T.Y.B.Sc. Statistics

Paper –I

Unit-II- Inequalities and Law of Large Numbers

INTRODUCTION

We intuitively feel it is rare for an observation to deviate greatly from the expected value. Markov’s inequality and Chebyshev’s inequality place this intuition on firm mathematical ground.

In probability theory, Markov's inequality gives an upper bound for the probability that a non-negative function of a random variable is greater than or equal to some positive constant. It is named after the Russian mathematician Andrey Markov, although it appeared earlier in the work of Pafnuty Chebyshev (Markov's teacher), and many sources, especially in analysis, refer to it as Chebyshev's inequality (sometimes, calling it the first Chebyshev inequality, while referring to Chebyshev's inequality as the second Chebyshev inequality) or Bienaymé's inequality.

Chebyshev's inequality, also known as Chebyshev's theorem, makes a fairly broad but useful statement about data dispersion for almost any data distribution. This theorem states that no more than $1 / k^2$ of the distribution's values will be more than k standard deviations away from the mean. Looked at another way, $1 - (1 / k^2)$ of the distribution's values will lie within k standard deviations of the mean.

The importance of Markov’s and Cheybyshev’s inequalities is that they enable us to derive bounds on probabilities when only the mean, or both the mean and the variance are known but the probability distribution are unknown. Of course, if the actual distribution were known, then the desired probabilities could be exactly computed and we would not need to use these bounds.

MARKOV'S INEQUALITY

Let X be a nonnegative random variable such that (x) ≥ 0 for all x ∈ R. Assume that E (X) exists, then for any value a > 0

P{ X ≥ a} ≤ $\frac{E(X)}{a}$

Proof:

We give a proof for the case where X is continuous with density f.

Define A1 = {x | (x) ≥ a} and A2 = {x | (x) < a}.

E [(X)] = $\int_0^\infty (x)\ f(x)dx$

= $\int_{A_1}(x)f(x)dx + \int_{A_2}(x)f(x)dx$

= $\int_0^a (x)\ f(x)dx$ + $\int_a^\infty (x)\ f(x)dx$

$\geq \int_a^\infty (x)\ f(x)dx$

$\geq \int_a^\infty a\ f(x)dx$

= a $\int_a^\infty\ f(x)dx$

= a $P\big((X) \geq a\big)$

⇒ P{(X) ≥ a} ≤ $\frac{E(X)}{a}$……………Eq. 1

We give a proof for the case where X is discrete with pmf f (x).

E [(X)] = $\sum_0^\infty (x)\ f(x)$

= $\sum_0^a (x)\ f(x)$ + $\sum_a^\infty (x)\ f(x)$

$\geq \sum_a^\infty (x)\ f(x)$

$\geq \sum_a^\infty a\ f(x)$

= a $\sum_a^\infty\ f(x)$

= a $P\big((X) \geq a\big)$ ⇒ P{(X) ≥ a} ≤ $\frac{E(X)}{a}$

Corollary:

1. Let g(X) be a nonnegative function of an RV X.

If E (g(X)) exists, then for every a > 0, P{g(X) ≥ a} ≤ $\frac{Eg(X)}{a}$

2. Let g (X) is = $|X|^r$ and a = k^r

Where r > O and K > 0. Then P($|X|^r \geq k^r$) $\leq (E|X|^r)/k$^r

Examples:

1. A biased coin, which lands heads with probability1/10 each time it is flipped, is flipped 200 times consecutively. Give an upper bound on the probability that it lands heads at least 120 times.

Answer: The number of heads is a binomially distributed r.v., X, with parameters p=1/10 and n=200.

