Sumários de Variávveis Aleatórias Discretas

Prof. Leandro Chaves Rêgo

Departamento de Estatística e Matemática Aplicada - UFC

Fortaleza, 10 de dezembro de 2021

Moda

A *moda* de uma distribuição de probabilidade discreta (também conhecida como moda da variável aleatória discreta) é o valor x no qual a função probabilidade de massa assume seu valor máximo. A moda não é necessariamente única e pode ter diversos valores todos sendo pontos de máximo da função de probabilidade de massa.

Exemplo

Considere uma variável aleatória X tal que: $P(X=-2)=0,25$, $P(X=0)=0,5$ e $P(X=2)=0,25$. Então,

$$Moda(X)=0.$$

■

Mediana

Para qualquer tipo de variável aleatória, sua *mediana* é definida como sendo qualquer número m que satisfaz as seguintes desigualdades:

$$P(X \leq m) \geq 1/2 \text{ e } P(X \geq m) \geq 1/2.$$

Então, uma mediana é um valor tal que a probabilidade da variável aleatória assumir um valor maior igual que ele é pelo menos $1/2$ e o mesmo ocorre com a probabilidade de ser menor ou igual que ele. Qualquer distribuição de probabilidade tem pelo menos uma mediana, mas esta pode não ser única.

Exemplo

Considere uma variável aleatória X tal que: $P(X = -2) = 0,2$, $P(X = 0) = 0,5$ e $P(X = 2) = 0,3$. Então, $Mediana(X) = 0$, pois $P(X \leq 0) = 0,75$ e $P(X \geq 0) = 0,75$. ■

Exemplo - Mediana

Exemplo

Considere uma variável aleatória X tal que: $P(X=-2)=0,25$, $P(X=1)=0,25$, $P(X=4)=0,25$, $P(X=5)=0,2$ e $P(X=6)=0,05$. Então, qualquer número no intervalo $[1,4]$ é uma mediana para X, pois para qualquer número $m \in [1,4]$, $P(X \leq m)=0,5$ e $P(X \geq m)=0,5$. Além disso, temos que -2, 1 e 4 são modas dessa distribuição. ■

Quantis

O quantil de uma distribuição é uma generalização do conceito de mediana. Para qualquer $0 < p < 1$, o quantil p de uma distribuição de probabilidade de uma variável aleatória é qualquer valor q tal que:

$$P(X \leq q) \geq p \text{ e } P(X \geq q) \geq 1 - p.$$

Deste modo, a mediana é o quantil $1/2$ de uma distribuição. Alguns quantis recebem nomes especiais: 1^o-quartil ($p = 1/4$), 2^o-quartil ou mediana ($p = 1/2$), 3^o-quartil ($p = 3/4$), 95^o-percentil ($p = 0{,}95$) e 5^o-percentil ($p = 0{,}05$).

Quantis - Exemplo

Exemplo

Considere novamente a variável aleatória X tal que: $P(X = -2) = 0{,}25$, $P(X = 1) = 0{,}25$, $P(X = 4) = 0{,}25$, $P(X = 5) = 0{,}2$ e $P(X = 6) = 0{,}05$.
Determine:

(a) o quantil 0,2.

(b) o quantil 0,25.

(c) o quantil 0,6.

Esperança - Motivação

O conceito de esperança matemática ou valor esperado de uma variável aleatória X, ou a "média" é tão antigo quanto o próprio conceito de probabilidade. Na verdade, é até possível definir probabilidade em termos de esperança, mas esta não é uma maneira comum de se apresentar a teoria, segundo Fine (2012). As seguintes podem ser interpretações da esperança:

(a) Parâmetro m de uma medida de probabilidade, função de distribuição, ou função probabilidade de massa, também conhecido como média.

(b) Operador linear em um conjunto de variáveis aleatórias que retorna um valor típico da variável aleatória interpretado como uma medida de localização da variável aleatória.

(c) Média do resultado de repetidos experimentos independentes no longo prazo.

(d) Preço justo de um jogo com pagamentos descritos por X.

Esperança - Motivação

A definição de esperança pode ser motivada considerando o cálculo do resultado médio de 1000 lançamentos de um dado. Uma maneira de calcular este resultado médio seria somar todos os resultados e dividir por 1000. Uma maneira alternativa seria calcular a fração $p(k)$, $k = 1, \ldots, 6$ de todos os lançamentos que tiveram resultado igual a k e calcular o resultado médio através da soma ponderada:

$$1p(1) + 2p(2) + 3p(3) + 4p(4) + 5p(5) + 6p(6).$$

Quando o número de lançamentos torna-se grande as frações de ocorrência dos resultados tendem à probabilidade de cada resultado.

Esperança - Definição - Caso Discreto

Em geral, define-se a esperança de uma variável discreta como uma soma ponderada onde as probabilidades são os pesos de ponderação.

