

## Capítulo 5.

**Teorema 5.3.1** (Universalidade da Uniforme): Seja  $F$  uma CDF, uma função contínua e estritamente crescente no suporte da distribuição. Isto garante que a função inversa  $F^{-1}$  existe e é única, onde  $F^{-1} : (0, 1) \rightarrow \mathbb{R}$ . Os seguintes resultados valem:

1. Seja  $U \sim \text{Unif}(0, 1)$  e  $X = F^{-1}(U)$ . Então,  $X$  é uma v.a. com CDF  $F$ .
2. Seja  $X$  uma v.a. com CDF  $F$ . Então,  $F(X) \sim \text{Unif}(0, 1)$ .

**Prova:**

1. Tomando  $X = F^{-1}(U)$ , temos que

$$P(X \leq x) = P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x).$$

2. Seja  $U = F(X)$ , então

$$P(U \leq u) = P(F(X) \leq u) = P(X \leq F^{-1}(u)) = F(F^{-1}(u)) = u.$$

□

**Teorema** (Propriedades da Normal): Seja  $Z \sim N(0, 1)$  com PDF  $\varphi(z)$  e CDF  $\Phi(z)$ . Então, as seguintes propriedades valem:

- Simetria:  $\varphi(z) = \varphi(-z)$
- Simetria das caudas:  $\Phi(z) = 1 - \Phi(-z)$
- Simetria entre  $Z$  e  $-Z$ :  $\Phi_{-Z}(z) = \Phi_Z(z)$

**Prova:**

- A simetria é trivial, pois  $\varphi(z) = \left(\frac{1}{\sqrt{2\pi}}\right)e^{-\frac{z^2}{2}} = \left(\frac{1}{\sqrt{2\pi}}\right)e^{-\frac{(-z)^2}{2}} = \varphi(-z)$ .
- A simetria das caudas é dada por  $\Phi(z) = \int_{-\infty}^z \varphi(t)dt = \int_{-\infty}^z \varphi(-t)dt = -\int_{\infty}^{-z} \varphi(u)du = \int_{-z}^{\infty} \varphi(u)du = 1 - \int_{-\infty}^{-z} \varphi(u)du = 1 - \Phi(-z)$ .
- A simetria entre  $Z$  e  $-Z$  é dada por  $\Phi_{-Z}(z) = P(-Z \leq z) = P(Z \geq -z) = 1 - P(Z \leq -z) = 1 - \Phi(-z) = \Phi(z)$ .

□

**Definição 5.5.2** (Propriedade da não memória): Dizemos que uma v.a.  $X$  tem a propriedade da não memória se, para todo  $s, t > 0$ , vale

$$P(X > s + t \mid X > s) = P(X > t)$$

Note que se  $X \sim \text{Expo}(\lambda)$ , então  $X$  tem a propriedade da não memória. Pois

$$P(X > s + t \mid X > s) = \frac{P((X > s + t) \cap (X > s))}{P(X > s)} = \frac{P(X > s + t)}{P(X > s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P(X > t)$$

**Teorema 5.5.3:** Se  $X$  é uma v.a. contínua positiva com a propriedade da não memória, então  $X$  é uma v.a. exponencial.

**Prova 1:** Seja  $F$  a CDF de  $X$  e  $G(x) = P(X > x) = 1 - F(x)$ . Pela propriedade da não memória, temos

$$G(s+t) = G(s)G(t)$$

pois  $G(s+t) = P(X > s+t) = P(X > s+t \mid X > s)P(X > s) = P(X > t)P(X > s) = G(t)G(s)$ , a segunda igualdade decorre da lei da probabilidade total e de que  $P(X > s+t \mid X \leq s) = 0$ . Diferenciando em relação a  $s$ , temos

$$G'(s+t) = G'(s)G(t)$$

e quando  $s = 0$

$$G'(t) = G'(0)G(t)$$

resolvendo a equação diferencial, temos

$$G(t) = Ke^{-\lambda t}$$

onde  $\lambda = -G'(0)$ , e  $K = G(0) = 1$ . Portanto,  $X \sim \text{Expo}(\lambda)$ . □

