Probabilidade (PPGECD000000001)

Programa de Pós-Graduação em Estatística e Ciência de Dados (PGECD)

Sessão 7

Raydonal Ospina

Departamento de Estatística
Universidade Federal da Bahia
Salvador/BA

Esperança

Interpretação

1. Parâmetro μ de uma medida de probabilidade, função de distribuição, ou função probabilidade de massa, também conhecido como média.
2. Um 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.
3. média do resultado de repetidos experimentos independentes no longo prazo.
4. preço justo de um jogo com pagamentos descritos por X.

Esperança - Caso Discreto

Considere que um dado é lançado 1000 vezes. Uma maneira de calcular este resultado médio seria somar todos os resultados e dividir por 1000. Uma maneira alternativa seria calcular a fração $p(k)$ 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 se torna grande as frações de ocorrência dos resultados tendem a probabilidade de cada resultado.

Definição 1

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.

Exemplos

Exemplo 1 (Aleatória)

Se $X \in \{1, 2, \ldots, n\}$ for uma variável aleatória com distribuição de probabilidade aleatória com parâmetro n, temos que sua esperança é dada por:

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

Onde utilizamos a fórmula da soma dos primeiros n termos de uma progressão aritmética.

Exemplo 2

Bernoulli
Se $X \in \{0, 1\}$ for uma variável aleatória com distribuição de probabilidade Bernoulli com parâmetro p, temos que sua esperança é dada por:

$$EX = 0(1-p) + 1(p) = p.$$

Exemplo 3 (Binomial)

Se X for uma variável aleatória com distribuição de probabilidade Binomial com parâmetros n e p, temos que sua esperança é dada por:

$$EX = \sum_{k=0}^{n} k\binom{n}{k}p^k(1-p)^{n-k} = \sum_{k=1}^{n} k\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}$$

$$\sum_{k=1}^{n} n\frac{(n-1)!}{(k-1)!(n-k)!}p^k(1-p)^{n-k} = np\sum_{k=1}^{n}\binom{n-1}{k-1}p^{k-1}(1-p)^{n-k} = np.$$

Onde utilizamos o Teorema Binomial na última igualdade.

Exemplo 4 (Geométrica)

Se X for uma variável aleatória com distribuição de probabilidade Geométrica com parâmetro β, temos que sua esperança é dada por:

$$EX = \sum_{k=0}^{\infty} k(1-\beta)\beta^k = \sum_{k=1}^{\infty} k(1-\beta)\beta^k = \sum_{k=1}^{\infty}\sum_{j=1}^{k}(1-\beta)\beta^k = (1-\beta)\sum_{j=1}^{\infty}\sum_{k=j}^{\infty}\beta^k$$

$$= \sum_{j=1}^{\infty}\beta^j = \frac{\beta}{1-\beta}$$

Onde utilizamos a fórmula da soma infinita de uma progressão geométrica com razão β.

Exemplo 5 (Binomial Negativa)

Se X for uma variável aleatória com distribuição de probabilidade Binomial Negativa com parâmetros r e p, temos que sua esperança é dada por:

$$
\begin{aligned}
EX &= \sum_{k=r-1}^{\infty} k\binom{k}{r-1}p^r(1-p)^{k-r+1} \\
&= \left(\sum_{k=r-1}^{\infty}(k+1)\binom{k}{r-1}p^r(1-p)^{k-r+1}\right) - 1 \\
&= \left(\sum_{k=r-1}^{\infty}\frac{(k+1)k!}{(r-1)!(k-r+1)!}p^r(1-p)^{k-r+1}\right) - 1 \\
&= \frac{r}{p}\left(\sum_{k=r-1}^{\infty}\frac{(k+1)!}{r!(k+1-r)!}p^{r+1}(1-p)^{k+1-r}\right) - 1
\end{aligned}
$$

Substituindo $j = k+1$ e $s = r+1$ no somatório, temos

$$EX = \frac{r}{p}\left(\sum_{j=s-1}^{\infty}\frac{(j)!}{(s-1)!(j-s+1)!}p^s(1-p)^{j-s+1}\right) - 1 = \frac{r}{p} - 1$$

Onde utilizamos o fato que o somatório é igual soma da função probabilidade de massa de uma variável aleatória Binomial Negativa para todos os valores que tem probabilidade positiva, e portanto, é igual a 1.

Exemplo 6 (Poisson)

Se X for uma variável aleatória com distribuição de probabilidade Poisson com parâmetros λ, temos que sua esperança é dada por:

$$EX = \sum_{k=0}^{\infty} k \frac{e^{-\lambda}\lambda^k}{k!} = \sum_{k=1}^{\infty} k \frac{e^{-\lambda}\lambda^k}{k!} = \lambda \sum_{k=1}^{\infty} \frac{e^{-\lambda}\lambda^{k-1}}{(k-1)!} = \lambda.$$

Exemplo 7 (Zeta)

Se X for uma variável aleatória com distribuição de probabilidade Zeta com parâmetro $\alpha > 2$, temos que sua esperança é dada por:

$$EX = \sum_{k=1}^{\infty} k \frac{k^{-\alpha}}{\zeta(\alpha)} = \frac{1}{\zeta(\alpha)} \sum_{k=1}^{\infty} k^{-(\alpha-1)} = \frac{\zeta(\alpha-1)}{\zeta(\alpha)},$$

onde $\zeta(\alpha) = \sum_{k=1}^{\infty} k^{-\alpha}$.

Exemplo 8 (Hipergeométrica)

Se X for uma variável aleatória com distribuição de probabilidade Hipergeométrica com parâmetro N, D, n, temos que sua esperança é dada por:

$$EX = \sum_{k=0}^{n} k \frac{\binom{D}{k}\binom{N-D}{n-k}}{\binom{N}{n}} = \sum_{k=1}^{n} \frac{D!(N-D)!(N-n)!n!}{k!(D-k)!(n-k)!(N-D-n+k)!N!}$$

$$= \frac{nD}{N} \sum_{k=1}^{n} \frac{(D-1)!(N-D)!(N-n)!(n-1)!}{(k-1)!(D-k)!(n-k)!(N-D-n+k)!(N-1)!}$$

$$= \frac{nD}{N} \sum_{k=1}^{n} \frac{\binom{D-1}{k-1}\binom{N-D}{n-k}}{\binom{N-1}{n-1}}$$

Substituindo no somatório $D^* = D-1$, $k^* = k-1$, $n^* = n-1$ e $N^* = N-1$, temos

$$EX = \frac{nD}{N} \sum_{k^*=0}^{n^*} \frac{\binom{D^*}{k^*}\binom{N^*-D^*}{n^*-k^*}}{\binom{N^*}{n^*}} = \frac{nD}{N}.$$

Onde utilizamos o fato que o somatório é igual soma da função probabilidade de massa de uma variável aleatória Hipergeométrica para todos os valores que tem probabilidade positiva, e portanto, é igual a 1.

