Probabilistic Inequalities - I

Prof. Dan A. Simovici

UMB
1. Markov and Chebyshev Inequalities

2. Hoeffding’s Inequality
Markov Inequality

**Theorem**

Let $X$ be a non-negative random variable. For every $a \geq 0$ we have

$$P(X \geq a) \leq \frac{E(X)}{a}.$$
Proof in the discrete case

Suppose that

\[ X : \begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ p_1 & p_2 & \cdots & p_n \end{pmatrix}, \]

where \( x_1 < x_2 < \cdots < x_n \). Suppose further that

\[ x_1 < x_2 < \cdots < x_k < a \leq x_{k+1} < \cdots < x_n. \]

Then \( P(X \geq a) = p_{k+1} + \cdots + p_n \).

Since

\[ E(X) = x_1 p_1 + \cdots + x_k p_k + x_{k+1} p_{k+1} + \cdots + x_n p_n \]
\[ \geq x_{k+1} p_{k+1} + \cdots + x_n p_n \geq a(p_{k+1} + \cdots + p_n) \]
\[ = a P(X \geq a), \]

we obtain Markov Inequality.
Chebyshev Inequality

Recall that the variance of a random variable $X$ is the number $\text{var}(X) = E((X - E(X))^2)$. Equivalently, $\text{var}(X) = E(X^2) - (E(X))^2$.

**Theorem**

*We have*

$$P(|X - E(X)| \geq a) \leq \frac{\text{var}(X)}{a^2}.$$
Proof

The Markov Inequality applied to the random variable \( Y = (X - E(X))^2 \) and to \( a^2 \) is:

\[
P(Y \geq a^2) \leq \frac{E(Y)}{a^2}.
\]

This amounts to

\[
P((X - E(X))^2 \geq a^2) \leq \frac{E((X - E(X))^2)}{a^2}.
\]

This is equivalent to

\[
P(|X - E(X)| \geq a) \leq \frac{\text{var}(X)}{a^2},
\]

which is the Chebyshev’s Inequality.
Lemma

Let $L$ be the function defined as

$$L(x) = -xp + \log(1 - p + pe^x).$$

We have $L(x) \leq \frac{x^2}{8}$ for $x \geq 0$. 
Proof

We need to show that \( f(x) = \frac{x^2}{8} - L(x) \geq 0 \). Since \( L(0) = 0 \) we have \( f(0) = 0 \). Note that:

\[
\begin{align*}
    f'(x) &= \frac{x}{4} - p + \frac{pe^x}{1 - p + pe^x} \\
         &= \frac{x}{4} - p + 1 + \frac{p - 1}{1 - p + pe^x} \\
    f''(x) &= \frac{1}{4} - \frac{(p - 1)pe^x}{(1 - p + pe^x)^2} \\
         &= \frac{(1 - p - pe^x)^2}{4(1 - p + pe^x)^2}.
\end{align*}
\]

Note that \( f''(x) \geq 0 \) and \( f'(0) = 0 \).
Therefore, $f'$ is increasing and $f'(x) \geq 0$ for $x \geq 0$.
Since $f'(x) \geq 0$ and $f(0) = 0$, it follows that $x \geq 0$ implies $f(x) \geq 0$, which we need to prove.
Lemma

Let $X$ be a random variable that takes values in the interval $[a, b]$ such that $E(X) = 0$. Then, for every $\lambda > 0$ we have

$$E(e^{\lambda X}) \leq e^{\frac{\lambda^2(b-a)^2}{8}}.$$
Proof

Since $f(x) = e^{\lambda x}$ is a convex function, we have that for every $t \in [0, 1]$ and $x \in [a, b]$,

$$f(x) \leq (1 - t)f(a) + tf(b).$$

For $t = \frac{x-a}{b-a} \in [0, 1]$ we have $e^{\lambda x} \leq \frac{b-x}{b-a} e^{\lambda a} + \frac{x-a}{b-a} e^{\lambda b}$.