**Prova 2:** Usando o resultado da **prova 1**:  $G(s+t) = G(s)G(t)$ , podemos mostrar que  $G$  é uma função exponencial. Note que  $G(0) = 1$ , pois

$$G(0) = G(0+0) = G(0)G(0) = G(0)^2$$

E se  $G(0) = 0$ , então  $G(t) = 0$  para todo  $t$ , o que é absurdo, pois  $G(t) = P(X > t)$ . Portanto,  $G(0) = 1$ . Podemos encontrar  $G(2)$  da seguinte forma

$$G(2) = G(1+1) = G(1)G(1) = G(1)^2$$

De forma similar  $G(3)$  é

$$G(3) = G(1+2) = G(1)G(2) = G(1)G(1)^2 = G(1)^3$$

podemos provar por indução que  $G(n) = G(1)^n$ , para  $n$  inteiro positivo da seguinte forma

$$G(n) = G(1)^n$$

$$G(n+1) = G(n)G(1) = G(1)^n G(1) = G(1)^{n+1}.$$

Queremos estender essa propriedade para  $n$  racional, para isso observe que

$$G(1) = G\left(\underbrace{\frac{1}{n} + \dots + \frac{1}{n}}_{n \text{ termos}}\right),$$

então

$$G(1) = G\left(\frac{1}{n}\right)G\left(\frac{1}{n}\right)\dots G\left(\frac{1}{n}\right) = G\left(\frac{1}{n}\right)^n$$

e portanto

$$G\left(\frac{1}{n}\right) = G(1)^{\frac{1}{n}}$$

e para  $m$  inteiro positivo, temos

$$G\left(\frac{m}{n}\right) = G\left(\underbrace{\frac{1}{n} + \dots + \frac{1}{n}}_{m \text{ termos}}\right) = G\left(\frac{1}{n}\right) \dots G\left(\frac{1}{n}\right) = G\left(\frac{1}{n}\right)^m = G(1)^{\frac{m}{n}}$$

A extensão para  $x$  real positivo vem com a pré-requisito de um entendimento de análise real, portanto não será feito aqui. Portanto,  $G(x) = G(1)^x$ . Por fim, observe que

$$G(x) = G(1)^x = e^{\ln(G(1)^x)} = e^{x \ln(G(1))}$$

Chamando  $\lambda = -\ln(G(1))$ , temos que  $G(x) = e^{-\lambda x}$ , ou seja,  $X \sim \text{Expo}(\lambda)$ . □

## Capítulo 6.

**Teorema 6.1.4** (Caiu no teste uma parte): Seja  $X$  uma v.a. com média  $\mu$  e mediana  $m$ . Então

- O valor  $c$  que minimiza  $E(X - t)^2$  é  $t = \mu$ .
- O valor  $c$  que minimiza  $E|X - t|$  é  $t = m$ .

**Prova:** Seja  $f(t) = E(X - t)^2 = E(X^2 - 2Xt + t^2) = E(X^2) - 2tE(X) + t^2$ . Derivando em relação a  $t$ , temos

$$f'(t) = -2E(X) + 2t$$

e igualando a zero concluímos que

$$-2E(X) + 2t = 0 \Rightarrow t = E(X) = \mu.$$

E esse ponto é de mínimo, pois  $f''(t) = 2 > 0$ , ou por conta da função ser uma parábola para cima.

Já para  $f(t) = E|X - t|$ , não podemos derivar diretamente. Portanto, vamos provar que  $E|X - t| \geq E|X - m|$  para todo  $t$ . Podemos simplificar o problema da seguinte maneira

$$E|X - t| \geq E|X - m| \Rightarrow E(|X - t| - |X - m|) \geq 0.$$

Assuma que  $t > m$  (o caso  $t < m$  é similar). Então, para  $X \leq m$  temos

$$|X - t| - |X - m| = -(X - t) - (m - X) = t - m,$$

e se  $X > m$  temos

$$|X - t| - |X - m| = X - t - (X - m) = m - t.$$

Seja  $Y = |X - t| - |X - m|$ , então, pela lei da esperança total, temos

$$\begin{aligned}
E(Y) &= E(Y \mid X \leq m)P(X \leq m) + E(Y \mid X > m)P(X > m) \\
&= E(t - m \mid X \leq m)P(X \leq m) + E(m - t \mid X > m)P(X > m) \\
&= (t - m)P(X \leq m) + (m - t)P(X > m) \\
&= (t - m)P(X \leq m) - (t - m)(1 - P(X \leq m)) \\
&= 2(t - m)P(X \leq m) - (t - m) = (t - m)(2P(X \leq m) - 1).
\end{aligned}$$