As integrais de Riemman-Stieltjes e de Lebesgue-Stieltjes

- Antes de introduzirmos a definição geral da Esperança de uma variável aleatória qualquer, vamos estudar um pouco sobre as integrais de Riemann-Stieltjes e de Lebesgue-Stieltjes.
- Vamos relembrar a definição da integral de Riemann. Uma partição P do intervalo $[a, b]$ é uma sequência de pontos $\{x_1, \ldots, x_n\}$ tal que $a = x_1 < x_2 < \cdots < x_n = b$; a norma da partição P é definida como sendo $\max_{1 \leq i \leq n-1}(x_{i+1} - x_i)$.
- Suponha que φ seja uma função real qualquer definida no intervalo $[a, b]$. Diz-se que esta função é Riemann integrável se as somas de Riemann

$$\sum_{i=1}^{n-1} \varphi(y_i)(x_{i+1} - x_i),$$

onde $y_i \in [x_i, x_{i+1}]$, convergem quando a norma de P tende a zero e este limite é independente da escolha dos y_i's e da partição P. Se esta integral existe denota-se o limite por $\int_a^b \varphi(x)dx$.
- A integral de Riemann-Stieltjes é uma generalização de integral de Riemann.

Integral de Riemman-Stieltjes

Se φ é uma função contínua definida no intervalo $[a, b]$ e F é uma função de distribuição, define-se a integral de Riemann-Stieltjes de φ em $[a, b]$, em relação a F, como o limite de somas de Riemann da forma

$$\sum_{i=1}^{n-1} \varphi(y_i)[F(x_{i+1}) - F(x_i)],$$

em que $a = x_1 < x_2 < \cdots < x_n = b$, y_i é um ponto arbitrário de $[x_i, x_{i+1}]$, e toma-se o limite quando a norma de partição P tende a zero. Tal limite existe e é finito sob as condições descritas, e é representado por

$$\int_a^b \varphi(x) dF(x).$$

A função φ é chamada de integrando e F de integrador. O limite acima existe mesmo que F não seja uma função de distribuição basta que ela seja de variação limitada.

Definição 2 (Variação Total)

Define-se variação total de uma função f em $[a, b]$ pelo funcional:

$$V(f, [a, b]) = \sup\left\{ \sum_{i=1}^{n} |f(x_{i+1}) - f(x_i)| \right\},$$

em que o supremo é tomado sobre todas as possíveis partições do intervalo fechado $[a, b]$. Uma função é de variação limitada se $V(f, [a, b]) < \infty$.

- A integral de Rieman-Stieltjes sobre a reta é uma integral imprópria definida da mesma maneira que a integral imprópria de Riemann:

$$\int_{-\infty}^{\infty} \varphi(x)dF(x) = \lim_{a\to-\infty, b\to\infty} \int_a^b \varphi(x)dF(x),$$

  se o limite existe.
- Pode-se estender esta definição para outras funções φ além das contínuas.
- Para uma função qualquer φ, define-se $\int_a^b \varphi(x)dF(x)$ como sendo o limite das somas de Riemann descritas acima quando a norma da partição tende a zero, se este limite existe e é independente das escolhas dos y_i's e da partição P.
- O problema é que mesmo para funções bem simples este limite pode não existir como mostra o próximo exemplo:

Exemplo 9

Seja $F_0(x) = 1$ se $x \geq 0$, e $F_0(x) = 0$, caso contrário. Consideremos a integral de Riemann-Stieltjes de F_0 em $[-1, 1]$ em relação a F_0. Note que se zero não é um dos pontos da partição, de modo que $x_i < 0 < x_{i+1}$ para algum i, com $F_0(x_{i+1}) - F_0(x_i) = 1$, então o valor da soma de Riemann pode ser igual a 1 ou 0 dependendo de y_i ser maior que 0, ou não.

- Uma integral mais robusta que não sofre desta deficiência é a integral de Lebesgue-Stieltjes.
- A idéia da integral de Lebesgue-Stieltjes é particionar a imagem da função φ ao invés de particionar o seu domínio.
- Diz-se que uma partição P' é um refinamento de P se $P \subseteq P'$, ou seja, quando os intervalos da partição P são particionados na partição P'.
- Suponha que φ seja **não negativa e mensurável em relação a σ-álgebra de Borel**. Seja λ uma medida nos reais, ou seja, uma função cujo domínio é a σ-álgebra de Borel que tem como imagem do conjunto vazio o zero, é não-negativa e é σ-aditiva.
- Dada uma sequência $\{P_1, P_2, \ldots\}$ de partições de $[0, \infty)$ onde $P_n = \{y_1, y_2, \ldots, y_n\}$, $y_n \to \infty$, P_{i+i} é um refinamento de P_i, e a norma de P_n tende a zero quando $n \to \infty$, define-se a soma de Lebesgue em relação a partição P_n como sendo,

$$\sum_{i=1}^{n-1} y_i \lambda(\{x : y_i \leq \varphi(x) < y_{i+1}\}) + y_n \lambda(\{x : \varphi(x) \geq y_n\}).$$

- A integral de Lebesgue-Stieltjes de φ em relação a λ é definida como sendo igual ao limite das somas de Lebesgue, quando $n \to \infty$. Dadas as condições acima este limite sempre existe (pode ser $+\infty$) e é denotado por $\int \varphi d\lambda$.

- Para uma função mensurável φ qualquer, podemos escrever $\varphi = \varphi^+ - \varphi^-$, onde $\varphi^+ = \max(\varphi, 0)$, a parte positiva de φ, e $\varphi^- = -\min(\varphi, 0)$, o módulo da parte negativa de φ, são funções não-negativas e portanto possuem integral de Lebesgue-Stieltjes.
- Se φ^+ ou φ^- possui integral de Lebesgue-Stieltjes finita em relação a μ, define-se a integral de Lebesgue-Stieltjes de φ em relação a λ como sendo

$$\int \varphi d\lambda = \int \varphi^+ d\lambda - \int \varphi^- d\lambda.$$

- Se λ for uma medida de probabilidade em $(\mathbb{R}, \mathcal{B})$ e F for a distribuição de probabilidade acumulada associada a variável aleatória $X(\omega) = \omega$, então escreve-se

$$\int \varphi(x) dF(x)$$

(ou simplesmente, $\int \varphi dF$) para denotar $\int \varphi d\lambda$.