Definição

Se X é uma variável aleatória discreta assumindo valores $\{x_1, x_2, x_3, \ldots\}$ com probabilidade $\{p_1, p_2, p_3, \ldots\}$, respectivamente, então sua esperança é dada pela fórmula

$$EX = \sum_{i:x_i<0} x_i p_i + \sum_{i:x_i\geq 0} x_i p_i,$$

desde que pelo menos um dos somatórios seja finito. Em caso os dois somatórios não sejam finitos, a esperança não existe. Caso EX seja finita, diz-se que X é integrável, e pode-se calculá-la pela fórmula:

$$EX = \sum_i x_i p_i.$$

Exemplo

Exemplo

Considere uma variável aleatória X tal que: $P(X = -2) = 0.25$, $P(X = 0) = 0.5$ e $P(X = 2) = 0.25$. Então,

$$E(X) = -2(0.25) + 0(0.5) + 2(0.25) = 0.$$

■

Exemplo

Seja uma variável aleatória X tal que: $P(X = -a) = P(X = a) = 1/3$ e $P(X = b) = 1/3$. Então,

$$E(X) = -a(1/3) + a(1/3) + b(1/3) = b/3.$$

■

Note então que muitas variáveis aleatórias diferentes podem ter o mesmo valor esperado ou esperança. (é só variar o valor de a no exemplo anterior.)

Exemplo

Exemplo

Se $X \in \{1, 2, \ldots, n\}$ for uma variável aleatória com distribuição de probabilidade aleatória com parâmetro n,

$$P(X = k) = \frac{1}{n}, \; k = 1, 2, \ldots, n,$$

sua esperança é dada por:

$$E(X) = \sum_{k=1}^{n} kp(k) = \sum_{k}^{n} k\frac{1}{n} = \frac{1}{n}\sum_{k}^{n} k = \frac{1}{n}\frac{n(n+1)}{2} = \frac{n+1}{2}.$$

■

Observação

Deve-se observar a analogia entre o valor esperado de uma variável aleatória e o conceito de centro de gravidade em Mecânica. Se um objeto tem massa distribuída sobre a reta, em pontos discretos, $x_1, x_2, \ldots$, e se $p(x_i)$ for a massa do ponto x_i, então vemos que $\sum_{i=1}^{\infty} x_i p(x_i)$ representa o centro de gravidade do objeto em relação a origem.

Esperança Matemática de Funções e Vetores de Variáveis Aleatórias

Se X for uma variável aleatória e se $Y = H(X)$, então Y também será uma variável aleatória. Consequentemente, pode-se calcular $E(Y)$. Existem duas maneiras equivalentes de calcular $E(Y)$, quer a variável seja discreta, quer seja contínua: (i) primeiro, encontrar a lei de probabilidade da variável $Y = H(X)$ pelos métodos já vistos anteriormente para, em seguida, calcular a esperança da variável Y; (ii) calcular a esperança de Y diretamente usando a função $H(X)$. Se $\vec{X}$ for um vetor aleatório e $Y = H(\vec{X})$ os comentários anteriores valem, isto é, pode-se calcular $E(Y)$ a partir do cálculo inicial de $f_Y(y)$ ou calcular $E(Y)$ usando $H(\vec{X})$.

Caso Discreto

Definição

Seja X uma variável aleatória discreta e seja $Y = H(X)$. Se Y assumir os seguintes valores $y_1, y_2, \ldots$ e se $P(Y = y_i) = p(y_i)$, define-se:

$$E(Y) = \sum_{i=1}^{\infty} y_i p(y_i).$$

Exemplo

Exemplo

Suponha X tal que $P(X=k)=\binom{2}{k}(\frac{1}{2})^k(\frac{1}{2})^{2-k}$, para $k=0,1,2$, e $Y=H(X)=X+1$. Calcular $E(Y)$.

Solução. O contradomínio de Y é $R_Y=\{1,2,3\}$ porque $X=0\Rightarrow Y=1$, $X=1\Rightarrow Y=2$ e $X=2\Rightarrow Y=3$. Portanto, $P(Y=1)=P(X=0)=\frac{1}{4}$, $P(Y=2)=P(X=1)=\frac{2}{4}$ e $P(Y=3)=P(X=2)=\frac{1}{4}$. Assim,

$$E(Y)=1\times\frac{1}{4}+2\times\frac{2}{4}+3\times\frac{1}{4}=2.$$

$\blacksquare$

Método Alternativo

Conforme visto no capítulo anterior pode-se determinar as probabilidades $p(y_i)$ dado que sabe-se a distribuição de X. No entanto, é possível encontrar $E(Y)$ sem, preliminarmente, encontrar a distribuição de probabilidade de Y, partindo-se apenas do conhecimento da distribuição de probabilidade de X, conforme mostra o seguinte teorema.

