# Notes on Fundamental Inequalities

Claudio Agostinelli[*], Giacomo Pasini[†]

October 2005

## Contents

**1 Moments** **2**

**2 Base (Generalized Chebyshev) Inequality** **4**

**3 Markov Inequality** **4**

**4 Chebyshev Inequality** **5**

**5 Convex Functions** **5**

**6 Jensen Inequality** **6**

**7 Chauchy–Schwarz Inequality** **6**

[*]Department of Statistics, University of Venice

[†]Department of Economic Science, University of Venice

# 1 Moments

Given a measure space $(\Omega, \mathcal{F}, \nu)$ define the Expected Value as

$$E[h(x)] = \int_{\Omega} h(X(\omega)) d\nu(\omega) \tag{1}$$

We define the following:

- k–th moment of $X$:

$$E\left[X^k\right] \qquad \text{with } k \geq 1,\ k \in \mathbb{N} \tag{2}$$

- k–th absolute moment:

$$E\left[|X|^k\right] \qquad \text{with } k \geq 0,\ k \in \mathbb{N} \tag{3}$$

- norm of order k:

$$\| X \|_k = \left(E\left[|X|^k\right]\right)^{1/k} \text{with } k \geq 0,\ k \in \mathbb{N} \tag{4}$$

- k–th central moment:

$$\mu_k = E\left[(X-\mu)^k\right] \qquad \text{with } k \geq 1,\ k \in \mathbb{N} \tag{5}$$

  where $\mu = E[X^1]$ (the mean) is the moment of order 1.

Moments and Norms are defined as long as the correspondent integrals exist.

PROPOSITION:

Given $\nu(\Omega) < \infty$, if $E[|X|^r]$ exists and is finite then also $E[|X|^s]$ is finite $\forall\ 0 \leq s \leq r$. Further on, also $E\left[X^k\right]$ exists and it's finite $\forall\ 0 \leq k \leq r$.

Proof:

To prove the proposition we need the following disequality:

$$|x|^s \leq 1 + |x|^r \qquad \forall\ 0 \leq s \leq r \quad r, s \in \mathbb{N} \tag{6}$$

w.l.g. we consider only positive values of $x$ (the absolute value is a symmetric function).

a) $0 \leq x \leq 1$

$\Longrightarrow 0 \leq x^s \leq 1 \qquad \forall s$

$\Longrightarrow x^s \leq 1 + A \qquad$ where A is any positive quantity

$\Longrightarrow x^s \leq 1 + |x|^r, \qquad 0 \leq r \leq s$

b) $x > 1$

$\Longrightarrow x^t \geq 1$ where $t = \frac{r}{s} \geq 1$

$\Longrightarrow x^r \geq x^s$

$\Longrightarrow 1 + x^r \geq x^s$

Therefore inequality (6) is proved.

Apply the Expected Value to both sides of inequality (6):

$$E\left[|X|^s\right] \leq 1 + E\left[|X|^r\right] \qquad \forall\ 0 < s \leq r \tag{7}$$

The inequality is maintained since the Expected Value is a linear function. From (7), if $E\left[|X|^r\right]$ exists and is finite, also $E\left[|X|^s\right]$ is finite. Q.E.D.

The proof of the second part of the proposition is omitted.

EXAMPLE:
In a probability space ($\nu(\Omega) = 1$)

$$\sigma^2 = E\left[(X-\mu)^2\right] < \infty \Longleftrightarrow E\left[X^2\right] < \infty$$

Proof:

$\Longrightarrow$

It's easy to prove that $\sigma^2 = E\left[X^2\right] - \mu^2$.

$0 < \sigma^2 < \infty \Longleftrightarrow 0 < E\left[X^2\right] - \mu^2 < \infty$

Since $\mu^2$ is positive, $E\left[X^2\right]$ must be finite.

$\Longleftarrow$

By the proposition, $E[X^2] < \infty \Longrightarrow E[X^1] = \mu < \infty$. Since $\sigma^2 = E[X^2] - E[X]^2$ the result follows.