Como  $P(X \leq m) \geq \frac{1}{2}$ , temos que  $2P(X \leq m) - 1 \geq 0$ , e portanto  $E(Y) \geq 0$ . Concluindo que  $E|X - t| \geq E|X - m|$  para todo  $t$ .  $\square$

## Capítulo 7.

**Teorema 7.1.20:** Seja  $f_{xy}$  a PDF conjunta de  $X$  e  $Y$  tal que

$$f_{xy}(x, y) = g(x)h(y)$$

para todo  $x$  e  $y$ , onde  $g(x)$  e  $h(y)$  são funções não negativas. Então  $X$  e  $Y$  são independentes. Se  $g$  ou  $h$  for uma PDF válida, então a outra também é, e a PDF conjunta é o produto das marginais.

**Prova:** define

$$c = \int_{-\infty}^{\infty} h(y)dy > 0$$

podemos reescrever a PDF conjunta como

$$f_{xy}(x, y) = g(x)h(y) = cg(x)\frac{h(y)}{c}$$

então a PDF marginal de  $X$  é

$$f_X = \int_{-\infty}^{\infty} f_{xy}(x, y)dy = \int_{-\infty}^{\infty} cg(x)\frac{h(y)}{c}dy = cg(x) \int_{-\infty}^{\infty} \frac{h(y)}{c}dy = cg(x).$$

Segue que  $\int_{-\infty}^{\infty} g(x)dx = 1$  já que  $f_X$  é uma PDF válida. Analogamente,  $\frac{h(y)}{c}$  é a Pdf marginal de  $Y$ . Portanto,  $cg(x)$  e  $\frac{h(y)}{c}$  são PDFs válidas, o que conclui que  $X$  e  $Y$  são independentes.  $\square$

**Teorema 7.3.2** (independente implica corr = 0): Sejam  $X$  e  $Y$  v.a. independentes. Então,  $\text{corr}(X, Y) = 0$ .

**Prova:** Como a fórmula da correlação é

$$\text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}$$

Basta mostrar que a covariância é zero. Como  $X$  e  $Y$  são independentes, temos que

$$\text{cov}(X, Y) = E(XY) - E(X)E(Y) = E(X)E(Y) - E(X)E(Y) = 0$$

e a prova de que  $E(XY) = E(X)E(Y)$  é, no caso contínuo,

$$\begin{aligned}
E(XY) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{xy}(x, y) dx dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{X(x)} f_{Y(y)} dx dy \\
&= \int_{-\infty}^{\infty} x f_{X(x)} dx \int_{-\infty}^{\infty} y f_{Y(y)} dy = E(X)E(Y)
\end{aligned}$$

E no caso discreto

$$\begin{aligned}
E(XY) &= \sum_x \sum_y xy f_{xy}(x, y) \\
&= \sum_x \sum_y xy f_{X(x)} f_{Y(y)} \\
&= \sum_x x f_{X(x)} \sum_y y f_{Y(y)} = E(X)E(Y)
\end{aligned}$$

□

**Observação:** A recíproca não é verdadeira, ou seja,  $\text{corr}(X, Y) = 0$  não implica independência.

**Teorema 7.5.2:** Dentro de uma distribuição normal multivariada (MVN), não correlacionadas implica independência. Ou seja, se  $\mathbf{X} \sim (\mathbf{X}_1, \mathbf{X}_2)$ , onde  $\mathbf{X}_1$  e  $\mathbf{X}_2$  são subvetores de  $\mathbf{X}$ , e  $\text{corr}(\mathbf{X}_1, \mathbf{X}_2) = 0$ , então  $\mathbf{X}_1$  e  $\mathbf{X}_2$  são independentes.