Applying the expectation we obtain:

$$E(e^{\lambda X}) \leq \frac{b - E(X)}{b - a} e^{\lambda a} + \frac{E(X) - a}{b - a} e^{\lambda b}$$

$$= \frac{b}{b - a} e^{\lambda a} - \frac{a}{b - a} e^{\lambda b},$$

because $E(X) = 0$. 
Proof (cont’d)

If \( h = \lambda(b - a) \), \( p = \frac{-a}{b-a} \) and \( L(h) = -hp + \log(1 - p + pe^h) \), then
\[
-hp = \lambda a, \quad 1 - p = 1 + \frac{a}{b-a} = \frac{b}{b-a}, \quad \text{and}
\]

\[
e^{L(h)} = e^{-hp}(1 - p + pe^h)
\]
\[
= e^{\lambda a} \left( \frac{b}{b-a} - \frac{a}{a-b} e^{\lambda(b-a)} \right)
\]
\[
= \frac{b}{b-a} e^{\lambda a} - \frac{a}{a-b} e^{\lambda b}.
\]

This implies

\[
\frac{b}{b-a} e^{\lambda a} - \frac{a}{b-a} e^{\lambda b} = e^{L(h)} \leq e^{\frac{\lambda^2(b-a)^2}{8}}
\]

because we have shown that \( L(h) \leq \frac{h^2}{8} = \frac{\lambda^2(b-a)^2}{8} \). This gives the desired inequality.
Hoeffding’s Theorem

Theorem

Let \((Z_1, \ldots, Z_m)\) be a sequence of iid random variables and let

\[ \tilde{Z} = \frac{1}{m} \sum_{i=1}^{m} Z_i. \]

Assume that

\[ E(\tilde{Z}) = \mu \text{ and that } P(a \leq Z_i \leq b) = 1 \]

for \(1 \leq i \leq m\). Then, for every \(\epsilon > 0\) we have

\[ P(|\tilde{Z} - \mu| > \epsilon) \leq 2e^{-\frac{2m\epsilon^2}{(b-a)^2}}. \]
Proof

Let $X_i = Z_i - E(Z_i) = Z_i - \mu$ and $\tilde{X} = \frac{1}{m} \sum_{i=1}^{m} X_i$.

Note that $E(X_i) = 0$ for $1 \leq i \leq m$, which implies $E(\tilde{X}) = 0$.

Thus,

$$\tilde{Z} - \mu = \left( \frac{1}{m} \sum_{i=1}^{m} Z_i \right) - \mu = \frac{1}{m} \sum_{i=1}^{m} (Z_i - \mu)$$

$$= \frac{1}{m} \sum_{i=1}^{m} X_i = \tilde{X}$$

and

$$P(|\tilde{Z} - \mu| > \epsilon) = P(|\tilde{X}| > \epsilon) = P(\tilde{X} > \epsilon) + P(\tilde{X} < -\epsilon).$$
Let $\epsilon$ and $\lambda$ be two positive numbers. Note that $P(\tilde{X} \geq \epsilon) = P(e^{\lambda \tilde{X}} \geq e^{\lambda \epsilon})$. By Markov Inequality,

$$P(e^{\lambda \tilde{X}} \geq e^{\lambda \epsilon}) \leq \frac{E(e^{\lambda \tilde{X}})}{e^{\lambda \epsilon}}.$$ 

Since $X_1, \ldots, X_m$ are independent, we have

$$E(e^{\lambda \tilde{X}}) = E \left( \prod_{i=1}^{m} e^{\frac{\lambda X_i}{m}} \right) = \prod_{i=1}^{m} E(e^{\frac{\lambda X_i}{m}}).$$
Proof (cont’d)

By Lemma 2, for every $i$ we have

$$E \left( e^{\frac{\lambda x_i}{m}} \right) \leq e^{\frac{\lambda^2 (b-a)^2}{8m^2}}.$$

Therefore,

$$P(\tilde{X} \geq \epsilon) \leq e^{-\lambda \epsilon} \prod_{i=1}^{m} e^{\frac{\lambda^2 (b-a)^2}{8m^2}} = e^{-\lambda \epsilon} e^{\frac{\lambda^2 (b-a)^2}{8m}}.$$

Choosing $\lambda = \frac{4m \epsilon}{(b-a)^2}$ yields

$$P(\tilde{X} \geq \epsilon) \leq e^{-\frac{2m \epsilon^2}{(b-a)^2}}.$$

The same arguments applied to $-\tilde{X}$ yield $P(\tilde{X} \leq -\epsilon) \leq e^{-\frac{2m \epsilon^2}{(b-a)^2}}$. 
By applying the union property of probabilities we have

\[ P(|\tilde{X}| > \epsilon) = P(\tilde{X} > \epsilon) + P(\tilde{X} < -\epsilon) \leq 2e^{-\frac{2m\epsilon^2}{(b-a)^2}}. \]