Basic Inequalities in Probability[1]

Standard Notations: $\mu_X = \mathbb{E}X$ (expected value), $\sigma_X^2 = \mathbb{E}(X-\mu_X)^2$ (variance).

FUNDAMENTAL INEQUALITIES.

(1) RE-ARRANGEMENT: for every ream numbers $a_1 \le a_2 \le \cdots \le a_n$ and $b_1 \le b_2 \le \cdots \le b_n$, and every permutation τ of the set $\{1, 2, \ldots, n\}$,

$$a_1b_n + a_2b_{n-1} + \ldots + a_nb_1 \le a_1b_{\tau(1)} + a_2b_{\tau(2)} + \ldots + a_nb_{\tau(n)} \le a_1b_1 + a_2b_2 + \ldots + a_nb_n.$$

Proof: induction can work.

Example: if $x, y, z > 0$, then $\frac{x}{y+z} + \frac{y}{x+z} + \frac{z}{x+y} \ge \frac{3}{2}$. Indeed, with no loss of generality, assume that $x \le y \le z$, so that $y+z \ge x+z \ge x+y$ and $1/(y+z) \le 1/(x+z) \le 1/(x+y)$. Now take $a_1 = 1/(y+z), a_2 = 1/(x+z), a_3 = 1/(x+y)$, $b_1 = x, b_2 = y, b_3 = z$, and note that $3 = a_1(b_2+b_3) + a_2(b_1+b_3) + a_3(b_1+b_2) \le 2(a_1b_1 + a_2b_2 + a_3b_3)$.

(2) POWER MEAN: If $a_1 > 0, a_2 > 0, \ldots, a_n > 0$, and

$$M_p = \begin{cases} \left(\frac{1}{n}\sum_{k=1}^n a_k^p\right)^{1/p}, & p \ne 0, \pm\infty, \\ M_0 = (a_1a_2\cdots a_n)^{1/n}, & p = 0, \\ M_{+\infty} = \max(a_1, \ldots, a_n), & p = +\infty, \\ M_{-\infty} = \min(a_1, \ldots, a_n), & p = -\infty, \end{cases}$$

then

$$\lim_{p\to 0} M_p = M_0,\ \lim_{p\to -\infty} M_p = M_{-\infty},\ \lim_{p\to +\infty} M_p = M_{+\infty},$$

and the function $p \to M_p$ is *strictly increasing* unless $a_1 = a_2 = \ldots = a_n$.

Proof: induction can work.

Special names: M_1 is arithmetic mean (AG), M_0 is geometric mean (GM), M_{-1} is harmonic mean (HM). The (famous) AG/GM/HM inequality, $M_1 \ge M_0 \ge M_{-1}$, is a particular case of the power mean inequality.

CONCENTRATION INEQUALITIES.

(1) MARKOV [1880]: if $Y > 0$, then

$$\mathbb{P}(Y \ge a) \le \frac{\mathbb{E}Y}{a}.$$

Proof. $\mathbb{E}Y \ge \mathbb{E}YI_{Y\ge a} \ge a\mathbb{E}I_{Y\ge 0} = a\mathbb{P}(Y \ge a)$.

(2) CHEBYSHEV [1865]: with $\mu_X = \mathbb{E}X$, $\sigma_X^2 = \mathbb{E}(X-\mu_X)^2$,

$$\mathbb{P}(|X-\mu_X| \ge a) \le \frac{\sigma_X^2}{a^2}.$$

Proof. Apply Markov with $Y = (X-\mu_X)^2$.

VARIATIONS.

- STANDARTIZED:

$$\mathbb{P}(|X-\mu_X| \ge k\sigma_X) \le \frac{1}{k^2}.$$

- CANTELLI [1928]:

$$\mathbb{P}(X-\mu_X \ge a) \le \frac{\sigma_X^2}{\sigma_X^2 + a^2}.$$

Proof. For $t > 0$, by Markov, $\mathbb{P}(X-\mu_X+t > a+t) = \mathbb{P}\big((X-\mu_X+t)^2 > (a+t)^2\big) \le (\sigma_X^2+t^2)/(a+t)^2$. Direct computations show that the right hand side is minimized by taking $t = \sigma_X^2/a$.

- VYSOCHANSKIJ-PETUNIN [1980]: if X is *unimodal*, then

$$\mathbb{P}(|X-\mu_X| > k\sigma_X) \le \frac{4}{9k^2}.$$

[1]Sergey Lototsky, USC

(3) CHERNOFF [1955]: If $M_X(t) = \mathbb{E}e^{tX}$ exists for all $t > 0$, and $a > 0$, then

$$\mathbb{P}(X \geq a) \leq e^{\ln M_X(t) - at},$$

with subsequent minimization of the right-hand side with respect to $t > 0$.
Proof. Use Markov with $Y = e^{tX}$.
Example. If X is standard normal, then $M_X(t) = e^{t^2/2}$, so that $(t^2/2) - at \geq -a^2/2$, with the lower bound achieved for $t = a/2$, and therefore $\mathbb{P}(X \geq a) \leq e^{-a^2/2}$.