- Em geral, usa-se a notação $\int \varphi(x)dF(x)$ não somente para funções de distribuição, mas para qualquer função F que pode ser escrita como a diferença de duas funções monótonas não-decrescentes, limitadas e contínuas à direita.
- Se G for uma função monótona não-decrescente, limitada e contínua à direita, então dado um intervalo qualquer $I = [x_1, x_2]$, defina $\nu(I) = G(x_2) - G(x_1)$, então usa-se a notação $\int \varphi(x)dG(x)$ para denotar a integral $\int \varphi(x)d\nu$, onde ν é a única medida que satisfaz $\nu(I) = G(x_2) - G(x_1)$ para todo intervalo I.
- Desta forma, se $F = G_1 - G_2$, onde G_1 e G_2 são funções monótonas não-decrescentes, limitadas e contínuas à direita, então $\int \varphi(x)dF(x)$ é utilizado para denotar
$$\int \varphi(x)dG_1(x) - \int \varphi(x)dG_2(x)$$
.
- Dada um intervalo qualquer $[a, b]$, define-se a integral de Lebesgue-Stieltjes de φ em relação a λ no intervalo $[a, b]$ como sendo $\int \varphi I_{[a,b]} d\lambda$ e denota-se por $\int_a^b \varphi d\lambda$.

Propriedades da Integral de Lebesgue-Stieltjes

P1. Quando o integrando é contínuo, a integral de Lebesgue-Stieltjes torna-se uma integral de Riemman-Stieltjes.

P2. $\int_a^b dF = F(b) - F(a)$. Análoga ao teorema fundamental do cálculo: $\int_a^b \varphi'(x)dx = \varphi(b) - \varphi(a)$, onde $\varphi(x)$ é a derivada de φ.

P3. Linearidade no integrando e no integrador. Se $\varphi(x) = \alpha f(x) + \beta g(x)$, temos

$$\int \varphi dF = \alpha \int f dF + \beta \int g dF,$$

e para $H(x) = \alpha F(x) + \beta G(x)$, temos

$$\int \varphi dH = \alpha \int \varphi dF + \beta \int \varphi dG.$$

P4. Aditividade. Se $-\infty \leq a < b < c \leq \infty$, então

$$\int_a^c \varphi dF = \int_a^b \varphi dF + \int_b^c \varphi dF.$$

P5. Se F for a função de distribuição de uma variável aleatória discreta, ou seja, se

$$F(x) = \sum_{i=1}^{\infty} p_i U(x - x_i),$$

onde $P(X = x_i) = p_i$ e $\sum_{i=1}^{\infty} p_i = 1$, então

$$\int \varphi dF = \sum_{i=1}^{\infty} p_i \varphi(x_i).$$

P6. Se F for a função de distribuição de uma variável aleatória contínua, tendo densidade f, temos $\frac{dF(x)}{dx} = f(x)$ em quase toda parte, e consequentemente,

$$\int \varphi(x)dF(x) = \int \varphi(x)f(x)dx.$$

P7. No caso de uma distribuição geral F, vimos que F pode ser decomposta em suas partes discreta, contínua e singular da seguinte forma $F = F_d + F_{ac} + F_s$, então por linearidade do integrador:

$$\int \varphi(x)dF(x) = \int \varphi(x)dF_d(x) + \int \varphi(x)dF_{ac}(x) + \int \varphi(x)dF_s(x).$$

Se a parte singular for nula, $F_s(x) = 0, \forall x$, então:

$$\int \varphi(x)dF(x) = \sum_i \varphi(x_i)p_i + \int \varphi(x)f(x)dx,$$

onde p_i é o salto de F em x_i e f é a derivada de F.

Esperança - Caso Geral

- Agora motivar a definição da Esperança no caso geral.
- Consideremos uma sequência $\{P_1, P_2, \ldots\}$ de partições de $[0, \infty)$ onde $P_n = \{y_1, y_2, \ldots, y_n\}$, $y_n \to \infty$, P_{i+i} é um refinamento de P_i, e a norma de P_n tende a zero quando $n \to \infty$.
- Dada uma variável aleatória não-negativa qualquer X e uma partição P_n desta sequência, definamos uma outra variável aleatória Y discreta que aproxima X assumindo o valor y_i quando $y_i \leq X < y_{i+1}$ e $Y = y_n$ se $X \geq y_n$, ou seja, $Y = \sum_{i=1}^{n-1} y_i I_{[y_i \leq X < y_{i+1}]} + y_n I_{[X \geq y_n]}$.
- Como Y é discreta temos que sua esperança é dada por

$$\begin{aligned} EY &= \sum_{i=1}^{n} y_i P(Y = y_i) \\ &= \sum_{i=1}^{n-1} y_i P(y_i \leq X < y_{i+1}) + y_n P(X \geq y_n). \end{aligned}$$

- Note que esta esperança é uma soma de Lebesgue em relação a partição P_n com integrando X e função integradora dada pela medida de probabilidade P.
- Note que a medida que pegamos partições mais refinadas na sequência, Y se torna cada vez uma melhor aproximação para X.
- Uma vez que os valores de X e Y ficam cada vez mais próximos é intuitivo requerer que nossa definição de esperança (média) EX seja igual ao limite de EY quando $n \to \infty$, ou seja

$$\begin{aligned} EX &= \lim_{n\to\infty} \sum_{i=1}^{n} y_i P(Y = y_i) \\ &= \lim_{n\to\infty} \sum_{i=1}^{n-1} y_i P(y_i \leq X < y_{i+1}) + y_n P(X \geq y_n) \\ &= \int X dP. \end{aligned}$$

Logo, EX é definida como sendo a integral de Lebesgue-Stieltjes de X em relação a medida de probabilidade P, ou similarmente, $EX = \int XdF$, onde F é a função de distribuição acumulada de X. No caso geral, temos a seguinte definição

Definição 3

Se X é uma variável aleatória com função de distribuição F, então sua esperança é dada pela fórmula

$$EX = \int XdF = \int_{-\infty}^{0} XdF + \int_{0}^{\infty} XdF,$$

desde que pelo menos uma das integrais seja finita. Em caso as duas integrais não sejam finitas, a esperança não existe. Caso EX seja finita, diz-se que X é integrável.

Pela Propriedade P7 da integral de Lebesgue-Stieltjes, temos que se $F = F_d + F_{ac} + F_s$, então

$$EX = \int XdF = \sum_i x_i p_i + \int xf(x)dx + \int xdF_s(x),$$

onde p_i é o salto de F em x_i e f é a derivada de F. Como a parte singular costuma ser nula, na prática a esperança reduz-se a uma série e/ou uma integral imprópria, usualmente de Riemann se f for integrável no sentido de Riemann.