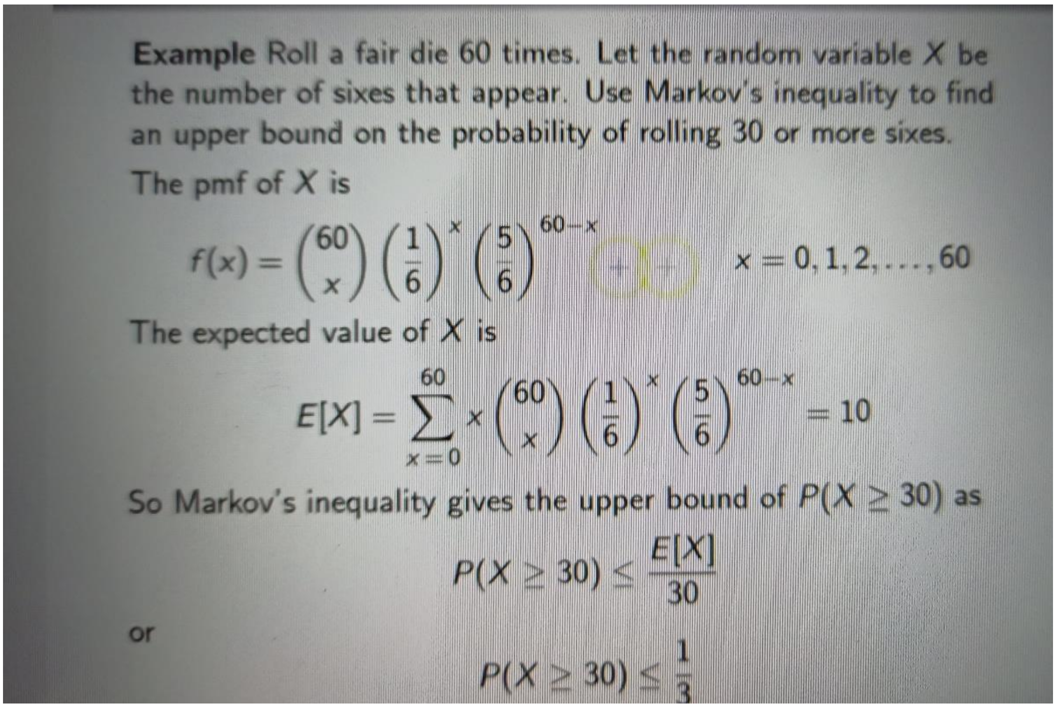
Thus, the expected number of heads is

E(X) =np=200·(1/10) =20.

By Markov Inequality, the probability of at least 120 heads is

P(X≥120) ≤ E(X)/120 = 20/120 = 1/6

2.



Example Roll a fair die 60 times. Let the random variable X be the number of sixes that appear. Use Markov's inequality to find an upper bound on the probability of rolling 30 or more sixes.

The pmf of X is

$$f(x) = \binom{60}{x}\left(\frac{1}{6}\right)^x\left(\frac{5}{6}\right)^{60-x} \qquad x = 0, 1, 2, \ldots, 60$$

The expected value of X is

$$E[X] = \sum_{x=0}^{60} x\binom{60}{x}\left(\frac{1}{6}\right)^x\left(\frac{5}{6}\right)^{60-x} = 10$$

So Markov's inequality gives the upper bound of $P(X \geq 30)$ as

$$P(X \geq 30) \leq \frac{E[X]}{30}$$

or

$$P(X \geq 30) \leq \frac{1}{3}$$

3.

Example 6.20

Let $X \sim Binomial(n, p)$. Using Markov's inequality, find an upper bound on $P(X \geq \alpha n)$, wh… $p < \alpha < 1$. Evaluate the bound for $p = \frac{1}{2}$ and $\alpha = \frac{3}{4}$.

Solution

Note that X is a nonnegative random variable and $EX = np$. Applying Markov's inequality … obtain

$$P(X \geq \alpha n) \leq \frac{EX}{\alpha n} = \frac{pn}{\alpha n} = \frac{p}{\alpha}.$$

For $p = \frac{1}{2}$ and $\alpha = \frac{3}{4}$, we obtain

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

4.

Let $X \sim Exponential(\lambda)$. Using Markov's inequality find an upper bound for $P(X \geq a)$, where $a > 0$. Compare the upper bound with the actual value of $P(X \geq a)$.