**Prova:** A prova será feita para o caso bivariado, onde a prova em maiores dimensões é análoga. Seja  $\mathbf{X} = (X, Y)$ , onde  $W = sX + tY$  é uma normal univariada para todo  $s, t$ . Lembrando que a MGF de uma normal é

$$M_W(t_1, \dots, t_n) = \exp\left(t_1 E(X_1) + \dots + t_k E(X_k) + \frac{1}{2} \text{Var}(t_1 X_1 + \dots + t_k X_k)\right)$$

Então, a MGF de  $W$  é

$$M_W(s, t) = \exp\left(sE(X) + tE(Y) + \frac{1}{2} \text{Var}(sX + tY)\right)$$

Chamando  $E(X) = \mu_X$ ,  $E(Y) = \mu_Y$ ,  $\text{Var}(X) = \sigma_X^2$ ,  $\text{Var}(Y) = \sigma_Y^2$  e  $\text{corr}(X, Y) = \rho$ , temos

$$M_W(s, t) = \exp\left(s\mu_X + t\mu_Y + \frac{1}{2}(s^2\sigma_X^2 + t^2\sigma_Y^2 + 2st\rho)\right)$$

Se  $\rho = 0$ , então

$$\begin{aligned}
M_W(s, t) &= \exp\left(s\mu_X + t\mu_Y + \frac{1}{2}(s^2\sigma_X^2 + t^2\sigma_Y^2)\right) \\
&= \exp\left(s\mu_X + \frac{1}{2}s^2\sigma_X^2\right) \exp\left(t\mu_Y + \frac{1}{2}t^2\sigma_Y^2\right)
\end{aligned}$$

Porém isto é a MGF de  $X$  e  $Y$  separadamente, o que implica que  $X$  e  $Y$  são independentes. Portanto,  $\text{corr}(X, Y) = 0$  implica independência (em multivariadas)

□

## Capítulo 8.

**Teorema 8.2.1** (Convolução): Sejam  $X$  e  $Y$  v.a.s independentes discretas, então a PMF da sua soma  $T = X + Y$  é

$$P(T = t) = \sum_x P(Y = t - x)P(X = x)$$

$$P(T = t) = \sum_y P(X = t - y)P(Y = y)$$

E para o caso contínuo, a PDF da soma é

$$f_{T(t)} = \int_{-\infty}^{\infty} f_Y(t - x)f_X(x)dx$$

$$f_{T(t)} = \int_{-\infty}^{\infty} f_X(t - y)f_Y(y)dy$$

**Prova:** Para o caso discreto, temos

$$\begin{aligned} P(T = t) &= P(X + Y = t) \stackrel{\text{LOTP}}{=} \sum_x P(X + Y = t | X = x)P(X = x) \\ &= \sum_x P(Y = t - x | X = x)P(X = x) \\ &= \sum_x P(Y = t - x)P(X = x) \end{aligned}$$

De forma análoga, pode-se provar para  $P(T = t) = \sum_y P(X = t - y)P(Y = y)$ .

Note que a terceira igualdade é verdadeira pois  $X$  e  $Y$  são independentes. Para o caso contínuo:

$$\begin{aligned} f_{T(t)} &= P(T \leq t) = P(X + Y \leq t) \\ &\stackrel{\text{LOTP}}{=} \int_{-\infty}^{\infty} P(X + Y \leq t | X = x)f_X(x)dx \\ &= \int_{-\infty}^{\infty} P(Y \leq t - x)f_X(x)dx \\ &= \int_{-\infty}^{\infty} f_Y(t - x)f_X(x)dx \end{aligned}$$

Analogamente para  $f_{T(t)} = \int_{-\infty}^{\infty} f_X(t - y)f_Y(y)dy$ . □

**Teorema 8.4.3** (Gamma = soma expo): Sejam  $X_1, \dots, X_n$  v.a.s independentes com  $X_i \sim \text{Expo}(\lambda)$ . Então

$$X_1 + \dots + X_n \sim \text{Gamma}(n, \lambda)$$

**Prova:** A MGF de  $X_i$  é  $M(t) = \frac{\lambda}{\lambda - t}$ , então a MGF da soma é

$$M_{X_1+\dots+X_n}(t) = M_{X_1}(t) \times \dots \times M_{X_n}(t) = \left( \frac{\lambda}{\lambda-t} \right)^n$$

Para  $t < \lambda$ . Seja  $Y \sim \text{Gamma}(n, \lambda)$ , então a MGF de  $Y$  é

$$\begin{aligned} M_Y(t) &= E(e^{tY}) = \int_0^\infty \frac{1}{\Gamma(n)} e^{ty} (\lambda y)^n e^{-\lambda y} \frac{dy}{y} \\ &= \lambda^n \int_0^\infty \frac{1}{\Gamma(n)} y^n e^{-(\lambda-t)y} \frac{dy}{y} \\ &= \frac{\lambda^n}{(\lambda-t)^n} \int_0^\infty \frac{1}{\Gamma(n)} ((\lambda-t)y)^n e^{-(\lambda-t)y} \frac{dy}{y} \\ &= \frac{\lambda^n}{(\lambda-t)^n} \end{aligned}$$

