Inequality Extensions

SI252 Reinforcement Learning

School of Information Science and Technology
ShanghaiTech University

March 11, 2020

Cauchy-Schwarz

Theorem

For any r.v.s X and Y with finite variances,

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

Example: Second Moment Method

Let X be a nonnegative random variable, then

$X \geq 0$ r.v.

$$\Pr(X = 0) \leq \frac{\mathrm{Var}(X)}{\mathbb{E}(X^2)}.$$

Examples:

① X : number of questions that Fred gets wrong on an exam

$P(X=0) = P\{$ Fred gets a perfect score $\}$

② X : number of pairs of people at a party with the same birth

$P(X=0) = P\{$ No birthday matches $\}$

Proof: Since $X \geq 0$

$X = X I\{X>0\}$ ✓ $\quad I\{X>0\}$: indicator of event $X>0$

If $X>0$: $X I\{X>0\} = X \cdot 1 = X$

If $X=0$: $X I\{X>0\} = X \cdot 0 = 0$ $\quad I^2\{\cdot\} = I\{\cdot\}$

$E[I\{A\}] = P_r(A)$

By Cauchy-Schwartz inequality,

$$E(X) = E(X I\{X>0\}) \leq \sqrt{E(X^2) \cdot E[I^2\{X>0\}]}$$

$$= \sqrt{E(X^2) \cdot E[I\{X>0\}]}$$

$$= \sqrt{E(X^2) \cdot P(X>0)}$$

$$E^2(X) \leq E(X^2) \cdot P(X>0) \Rightarrow P(X>0) \geq \frac{E^2(X)}{E(X^2)}$$

$$P(X=0) = 1 - P(X>0) \leq 1 - \frac{E^2(X)}{E(X^2)} = \frac{E(X^2) - E^2(X)}{E(X^2)} = \frac{Var(X)}{E(X^2)}$$

Example: Application of Second Moment Method

Covariance : rvs X and Y

$Cov(X,Y) = E((X-EX)(Y-EY)) = E(XY) - E(X)E(Y)$

Correlation : rvs X and Y

$Corr(X, Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$ $\quad E(I_j) = Pr(I_j = 1) \in (0,1)$

Assume $X = I_1 + \cdots + I_n$, where the I_j are uncorrelated indicator r.v.s. Let $\boxed{p_j = \mathbb{E}(I_j)}$. Upper bound of $\Pr(X = 0)$?

Uncorrelated : X, Y uncorrelated $\Longleftrightarrow$ $Cov(X,Y) = 0$ or $Corr(X,Y) = 0$.

Property : If rvs. X and Y uncorrelated, then $Var(X+Y) = Var(X) + Var(Y)$

Proof : $Cov(X+Y, Z) = E((X+Y)Z) - E(X+Y)E(Z)$

$= E(XZ) + E(YZ) - (E(X)E(Z) + E(Y)E(Z))$

$= Cov(X,Z) + Cov(Y,Z)$

$$\mathrm{Var}(X+Y) = E[(X+Y)^2] - E^2(X+Y)$$
$$= \mathrm{Cov}(X+Y, X+Y)$$
$$= \mathrm{Cov}(X,X) + \mathrm{Cov}(X,Y) + \mathrm{Cov}(Y,X) + \mathrm{Cov}(Y,Y)$$
$$= \mathrm{Var}(X) + 0 + 0 + \mathrm{Var}(Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

$$\mathrm{Cov}(X,X) = \mathrm{Var}(X)$$

$$P(X=0) \le \frac{\mathrm{Var}(X)}{E(X^2)}$$

$p_j \in (0,1)$ $\quad \mu = \sum_{j=1}^{n} p_j > \sum_{j=1}^{n} p_j^2 = c$

$\mu - c > 0$

$\mu > \mu - c$ $\quad \dfrac{\mu^2}{\mu - c} > \dfrac{\mu(\mu-c)}{\mu-c} = \mu$

Since $\{I_j\}$ are uncorrelated, we have

$$\mathrm{Var}(X) = \mathrm{Var}\left(\sum_{j=1}^{n} I_j\right) = \sum_{j=1}^{n} \mathrm{Var}(I_j) = \sum_{j=1}^{n} [E(\underbrace{I_j^2}_{I_j}) - E^2(I_j)]$$
$$= \sum_{j=1}^{n} [E(I_j) - E^2(I_j)] = \sum_{j=1}^{n} (p_j - p_j^2) = \mu - c$$

$$\mu \triangleq E(X) = \sum_{j=1}^{n} E(I_j) = \sum_{j=1}^{n} p_j \qquad c \triangleq \sum_{j=1}^{n} p_j^2$$

$$E(X^2) = \mathrm{Var}(X) + (EX)^2 = \mu - c + \mu^2$$

$$P(X=0) \le \frac{\mu - c}{\mu - c + \mu^2} = \frac{1}{1 + \frac{\mu^2}{\mu - c}} < \frac{1}{1+\mu}$$