In the following sections we will limit our attention to the case $\nu(\Omega) = 1$, i.e. we restrict ourselves to probability spaces.

## 2 Base (Generalized Chebyshev) Inequality

Let $g \geq 0$ be an even, non–decreasing function in $\mathbb{R}^+$. Let $(\Omega, \mathcal{F}, \nu)$, $\nu(\Omega) = 1$ be a probability space. If $g$ is $\nu$–measurable

$$\mathrm{P}\left(|X| \geq \lambda\right) \leq \frac{E\left[g(X)\right]}{g(\lambda)} \qquad \forall\ \lambda > 0 \tag{8}$$

Proof:

$$\begin{aligned}
E\left[g(X)\right] &= \int_{\Omega} g(x) d\nu(x) \\
&= \int_{|x|\geq\lambda} g(x) d\nu(x) + \int_{|x|<\lambda} g(x) d\nu(x) \\
&\geq \int_{|x|\geq\lambda} g(x) d\nu(x) \\
&\geq g(\lambda) \int_{|x|\geq\lambda} 1 d\nu(x) \\
&= g(\lambda) \mathrm{P}\left(|X| \geq \lambda\right)
\end{aligned}$$

Q.E.D.

The only non–trivial step is the third, that is given by the fact that $g$ is positive and non-decreasing.

## 3 Markov Inequality

$$\mathrm{P}(|X| \geq \lambda) \leq \frac{E\left[|X|^r\right]}{\lambda^r} \qquad \forall\ \lambda > 0 \tag{9}$$

Proof:
The result follows from the Base inequality setting $g : x \mapsto |x|^r$.

## 4 Chebyshev Inequality

$$\mathrm{P}(|X-\mu| \geq \lambda) \leq \frac{\sigma^2}{\lambda^2} \qquad \forall\ \lambda > 0 \tag{10}$$

Proof:

Remember that $\mu = E[X]$ e $\sigma^2 = E\left[(X-\mu)^2\right]$.

The result follows from the Base inequality setting $g : (x-\mu) \mapsto (x-\mu)^2$

## 5 Convex Functions

- A real valued function $f : I \mapsto \mathbb{R}$ is convex if

$$\begin{gathered} f\left(\alpha x + (1-\alpha)y\right) \leq \alpha f(x) + (1-\alpha) f(y) \\ \forall\ x, y \in I \\ \forall\ \alpha \in [0,1] \end{gathered} \tag{11}$$

- If $f$ is a continuous function, an equivalent definition is

$$\forall\ c \in I,\ \exists b(\cdot) : f(z) \geq f(c) + b(z-c) \quad \forall\ z \in I$$

  Note that the function $b$ is not resctricted, can be either a positive or a negative–valued function.

- If $f$ is also differentiable, an equivalent definition is $f'(x)$ is an increasing function in $x$.

- If $f$ is twice differentiable, $f$ is convex is equivalent to $f''(x) \geq 0\ \forall x \in I$

THEOREM: Supporting Hyperplane

Let $f : I \mapsto \mathbb{R}$ be a convex function. Then $\forall x_0 \in I$ exists a linear function $l(x) = a_0 + b_0 x$ such that $f(x) \geq l(x)\ \forall x \neq x_0$, and $f(x_0) = l(x_o)$. The function $l(x)$ is called supporting hyperplane[1].

[1] it's called hyperplane because the theorem holds also for convex functions in $\mathbb{R}^n$

# 6 Jensen Inequality

Given a probability space $(\Omega, \mathrm{F}, \nu)$, $\nu(\Omega) = 1$, let $g : (a, b) \mapsto \mathbb{R}$ be a convex $\nu$–measurable function, where $-\infty \leq a < b \leq +\infty$.