Teorema

Seja X uma variável aleatória discreta assumindo os valores $x_1, x_2, \ldots$ e seja $Y = H(X)$. Se $p(x_i) = P(X = x_i)$, então

$$E(Y) = E(H(X)) = \sum_{i=1}^{\infty} H(x_i)p(x_i).$$

Prova

Prova: Reordenando o somatório $\sum_{i=1}^{\infty} H(x_i)p(x_i)$, e agrupando os termos onde x_i tem a mesma imagem de acordo com a função H, ou seja, sejam $x_{i1}, x_{i2}, \ldots$, todos os valores x_i tal que $H(x_{ij}) = y_i$ para $j \geq 1$, onde $y_1, y_2, \ldots$ são os possíveis valores de Y, tem-se,

$$\sum_{i=1}^{\infty} H(x_i)p(x_i) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} H(x_{ij})p(x_{ij}) = \sum_{i=1}^{\infty} y_i \sum_{j=1}^{\infty} p(x_{ij}) = \sum_{i=1}^{\infty} y_i p(y_i) = E(Y).$$

$\blacksquare$

Exemplo

Exemplo

Sejam X e Y como no exemplo anterior. Calculando $E(Y)$ sem encontrar a distribuição de probabilidade de Y.

Solução.

$$
\begin{aligned}
E(Y) &= E(X+1) = \sum_{k=0}^{2}(k+1)P(X=k) \\
&= \sum_{k=0}^{2} kP(X=k) + \sum_{k=0}^{2} P(X=k) = 1+1=2.
\end{aligned}
$$

Lembrando que $\sum_{k=0}^{2} P(X=k) = 1$.

Definição

Definição

Se $\vec{X}$ um vetor aleatório com distribuição de probabilidade $p_{\vec{X}}$ e $Y = H(\vec{X})$ assume os valores $y_1, y_2, \dots,$ de forma que $P(Y = y_i) = p(y_i)$, então,

$$E(Y) = E(H(\vec{X})) = \sum_i y_i p_{\vec{Y}}(y_i).$$

Exemplo

Exemplo

Seja $\vec{X} = (X_1, X_2)$ com $R_{\vec{X}} = \{(1,1),(1,2),(2,1),(2,2)\}$, $P(X_1 = i, X_2 = j) = \frac{1}{4}$, para $i, j = 1, 2$ e $Y = H(\vec{X}) = X_1 + X_2$. Calcular $E(Y) = E(H(\vec{X}))$.

Exemplo

Solução. $R_Y = \{2, 3, 4\}$ e $P(Y = 2) = P(X_1 = 1, X_2 = 1) = \frac{1}{4}$,
$P(Y = 3) = P(X_1 = 1, X_2 = 2) + P(X_1 = 2, X_2 = 1) = \frac{2}{4}$ e
$P(Y = 4) = P(X_1 = 2, X_2 = 2) = \frac{1}{4}$. Logo,

$$E(Y) = 2 \times \frac{1}{4} + 3 \times \frac{2}{4} + 4 \times \frac{1}{4} = 3.$$

■

Método Alternativo

Teorema

Seja $\vec{X}$ um vetor aleatório com distribuição de probabilidade $p_{\vec{X}}(\vec{x}_i) = P(X_1 = x_{1i}, X_2 = x_{2j}, \ldots, X_n = x_{nk})$ e $Y = H(\vec{X})$. Então,

$$E(Y) = E(H(\vec{X})) = \sum_n \ldots \sum_j \sum_i H(\vec{x}_i) p_{\vec{X}}(\vec{x}_i),$$

em que os $\vec{x}_i$ são os valores assumidos pelo vetor aleatório $\vec{X}$.

Exemplo

Exemplo

Suponha $\vec{X}$ como no Exemplo anterior.

Solução. Então, o cálculo de $E(Y)$ é dado por

$$
\begin{aligned}
E(Y) &= E(X_1 + X_2) = \sum_j \sum_i (i+j)P(X_1 = i, X_2 = j) \\
&= \sum_j \sum_i (iP(X_1 = i, X_2 = j) + jP(X_1 = i, X_2 = j)) \\
&= \sum_j \sum_i (iP(X_1 = i, X_2 = j)) + \sum_j \sum_i (jP(X_1 = i, X_2 = j) \\
&= \sum_j (\sum_i (iP(X_1 = i, X_2 = j))) + \sum_j (\sum_i (jP(X_1 = i, X_2 = j))
\end{aligned}
$$

Exemplo

$$
\begin{aligned}
EY &= \sum_j (P(X_1 = 1, X_2 = j) + 2P(X_1 = 2, X_2 = j)) \\
&\quad + \sum_j (jP(X_1 = 1, X_2 = j) + jP(X_1 = 2, X_2 = j)) \\
&= P(X_1 = 1, X_2 = 1) + P(X_1 = 1, X_2 = 2) + 2(P(X_1 = 2, X_2 = 1) \\
&\quad + P(X_1 = 2, X_2 = 2)) + P(X_1 = 1, X_2 = 1) + P(X_1 = 2, X_2 = 1) + \\
&\quad 2(P(X_1 = 2, X_2 = 2) + P(X_1 = 2, X_2 = 2)) = 3.
\end{aligned}
$$

■

Propriedades

As seguintes propriedades são aplicaçẽs imediatas da definição de esperança:

(i) $P(X = c) = 1 \Rightarrow E(X) = c$.