Exemplo 10 (Variável Mista)

Considere uma variável aleatória Y com função de distribuição F, tal que

$$F(x) = \begin{cases} 0 & \text{, se } x < 0 \\ x & \text{, se } 0 \leq x < 1/2 \\ 1 & \text{, se } x \geq 1/2. \end{cases}$$

Decompondo em parte discreta e contínua tem-se

$$F_d(x) = \begin{cases} 0 & \text{, se } x < 1/2 \\ 1/2 & \text{, se } x \geq 1/2, \end{cases}$$

e

$$F_{ac}(x) = \begin{cases} 0 & \text{, se } x < 0 \\ x & \text{, se } 0 \leq x < 1/2 \\ 1/2 & \text{, se } x \geq 1/2. \end{cases}$$

Portanto,

$$EY = \frac{1}{2}P(Y = \frac{1}{2}) + \int_0^{1/2} y dy = \frac{1}{4} + \frac{1}{8} = \frac{3}{8}.$$

Exemplo 11 (Uniforme)

Se $X \sim U(a, b)$, então X possui densidade igual a $f(x) = \frac{1}{b-a}$ se $x \in (a, b)$, e $f(x) = 0$, caso contrário. Logo, temos que sua esperança é dada por:

$$EX = \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2}.$$

Exemplo 12 (Exponencial)

Se $X \sim Exp(\lambda)$, então X possui densidade igual a $f_X(x) = \lambda e^{-\lambda x} U(x)$. Logo, temos que sua esperança é dada por:

$$\begin{aligned} EX &= \int_0^\infty x\lambda e^{-\lambda x} dx = -xe^{-\lambda x}|_0^\infty + \int_0^\infty e^{-\lambda x} dx \\ &= \frac{-e^{-\lambda x}}{\lambda}|_0^\infty = \frac{1}{\lambda}. \end{aligned}$$

Exemplo 13 (Normal)

Se $X \sim \mathcal{N}(\mu, \sigma)$, então X possui densidade igual a $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$. Logo, temos que sua esperança é dada por:

$$EX = \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx.$$

Fazendo a mudança de variável $y = \frac{x-\mu}{\sigma}$, temos

$$EX = \int_{-\infty}^{\infty} \frac{\sigma y + \mu}{\sqrt{2\pi}} e^{\frac{-y^2}{2}} dy = \int_{-\infty}^{\infty} \frac{\sigma y}{\sqrt{2\pi}} e^{\frac{-y^2}{2}} dy + \int_{-\infty}^{\infty} \frac{\mu}{\sqrt{2\pi}} e^{\frac{-y^2}{2}} dy = 0 + \mu = \mu.$$

Exemplo 14 (Cauchy)

Se $X \sim Cauchy(a)$, então X possui densidade igual a $f_X(x) = \frac{1}{\pi} \cdot \frac{a}{a^2+x^2}$. Neste caso X não é integrável, ou seja EX não está definida, pois:

$$\int_{-\infty}^{0} \frac{x}{\pi} \cdot \frac{a}{a^2+x^2} dx = \frac{a}{2\pi} \log(a^2+x^2)|_{-\infty}^{0} = -\infty, \text{ e}$$

$$\int_{0}^{\infty} \frac{x}{\pi} \cdot \frac{a}{a^2+x^2} dx = \frac{a}{2\pi} \log(a^2+x^2)|_{0}^{\infty} = \infty.$$

Interpretação Geométrica da Esperança

Por definição, $EX = \int xdF(x)$, ou seja, EX é a integral da diferencial xdF. Mas xdF é uma diferencial de área.

- Para $x > 0$, xdF é uma diferencial da área da região compreendida entre as curvas $x = 0$, $y = 1$, e $y = F(x)$ no plano Euclideano, cuja área total é dada por $\int_0^\infty (1 - F(x))dx$.
- Para $x < 0$, $-xdF$ é uma diferencial da área da região compreendida entre as curvas $x = 0$, $y = 0$, e $y = F(x)$ no plano Euclideano, cuja área total é dada por $\int_{-\infty}^0 F(x)dx$.
- Logo, $EX = \int_0^\infty (1 - F(x))dx - \int_{-\infty}^0 F(x)dx$.

Formalmente, podemos provar isso da seguinte maneira. A prova será dividida em duas etapas:

(a) $\int_0^\infty x dF(x) = \int_0^\infty (1 - F(x))dx$ e

(b) $\int_{-\infty}^0 x dF(x) = -\int_{-\infty}^0 F(x)dx$.

Comecemos provando (b). Utilizando integração por partes, temos que $\forall a < 0$,

$$\int_a^0 x dF(x) = -aF(a) - \int_a^0 F(x)dx = \int_a^0 [F(a) - F(x)]dx.$$

Como $F(a) \geq 0$ e $a < 0$, temos

$$\int_a^0 x dF(x) \geq -\int_a^0 F(x)dx.$$

Como a desigualdade é válida para todo $a < 0$, temos que tomando o limite quando $a \to -\infty$

$$\int_{-\infty}^0 x dF(x) \geq -\int_{-\infty}^0 F(x)dx.$$

Por outro lado, seja $\lambda < 0$. Se $a < \lambda$, então

$$\int_a^0 [F(a) - F(x)]dx \leq \int_\lambda^0 [F(a) - F(x)]dx$$

$$= F(a)(-\lambda) - \int_\lambda^0 F(x)dx.$$

Portanto, tomando o limite quando $a \to -\infty$, temos que

$$\int_{-\infty}^{0} xdF(x) \leq -\int_{\lambda}^{0} F(x)dx.$$

Como isto é válido para todo $\lambda < 0$, tomando o limite quando $\lambda \to -\infty$, temos

$$\int_{-\infty}^{0} xdF(x) \leq -\int_{-\infty}^{0} F(x)dx,$$

como queríamos demonstrar. Para parte (a), utilizando integração por partes, temos que $\forall b > 0$,

$$\int_{0}^{b} xdF(x) = bF(b) - \int_{0}^{b} F(x)dx = \int_{0}^{b} [F(b) - F(x)]dx.$$

Como $F(b) \leq 1$ e $1 - F(x) \geq 0$, temos

$$\int_{0}^{b} xdF(x) = \int_{0}^{b} [F(b) - F(x)]dx \leq \int_{0}^{\infty} [1 - F(x)]dx.$$

Como a desigualdade é válida para todo $b > 0$, temos que tomando o limite quando $b \to \infty$

$$\int_{0}^{\infty} xdF(x) \leq \int_{0}^{\infty} [1 - F(x)]dx.$$

Por outro lado, seja $\lambda > 0$. Se $b > \lambda$, então

$$\int_0^b [F(b) - F(x)]dx \geq \int_0^\lambda [F(b) - F(x)]dx$$
$$= \int_0^\lambda [F(b) - 1]dx + \int_0^\lambda [1 - F(x)]dx$$
$$= \lambda[F(b) - 1] + \int_0^\lambda [1 - F(x)]dx,$$

e portanto, tomando o limite quando $b \to \infty$, temos que

$$\int_0^\infty x dF(x) \geq \int_0^\lambda [1 - F(x)]dx.$$

Como isto é válido para todo $\lambda > 0$, tomando o limite quando $\lambda \to \infty$, temos

$$\int_0^\infty x dF(x) \geq \int_0^\infty [1 - F(x)]dx,$$

como queríamos demonstrar.

