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 $a \geqslant 0$ we have

$$P(X \geqslant a) \leqslant \frac{E(X)}{a}.$$

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

$$x_1 < x_2 < \cdots x_k < a \leqslant x_{k+1} < \cdots < x_n.$$

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

$$\begin{aligned} E(X) &= x_1 p_1 + \cdots + x_k p_k + x_{k+1} p_{k+1} + \cdots + x_n p_n \\ &\geqslant x_{k+1} p_{k+1} + \cdots + x_n p_n \geqslant a(p_{k+1} + \cdots + p_n) \\ &= aP(X \geqslant a), \end{aligned}$$

we obtain Markov Inequality.

Chebyshev Inequality

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

Theorem

We have

$$P(|X - E(X)| \geqslant a) \leqslant \frac{var(X)}{a^2}.$$

Proof

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

$$P(Y \geqslant a^2) \leqslant \frac{E(Y)}{a^2}.$$

This amounts to:

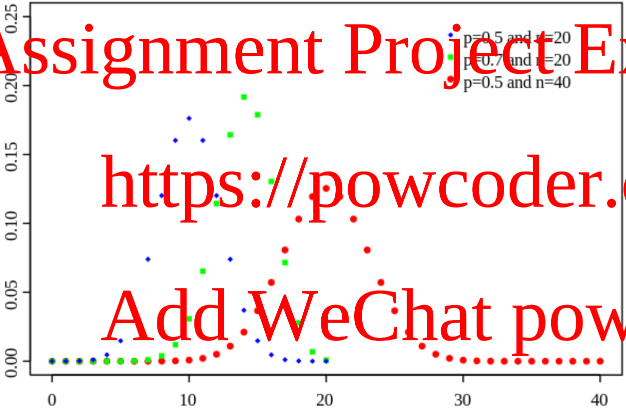
$$P((X - E(X))^2 \geqslant a^2) \leqslant \frac{E((X - E(X))^2)}{a^2}.$$

This is equivalent to

$$P(|X - E(X)| \geqslant a) \leqslant \frac{var(X)}{a^2},$$

which is the Chebyshev's Inequality.

The probability distribution of a binomial variable:



Example

If X is a binomial variable,

$$X : \begin{pmatrix} 0 & 1 & \cdots & k & \cdots & n \\ q^n & \binom{n}{1} q^{n-1} p & \cdots & \binom{n}{k} q^{n-k} p^k & \cdots & p^n \end{pmatrix},$$

we have

$$P(|X - np| \geqslant a) \leqslant \frac{npq}{a^2},$$

which can be written also as

$$P((X > np + a) \vee (X < np - a)) \leqslant \frac{npq}{a^2}.$$

Since the distribution is symmetric relative to np this is equivalent to

$$P(X > np + a) \leqslant \frac{npq}{2a^2} \text{ and } P(X < np - a) \leqslant \frac{npq}{2a^2}.$$

Lemma

Let L be the function defined as

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

We have $L(x) \leqslant \frac{x^2}{8}$ for $x \geqslant 0$.

Proof

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

$$
\begin{aligned}
f'(x) &= \frac{x}{4} - p + \frac{pe^x}{1 - p + pe^x} \\
&= \frac{x}{4} - p + 1 + \frac{p - 1}{1 - p + pe^x} \\
f''(x) &= \frac{1}{4} - \frac{(p-1)pe^x}{(1 - p + pe^x)^2} \\
&= \frac{(1 - p - pe^x)^2}{4(1 - p + pe^x)^2}.
\end{aligned}
$$

Note that $f''(x) \geqslant 0$ and $f'(0) = 0$.

Proof (cont'd)

Therefore, f' is increasing and $f'(x) \geqslant 0$ for $x \geqslant 0$.
Since $f'(x) \geqslant 0$ and $f(0) = 0$, it follows that $x \geqslant 0$ implies $f(x) \geqslant 0$, 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

$$E[e^{\lambda X}] \leqslant e^{\frac{\lambda^2 (b-a)^2}{8}}.$$

Proof

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

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

For $t = \frac{x-a}{b-a} \in [0, 1]$ we have $e^{\lambda x} \leqslant \frac{b-x}{b-a} e^{\lambda a} + \frac{x-a}{b-a} e^{\lambda b}$.
Applying the expectation we obtain:

$$\begin{aligned} E(e^{\lambda X}) &\leqslant \frac{b - E(X)}{b - a} e^{\lambda a} + \frac{E(X) - a}{b - a} e^{\lambda b} \\ &= \frac{b}{b - a} e^{\lambda a} - \frac{a}{b - a} e^{\lambda b}, \end{aligned}$$

