MACHINE LEARNING
Probabilistic Inequalities - Supplement

Prof. Dan A. Simovici

UMB

Theorem

We have the following inequalities:

$$1 + a \leqslant e^a \text{ for } a \in \mathbb{R}, \quad (1)$$

$$\left(1 + \frac{b}{x}\right) \leqslant e^b \text{ for } b \in \mathbb{R} \text{ and } x > 0, \quad (2)$$

$$-\frac{1}{2}\epsilon^2 \leqslant \epsilon - (1+\epsilon)\ln(1+\epsilon) \leqslant -\frac{1}{3}\epsilon^2 \text{ for } 0 \leqslant \epsilon \leqslant 1 \quad (3)$$

$$-\frac{1}{2}\epsilon^2 \geqslant \epsilon - (1+\epsilon)\ln(1+\epsilon) \text{ for } -1 < \epsilon \leqslant 0. \quad (4)$$

Proof of Inequality (1)

Let $g(a) = e^a - (1 + a)$ for $a \in \mathbb{R}$. Then $g''(a) = e^a > 0$ for all $a \in \mathbb{R}$, while $g'(0) = 0$. Hence $g(a) \geqslant g(0) = 0$ for all $a \in \mathbb{R}$, which concludes the proof.

Proof of Inequality (2)

Putting $a = \frac{b}{x}$ and raising both sides of Inequality (1) to the x^{th} power yields (2).

Proof of Inequality (3)

For $q \in \{0, 1\}$ let

$$f_q(\epsilon) = \epsilon - (1+\epsilon)\ln(1+\epsilon) + \frac{1}{2}\epsilon^2 - \frac{1}{6}q\epsilon^3$$

and $-1 < \epsilon \leqslant 1$. We have:

$$\begin{aligned} f_q'(\epsilon) &= -\ln(1+\epsilon) + \epsilon - \frac{1}{2}q\epsilon^2 \\ f_q''(\epsilon) &= -\frac{1}{1+\epsilon} + 1 - q\epsilon \\ f_q'''(\epsilon) &= \frac{1}{(1+\epsilon)^2} - q. \end{aligned}$$

Proof of Inequalities (3) and (4) cont'd

We have the following sign variations if f_q and its derivatives:

q	0			1		
	$\epsilon < 0$	$\epsilon = 0$	$\epsilon > 0$	$\epsilon < 0$	$\epsilon = 0$	$\epsilon > 0$
$f_q'''(\epsilon)$	$+$	$+$	$+$	$+$	0	$-$
$f_q''(\epsilon)$	$-$	0	$+$	$-$	0	$-$
$f_q'(\epsilon)$	$+$	0	$+$	$+$	0	$-$
$f_q(\epsilon)$	$-$	0	$+$	$-$	0	$-$

The sign variation of f_0 implies Inequality (4) and the left half of (3). Also, from the sign variation of f_1 we get for $0 \leqslant \epsilon \leqslant 1$:

$$\epsilon - (1+\epsilon) \leqslant (1+\epsilon) \leqslant -\epsilon^2 \left(\frac{1}{2} - \frac{1}{6}\epsilon \right) \leqslant \frac{1}{3}\epsilon^2.$$

Let $n \in \mathbb{N}$ and let $p_1, \ldots, p_n \in \mathbb{R}$ with $0 \leqslant p_i \leqslant 1$. Define $p = \frac{p_1 + \cdots + p_n}{n}$ and $m = np$ and let $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ be independent, 0-1 random variables with

$$P(X_i = 1) = p_i \text{ and } P(Y_i = 1) = p \text{ for } 1 \leqslant i \leqslant n.$$

We are interested in the behavior of the random variables

$$\begin{aligned} S &= X_1 + \cdots + X_n, \text{ and} \\ S' &= Y_1 + \cdots + Y_n. \end{aligned}$$

Theorem

For $S = X_1 + \cdots + X_n$ *we have:*

$$\begin{aligned} P(S \leqslant (1-\epsilon)m) &\leqslant e^{-\frac{\epsilon^2 m}{2}} \\ &\leqslant \left(\frac{e^{\epsilon}}{(1+\epsilon)^{1+\epsilon}}\right)^m \text{ for } 0 \leqslant \epsilon \leqslant 1, \end{aligned}$$

and

$$P(S \geqslant (1+\epsilon)m) \leqslant \left(\frac{e}{1+\epsilon}\right)^{(1+\epsilon)m}.$$

Proof

Let $\epsilon \geqslant 0$ and $t \geqslant 0$. Then

$$\begin{aligned} & P(S \geqslant (1+\epsilon)m) \\ \leqslant \quad & e^{-t(1+\epsilon)m} e^{t(1+\epsilon)m} P(s^{tS} \geqslant e^{t(1+\epsilon)m}) \\ \leqslant \quad & e^{-t(1+\epsilon)m} E(e^{tS}). \end{aligned}$$

Proof (cont'd)

Since $X_1, \ldots, X_n$ are independent, by Inequality (1) we further get:

$$\begin{aligned} E(e^{tS}) &= E(e^{t(X)1+\cdots+X_n)}) = E(e^{tX_1}\cdots e^{tX_n}) \\ &= \prod_{i=1}^{n} E(e^{tX_i}) = \prod_{i=1}^{n}(p_i e^t + (1-p_i)) \\ &= \prod_{i=1}^{n}(1 + p_i(e^t - 1)) = e^{\left(\sum_{i=1}^{n} p_i(e^t-1)\right)} \\ &= e^{m(e^t-1)}. \end{aligned}$$