Na segunda igualdade apenas foi retirado o  $\lambda^n$  da integral. E na terceira foi multiplicado e dividido por  $(\lambda-t)^n$  e nota-se que a integral restante é a PDF de uma v.a.  $\text{Gamma}(n, \lambda-t)$ . Portanto, como as MGFs são iguais, temos que  $X_1 + \dots + X_n \sim \text{Gamma}(n, \lambda)$ .  $\square$

## Capítulo 9.

**Teorema 9.1.5** (lei da esperança total): Seja  $A_1, \dots, A_n$  uma partição do espaço amostral, onde  $P(A_i) > 0$  para todo  $i$ , e  $X$  uma v.a. Então

$$E(X) = \sum_{i=1}^n E(X|A_i)P(A_i)$$

**Prova:** Pelo **Teorema 9.3.7** (Lei de Adão), temos que

$$E(E(X|Y)) = E(X)$$

Considerando  $Y$  uma v.a. discreta, e  $g(y) = E(X|Y = y)$ , então

$$E(E(X|Y)) = E(g(Y)) = \sum_y g(y)P(Y = y) = \sum_y E(X|Y = y)P(Y = y)$$

Fazendo a relação  $A_i = \{Y = i\}$ , temos

$$E(E(X|Y)) = \sum_y E(X|Y = y)P(Y = y) = \sum_{i=1}^n E(X|A_i)P(A_i) = E(X)$$

Portanto,  $E(X) = \sum_{i=1}^n E(X|A_i)P(A_i)$ . Para o caso contínuo, a prova é análoga.  $\square$

**Teorema 9.3.7** (Lei de Adão): Para quaisquer v.a.s  $X$  e  $Y$  vale

$$E(E(Y|X)) = E(Y)$$

**Prova:** Para  $X$  e  $Y$  discretas e  $g(x) = E(Y|X)$ , temos

$$\begin{aligned}
E(g(X)) &= \sum_x g(x)P(X = x) \\
&= \sum_x E(Y|X = x)P(X = x) \\
&= \sum_x \left( \sum_y yP(Y = y|X = x) \right) P(X = x)
\end{aligned}$$

Lembre-se que  $P(A \cap B) = P(A|B)P(B)$ , então

$$\begin{aligned}
E(g(X)) &= \sum_x \sum_y yP(Y = y|X = x)P(X = x) \\
&= \sum_y y \sum_x P(Y = y \cap X = x) \\
&= \sum_y yP(Y = y) = E(Y)
\end{aligned}$$

Portanto,  $E(E(Y|X)) = E(Y)$ . Para o caso contínuo, a prova é análoga.  $\square$

**Teorema 9.5.4** (Lei de Eva): Para quaisquer v.a.s  $X$  e  $Y$  vale

$$\text{Var}(Y) = E(\text{Var}(Y|X)) + \text{Var}(E(Y|X))$$

O nome da lei vem de que a ordem de esperanças e variâncias é EVVE, onde em inglês EVE é o nome de Eva.

**Prova:** Seja  $g(X) = E(Y|X)$ , então pela lei de Adão  $E(Y) = E(E(Y|X)) = E(g(X))$ . Então

$$\begin{aligned}
E(\text{Var}(Y|X)) &= E(E(Y^2|X) - E(Y|X)^2) \\
&= E(E(Y^2|X)) - E(E(Y|X))^2 \\
&= E(Y^2) - E(g(X))^2
\end{aligned}$$

e

$$\begin{aligned}
\text{Var}(E(Y|X)) &= \text{Var}(g(X)) \\
&= E(g(X)^2) - E(g(X))^2 \\
&= E(g(X)^2) - E(Y)^2
\end{aligned}$$

Por fim, somando as duas equações, temos

$$\begin{aligned}
E(\text{Var}(Y|X)) + \text{Var}(E(Y|X)) &= E(Y^2) - E(g(X))^2 + E(g(X)^2) - E(Y)^2 \\
&= E(Y^2) - E(Y)^2 = \text{Var}(Y)
\end{aligned}$$