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 $-hp = \lambda a$, $1 - p = 1 + \frac{a}{b-a} = \frac{b}{b-a}$, and

$$\begin{aligned} e^{L(h)} &= e^{-hp}(1 - p + pe^h) \\ &= e^{\lambda a}\left(\frac{b}{b-a} - \frac{a}{a-b} e^{\lambda(b-a)}\right) \\ &= \frac{b}{b-a} e^{\lambda a} - \frac{a}{a-b} e^{\lambda b}. \end{aligned}$$

This implies

$$\frac{b}{b-a} e^{\lambda a} - \frac{a}{b-a} e^{\lambda b} = e^{L(h)} \leqslant e^{\frac{\lambda^2(b-a)^2}{8}}$$

because we have shown that $L(h) \leqslant \frac{h^2}{8} = \frac{\lambda^2(b-a)^2}{8}$. 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

$$E(\tilde{Z}) = \mu \text{ and that } P(a \leqslant Z_i \leqslant b) = 1$$

for $1 \leqslant i \leqslant m$. Then, for every $\epsilon > 0$ we have

$$P(|\tilde{Z} - \mu| > \epsilon) \leqslant 2e^{-\frac{2m\epsilon^2}{(b-a)^2}}.$$

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 $1 \leqslant i \leqslant m$, which implies $E(\tilde{X}) = 0$.

Thus,

$$
\begin{aligned}
\tilde{Z} - \mu &= \left(\frac{1}{m}\sum_{i=1}^{m} Z_i\right) - \mu = \frac{1}{m}\sum_{i=1}^{m}(Z_i - \mu) \\
&= \frac{1}{m}\sum_{i=1}^{m} X_i = \tilde{X}
\end{aligned}
$$

and

$$
\begin{aligned}
P(|\tilde{Z} - \mu| > \epsilon) &= P(|\tilde{X}| > \epsilon) \\
&= P(\tilde{X} > \epsilon) + P(\tilde{X} < -\epsilon).
\end{aligned}
$$

Proof (cont'd)

Let ϵ and λ be two positive numbers. Note that $P(\tilde{X} \geqslant \epsilon) = P(e^{\lambda \tilde{X}} \geqslant e^{\lambda\epsilon})$. By Markov Inequality,

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

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

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

Therefore,

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

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

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

The same arguments applied to $-\tilde{X}$ yield $P(\tilde{X} \leqslant -\epsilon) \leqslant e^{-\frac{2m\epsilon^2}{(b-a)^2}}$.

By applying the union property of probabilities we have

$$
\begin{aligned}
P(|\tilde{X}| > \epsilon) &= P(\tilde{X} > \epsilon) + P(\tilde{X} < -\epsilon) \\
&\leqslant 2e^{-\frac{2m\epsilon^2}{(b-a)^2}}.
\end{aligned}
$$

A special case of Hoeffding's Theorem

Theorem

Let $X_1, \ldots, X_m$ be m independent and identially distributed Bernoulli random variables with $P(X_i = 1) = p$ for $1 \leqslant i \leqslant m$. Let $S = X_1 + \cdots + X_m$ be the binomial variable indicating the total number of succeses, so $E[S] = pm$. For $\epsilon \in [0, 1]$ we have:

$$P\left((S > m(p+\epsilon)) \vee (S < m(p-\epsilon))\right) \leqslant 2e^{-2m\epsilon^2}$$

Proof

Note that $E[S] = mp$. Since $\tilde{Z} = \frac{S}{m}$ it follows that $E[\tilde{Z}] = p$. Then, for $\epsilon > 0$, $a = 0$, and $b = 1$ we have

$$P(|\tilde{Z} - p| > \epsilon) \leqslant 2e^{-2m\epsilon^2},$$

or

$$P(|S - mp| > m\epsilon) \leqslant 2e^{-2m\epsilon^2},$$

Thus,

$$P\left((S > m(p+\epsilon)) \vee (S < m(p-\epsilon)\right) \leqslant 2e^{-2m\epsilon^2}.$$

In new of the symmetry of S with respect to mp, the previous inequality amounts to

$$P(S > m(p+\epsilon)) \leqslant e^{-2m\epsilon^2},$$

and

$$P(S < m(p-\epsilon)) \leqslant e^{-2m\epsilon^2}.$$