tag{70}$$

a variância

$$Var[X] = \frac{r}{\lambda^2}. \tag{71}$$

e o desvio padrão

$$\sigma_X = \sqrt{\frac{r}{\lambda^2}}. \tag{72}$$

□

Demonstração. Para determinar a esperança temos

$$\begin{aligned}E[X] &= \int_0^{\infty} x\frac{\lambda}{\Gamma(r)}(\lambda x)^{r-1}e^{-\lambda x}dx \\ &= \int_0^{\infty} \frac{1}{\Gamma(r)}(\lambda x)^{r}e^{-\lambda x}dx \\ &= \frac{1}{\Gamma(r)}\int_0^{\infty} t^{r}e^{-t}\frac{1}{\lambda}dx, \qquad (t = \lambda x) \\ &= \frac{\Gamma(r+1)}{\lambda\Gamma(r)}, \qquad \text{(Função gama, expressão (69))} \\ &= \frac{r\Gamma(r)}{\lambda\Gamma(r)} = \frac{r}{\lambda}.\end{aligned}$$

Para a variância, temos

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

em que

$$\begin{aligned}
E[X^2] &= \int_0^\infty x^2 \frac{\lambda}{\Gamma(r)}(\lambda x)^{r-1} e^{-\lambda x} dx \\
&= \int_0^\infty \frac{\lambda}{\lambda} x^2 \frac{\lambda}{\Gamma(r)}(\lambda x)^{r-1} e^{-\lambda x} dx \\
&= \frac{1}{\lambda} \int_0^\infty \frac{1}{\Gamma(r)}(\lambda x)^{(r+2)-1} e^{-\lambda x} dx \\
&= \frac{1}{\lambda\Gamma(r)} \int_0^\infty t^{(r+2)-1} e^{-t} \frac{1}{\lambda} dx, \qquad (t = \lambda x) \\
&= \frac{\Gamma(r+2)}{\lambda^2\Gamma(r)}, \qquad \text{(Função gama, expressão (69))} \\
&= \frac{r(r+1)\Gamma(r)}{\lambda\Gamma(r)} = \frac{r(r+1)}{\lambda}.
\end{aligned}$$

Logo,

$$Var[X] = \frac{r(r+1)}{\lambda} - \frac{r^2}{\lambda^2} = \frac{r}{\lambda^2},$$

e o desvio padrão é de imediato a expressão (72). □

Se considerarmos na distribuição gama $r = \nu/2$ e $\lambda = 1/2$ resulta na distribuição qui-quadrado com ν graus de liberdade.

Definição 32 (Distribuição Qui-quadrado). *Uma variável aleatória contínua X tem distribuição qui-quadrado se sua função densidade de probabilidade é definida por:*

$$f(x) = \begin{cases} \frac{1}{\Gamma(\frac{\nu}{2})} \left(\frac{1}{2}\right)^{\frac{\nu}{2}} x^{\frac{\nu}{2}-1} e^{-\frac{1}{2}x}, & x > 0 \\ 0, & x \geq 0, \end{cases} \tag{73}$$

com ν graus de liberdade, sendo $\nu > 0$. □

Teorema 20 (Esperança, Variância e Desvio padrão da distribuição Qui-quadrado). *Se a variável aleatória X possui distribuição Qui-quadrado expressa em* (73), *então a esperança de X é dada por*

$$E[X] = \nu, \tag{74}$$

a variância

$$Var[X] = 2\nu. \tag{75}$$

e o desvio padrão

$$\sigma_X = \sqrt{2\nu}. \tag{76}$$

□

Demonstração. Basta usar os resultados do Teorema 19 para $r = \nu/2$ e $\lambda = 1/2$. □

5.2.1 Distribuição Qui-quadrado

Seja $Z_1, Z_2, \ldots, Z_\nu$, variáveis aleatórias independentes com distribuição normal padrão. Então podemos definir

$$\chi^2 = Z_1^2 + Z_2^2 + \ldots + Z_\nu^2 = \sum_{i=1}^{\nu} Z_i^2, \tag{77}$$

tem distribuição de qui-quadrado com ν graus de liberdade.