If $E[X]$ exists and it's finite (and therefore $E[X] \in (a, b)$), then

$$g\left(E[X]\right) \leq E\left[g(X)\right] \tag{12}$$

For strictly convex functions (12) holds with equality iff $X = E[X]$ *a.e.*

Proof:

Let $l(\cdot)$ be the supporting hyperplane of $g(\cdot)$ in $x_0 \equiv E[X]$. Then

$$\begin{aligned} E[g(X)] &\geq E[l(X)] \\ &= l(E[X]) \qquad \text{since } l \text{ is a linear function and } \nu(\Omega) = 1 \\ &= g(E[X]) \qquad \text{since } g(\cdot) = l(\cdot) \text{ in } x_0 \equiv E[X] \end{aligned}$$

$g(X) - l(X) \geq 0$, and therefore $E[g(X)] = E[l(X)]$ iff $g(X) = l(X)$ *a.e.* $\Longleftrightarrow$ $X = E[X]$ *a.e.*.

Q.E.D.

# 7 Chauchy–Schwarz Inequality

If the $E[X^2]$, $E[Y^2]$ exist and are finite, then also $E[|XY|]$ exists and is finite. Moreover,

$$E[|XY|]^2 \leq E\left[X^2\right] E\left[Y^2\right] \tag{13}$$

Proof:

(a) Existence:

$$\begin{aligned} (X - Y)^2 &\geq 0 \\ X^2 - 2|XY| + Y^2 &\geq 0 \\ 2|XY| &\leq X^2 + Y^2 \\ |XY| &\leq X^2 + Y^2 \\ E[|XY|] &\leq E[X^2] + E[Y^2] \end{aligned}$$

The transformed r.v. $|XY|$ can only assume positive values, therefore $E[|XY|] = \int_\Omega |XY(\omega)| d\nu(\omega)$ exists. From the last inequality (Triangular Inequality), finiteness of $E[X^2]$ and $E[Y^2]$ imply finiteness of $E[|XY|]$. By the proposition proved in the first section, existence and finiteness of $E[|XY|]$ imply existence and finiteness for $E[XY]$

(b) Chaucy–Schwartz inequality:

Given that $E[|XY|]^2 = E[XY]^2$, we prove $E[XY]^2 \leq E\left[X^2\right] E\left[Y^2\right]$.

$\forall z \in \mathbb{R}$ we have

$$0 \leq E\left[(zX+Y)^2\right] = z^2 E[X^2] + 2zE[XY] + E[Y^2]$$

The last equality comes from the linearity of $E[\cdot]$. The discriminant of the quadratic equation in $z$ is

$$\sqrt{4\left(E[XY]^2 - E[X^2]E[Y^2]\right)},$$

And since the equation is always non–negative,

$$E[XY]^2 - E[X^2]E[Y^2] \leq 0$$

Q.E.D.

EXAMPLE: Inequalities for correlation coefficient

If $X,\ Y$ are r.v. with finite (and positive) variance, then

$$-1 \leq \rho \leq 1 \qquad \text{where } \rho = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

Proof:

Let $X' = X - E[X]$, $Y' = Y - E[Y]$, and apply the Chaucy–Schwartz inequality to the transformed variables:

$$\begin{aligned} E\Big[\big|(X-E[X])(Y-E[Y])\big|\Big]^2 &\leq E\left[(X-E[X])^2\right] E\left[(Y-E[Y])^2\right] \\ E\Big[\big|(X-E[X])(Y-E[Y])\big|\Big] &\leq \sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)} \end{aligned}$$

$|\cdot|$ is a convex function, by Jensen inequality $|E[Z]| \leq E[|Z|]$:

$$\left|E[(X-E[X])(Y-E[Y])]\right| \leq E\Big[\big|(X-E[X])(Y-E[Y])\big|\Big] \leq \sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}$$

$$\begin{aligned} |\mathrm{Cov}(X,Y)| &\leq \sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)} \\ |\rho| &\leq 1 \\ -1 \leq \rho &\leq 1 \end{aligned}$$

Q.E.D.