Esperança de Funções de Variáveis Aleatórias

Caso Discreto

Como vimos anteriormente, se X for uma variável aleatória discreta e se $Y = H(X)$, então Y também será uma variável aleatória discreta. Consequentemente, pode-se calcular EY. Existem duas maneiras de calcular EY que são equivalentes.

Definição 4

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

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

Conforme vimos anteriormente podemos determinar as probabilidades $p(y_i)$ dado que sabemos a distribuição de X. No entanto, podemos encontrar EY sem preliminarmente encontrarmos a distribuição de probabilidade de Y, partindo-se apenas do conhecimento da distribuição de probabilidade de X, conforme mostra o seguinte teorema.

Teorema 1 (Caso Discreto)

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

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

Demonstração.

Vamos re-ordenar o somatório $\sum_{i=1}^{\infty} H(x_i)p(x_i)$, 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. Desse modo podemos reescrever

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

$$= \sum_{i=1}^{\infty} y_i \sum_{j=1}^{\infty} p(x_{ij}) = \sum_{i=1}^{\infty} y_i p(y_i) = EY.$$

$\square$

Exemplo 15

Suponha que X é uma variável aleatória Poisson com parâmetro λ. Seja $Y = X^2$, vamos calcular EY. Utilizando o Teorema 1, temos

$$
\begin{aligned}
EY &= \sum_{k=0}^{\infty} k^2 e^{-\lambda} \frac{\lambda^k}{k!} \\
&= \sum_{k=1}^{\infty} k^2 e^{-\lambda} \frac{\lambda^k}{k!} = \sum_{k=1}^{\infty} k(k-1) e^{-\lambda} \frac{\lambda^k}{k!} + \sum_{k=1}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} \\
&= \lambda^2 \sum_{k=2}^{\infty} e^{-\lambda} \frac{\lambda^{k-2}}{(k-2)!} + \lambda = \lambda^2 + \lambda.
\end{aligned}
$$

Caso Discreto Vetorial

Também podemos estender este resultado para o caso de uma função real de um vetor aleatório. Neste caso, se $Y = H(\vec{X})$, temos que $EY = \sum_i H(\vec{x}_i) p_{\vec{X}}(\vec{x}_i)$, onde os $\vec{x}_i$ são os valores assumidos pelo vetor aleatório $\vec{X}$.

Caso Geral

No caso de uma variável aleatória qualquer X também podemos calcular a esperança de uma função $Y = \varphi(X)$ de forma similar.

Teorema 2

Seja X uma variável aleatória qualquer, $Y = \varphi(X)$ uma outra variável aleatória, então

$$EY = \int y dF_Y(y) = \int \varphi(x) dF_X(x),$$

desde que estas integrais existam.

Demonstração.

A prova no caso geral envolve Teoria da Medida e será omitida. $\square$

Caso Vetorial

Uma fórmula análoga também é válida quando consideramos funções de vetores aleatórios.

Teorema 3

Seja $\vec{X}$ um vetor aleatório e $Y = \varphi(\vec{X})$ uma variável aleatória. Então,

$$EY = \int y dF_Y(y) = \int \varphi dF_{\vec{X}}.$$

Propriedades da Esperança

As seguintes propriedades são aplicações imediatas da definição de esperança:

1. $P(X = c) = 1 \Rightarrow EX = c$.
2. $P(X \geq 0) = 1 \Rightarrow EX \geq 0$.
3. $E(aX) = aEX$, onde $a \in \mathbb{R}$ (segue da esperança de uma função de v. a.)

4. $E(X+Y) = EX + EY$. No caso discreto, note que

$$E(X+Y) = \sum_i \sum_j (x_i + y_j) p(x_i, y_j) = \sum_i x_i \sum_j p(x_i, y_j) + \sum_i \sum_j y_j p(x_i, y_j)$$

$$= \sum_i x_i p(x_i) + \sum_j y_j \sum_i p(x_i, y_j) = EX + \sum_j y_j p(y_j) = EX + EY.$$

No caso geral, temos que

$$E(X+Y) = E(\varphi(X,Y)) = \int \int (x+y) dF_{X,Y}(x,y),$$

e pela linearidade da integral obtemos

$$E(X+Y) = \int \int x dF_{X,Y}(x,y) + \int \int y dF_{X,Y}(x,y) = EX + EY.$$

Corolário 1

$E(\sum_i^n a_i X_i) = \sum_i^n a_i EX_i$.

Demonstração.

Aplicação das duas propriedades anteriores e indução matemática. □

5. $P(X \geq Y) = 1 \Rightarrow EX \geq EY$. Propriedade 5 segue da propriedades 2 e do corolário anterior, pois

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

o que, pela propriedade 2, implica que $E(X - Y) \geq 0$. Pelo corolário, temos que $E(X - Y) = EX - EY$, ou seja podemos concluir que $EX - EY \geq 0$.

6. Se $\{X_1, \ldots, X_n\}$ são variáveis aleatórias mutuamente independentes, então

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

Provaremos esta propriedade nos casos discreto e contínuo. No caso discreto, note que

$$E(\prod_{i=1}^{n} X_i) = \sum_{i_1} \cdots \sum_{i_n} x_{i_1} \cdots x_{i_n} p(x_{i_1}, \ldots, x_{i_n})$$

$$= \sum_{i_1} \cdots \sum_{i_n} x_{i_1} \cdots x_{i_n} \prod_{j=1}^{n} p(x_{i_j})$$

$$= \sum_{i_1} x_{i_1} p(X_{i_1}) \cdots \sum_{i_n} x_{i_n} p(x_{i_n}) = \prod_{i=1}^{n} EX_i.$$

No caso contínuo, temos que $f_{\vec{X}}(\vec{x}) = \prod_{i=1}^{n} f_{X_i}(x_i)$, logo

$$\begin{aligned} E(\prod_{i=1}^{n} X_i) &= \int \cdots \int x_1 \cdots x_n f_{\vec{X}}(\vec{x}) dx_1 \cdots dx_n \\ &= \int \cdots \int \prod_{i=1}^{n} x_i f_{X_i}(x_i) dx_1 \cdots dx_n \\ &= \prod_{i=1}^{n} \int x_i f_{X_i}(x_i) dx_i = \prod_{i=1}^{n} EX_i. \end{aligned}$$

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

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

.