(ii) $P(X \geq 0) = 1 \Rightarrow E(X) \geq 0$.

(iii) $E(aX) = aE(X)$, onde a um número real qualquer.

A propriedade (iii) segue facilmente da expressão da esperança de uma função de variável aleatória.

Propriedades

(iv) $E(X+Y) = E(X) + E(Y)$.

No caso discreto,

$$
\begin{aligned}
E(X+Y) &= \sum_i \sum_j (x_i + y_j) p(x_i, y_j) = \sum_i \sum_j x_i p(x_i, y_j) + \sum_i \sum_j y_j p(x_i, y_j) \\
&= \sum_i x_i \sum_j p(x_i, y_j) + \sum_j y_j \sum_i p(x_i, y_j) \qquad (1) \\
&= \sum_i y_i p(x_i) + \sum_j y_j p(y_j) = E(X) + E(Y).
\end{aligned}
$$

Propriedades

(v) $E(\sum_i^n a_i X_i) = \sum_i^n a_i E(X_i)$.
Para provar esta propriedade basta usar as duas últimas propriedades e indução matemática.

(vi) $P(X \geq Y) = 1 \Rightarrow E(X) \geq E(Y)$.
Esta segue das Propriedades (ii) e (v), pois

$$P(X \geq Y) = P(X - Y \geq 0),$$

o que, pela Propriedade (ii), implica que $E(X - Y) \geq 0$. Pela Propriedade (v), $E(X - Y) = E(X) - E(Y)$, ou seja pode-se concluir que $E(X) - E(Y) \geq 0$.

Propriedades

(vii) Se $\{X_1, \ldots, X_n\}$ são variáveis aleatórias mutuamente independentes, então

$$E(\prod_{i=1}^{n} X_i) = \prod_{i=1}^{n} E(X_i).$$

No caso discreto,

$$\begin{aligned} E(\prod_{i=1}^{n} X_i) &= \sum_{i_1} \cdots \sum_{i_n} x_{i_1} \ldots x_{i_n} p(x_{i_1}, \ldots, x_{i_n}) \\ &= \sum_{i_1} \cdots \sum_{i_n} x_{i_1} \ldots x_{i_n} \prod_{j=1}^{n} p(x_{i_j}) \\ &= \sum_{i_1} x_{i_1} p(x_{i_1}) \ldots \sum_{i_n} x_{i_n} p(x_{i_n}) \\ &= \prod_{i=1}^{n} E(X_i). \end{aligned}$$

Propriedades

De maneira análoga, pode-se provar a seguinte generalização deste resultado: Se $\{X_1, \ldots, X_n\}$ são variáveis aleatórias mutuamente independentes, então

$$E(\prod_{i=1}^{n} G(X_i)) = \prod_{i=1}^{n} E(G(X_i)).$$

Propriedades

(viii) Se Y for uma variável aleatória que assume valores inteiros não-negativos, então

$$E(Y) = \sum_{k=1}^{\infty} kP(Y = k) = \sum_{k=1}^{\infty} \sum_{j=1}^{k} P(Y = k),$$

trocando a ordem dos somatórios:

$$E(Y) = \sum_{j=1}^{\infty} \sum_{k=j}^{\infty} P(Y = k) = \sum_{j=1}^{\infty} P(Y \geq j).$$

Propriedades

(ix) **Desigualdade de Jensen.** Seja φ uma função mensurável e convexa definida na reta. Se X é integrável, então $E(\varphi(X)) \geq \varphi(E(X))$.

Prova: Pela convexidade de φ, dado algum ponto $(x_0, \varphi(x_0)$ do gráfico de φ, existe uma reta que passa por esse ponto e fica sempre abaixo do gráfico de φ, ou seja, existe algum λ tal que

$$\varphi(x) \geq \varphi(x_0) + \lambda(x - x_0), \forall x.$$

Logo, pela monotonicidade e linearidade da esperança,

$$E\varphi(X) \geq \varphi(x_0) + \lambda(E(X) - x_0).$$

Em particular, para $x_0 = EX$, tem-se $E\varphi(X) \geq \varphi(E(X))$. ∎

Esperança Condicional

O conceito da esperança condicional de X dado $Y = y$, $E(X \mid Y = y)$, deve ser entendido como uma média da variável X quando um valor de Y está fixado.

Definição

Se (X, Y) for um vetor aleatório discreto, então a esperança condicional de X dado $Y = y_j$ é

$$E(X \mid Y = y_j) = \sum_i x_i P(X = x_i \mid Y = y_j),$$

desde que a série seja absolutamente convergente e onde y_j é um valor fixo no contradomínio de Y.

Exemplo

Exemplo

Considere que a distribuição de probabilidade conjunta do vetor aleatório (X, Y) é dada por

$$p(1,1) = \frac{1}{9}, \quad p(2,1) = \tfrac{1}{3}, \quad p(3,1) = \frac{1}{9},$$
$$p(1,2) = \frac{1}{9}, \quad p(2,2) = 0, \quad p(3,2) = \frac{1}{18},$$
$$p(1,3) = 0, \quad p(2,3) = \tfrac{1}{6}, \quad p(3,3) = \frac{1}{9}.$$

Calcular $E(X \mid Y = j)$, para $j = 1, 2, 3$.