Solution

If $X \sim Exponential(\lambda)$, then $EX = \frac{1}{\lambda}$, using Markov's inequality

$$P(X \geq a) \leq \frac{EX}{a} = \frac{1}{\lambda a}.$$

The actual value of $P(X \geq a)$ is $e^{-\lambda a}$, and we always have $\frac{1}{\lambda a} \geq e^{-\lambda a}$.

5. Suppose that the average grade on the upcoming Math 20 exam is 70%. Give an upper bound on the proportion of students who score at least 90%.

Solution: $P(X \geq 90) \leq E(X)/90 = 7/9$ so at most 77.8% of students can possible score this high.

6. Let X be a positive random variable whose expected value is E(X) =10. Find a lower bound to the probability P(X < 20)

Solution

First of all, we need to use the formula for the probability of a complement:

$$P(X < 20) = 1 - P(X \geq 20)$$

Now, we can use Markov's inequality:

$$P(X \geq 20) \leq \frac{E[X]}{20} = \frac{10}{20} = \frac{1}{2}$$

Multiplying both sides of the inequality by, we obtain Adding to both sides of the inequality, we obtain

$$1 - P(X \geq 20) \geq 1 - \frac{1}{2} = \frac{1}{2}$$

Thus, the lower bound is

$P(X< 20) \geq \frac{1}{2}$

7. For $f(x) = e^{-x}$ for $x > 0$,

E[X] = 1

Markov's inequality gives

c = 1: $P(X \geq 1) \leq E[X]/c = 1$
c = 10: $P(X \geq 10) \leq E[X]/10 = 1/10$

CHEBYSHEV'S INEQUALITY

https://www.youtube.com/watch?v=RZj2stql-L4

If X is a random variable with mean μ and variance σ^2, then for any value k > 0,

$$\mathbf{P\{|X - \mu| \geq k\} \leq \frac{\sigma^2}{k^2}} \text{...........} \mathbf{Eq. 1}$$

Proof: Since $(X - \mu)^2$ is a nonnegative random variable, we can apply Markov's inequality $[P\{ g(X) \geq a\} \leq \frac{E(g(X))}{a}]$ with $a = k^2$ and $g(x) = (X - \mu)^2$ to obtain

$$P\{|X-\mu|^2 \geq k^2\} \leq \frac{E(X-\mu)^2}{k^2}$$

But since $(X - \mu)^2 \geq k^2$ if and only if $|X - \mu| \geq k$, Equation 3 is equivalent to

$$P\{|X-\mu| \geq k\} \leq \frac{E(X-\mu)^2}{k^2}$$

$$\Rightarrow \mathbf{P\{|X-\mu| \geq k\} \leq \frac{\sigma^2}{k^2}} \qquad [\text{Since } E(X-\mu)^2 = \sigma^2]$$

Note: 1. In particular, if we take $g(X) = (X-\mu)^2$, $a = k^2\sigma^2$

we get Chebychev-Bienayme inequality:

$$\mathbf{P\{|X-\mu| \geq k\sigma\} \leq \frac{1}{k^2}} \qquad \textbf{Eq. 2}$$

Note: 2. $P\{|X-\mu| \geq k\sigma\} \leq \frac{1}{k^2}$

$$\Rightarrow -P\{|X-\mu| \geq k\sigma\} \geq -\frac{1}{k^2}$$

$$\Rightarrow 1 - P\{|X-\mu| \geq k\sigma\} \geq 1 - \frac{1}{k^2}$$

$$\Rightarrow \mathbf{P\{|X-\mu| < k\sigma\} \geq 1 - \frac{1}{k^2}} \qquad \textbf{Eq. 3}$$

Note: 3. $|X-\mu| < k\sigma$

$$\Rightarrow -k\sigma < X - \mu < +k\sigma$$

$$\Rightarrow \mu - k\sigma < X < \mu + k\sigma$$

So, $P\{|X-\mu| < k\sigma\} \geq 1 - \frac{1}{k^2}$

$$\Rightarrow \boldsymbol{P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2}} \qquad \textbf{Eq. 4}$$