7. Se Y for uma variável aleatória que assume valores inteiros não-negativos, temos que

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

trocando a ordem dos somatórios:

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

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

Demonstração.

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, temos

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

Em particular, para $x_0 = EX$, temos $E\varphi(X) \geq \varphi(EX)$. □

Integrabilidade

O próximo Lema estabelece um critério para integrabilidade de variáveis aleatórias.

Lema 1

Seja X uma variável aleatória qualquer. Então,
$\sum_{n=1}^{\infty} P(|X| \geq n) \leq E|X| \leq 1 + \sum_{n=1}^{\infty} P(|X| \geq n)$*, e, portanto, X é integrável se, e somente se,*

$$\sum_{n=1}^{\infty} P(|X| \geq n) < \infty.$$

Demonstração.

Se $x \geq 0$, seja $\lfloor x \rfloor$ a parte inteira de x. Então, a variável aleatória $\lfloor |X| \rfloor$ assume o valor k quando $k \leq |X| < k + 1$ e $0 \leq \lfloor |X| \rfloor \leq |X| \leq \lfloor |X| \rfloor + 1$, então pela monotonicidade e linearidade da esperança temos: $0 \leq E\lfloor |X| \rfloor \leq E|X| \leq 1 + E\lfloor |X| \rfloor$.
Como $\lfloor |X| \rfloor$ é uma variável aleatória que só assume valores inteiros não-negativos, temos

$$E\lfloor |X| \rfloor = \sum_{n=1}^{\infty} P(\lfloor |X| \rfloor \geq n) = \sum_{n=1}^{\infty} P(|X| \geq n),$$

logo

$$\sum_{n=1}^{\infty} P(|X| \geq n) \leq E|X| \leq 1 + \sum_{n=1}^{\infty} P(|X| \geq n).$$

$\square$

Se $X^+ = \max(X,0)$ e $X^- = -\min(X,0)$, temos que $X = X^+ - X^-$ e $|X| = X^+ + X^-$. Por definição, temos que $EX < \infty$ se, e somente se, $EX^+ < \infty$ e $EX^- < \infty$. Portanto, vemos que $EX < \infty$ se, e somente se, $E|X| < \infty$. De forma análoga, pode-se concluir que $E\varphi(X) < \infty$ se, e somente se, $E|\varphi(X)| < \infty$ para qualquer função mensurável φ. O próximo teorema nos dá um outro critério para integrabilidade de uma variável aleatória.

Teorema 4

Sejam X e Y variáveis aleatórias tais que $Y \geq 0$, Y é integrável, e $|X| < Y$. Então, X é integrável.

Demonstração.

Note que $0 \leq |X| \leq Y$ implica que $0 \leq E|X| \leq EY$. Portanto, se $EY < \infty$, temos que $E|X| < \infty$, o que por sua vez implica que $EX < \infty$. □

Momentos

Momentos dão informações parciais sobre a medida de probabilidade P, a função de distribuição acumulada, ou a função probabilidade de massa de uma variável aleatória X. Momentos de X são esperanças de potências de X.

Definição 5

Para qualquer inteiro não-negativo n, o n-ésimo momento da variável aleatória X é EX^n, se esta esperança existe.

Vimos anteriormente que o segundo momento de uma variávelaleatória Poisson com parâmetro λ é dado por: $\lambda^2 + \lambda$. Vamos agora calcular o segundo momento de uma variável aleatória X Binomial com parâmetros n e p:

$$EX^2 = \sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^{n} k^2 \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

$$= \sum_{k=1}^{n} k(k-1) \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} + \sum_{k=1}^{n} k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

$$= n(n-1)p^2 \sum_{k=2}^{n} \frac{(n-2)!}{(k-2)!(n-k)!} p^{k-2} (1-p)^{n-k} + np$$

$$= n(n-1)p^2 \sum_{j=0}^{m} \frac{(m)!}{(j)!(m-j)!} p^j (1-p)^{m-j} + np = n(n-1)p^2 + np.$$

Existência de momentos

Teorema 5

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

Demonstração.

Por hipótese, temos que $E|X^k| < \infty$, 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, temos que $|X^j|$ também é integrável. $\square$

Teorema 6

Se $j < k$, então

$$|EX^k| < \infty \Rightarrow |EX^j| < \infty$$

Demonstração

Vamos provar apenas para o caso em que X é uma variável aleatória discreta. Por hipótese, temos que existem constantes finitas c_1 e c_2 tais que:

$$c_1 > \sum_{i:x_i \geq 0} x_i^k p(x_i), \text{ e } c_2 > |\sum_{i:x_i < 0} x_i^k p(x_i)|.$$

Então, para $j < k$

$$c_1 > \sum_{i:x_i \geq 0} x_i^k p(x_i) \geq \sum_{i:x_i \geq 1} x_i^k p(x_i) \geq \sum_{i:x_i \geq 1} x_i^j p(x_i)$$

$$\geq \sum_{i:x_i \geq 0} x_i^j p(x_i) - \sum_{i:0 \leq x_i \leq 1} x_i^j p(x_i) \geq \sum_{i:x_i \geq 0} x_i^j p(x_i) - \sum_{i:0 \leq x_i \leq 1} p(x_i) = \sum_{i:x_i \geq 0} x_i^j p(x_i) - 1$$

Portanto,

$$0 \leq \sum_{i:x_i \geq 0} x_i^j p(x_i) < c_1 + 1 < \infty.$$

Um cálculo similar prova que $\sum_{i:x_i<0} x_i^j p(x_i)$ é finito.

Portanto, temos que *momentos de ordem superiores finitos implicam momentos de ordem inferiores finitos.*

Vamos agora enunciar dois teoremas importantes que tratam da convergência de esperanças de variáveis aleatórias. Neste caso, estaremos tratando de convergência pontual de variáveis aleatórias, ou seja, $X_n \to X$ se, e somente se, $X_n(\omega) \to X(\omega)$ para todo $\omega \in \Omega$. Mais adiante, veremos outras noções de convergência de variáveis aleatórias.

Convergência Monótona e Convergência Dominada

Teorema 7

Teorema da Convergência Monótona. Sejam $X, X_1, X_2, \ldots$ *variáveis aleatórias. Se* $0 \leq X_n \uparrow X$*, então,* $EX_n \uparrow EX$.