Uma variável aleatória X tem distribuição qui-quadrado se sua densidade for considerada

$$f(x) = \frac{1}{\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{1}{2}\right)^{\frac{\nu}{2}} x^{\frac{\nu}{2}-1} e^{-\frac{1}{2}x}, \quad x > 0. \tag{78}$$

A esperança é dada por:

$$E[X] = \nu, \tag{79}$$

e a variância,

$$Var[X] = 2\nu. \tag{80}$$

Considerando $S^2 = \sum\limits_{i=1}^{n}(X - \overline{X})^2/(n-1)$ a variância amostra e σ^2, a variância populacional, então

$$\chi = \frac{(n-1)S^2}{\sigma^2},$$

tem distribuição de qui-quadrado com $n-1$ graus de liberdade.

5.2.2 Distribuição t de Student

Seja Z uma variável aleatória com distribuição normal padrão $N(0,1)$, e U uma variável aleatória com distribuição de qui-quadrado com ν graus de liberdade. Considerando que essas duas variáveis são independentemente distribuídas, podemos então definir X sendo,

$$X = \frac{Z}{\sqrt{\frac{U}{\nu}}}, \tag{81}$$

tem distribuição t de Student com ν graus de liberdade.

A função densidade de X é dada por:

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\left(1 + \frac{\nu^2}{2}\right)^{\frac{\nu+1}{2}}}, \quad -\infty \leq x \leq \infty. \tag{82}$$

A esperança e a variância são respectivamente,

$$E[X] = 0 \quad \text{para } \nu > 1, \tag{83}$$

e

$$Var(X) = \frac{\nu}{\nu - 2} \quad \text{para } \nu > 2. \tag{84}$$

Agora, seja uma amostra $X_1, X_2, \ldots, X_n$ com distribuição normal com média μ e variância σ^2, sendo $\overline{X}$, a média amostral. Então definimos

$$Z = \frac{(\overline{X} - \mu)}{\frac{\sigma}{\sqrt{n}}}, \tag{85}$$

com distribuição aproximadamente normal padrão quando n tende ao infinito (TEOREMA DO LIMITE CENTRAL). E sendo $U = \frac{(n-1)S^2}{\sigma^2}$ com distribuição qui-quadrado com $n - 1$ graus de liberdade, então

$$\begin{aligned} T &= \frac{Z}{\sqrt{\frac{U}{n-1}}} = \frac{\frac{\frac{(\overline{X}-\mu)}{\sigma}}{\sqrt{n}}}{\sqrt{\frac{\frac{(n-1)S^2}{\sigma^2}}{(n-1)}}} = \frac{\frac{\frac{(\overline{X}-\mu)}{\sigma}}{\sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2}\frac{1}{(n-1)}}}, \\ &= \frac{\frac{\frac{(\overline{X}-\mu)}{\sigma}}{\sqrt{n}}}{\sqrt{\frac{S^2}{\sigma^2}}} = \frac{\frac{\frac{(\overline{X}-\mu)}{\sigma}}{\sqrt{n}}}{\frac{S}{\sigma}} = \frac{(\overline{X}-\mu)}{\frac{\sigma}{\sqrt{n}}}\frac{\sigma}{S} = \frac{(\overline{X}-\mu)}{\frac{S}{\sqrt{n}}}, \end{aligned}$$

tem distribuição t de Student com $n - 1$ graus de liberdade.

5.2.3 Distribuição F

Seja U uma variável aleatória com distribuição de qui-quadrado com m graus de liberdade e V uma variável aleatória com distribuição de qui-quadrado com n graus de liberdade, sendo U e V independentes, então a variável

$$X = \frac{U/m}{V/n}, \tag{86}$$

tem distribuição F com m e n graus de liberdade, cuja função densidade de probabilidade é dada por:

$$f_X(x) = \frac{\Gamma[(m+n)/2]}{\Gamma(m/2)\Gamma(n/2)} \times \left(\frac{m}{n}\right)^{m/2} \times \frac{x^{(m-2)/2}}{[1 + (m/n)x]^{(m+n)/2}} I_{(0,\infty)}(x), \qquad x > 0. \tag{87}$$

A esperança é dada por

$$E[X] = \frac{n}{n-2}, \tag{88}$$

e a variância por

$$Var[X] = \frac{2n^2[m+n-2]}{m(n-2)^2(n-4)}. \tag{89}$$

5.2.4 Relação entre variáveis aleatórias discretas e variáveis aleatórias contínuas

Ver Magalhães p.100; p. 105; p. 156-157