Putting $t = \ln(1+\epsilon)$ yields:

$$P(S \geqslant (1+\epsilon)m) \leqslant (1+\epsilon)^{-(1+\epsilon)m} e^{m\epsilon},$$

and hence

$$P(S \geqslant (1+\epsilon)m) \leqslant \left(\frac{e^{\epsilon}}{(1+\epsilon)^{1+\epsilon}}\right)^{m}. \quad (5)$$

Proof (cont'd)

The Inequality (3)

$$\epsilon - (1+\epsilon)\ln(1+\epsilon) \leqslant -\frac{1}{3}\epsilon^2$$

is equivalent to

$$\frac{e^{\epsilon}}{(1+\epsilon)^{1+\epsilon}} \leqslant e^{\frac{-\epsilon^2}{3}},$$

hence

$$P(S \geqslant (1+\epsilon)m) \leqslant e^{-\frac{\epsilon^2 m}{3}}$$

for $0 < \epsilon \leqslant 1$.

$$\begin{aligned} E(e^{-tS}) &= \prod_{i=1}^{n} E(e^{-tX_i}) \\ &= \prod_{i=1}^{n} \left(p_i e^{-t} + (1 - p_i)\right) \\ &= \prod_{i=1}^{n} \left(1 - p_i(1 - e^{-t})\right) \\ &\leqslant \prod_{i=1}^{n} e^{-p(1-e^{-t})} \\ &= e^{-(1-e^{-t})\sum_{i=1}^{n} p_i} = e^{-m(1-e^{-t})}. \end{aligned}$$

Putting $t = -\ln(1 - \epsilon)$ yields

$$P(S \leqslant (1 - \epsilon)m) \leqslant \left(\left(\frac{1}{1 - \epsilon} e^{-\epsilon}\right)^{1-\epsilon}\right)^{m}.$$

The left part of (3):

$$-\frac{1}{2}\epsilon^2 \leqslant \epsilon - (1+\epsilon)\ln(1+\epsilon),$$

for $0 \leqslant \epsilon \leqslant 1$, the Inequality (4):

$$-\frac{1}{2}\epsilon^2 \geqslant \epsilon - (1+\epsilon)\ln(1+\epsilon) \text{ for } -1 < \epsilon \leqslant 0$$

for $-1 < \epsilon \leqslant 0$, and the continuity at $\epsilon = 1$ imply

$$P(S \leqslant (1-\epsilon)m) \leqslant e^{-\frac{\epsilon^2 m}{2}} \leqslant \left(\frac{e^{\epsilon}}{(1+\epsilon)^{(1+\epsilon)}}\right)^m$$

for $0 \leqslant \epsilon \leqslant 1$.

We also have:

$$P(S \geqslant (1+\epsilon)m) \leqslant \left(\frac{e}{1+\epsilon}\right)^{(1+\epsilon)m}$$

In particular, for $r \leqslant 6m$ we have

$$P(S \geqslant r) \leqslant 2^{-r}.$$

Let $S' = Y_1 + \cdots + Y_n$, where $P(Y_i = 1) = p$ for $1 \leqslant i \leqslant n$. Also, let $0 < a < 1$, $a \geqslant p$ and $t \geqslant 0$. Choosing ϵ such that $(1+\epsilon)m = an$, by the previous calculations we get

$$P(S' \geqslant an) \leqslant e^{-tan} \left(pe^t + (1-p)\right)^n .$$

For $t = \ln \frac{a(1-p)}{p(1-a)}$, this becomes

$$\begin{aligned} P(S' \geqslant an) &\leqslant \left(\frac{p(1-a)}{a(1-p)}\right)^{an} \left(\frac{a(1-p)}{1-a} + (1-p)\right)^n \\ &= \left(\frac{p(1-a)}{a(1-p)}\right)^{an} \left(\frac{1-p}{1-a}\right)^n , \end{aligned}$$

which implies:

$$P(S' \geqslant an) \leqslant \left(\left(\frac{p}{a}\right)^a \left(\frac{1-p}{1-a}\right)^{1-a}\right)^n \qquad (6)$$

for $0 < a < 1$ and $a \geqslant p$.

Introduce the notation

$$S' \succeq k = \begin{cases} S' \geqslant k & \text{if } k \geqslant pn, \\ S' \leqslant k & \text{if } k < pn. \end{cases}$$

Note that the right member of Inequality (6) is invariant under a simultaneous interchange of a with $1 - a$ and p with $1 - p$ we then get by considering the random variabe $n - S'$,

$$P(S' \succeq an) \leqslant \left(\left(\frac{p}{a}\right)^a \left(\frac{1-p}{1-a}\right)^{1-a} \right)^n$$

for $0 < a < 1$.

Since by Inequality (2) we have

$$\left(\frac{1-p}{1-a}\right)^{1-a} = \left(1 + \frac{1-p}{1-a}\right)^{1-a} \leqslant e^{a-p}$$

we also have

$$P(S' \succeq an) \leqslant \left[\left(\frac{p}{a}\right)^{a} r^{a-p}\right]^{n}$$

when $0 < a \leqslant 1$.
Putting $na = k$ we obtain

$$P(S' \succeq k) \leqslant \left(\frac{np}{k}\right)^{k} \left(\frac{n-np}{n-k}\right)^{n-k} \leqslant \left(\frac{np}{k}\right)^{k} e^{k-np}$$

for $0 < k < n$.