Wednesday 03/08/2023

Markov's inequality : X is non-negative

$$P[X \geq \varepsilon] \leq \frac{E(X)}{\varepsilon}$$

Proof

$$E(X) = \sum_{x} x \cdot p(x)$$

$$\geq \sum_{x \in A} x \cdot p(x) \qquad \text{for any } A$$

Define $A = \{x \geq \varepsilon\}$

$$\geq \sum_{x \in A} \varepsilon \cdot p(x) \qquad (\text{on } A,\ x \geq \varepsilon)$$

$$= \varepsilon \sum_{x \in A} p(x)$$

$$= \varepsilon \; P(A)$$

Exp: $X \geq 0$

$E(X) = 100$

$P[X \geq 1000] \leq \frac{100}{1000} = 0.1$

We have proved:

$$\varepsilon \cdot P(\{X \geq \varepsilon\}) \leq E(X)$$

$$P(\{X \geq \varepsilon\}) \leq \frac{E(X)}{\varepsilon}$$

Monday, 03/13

Problem :

Let X be a non-negative random variable

Assume that $E(X^2) = 1$

Give an upper bound of

$$P[X \geq 10]$$

$$= P[X^2 \geq 100] \underset{(\text{Markov for } X^2)}{\leq} \frac{E[X^2]}{100} = \frac{1}{100}$$

Theorem :

X non-negative

$$P[X \geq \varepsilon] \leq \frac{E[X^2]}{\varepsilon^2}$$

Problem:

If X is non-negative

then $P[X \geq \varepsilon] \leq \frac{E[X^3]}{\varepsilon^3}$

Ex: If $X \geq 0$
and $E[X^3] = 1$

then $P[X \geq 10] \leq \frac{1}{1000}$

Exponential bound

If X is non-negative
and $E[e^X] = 10$

then $P[X \geq 10] = P[e^X \geq e^{10}]$

$\leq \frac{E[e^X]}{e^{10}} = \frac{10}{e^{10}} \approx 5 \times 10^{-4}$

Theorem:

If X is non-negative, $t > 0$

then $P[X \geq \varepsilon] \leq \frac{E[e^{tX}]}{e^{t\varepsilon}}$

$P[tX \geq t\varepsilon]$

$P[e^{tX} \geq e^{t\varepsilon}]$

Markov for e^{tX}

Moment-generating function

Theorem 3 (on the note) Hoeffding's lemma

If the r.v. X is bonded betwen $[a, b]$
and
$$E(X) = 0$$

Then: For all $t > 0$

$$E[e^{tX}] \leq \exp\left(\frac{t^2(b-a)^2}{8}\right)$$

Ex: If $X \in [-1, 1]$ (a, b)

and $E(X) = 0$

Give me an upper bound of:

$$P[X \geq 0.5]$$

Pick $t > 0$

$$P[X \geq 0.5] = P[tX \geq t \times 0.5]$$

$$= P[e^{tX} \geq e^{0.5t}]$$

$$\leq \frac{E[e^{tX}]}{e^{0.5 \times t}} = \frac{\exp\left(\frac{t^2}{2}\right)}{\exp(0.5 \times t)}$$

$$= \exp\left(\frac{t^2}{2} - \frac{t}{2}\right)$$

If $t = 1 \longrightarrow \exp(0) = 1$

$t = \frac{1}{2} \longrightarrow \exp\left(-\frac{1}{8}\right)$

Choose $t = \frac{1}{2}$

$$P[X \geq 0.5] \leq \exp\left(\frac{-1}{8}\right) \approx 0.88$$

Minimize $f(t) = \frac{t^2}{2} - \frac{t}{2}$

$$f'(t) = t - \frac{1}{2} = 0 \longrightarrow t = \frac{1}{2}$$

Review of probability : Independence

(1) z_1 and z_2 are ind. if and only if

$\forall a, b$

$$P[z_1 = a, z_2 = b] = P[z_1 = a] \times P[z_2 = b]$$

(2) If z_1 and z_2 are ind.

the $f(z_1)$ and $g(z_2)$ are also independent

for any functions f and g

(3) If z_1 and z_2 are ind.

$$E[f(z_1)\, g(z_2)] = E[f(z_1)]\, E[g(z_2)]$$

for any functions f and g

Theorem 4 (Hoeffding's inequality)

If $z_1, z_2, \dots, z_n \underset{i.i.d}{\sim} P_z$

(independently and identically distributed)

and $z \in [a, b]$

then

$$P\left[\left| \frac{z_1 + z_2 + \dots + z_n}{n} - E(z) \right| \geq \varepsilon \right] \leq 2 \exp\left(\frac{-n\varepsilon^2}{2(b-a)^2} \right)$$

If $z_1, z_2, \ldots, z_n \underset{i.i.d}{\sim} P_z$

(independently and identically distributed)

and $z \in [a, b]$ $\quad E(z) = 0$

then

$$P\left[\left| \frac{z_1 + z_2 + \ldots + z_n}{n} \right| \geq \varepsilon \right]$$

$$\leq 2 \exp\left(\frac{-n\varepsilon^2}{2(b-a)^2} \right)$$

$$X = \frac{z_1 + z_2 + \ldots + z_n}{n}$$

First step :

$$E[\exp(tX)] \qquad \text{for a fixed } t > 0$$

$$= E\left[\exp\left(\frac{t}{n} z_1 + \frac{t}{n} z_2 + \ldots + \frac{t}{n} z_n \right) \right]$$

$$= E\left[\exp\left(\frac{t}{n} z_1\right) \exp\left(\frac{t}{n} z_2\right) \ldots \exp\left(\frac{t}{n} z_n\right) \right]$$