Exemplo

Solução. Calcula-se a marginal de Y, em seguida as condicionais e por fim a esperança pedida.

$$P(Y=1)=\frac{5}{9},\ P(Y=2)=\frac{3}{18},\ P(Y=3)=\frac{5}{18}.$$

$$\begin{aligned}
&P(X=1\mid Y=1)=\frac{1}{5}, \quad P(X=1\mid Y=2)=\tfrac{2}{3}, \quad P(X=1\mid Y=3)=0,\\
&P(X=2\mid Y=1)=\frac{3}{5}, \quad P(X=2\mid Y=2)=0, \quad P(X=2\mid Y=3)=\frac{3}{5},\\
&P(X=3\mid Y=1)=\frac{1}{5}, \quad P(X=3\mid Y=2)=\tfrac{1}{3}, \quad P(X=3\mid Y=3)=\frac{2}{5}.
\end{aligned}$$

Logo,

$$E(X\mid Y=1)=\sum_{i=1}^{3}x_iP(X=x_i\mid Y=1)=1\times\frac{1}{5}+2\times\frac{3}{5}+3\times\frac{1}{5}=2.$$

De forma similar são calculadas $E(X\mid Y=2)$ e $E(X\mid Y=3)$. ■

Momentos

Informações parciais sobre a lei de probabilidade de uma variável aleatória X podem ser obtidas através de esperanças matemáticas de potências de X, as quais são chamadas de *momentos de* X.

Definição

Para qualquer inteiro não-negativo n, o n-ésimo momento da variável aleatória X é

$$E(X^n),$$

se esta esperança existe.

Este momento é usualmente denominado de momento em torno do zero, uma vez que poderia ser escrito como $E((X-0)^n)$. Observe que quando $n=1$ tem-se a esperança matemática de X.

Exemplo

Exemplo

Seja X tal que $P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$, $k=0,1,\ldots,n$. Então, o segundo momento de X, $E(X^2)$ é:

$$
\begin{aligned}
E(X^2) &= \sum_{k=0}^{n} k^2 \binom{n}{k} p^k(1-p)^{n-k} \\
&= \sum_{k=1}^{n} k^2 \frac{n!}{k!(n-k)!} p^k(1-p)^{n-k} \\
&= \sum_{k=1}^{n} k(k-1) \frac{n!}{k!(n-k)!} p^k(1-p)^{n-k} + \sum_{k=1}^{n} k \frac{n!}{k!(n-k)!} p^k(1-p)^{n-k} \\
&= n(n-1)p^2 \sum_{k=2}^{n} \frac{(n-2)!}{(k-2)!(n-k)!} p^{k-2}(1-p)^{n-k} + np \\
&= n(n-1)p^2 \sum_{j=0}^{m} \frac{(m)!}{(j)!(m-j)!} p^j(1-p)^{m-j} + np = n(n-1)p^2 + np.
\end{aligned}
$$

■

Relaçao entre ordens diferentes

Teorema

Se o k-ésimo momento de uma variável aleatória for finito, então todos os momentos de ordem menores do que k também serão finitos.

Prova: Por hipótese, $E(|X^k|) < \infty$, logo $E(1 + |X^k|) < \infty$. Como para qualquer j tal que $0 < j < k$, $|X^j| \leq 1 + |X^k|$, e $1 + |X^k|$ é integrável, tem-se que $|X^j|$ também é integrável, isto é $E(|X^j|) < \infty$. ∎

Momentos Centrais

Definição

Se X é uma variável aleatória seu n-ésimo momento central em torno de $E(X)$ é

$$E(X - E(X))^n,$$

se esta esperança existir.

O primeiro momento central em torno da média é zero, pois

$$E(X - E(X)) = E(X) - E(E(X)) = E(X) - E(X) = 0.$$

Variância

Definição

O segundo momento central é a *variância*, $V(X)$, e é dado por

$$V(X) = E(X - E(X))^2.$$

A variância também pode ser calculada por:

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

Exemplo

Exemplo

Considerando a variável do Exemplo **??**, tem-se que

$$V(X) = E(X^2) - E(X)^2 = \frac{1}{2} - (\frac{2}{3})^2 = \frac{1}{6}.$$

■

Momentos Centrais e Não-Centrais

Do Teorema Binomial tem-se:

$$(X - E(X))^n = \sum_{k=0}^{n} \binom{n}{k} X^k (-E(X))^{n-k}.$$

Logo, da linearidade da esperança,

$$E(X - E(X))^n = \sum_{k=0}^{n} \binom{n}{k} E(X^k)(-E(X))^{n-k}$$

porque $E((-E(X))^{n-k}) = (-E(X))^{n-k}$, pois tem-se a esperança de uma constante e

$$E(X^n) = E(X - E(X) + E(X))^n = \sum_{k=0}^{n} \binom{n}{k} (E(X))^{n-k} E(X - E(X))^k.$$

Portanto, o n-ésimo momento é finito se e somente se o n-ésimo momento central for finito.