Teorema 8

Teorema da Convergência Dominada. Sejam $Y, X, X_1, X_2, \ldots$ *variáveis aleatórias. Considere que* Y *seja integrável,* $|X_n| \leq Y$ *e* $X_n \to X$*. Assim* X *e* X_n *são integráveis e* $EX_n \to EX$.

Nota 1

Nem sempre $X_n \to X \Rightarrow EX_n \to EX$.

Definição 6 (Momentos Centrais)

Se X *é uma variável aleatória seu n-ésimo momento central é:* $E(X - EX)^n$*, se esta esperança existir.*

Note que o primeiro momento central é zero, pois $E(X-EX) = EX - EEX = EX - EX = 0$. O segundo momento central é conhecido como *variância* e denota-se por $VarX$. A variância pode ser também calculada por:

$$\begin{aligned} VarX &= E(X-EX)^2 = E(X^2 - 2XEX + (EX)^2) \\ &= EX^2 - 2E(XEX) + E((EX)^2) \\ &= EX^2 - 2(EX)^2 + (EX)^2 = EX^2 - (EX)^2. \end{aligned}$$

Existência

Do Teorema Binomial e da linearidade da esperança, temos

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

e

$$\begin{aligned} EX^n &= E(X - EX + EX)^n \\ &= \sum_{k=0}^{n} \binom{n}{k} (EX)^{n-k} E(X-EX)^k. \end{aligned}$$

Como um corolário, temos que o n-ésimo momento central é finito se, e somente se, o n-ésimo momento é finito.

Exemplo 16

Considere uma variável aleatória X tal que

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

$$EX = m, EX^2 = \frac{1}{2}[2m^2 + 2a^2] = m^2 + a^2, VarX = a^2.$$

Este exemplo, mostra que podemos encontrar uma variável aleatória bem simples possuindo qualquer esperança e variância predeterminadas.

Desvio-Padrão

O *desvio-padrão* σ de uma variável aleatória X é definido como a raiz quadrada da variância, $\sigma(X) = \sqrt{VarX}$.

Propriedades da Variância e de outros Momentos

As seguintes propriedades da variância são consequências imediatas de sua definição.

1. $VarX \geq 0$.
2. Se $X = c$, $Var(X) = 0$.

Demonstração.

Temos que $EX = c$, logo $Var(X) = E(X - c)^2 = E(0) = 0$. $\square$

3. $Var(X + a) = VarX$, onde a é uma constante real.

Demonstração.

$$
\begin{aligned}
&Var(X + a) = E(X + a)^2 - (E(X + a))^2 \\
&= EX^2 + 2aEX + a^2 - (EX)^2 - 2aEX - a^2 \\
&= EX^2 - (EX)^2 = VarX.
\end{aligned}
$$

$\square$

4. $Var(aX) = a^2 VarX$

Demonstração.

$$
\begin{aligned}
Var(aX) &= E(aX)^2 - (E(aX))^2 \\
&= a^2 EX^2 - a^2(EX)^2 = a^2 VarX.
\end{aligned}
$$

□

5. Se X e Y forem variáveis aleatórias mutuamente independentes, então $Vax(X + Y) = VarX + VarY$.

Demonstração.

$$
\begin{aligned}
Var(X+Y) &= E(X+Y)^2 - [E(X+Y)]^2 \\
&= E(X^2 + 2XY + Y^2) - (EX)^2 - 2EXEY - (EY)^2 \\
&= EX^2 + EY^2 - (EX)^2 - (EY)^2 + 2E(XY) - 2EXEY \\
&= VarX + VarY
\end{aligned}
$$

□

6. Se $X_1, \ldots, X_n$ são variáveis aleatórias independentes, então $Var(X_1 + \cdots X_n) = VarX_1 + \cdots + VarX_n$. Esta propriedade segue da propriedade anterior e de uma aplicação de indução matemática.
7. **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 $P(X \in A) \leq \min(1, Eg(X))$.

Demonstração.

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

$$\min(1, Eg(X)) \geq P(X \in A).$$

□

Corolário 2

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

Demonstração.

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 3

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

Demonstração.

$P(Z \geq \frac{1}{n}) \leq nEZ = 0$. 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 $P(X = EX) = 1$, i.e., X é constante com probabilidade 1.

Corolário 4 (Desigualdade (Original) de Chebyshev.)

Seja X uma variável aleatória, então $P(|X - EX| \geq \epsilon) \leq VarX/\epsilon^2$.

Demonstração.

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, $P(X \in A) = P(|X| \geq \epsilon) \leq EX^2/\epsilon^2$. Substituindo X por $X - EX$, temos $P(|X - EX| \geq \epsilon) \leq VarX/\epsilon^2$. □

Note que a desigualdade de Chebyshev converte conhecimento sobre um momento de segunda ordem ou uma variância numa cota superior para a probabilidade da cauda de uma variável aleatória.

8. Se X e Y são variáveis aleatórias em $(\Omega, \mathcal{A}, P)$ tais que $E|X^t| < \infty$ e $E|Y^t| < \infty$, então $E|X+Y|^t < \infty$.

Demonstração.

$|X+Y| \leq |X| + |Y| \leq 2\max(|X|, |Y|)$. Portanto,
$|X+Y|^t \leq 2^t \max(|X|^t, |Y|^t) \leq 2^t(|X|^t + |Y|^t)$. Logo,
$E|X+Y|^t \leq 2^t(E|X|^t + E|Y|^t) < \infty$. □

Como $E|X|^t < \infty$ obviamente implica $E|aX|^t < \infty$, $\forall a \in \mathbb{R}$, esta propriedade diz que a classe de variáveis aleatórias em $(\Omega, \mathcal{A}, P)$ possuidoras do t-ésimo momento finito é um espaço vetorial ou espaço linear.

9. $VarX = E(X-\mu)^2 = \min_{c\in\mathbb{R}} E(X-c)^2$.

Demonstração.

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

logo

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

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

Momentos Conjuntos

Podemos definir a noção de momento quando lidamos com vetores aleatórios.

Definição 7

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.

No caso bidimensional, temos que a correlação e a covariância são momentos conjuntos que são medidas do grau de dependência linear entre duas variáveis.

Definição 8

A correlação entre duas variáveis aleatórias X e Y é dada por EXY se esta esperança existe. A covariância entre elas é dada por
$Cov(X, Y) = E[(X - EX)(Y - EY)] = EXY - (EX)(EY)$.

Note que $Cov(X, X) = VarX$. Pela prova da propriedade 5 de variância, vemos que a relação $Var(X + Y) = VarX + VarY + 2Cov(X, Y)$ é válida.
Duas varáveis são *não-correlacionadas* se $Cov(X, Y) = 0$. Como já provamos que se X e Y são independentes, então $EXY = EXEY$. Temos que se X e Y são independentes, elas necessariamente são não-correlacionadas. O contrário nem sempre é verdadeiro como o próximo exemplo ilustra.