(4) PALEY-ZYGMUND [1932]: if $Y > 0$ and $0 < \theta < 1$, then

$$\mathbb{P}(Y > \theta\mu_Y) \geq (1-\theta)^2 \frac{\mu_Y^2}{\sigma_Y^2 + \mu_Y^2}.$$

Proof. Keeping in mind that $\mathbb{E}Y^2 = \sigma_Y^2 + \mu_Y^2$, $\mu_Y = \mathbb{E}YI_{Y \leq \theta\mu_Y} + \mathbb{E}YI_{Y > \theta\mu_Y}$. Then $\mathbb{E}YI_{Y < \theta\mu_Y} \leq \theta\mu_Y$ (obviously), and $\mathbb{E}YI_{Y > \theta\mu_Y} \leq \sqrt{\mathbb{E}Y^2}\sqrt{\mathbb{P}(Y > \theta\mu_Y)}$ (Cauchy-Schwarz; see below).

MOMENT INEQUALITIES.

(1) JENSEN [1905]: If $g = g(x),\ x \in \mathbb{R}$, is *convex*, then

$$\mathbb{E}g(X) \geq g(\mathbb{E}X) \quad (\text{e.g. } \mathbb{E}e^X \geq e^{\mu_X}.)$$

If $f = f(x)$ is *concave*, then

$$\mathbb{E}f(X) \leq f(\mathbb{E}X) \quad (\text{e.g. } X > 0 \ \Rightarrow\ \mathbb{E}\ln X \leq \ln \mu_X.)$$

Proof: g convex $\Leftrightarrow$ for every $x_0 \in \mathbb{R}$ there is a number $C \in \mathbb{R}$ such that $g(x) \geq g(x_0) + C(x - x_0)$; if g' exists, then $C = g'(x_0)$ [the graph of g is above the tangent line at x_0]. Now put $x = X, x_0 = \mu_X$, and take expected value on both sides.

(2) LYAPUNOV [1900]: if $0 < p < r$, then

$$\big(\mathbb{E}|X|^p\big)^{1/p} \leq \big(\mathbb{E}|X|^r\big)^{1/r}.$$

Proof: Use Jensen with $g(x) = |x|^{r/p}$ and $|X|^p$ instead of X.

(3) HÖLDER [1885]: if $p > 1, q > 1$, and $(1/p) + (1/q) = 1$, then

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

The inequality is strict unless $X = cY$ for some non-random number c.
Proof. Using concavity of the log function, argue that $ab \leq (a^p/p) + (b^q/q), a, b > 0$. Then set $a = |X|/\big(\mathbb{E}|X|^p\big)^{1/p}$, $b = |Y|/\big(\mathbb{E}|Y|^q\big)^{1/q}$, and take expectation on both sides. Note that the Hölder inequality is trivial if $\mathbb{E}|X|^p = 0$ and/or $\mathbb{E}|Y|^q = 0$.

(4) CAUCHY-BUNYAKOVSKY-SCHWARZ [1820→1855→1885]:

$$\mathbb{E}|XY| \leq \sqrt{\mathbb{E}X^2}\sqrt{\mathbb{E}Y^2}.$$

Proof: take $p = q = 2$.

(5) MINKOWSKI [1900]: if $p \geq 1$, then

$$\big(\mathbb{E}|X+Y|^p\big)^{1/p} \leq \big(\mathbb{E}|X|^p\big)^{1/p} + \big(\mathbb{E}|Y|^p\big)^{1/p}$$

(that is, the functional $X \mapsto \big(\mathbb{E}|X|^p\big)^{1/p}$ satisfies the *triangle inequality* and thus defines a *norm* on the space of random variables with finite p-th moment.)
Proof. $p = 1$ is obvious. For $p > 1$, take $q = p/(p-1)$ [so that $(1/p) + (1/q) = 1$], note that $|X+Y|^p = |X+Y| \cdot |X+Y|^{p-1} \leq |X| \cdot |X+Y|^{p-1} + |Y| \cdot |X+Y|^{p-1}$, and then, by Hölder, $\mathbb{E}\big(|X| \cdot |X+Y|^{p-1}\big) \leq \big(\mathbb{E}|X|^p\big)^{1/p}\big(\mathbb{E}|X+Y|^p\big)^{1/q}$, $\mathbb{E}\big(|Y| \cdot |X+Y|^{p-1}\big) \leq \big(\mathbb{E}|Y|^p\big)^{1/p}\big(\mathbb{E}|X+Y|^p\big)^{1/q}$. It remains to combine the inequalities: $\mathbb{E}|X+Y|^p \leq \Big(\big(\mathbb{E}|X|^p\big)^{1/p} + \big(\mathbb{E}|Y|^p\big)^{1/p}\Big)\big(\mathbb{E}|X+Y|^p\big)^{1/q}$ and then simplify, keeping in mind that $1 - (1/q) = 1/p$.

THE PEOPLE.

A.-L. Cauchy (1789–1857), French.
P. L. Chebyshev (1821–1894), Russian.
A. A. Markov (1856–1922), Russian.
J. L. W. V. Jensen (1859–1925), Danish.
F. P. Cantelli (1875–1966), Italian.
R. E. A. C. Paley (1907–1933), English.
Y. I. Petunin (1937–2011), Soviet.
V. Ya. Bunyakovsky (1804–1889), Russian.
K. H. A. Schwarz (1843–1921), German.
A. M. Lyapunov (1857–1918), Russian.
H. Minkowski (1864–1909), German.
A. Zygmund (1900-1992), Polish-American.
H. Chernoff (b. 1923), American.