Exemplo

Exemplo

Considere uma variável aleatória X tal que

$$P(X = m - a) = P(X = m + a) = \frac{1}{2} \Rightarrow E(X^k) = \frac{1}{2}[(m-a)^k + (m+a)^k].$$

$$E(X) = m,$$

$$E(X^2) = \frac{1}{2}(2m^2 + 2a^2) = m^2 + a^2,$$

$$V(X) = a^2.$$

■

Desvio-Padrão

Definição

O *desvio-padrão* σ de uma variável aleatória X é definido como a raiz quadrada positiva da variância,

$$\sigma(X) = \sqrt{V(X)}.$$

A variância mede a distância entre os valores que a variável aleatória assume e sua esperança matemática. Entretanto, quando aplicada em problemas práticos surge uma dificuldade: a unidade de medida da esperança matemática difere da unidade de medida da variância. Por exemplo, se os dados estão em milissegundos, ms, a unidade de medida da média é ms, mas a da variância é ms^2. Por esta razão é comum adotar-se o desvio-padrão.

Exemplo

Exemplo

Suponha que a variável aleatória X assume os valores 0 e 10 com iguais probabilidade. Portanto,

$$E(X) = 5,\ E(X^2) = 50,\ V(X) = 25,\ \sigma = 5.$$

Suponha agora uma outra variável aleatória, Y, assumindo os valores 5 e 5 também com iguais probabilidade. Logo,

$$E(Y) = 5,\ E(X^2) = 25,\ V(X) = 0,\ \sigma = 0.$$

Este exemplo trivial enfatiza a importância de se usar o desvio-padrão. Observe que as médias são as mesmas mesmo para valores tão distintos que as variáveis assumem. ∎

Propriedades da Variância

(i) $V(X) \geq 0$.

Prova: Pela definição de variância. $\blacksquare$

(ii) Se $X = c$, $V(X) = 0$.

Prova: $E(X) = c$, logo $V(X) = E(X - c)^2 = E(0) = 0$. $\blacksquare$

(iii) $V(X + a) = V(X)$, onde a é uma constante real.

Prova:

$$\begin{aligned} V(X+a) &= E(X+a)^2 - (E(X+a))^2 \\ &= E(X^2) + 2aE(X) + a^2 - (E(X))^2 - 2aE(X) - a^2 \\ &= E(X^2) - (E(X))^2 = V(X). \end{aligned}$$

$\blacksquare$

Propriedades da Variância

(iv) $V(aX) = a^2 V(X)$

Prova:

$$V(aX) = E(aX)^2 - (E(aX))^2 = a^2 E(X)^2 - a^2 (EX)^2 = a^2 V(X).$$

∎

(v) Se X e Y forem variáveis aleatórias mutuamente independentes, então

$$V(X+Y) = V(X) + V(Y).$$

Prova:

$$\begin{aligned}
V(X+Y) &= E(X+Y)^2 - (E(X+Y))^2 \\
&= E(X^2 + 2XY + Y^2) - (E(X))^2 - 2E(X)E(Y) - (EY)^2 \\
&= E(X^2) - E(X)^2 + E(Y^2) - E(Y)^2 + 2(E(XY) - E(X)E(Y)) \\
&= E(X^2) + E(Y^2) - (E(X))^2 - (E(Y))^2 + 2E(XY) - 2E(X)E(Y) \\
&= V(X) + V(Y).
\end{aligned}$$

porque $E(XY) = E(X)E(Y)$. ∎

Propriedades da Variância

(vi) $V(X) = E(X-\mu)^2 = \min_{c\in\boldsymbol{R}} E(X-c)^2$.

Prova:

$$(X-c)^2 = (X-\mu+\mu-c)^2 = (X-\mu)^2+2(\mu-c)(X-\mu)+(\mu-c)^2,$$

logo

$$E(X-c)^2 = E(X-\mu)^2+2(\mu-c)(E(X)-\mu)+(\mu-c)^2 = V(X)+(\mu-c)^2.$$

Portanto, $E(X-c)^2 \geq E(X-\mu)^2, \forall c \in \boldsymbol{R}$. ∎

(vii) Se $X_1, \ldots, X_n$ são variáveis aleatórias independentes, então

$$V(X_1 + \ldots X_n) = V(X_1) + \ldots + V(X_n).$$

Esta propriedade segue da propriedade anterior e da aplicação de indução matemática.

A Desigualdade de Tchebychev

Teorema

Desigualdade de Chebyshev Generalizada. Dado um conjunto A e uma função $g(x)$ tal que $\forall x\ g(x) \geq I_A(x)$, tem-se que $P(X \in A) \leq \min(1, Eg(X))$.