Exemplo 17

Se X é uma variável aleatória tal que $P(X = -a) = P(X = a) = 1/2$ e $Y = X^2$, temos que $EXY = -a^3(1/2) + a^3(1/2) = 0$ e $EX = -a(1/2) + a(1/2) = 0$. Logo, $EXY = EXEY = 0$, ou seja, $Cov(X, Y) = 0$. Porém, X e Y não são independentes, pois Y é uma função de X.

Vejamos agora uma expressão para a variância da soma de n variáveis aleatórias.

Teorema 9

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

$$Var(X_1 + \cdots + X_n) = \sum_{i=1}^{n} VarX_i + 2\sum_{i<j} Cov(X_i, X_j).$$

Demonstração.

$$Var(X_1 + \cdots + X_n) = E(X_1 + \cdots + X_n - E(X_1 + \cdots + X_n))^2$$

$$= E(\sum_{i=1}^{n}(X_i - EX_i))^2 = E[\sum_{i=1}^{n}(X_i - EX_i)^2 + 2\sum_{i<j}(X_i - EX_i)(X_j - EX_j)]$$

$$= \sum_{i=1}^{n} Var(X_i) + 2\sum_{i<j} Cov(X_i, X_j).$$

Corolário 5

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

$$Var(X_1 + \cdots + X_n) = \sum_{i=1}^{n} VarX_i.$$

O próximo teorema trata de uma importante desigualdade em teoria da probabilidade:

Teorema 10

$$(E(XY))^2 \leq EX^2EY^2 \text{ e } (Cov(X,Y))^2 \leq VarXVarY.$$

Demonstração.

$(aX + Y)^2 \geq 0 \Rightarrow E(aX + Y)^2 \geq 0 \Rightarrow a^2EX^2 + 2aEXY + EY^2 \geq 0$. Observe 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, temos que

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

e temos a primeira desigualdade. A segunda desigualdade segue da primeira trocando X por $X - EX$ e Y por $Y - EY$ na expressão da primeira desigualdade. □

Coeficiente de Correlação

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

$$\rho(X,Y) = Cov(X,Y)/\sqrt{Var(X)Var(Y)}.$$

O teorema anterior provou que $|\rho(X,Y)| \leq 1$. O próximo teorema mostra que o módulo do coefieciente de correlação entre duas variáveis é igual a 1 se, e somente se, as variáveis são linearmente dependentes.

Teorema 11

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

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

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

Demonstração

Parte (a). Como $(((X-EX)/\sqrt{Var(X)}) - ((Y-EY)/\sqrt{Var(Y)}))^2 \geq 0$, temos que

$$0 \leq E\Big(\frac{X-EX}{\sqrt{Var(X)}} - \frac{Y-EY}{\sqrt{Var(Y)}}\Big)^2 = E\Big(\frac{X-EX}{\sqrt{Var(X)}}\Big)^2 + E\Big(\frac{Y-EY}{\sqrt{Var(Y)}}\Big)^2$$

$$-\frac{2}{\sqrt{Var(X)Var(Y)}}E((X-EX)(Y-EY))$$

$$= \frac{VarX}{VarX} + \frac{VarY}{VarY} - \frac{2Cov(X,Y)}{\sqrt{Var(X)var(Y)}} = 2 - 2\rho(X,Y).$$

Se $\rho(X, Y) = 1$, então

$$E\left(\frac{X - EX}{\sqrt{Var(X)}} - \frac{Y - EY}{\sqrt{Var(Y)}}\right)^2 = 0,$$

o que por sua vez implica que

$$P\left(\frac{X - EX}{\sqrt{Var(X)}} = \frac{Y - EY}{\sqrt{Var(Y)}}\right) = 1,$$

em outras palavras,

$$P(Y = EY + \frac{\sqrt{VarY}}{\sqrt{VarX}}(X - EX)) = 1.$$

A prova da parte (b) é análoga, substituindo o sinal “+” por “-” na expressão acima. Verifique os detalhes!

Desigualdade de Hölder

O próximo teorema apresenta uma nova relação entre momentos conjuntos de variáveis aleatórias. Ele é conhecido como *Desigualdade de Hölder*.

Teorema 12

Suponha que p e q satisfazem: $p > 1$, $q > 1$, *e* $\frac{1}{p} + \frac{1}{q} = 1$. *Então, se* $E(|X|^p) < \infty$ *e* $E(|Y|^q) < \infty$, *temos que*

$$E(|XY|) \leq (E|X|^p)^{1/p}(E|Y|^q)^{1/q}.$$

Demonstração

A prova da desigualdade de Hölder utiliza um argumento de convexidade. Como $|X|^p \geq 0$ (resp., $|Y|^q \geq 0$), já vimos que se $E|X|^p = 0$, então $P(X = 0) = 1$. Portanto, em ambos os casos $E(|XY|) = 0$ e a desigualdade de Hölder é válida. Considere então o caso em que o lado direito da desigualdade de Hölder é estritamente positivo.
Note que para $a > 0$ e $b > 0$, existe $s, t \in \mathbb{R}$ tal que

$$a = \exp(\frac{s}{p}) \text{ e } b = \exp(\frac{t}{q}).$$

Como a função exponencial é convexa e $p^{-1} + q^{-1} = 1$, temos por convexidade que

$$\exp(\frac{s}{p} + \frac{t}{q}) \leq p^{-1}\exp(s) + q^{-1}\exp(t),$$

ou pela definição de s, t

$$ab \leq p^{-1} a^p + q^{-1} b^q.$$

Agora substituindo a por $\frac{|X|}{(E(|X|^p))^{1/p}}$ e b por $\frac{|Y|}{(E(|Y|^q))^{1/q}}$, temos

$$\begin{aligned}
& \frac{|XY|}{(E(|X|^p))^{1/p}(E(|Y|^q))^{1/q}} \\
& \leq p^{-1}\left(\frac{|X|}{(E(|X|^p))^{1/p}}\right)^p + q^{-1}\left(\frac{|Y|}{(E(|Y|^q))^{1/q}}\right)^q.
\end{aligned}$$

Finalmente, tomando o valor esperado, temos

$$\begin{aligned}
& \frac{E|XY|}{(E(|X|^p))^{1/p}(E(|Y|^q))^{1/q}} \\
& \leq p^{-1}\left(\frac{E|X|^p}{(E(|X|^p))}\right) + q^{-1}\left(\frac{E|Y|^q}{(E(|Y|^q))}\right) \\
& \leq p^{-1} + q^{-1} = 1.
\end{aligned}$$