Example: Application of Second Moment Method

$I_{ij} \triangleq I\{\text{person } i \text{ and } j \text{ have near birthdays}\}$

d_i : i's birthday

$X \triangleq$ number of "near birthday" pairs $\quad X = \sum_{i \neq j} I_{ij}$

$P(X=0) \le \frac{1}{1+E(X)} \quad \underline{E(X) = E\left(\sum_{i \neq j} I_{ij}\right) = \sum_{i \neq j} E(I_{ij})} = \sum_{i \neq j} \frac{3}{365} = \binom{14}{2} \frac{3}{365}$

Suppose there are 14 people in a room. How likely is it that there are two people with the same birthday or birthdays one day apart?

$$E(I_{ij}) = P(I_{ij}=1) = P(i \text{ and } j \text{ have near birthdays})$$

$$= \sum_{t=1}^{365} P\left(d_i \in \{(t-1), t, (t+1)\} \% 365 \mid d_j = t\right) \underline{P(d_j = t)}$$

$$= \sum_{t=1}^{365} \frac{3}{365} \times \frac{1}{365} = \frac{3}{365}$$

$$P(X=0) \le \frac{1}{1+E(X)} < 0.573.$$

Markov's Inequality & Chebyshev's Inequality

Theorem (Markov's Inequality)

For any r.v. X and constant $a > 0$,

$\in [0,1]$

$$\bar{P}(|X| \geq a) \leq \frac{E|X|}{a}.$$

≤ 1

Theorem (Chebyshev's Inequality)

Let X have mean μ and variance σ^2. Then for any $a > 0$,

$$P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}.$$

Example: Coin Flipping

$n \geq 4$

① Using Chebyshev's inequality:

$X_i \triangleq I\{\text{the } i\text{-th coin flip is head}\}$ $X \triangleq$ number of heads

$X = X_1 + \cdots + X_n$ $X_i \sim Bern(\frac{1}{2})$ $E(X_i) = \frac{1}{2}$ $Var(X_i) = \frac{1}{4}$

$E(X) = nE(X_1) = \frac{n}{2}$ $Var(X) = n\,Var(X_1) = \frac{n}{4}$

Find bounds on the probability of having no more than $n/4$ heads or fewer than $3n/4$ heads in a sequence of n fair coin flips.

$$P(X \leq \frac{n}{4} \text{ or } X \geq \frac{3n}{4}) = P(|X - \underbrace{\tfrac{n}{2}}_{E(X)}| \geq \frac{n}{4}) = P(|X - E(X)| \geq \frac{n}{4})$$

$$\leq \frac{Var(X)}{(\frac{n}{4})^2} = \frac{4}{n}$$

② Using Markov's inequality:

$$\boxed{\begin{array}{l} X \sim Bin(n, \frac{1}{2}) \\ n - X \sim Bin(n, \frac{1}{2}) \end{array}}$$

$$P(X \geq \frac{3n}{4}) \leq \frac{E(X)}{\frac{3n}{4}} = \frac{n/2}{3n/4} = \frac{2}{3}$$

$$P(X \leq \frac{n}{4}) = P(n - X \geq \frac{3n}{4}) = P(X \geq \frac{3n}{4}) \leq \frac{2}{3}$$

$$P(X \geq \frac{3n}{4} \text{ or } X \leq \frac{n}{4}) \leq \frac{2}{3} + \frac{2}{3} = \frac{4}{3} > 1$$

Chernoff's Technique

Theorem

For any r.v. X and constant a,

$$P\left(X \geq a\right) \leq \inf_{t>0} \frac{E\left(e^{tX}\right)}{e^{ta}},$$

$$P\left(X \leq a\right) \leq \inf_{t<0} \frac{E\left(e^{tX}\right)}{e^{ta}}.$$

Example: Sum of Independent Bernoulli R.V.s

Let $X_1, \ldots, X_n$ be independent Bernoulli random variables such that $\Pr(X_i = 1) = p_i$, $\Pr(X_i = 0) = 1 - p_i$. Let $\cancel{X = n_i = 1 - X_i}$ and $\mu = \mathbb{E}(X)$. Then for $0 < \delta < 1$,

$X = \sum_{i=1}^{n} X_i$

$$\Pr\left(|X - \mu| \geq \delta\mu\right) \leq \boxed{2e^{-\mu\delta^2/3}}.$$

Moment Generating Function (MGF):

r.v. X $M(t) = E(e^{tX})$ $\in$ open interval $(-a, a)$

MGF of Bernoulli r.v. $X \sim Bern(p)$: $M(t) = E(e^{tX}) = pe^t + 1 - p$

MGF of a sum of independent r.v.s :