Prova: Pela monotonicidade da Esperança, temos que $Eg(X) \geq EI_A(X) = P(X \in A)$. Mas, como a cota superior pode exceder 1, temos que $\min(1, Eg(X)) \geq P(X \in A)$. ∎

A Desigualdade de Tchebychev

Corolário

Seja X uma variável aleatória, então para todo $\epsilon > 0$, $P(|X| \geq \epsilon) \leq \frac{E|X|}{\epsilon}$.

Prova: Escolha $A = \{x : |x| \geq \epsilon\}$ e $g(x) = \frac{|x|}{\epsilon}$. Note que $g(x) \geq I_A(x)$, então $P(|X| \geq \epsilon) \leq \frac{E|X|}{\epsilon}$. ∎

Corolário

Se $Z \geq 0$ e $EZ = 0$, então $P(Z = 0) = 1$.

Prova: $P(Z \geq \frac{1}{n}) \leq nEZ = 0$. Como $[Z > 0] = \cup_n [Z \geq \frac{1}{n}]$, temos que

$$P(Z > 0) = P(\cup_n [Z \geq \frac{1}{n}]) \leq \sum_n P(Z \geq \frac{1}{n}) = 0.$$

Portanto, $P(Z = 0) = 1 - P(Z > 0) = 1$. ∎

Note que este último corolário implica que, quando $Var(X) = 0$, ou seja $E(X - EX)^2 = 0$, temos que $P(X = EX) = 1$, ou seja X é constante com probabilidade 1.

A Desigualdade de Tchebychev

Corolário

Desigualdade (Original) de Chebyshev. *Seja X uma variável aleatória, então* $P(|X-EX|\geq\epsilon)\leq\frac{VarX}{\epsilon^2}$.

Prova: Escolha $A=\{x:|x|\geq\epsilon\}$ e $g(x)=\frac{x^2}{\epsilon^2}$. Note que $g(x)\geq I_A(x)$, então pelo teorema anterior, $P(X\in A)=P(|X|\geq\epsilon)\leq\frac{EX^2}{\epsilon^2}$. Substituindo X por $X-EX$, temos $P(|X-EX|\geq\epsilon)\leq\frac{VarX}{\epsilon^2}$. ∎

Esta desigualdade declara que a probabilidade da variável aleatória diferir da sua média por mais do que uma constante qualquer (ε) é menor ou igual do que $\frac{\sigma^2}{\varepsilon^2}$. Portanto, quanto menor a variância, mais agrupados em torno da média estão os dados e, consequentemente, maior a probabilidade de se obter um valor (dos dados) próximo à média.
A desigualdade de Tchebychev é geral no sentido de que não há qualquer hipótese sobre a lei de probabilidade de X.

Exemplo

Exemplo

Sabe-se que o número de itens produzidos por uma fábrica durante uma semana é uma variável aleaória com média 500 e variância 100. O que pode ser dito sobre a probabilidade de que a produção semanal esteja entre 400 e 600?

Exemplo

Solução. Pela desegualdade de Tchebychev,

$$P(\mid X - 500 \mid < 100) \geq 1 - \frac{1}{100} = \frac{99}{100}.$$

Portanto, a probabilidade pedida é de pelo menos 0.99. ■

Momentos Conjuntos

A noção de momentos conjuntos é definida no contexto de vetores aleatórios.

Definição

Seja $\vec{X} = (X_1, X_2, \ldots, X_k)$ um vetor aleatório k-dimensional. Então, os *momentos conjuntos* de $\vec{X}$ são da forma $E(\prod_{i=1}^{k} X_i^{j_i})$, onde j_i's são inteiros positivos, se esta esperança existir.

De forma análoga ao caso unidimensional pode-se definir também *momentos conjuntos centrais*.

Correlação e Covariância

No caso bidimensional a *correlação* e a *covariância* são momentos conjuntos; estes medem o grau de dependência linear entre duas variáveis.

Definição

A *correlação* entre duas variáveis aleatórias X e Y é dada por $E(XY)$.

Definição

A *covariância* entre duas variáveis aleatórias X e Y é dada por

$$cov(X, Y) = E((X - E(X))(Y - E(Y))) = E(XY) - E(X)E(Y).$$

Variância da Soma

Note que $cov(X, X) = V(X)$. Na prova da Propriedade (v) da variância aparece a expressão $E(XY) - E(X)E(Y)$, o que implica que em geral tem-se que,

$$V(X + Y) = V(X) + V(Y) + 2cov(X, Y).$$

Veremos agora como obter a variância da soma de n variáveis aleatórias.