$\square$

## Capítulo 10.

vazio por hora



## 11. Desigualdades

**Teorema 11.1** (Desigualdade de Markov): Seja  $X$  uma variável aleatória não negativa. Então, para todo  $a > 0$ ,

$$P(|X| \geq a) \leq \frac{E(|X|)}{a}$$

**Prova:** Seja  $Y = \frac{|X|}{a}$  e  $I_{Y \geq 1}$  a função indicadora de  $Y \geq 1$ . Temos que  $I_{Y \geq 1} = 1 \Leftrightarrow Y \geq 1$  e  $I_{Y \geq 1} = 0 \Leftrightarrow Y < 1$ . Isso implica que  $I_{Y \geq 1} \leq Y$ . Logo, aplicando a esperança em ambos os lados, temos

$$E(I_{Y \geq 1}) \leq E(Y) \Rightarrow P(Y \geq 1) \leq E(Y) \Rightarrow P\left(\frac{|X|}{a} \geq 1\right) \leq E\left(\frac{|X|}{a}\right) \Rightarrow P(X \geq a) \leq \frac{E(|X|)}{a}$$

□

**Teorema 11.2** (Desigualdade de Chebyshev): Seja  $X$  uma variável aleatória com média  $\mu$  e variância  $\sigma^2$ . Então, para todo  $a > 0$ ,

$$P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}$$

**Prova:** Seja  $Y = |X - \mu|^2$ , com isso temos

$$P(|X - \mu| \geq a) = P(|X - \mu|^2 \geq a^2) = P(Y \geq a^2)$$

Aplicando a desigualdade de Markov, temos

$$P(Y \geq a^2) \leq \frac{E(Y)}{a^2} = \frac{E(|X - \mu|^2)}{a^2} = \frac{\sigma^2}{a^2}$$

Portanto

$$P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}$$

□

**Teorema 11.3** (Desigualdade de Chernoff): Seja  $X$  uma variável aleatória com média  $\mu$ . Então, para todo  $a > 0$ ,

$$P(X \geq a) \leq e^{-ta} M_{X(t)} = \frac{E(e^{tX})}{e^{ta}}$$

onde  $M_{X(t)}$  é a função geradora de momentos de  $X$ .

**Prova:** Seja  $Y = e^{tX}$ , com isso temos

$$P(X \geq a) = P(e^{tX} \geq e^{ta}) = P(Y \geq e^{ta})$$

Aplicando a desigualdade de Markov, temos

$$P(Y \geq e^{ta}) \leq \frac{E(Y)}{e^{ta}} = \frac{E(e^{tX})}{e^{ta}}$$

Portanto

$$P(X \geq a) \leq e^{-ta} M_{X(t)} = \frac{E(e^{tX})}{e^{ta}}$$

□

**Definição 11.1** (convexidade e concavidade): Seja  $g$  uma função duas vezes diferenciável. Dizemos que  $g$  é convexa se  $g''(x) \geq 0$  para todo  $x$  e concava se  $g''(x) \leq 0$  para todo  $x$ .

**Teorema 11.4** (Desigualdade de Jensen): Seja  $X$  uma variável aleatória e  $g$  uma função convexa, então

$$g(E(X)) \leq E(g(X))$$

Se  $g$  é concava, a desigualdade é invertida, ou seja,

$$g(E(X)) \geq E(g(X))$$

onde a igualdade vale se, e somente se,  $g(X) = a + bX$  com probabilidade 1.

**Prova:** Se  $g$  é convexa, então  $g''(x) \geq 0$  para todo  $x$ . E seja  $a + bX$  a reta tangente a  $g$  em  $E(X)$ , então temos

$$g(X) \geq a + bX$$

aplicando a esperança em ambos os lados, temos

$$E(g(X)) \geq a + bE(X) = g(E(X))$$

Onde a igualdade vale pois a reta é tangente em  $E(X)$ .

Usando a mesma  $g$ , temos que  $h = -g$  é concava, usando a desigualdade encontrada acima, temos

$$E(-h(X)) \geq -h(E(X)) \implies -E(h(X)) \geq -h(E(X))$$

multiplicando por  $-1$  em ambos os lados, temos

$$E(h(X)) \leq h(E(X))$$

□