X, Y independent, $M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$

$$M_{X+Y}(t) = E[e^{t(X+Y)}] = E(e^{tX} \cdot e^{tY}) = \underbrace{E(e^{tX})}_{M_X(t)} \cdot \underbrace{E(e^{tY})}_{M_Y(t)}$$

$$P(|X-\mu| \geq \delta\mu) = P(X - \mu \geq \delta\mu) + P(X - \mu \leq -\delta\mu)$$

$$= \underbrace{P(X \geq (1+\delta)\mu)}_{①} + \underbrace{P(X \leq (1-\delta)\mu)}_{②}$$

① $P(X \geq (1+\delta)\mu)$ by Chernoff's technique,

$$P(X \geq (1+\delta)\mu) \leq \frac{E(e^{tX})}{e^{t(1+\delta)\mu}} = \frac{M_X(t)}{e^{t(1+\delta)\mu}} \quad \forall t > 0$$

Since $X = \sum_{i=1}^{n} X_i$, $\{X_i\}$ independent $\Rightarrow \underline{M_X(t) = \prod_{i=1}^{n} M_{X_i}(t)}$

$\forall t \in R$, $\underline{M_{X_i}(t)} = E(e^{tX_i}) = P_i e^t + (1-P_i) = 1 + P_i(e^t - 1) \leq \underline{e^{P_i(e^t-1)}}$

$1 + x \leq e^x \quad \forall x \in R$

Proof: (if $x>0$) (if $x<0$)

① Taylor's theorem: $\exists \theta \in (0,x)$ or $(x,0)$ s.t.

$e^x = 1 + x + \frac{1}{2} f''(\theta) x^2 = 1 + x + \frac{1}{2} e^\theta x^2 \geq 1 + x$.

② Convexity: $f(x) = e^x - x - 1 \quad \forall x \in R$

$f(x) \geq f(0) = 0 \quad f(x) \geq 0 \ \forall x \in R$

$$M_X(t) \leq \prod_{i=1}^{n} e^{P_i(e^t-1)} = e^{(e^t-1)\cdot\sum_{i=1}^{n} P_i}$$

Since $\mu \triangleq E(X) = E\left(\sum_{i=1}^{n} X_i\right)$

$= \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} P_i \quad \Rightarrow \underline{M_X(t) \leq e^{(e^t-1)\cdot\mu}}$

Thus, $P(X \ge (1+\delta)\mu) \le \frac{M_X(t)}{e^{t(1+\delta)\mu}} = \frac{e^{(e^t-1)\cdot\mu}}{e^{t(1+\delta)\mu}} \quad \forall t>0$

Let $t=\ln(1+\delta)$ $\quad \forall \delta \in (0,1)$ we have $t>0$

$e^t \to 1+\delta \qquad \underline{P(X\ge(1+\delta)\mu)} \le \underline{\left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^\mu \le e^{-\mu\delta^2/3}}$

$\checkmark \qquad \forall\delta\in(0,1)$

$(\cdot)^{\frac{1}{\mu}}$

$\Rightarrow \quad \frac{e^\delta}{(1+\delta)^{(1+\delta)}} \le e^{-\delta^2/3} \quad \forall\delta\in(0,1)$

$\ln(\cdot)$

$\Rightarrow \quad \delta-(1+\delta)\ln(1+\delta) \le -\delta^2/3 \quad \forall\delta\in(0,1)$

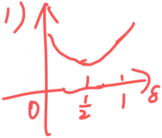
$\Rightarrow \quad f(\delta)=\delta-(1+\delta)\ln(1+\delta)+\delta^2/3 \le 0 \quad \forall\delta\in(0,1)$

$f'(\delta) = -\ln(1+\delta)+\frac{2}{3}\delta \qquad f''(\delta) = -\frac{1}{1+\delta}+\frac{2}{3}$

$f''(\delta)<0 \quad \forall\delta\,(0,\frac{1}{2}) \qquad f''(\delta)>0 \quad \forall\delta\in(\frac{1}{2},1)$

$f'(0)=0 \qquad f'(1)<0 \qquad f'(\delta)\le 0 \quad \forall\delta\in(0,1)$

$f(\delta) \le f(0)=0 \quad \forall\delta\,(0,1)$



Example: Revisit Example of Coin Flipping

$X_i \triangleq I\{\text{ i-th coin flip is head }\}$ $\quad X_i \sim Bern(\frac{1}{2})$ $\quad E(X_i) = \frac{1}{2}$

$X = X_1 + \cdots + X_n$ $\quad \mu = E(X) = nE(X_1) = \frac{n}{2}$

$P(X \le \frac{n}{4} \text{ or } X \ge \frac{3n}{4}) = P(|X - \frac{n}{2}| \ge \frac{n}{4}) = P(|X-\mu| \ge \delta\mu)$

$\delta = \frac{1}{2}$

Find bounds on the probability of having no more than $n/4$ heads or fewer than $3n/4$ heads in a sequence of n fair coin flips.

$\le 2e^{-\mu\delta^2/3} = 2e^{-\mu/12}$

$= 2e^{-\frac{n}{24}}$

Chebyshev : $\frac{4}{n}$

Markov : $\frac{4}{3} > 1$

$n \ge 4$

$n = 100$