Teorema

Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias tais que $V(X_i) < \infty$, então

$$V(X_1 + \ldots + X_n) = \sum_{i=1}^{n} V(X_i) + 2\sum_{i<j} cov(X_i, X_j).$$

Variância da Soma

Prova:

$$\begin{aligned} V(X_1+\cdots+X_n) &= E(X_1+\cdots+X_n - E(X_1+\cdots+X_n))^2 \\ &= E(\sum_{i=1}^{n}(X_i - E(X_i))^2 \\ &= E(\sum_{i=1}^{n}(X_i - E(X_i))^2 + 2\sum_{i<j}(X_i - E(X_i))(X_j - E(X_j))) \\ &= \sum_{i=1}^{n} V(X_i) + 2\sum_{i<j} cov(X_i, X_j). \end{aligned}$$

∎

Corolário

Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias tais que $V(X_i) < \infty$ e $Cov(X_i, X_j) = 0$ para $i \neq j$, então

$$V(X_1+\ldots+X_n) = \sum_{i=1}^{n} V(X_i).$$

Desigualdade de Cauchy-Schwarz

Teorema

Desigualdade de Cauchy-Schwarz

$$(E(XY))^2 \leq E(X^2)E(Y^2).$$

Prova:
$(aX + Y)^2 \geq 0 \Rightarrow E(aX + Y)^2 \geq 0 \Rightarrow a^2E(X^2) + 2aE(XY) + E(Y^2) \geq 0.$
Observa-se que esta equação do segundo grau em a não pode ter duas raízes reais diferentes, pois caso contrário essa expressão seria negativa para os valores entre as raízes. Então, utilizando a regra do discriminante,

$$4(EXY)^2 - 4EX^2EY^2 \leq 0,$$

o teorema está provado. ∎

Corolário

$(cov(X, Y))^2 \leq V(X)V(Y).$

Prova: Segue do teorema anterior trocando X por $X - E(X)$ e Y por $Y - E(Y)$. ∎

Coeficiente de Correlação

Definição

O *coeficiente de correlação* entre duas variáveis aleatórias X e Y é dado por

$$\rho(X,Y) = \frac{cov(X,Y)}{\sqrt{V(X)V(Y)}}.$$

Definição

Duas variáveis são *não-correlacionadas* se $cov(X,Y) = 0$.

Correlação e Independência

Como já foi provado que se X e Y são independentes, então $E(XY) = E(X)E(Y)$, se X e Y são independentes, elas necessariamente são não-correlacionadas. O contrário nem sempre é verdadeiro como o próximo exemplo ilustra.

Exemplo

Se X é uma variável aleatória tal que $P(X = -a) = P(X = a) = 1/3$, $P(X = 0) = 1/3$ e $Y = X^2$, então $E(XY) = -a^3(1/3) + a^3(1/3) + 0^3(1/3) = 0$ e $E(X) = -a(1/3) + a(1/3) + 0(1/3) = 0$. Logo, $E(XY) = E(X)E(Y) = 0$, ou seja, $Cov(X, Y) = 0$. Porém, X e Y não são independentes, pois

$$P(X = a, Y = 0) = 0 \neq P(X = a)P(Y = 0) = (1/3)^2.$$

■

Normalização do Coeficiente de Correlação

Corolário

$|\rho(X, Y)| \leq 1$.

Prova: Imediata a partir do (7.7). ∎

O próximo teorema mostra que o módulo do coeficiente de correlação entre duas variáveis é igual a 1 se, e somente se, as variáveis são linearmente dependentes com probabilidade 1.

Teorema

Sejam X e Y variáveis aleatórias com variâncias finitas e positivas. Então,

(i) $\rho(X, Y) = 1$ *se, e somente se,* $P(Y = aX + b) = 1$ *para algum* $a > 0$ *e* $b \in \mathbf{R}$.

(ii) $\rho(X, Y) = -1$ *se, e somente se,* $P(Y = aX + b) = 1$ *para algum* $a < 0$ *e* $b \in \mathbf{R}$.

Demonstração

Prova: Para a parte (i), como $(\frac{X-E(X)}{\sqrt{V(X)}} - \frac{Y-E(Y)}{\sqrt{V(Y)}})^2 \geq 0$, então,

$$\begin{aligned} 0 &\leq E(\frac{X-E(X)}{\sqrt{V(X)}} - \frac{Y-E(Y)}{\sqrt{V(Y)}})^2 \\ &= E(\frac{X-E(X)}{\sqrt{V(X)}})^2 + E(\frac{Y-E(Y)}{\sqrt{V(Y)}})^2 - \frac{2}{\sqrt{V(X)V(Y)}}E((X-E(X))(Y-E(Y))) \\ &= \frac{V(X)}{V(X)} + \frac{V(Y)}{V(Y)} - \frac{2Cov(X,Y)}{\sqrt{V(X)V(Y)}} = 2 - 2\rho(X,Y). \end{aligned}$$

Demonstração

Se $\rho(X, Y) = 1$, então

$$E(\frac{X - E(X)}{\sqrt{V(X)}} - \frac{Y - E(Y)}{\sqrt{V(Y)}})^2 = 0,$$

o que por sua vez implica que

$$P(\frac{X - E(X)}{\sqrt{V(X)}} = \frac{Y - E(Y)}{\sqrt{V(Y)}}) = 1,$$

em outras palavras,

$$P(Y = E(Y) + \frac{\sqrt{V(Y)}}{\sqrt{V(X)}}(X - E(X))) = 1.$$

A prova da parte (ii) é análoga, substituindo o sinal "+" por "-" na expressão acima.

∎