(independent)

$$= E\left[\exp\left(\frac{t}{n} z_1\right)\right] E\left[\exp\left(\frac{t}{n} z_2\right)\right] \ldots E\left[\exp\left(\frac{t}{n} z_n\right)\right]$$

(identically)

$$= \left(E\left[\exp\left(\frac{t}{n} z \right) \right] \right)^n \qquad \text{where } z \sim P_z$$

We know : $\quad E[z] = 0$

$$z \in [a, b]$$

which implies

$$E\left[e^{\frac{t}{n} z} \right] \leq \exp\left(\frac{t^2 (b-a)^2}{n^2 \, 8} \right)$$

In other words

$$E[\exp(tX)] \leq \left[\exp\left(\frac{t^2 (b-a)^2}{8n} \right) \right]$$

Using exponential Bound, we have:

$$P[X \geq \varepsilon] \leq \frac{E[e^{tX}]}{e^{t\varepsilon}}$$

$$= \frac{\exp\left(\frac{t^2(b-a)^2}{8n}\right)}{\exp(t\varepsilon)}$$

$$= \exp\left(\frac{t^2(b-a)^2}{8n} - t\varepsilon\right)$$

Mini-problem:

$$f(t) = \frac{t^2(b-a)^2}{8n} - t\varepsilon$$

$\varepsilon, a, b, n:$ const

Find t^* to minimize $f(t)$

Find $f(t^*)$

$$t^* = \frac{4n\varepsilon}{(b-a)^2}$$

$$f(t^*) = -\frac{2n\varepsilon^2}{(b-a)^2}$$

Conclusion:

$$P\left[\frac{Z_1 + Z_2 + \dots Z_n}{n} \geq \varepsilon\right] \leq \exp\left(-\frac{2n\varepsilon^2}{(b-a)^2}\right)$$

2nd step: Absolute value

$$|X| \geq \varepsilon \rightarrow X \geq \varepsilon$$

$$|X| \geq \varepsilon \rightarrow X \leq -\varepsilon \longleftrightarrow -X \geq \varepsilon$$

$$P\left[\frac{(-Z_1) + (-Z_2) + \dots + (-Z_n)}{n} \geq \varepsilon\right] \leq \exp\left(\frac{-2n\varepsilon^2}{(b-a)^2}\right)$$

If $z_1, z_2, \ldots, z_n \sim P_z$ $\quad E[z] = 0 \quad z \in [a, b]$

We have:

then

$$P\left[\left|\frac{z_1 + z_2 + \ldots + z_n}{n}\right| \geq \varepsilon\right] \leq 2\exp\left(-\frac{2n\varepsilon^2}{(b-a)^2}\right)$$

Now if I have:

$Y_1, Y_2, \ldots, Y_n \sim P_Y \qquad Y \in [a, b]$

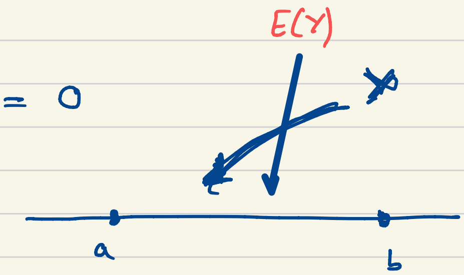
Define:

$$z = Y - E(Y)$$

then

(1) $E(z) = 0$

(2)



$$-(b-a) \leq z \leq b-a$$

$$P\left[\left|\frac{z_1 + z_2 + \ldots + z_n}{n}\right| \geq \varepsilon\right] \leq 2\exp\left(-\frac{n\varepsilon^2}{2(b-a)^2}\right)$$

$$P\left[\left|\frac{(Y_1 - EY) + (Y_2 - E(Y)) + \ldots + (Y_n - E(Y))}{n}\right| \geq \varepsilon\right]$$

$$P\left[\left|\frac{Y_1 + Y_2 + \ldots + Y_n - nE(Y)}{n}\right| \geq \varepsilon\right]$$

$$P\left[\left|\frac{Y_1 + \ldots + Y_n}{n} - E(Y)\right| \geq \varepsilon\right]$$

Statistical learning setting :

$$(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n) \sim P_{X,Y}$$

Hypothesis space $\mathcal{H}$

Loss function L

Empirical Risk Minimizer (ERM)

$$\min_{f \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{n} L(h(X_i), Y_i)$$

The optimal predictor is h^*

$$h^* = \arg\min R(h^*)$$

where

$$R(h) = E_{P_{X,Y}} [\, l(h(X), Y)]$$

Assumption :

① $0 \le L(y, y') \le c$ for all y, y'

② Finite hypothesis assumption:

$$\mathcal{H} = \{h_1, h_2, \ldots, h_m\}$$

Define : For a fixed h

$$z = L(h(X), Y)$$

We have

$$z_1, z_2, \ldots, z_n \overset{iid}{\sim} P_z$$

$$z \in [0, c]$$

$$P\left[\left| \frac{z_1 + z_2 + \ldots + z_n}{n} - E[z] \right| \geq \varepsilon \right] \leq 2 \exp\left(\frac{-n\varepsilon^2}{2c^2}\right)$$

$$P\left[\left| \frac{L(h(X_1), Y_1) + L(h(X_2), Y_2) + \ldots + L(h(X_n), Y_n)}{n} - E[\, l(h(X), Y)] \right| \geq \varepsilon \right]$$