Illustration of the Inequality

To illustrate the inequality, we will look at it for a few values of *K*:

- For K = 2 we have $1 - 1/K^2$ = 1 - 1/4 = 3/4 = 75%. So Chebyshev's inequality says that at least 75% of the data values of any distribution must be within two standard deviations of the mean.
- For K = 3 we have $1 - 1/K^2$ = 1 - 1/9 = 8/9 = 89%. So Chebyshev's inequality says that at least 89% of the data values of any distribution must be within three standard deviations of the mean.
- For K = 4 we have $1 - 1/K^2$ = 1 - 1/16 = 15/16 = 93.75%. So Chebyshev's inequality says that at least 93.75% of the data values of any distribution must be within two standard deviations of the mean.

Examples:

1. If P{X = 0} = 1- $\frac{1}{k^2}$, K > 1, constant,

and P{X = $\pm$ 1} = $\frac{1}{2k^2}$

prove that $P(|X| \geq 1) \leq 1/k^2$

Proof: E(X) = 0 $E(X^2) = \frac{1}{k^2}$ $\sigma = \frac{1}{k}$

P$(|X| \geq k\sigma) = P(|X| \geq 1) \leq 1/k^2$

2. Let X be distributed with pdf f (x) = 1 if 0 < x < 1, and = 0 otherwise. Then prove that $(|X - \frac{1}{2}| \leq 2\sqrt{\frac{1}{12}}) \geq 0.75$. Also compare with the actual value. Take k=2.

Proof: E(X) = $\frac{1}{2}$ $E(X^2) = \frac{1}{3}$ $Var(X) = \frac{1}{12}$

By Chebyshev's inequality, we get

$P(|X - \frac{1}{2}| \leq 2\sqrt{\frac{1}{12}}) \geq 1 - \frac{1}{4} = 0.75$

Actual value $= P\left(\left|X - \frac{1}{2}\right| \leq 2\sqrt{\frac{1}{12}}\right) = P(\frac{1}{2} - \frac{1}{\sqrt{3}} < X < \frac{1}{2} + \frac{1}{\sqrt{3}}) = 1$

3. Suppose that it is known that the number of items produced in a factory during a week is a random variable with mean 50.

(a) What can be said about the probability that this week's production will exceed 75?

(b) If the variance of a week's production is known to equal 25, then what can be said about the probability that this week's production will be between 40 and 60?

Solution: Let X be the number of items that will be produced in a week:

(a) By Markov's inequality

$P\{X > 75\} \leq \frac{E(X)}{75} = \frac{50}{75} = \frac{2}{3}$

(b) By Chebyshev's inequality

$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2}$

$P\{40 < X < 60\} \geq 1 - \frac{1}{k^2}$

Hence

$\mu - k\sigma = 40 \Rightarrow 50 - k \times 5 = 40 \Rightarrow 5k = 10 \Rightarrow k = 2$

So, $P\ (40 < X < 60) \geq 1 - \frac{1}{2^2} = 1 - \frac{1}{4} = \frac{3}{4} = 0.75$

and so the probability that this week's production will be between 40 and 60 is at least 0.75.

4. Suppose the value of the Canadian dollar in terms of the US dollar over a certain period is a random variable X with

mean μ = 0.98 and standard deviation σ = 0.05 .

What can be said of the probability that the Canadian dollar is valued between $0.88US and $1.08US?

Solution: By Chebyshev's inequality we have

$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2}$

P{0.88 < X < 1.08} $\geq 1 - \frac{1}{k^2}$

Hence

$\mu - k\sigma = 0.88 \Rightarrow 0.98 - k \times 0.05 = 0.88 \Rightarrow 0.05k = 0.1 \Rightarrow k = 2$

So, $P\ (0.88 < X < 1.08) \geq 1 - \frac{1}{2^2} = 1 - \frac{1}{4} = \frac{3}{4} = 0.75$

and so the probability that this week's production will be between 40 and 60 is at least 0.75.

5. The score of students taking an examination is a random variable with mean μ = 65 and standard deviation σ = 5 .

(a) What is the probability a student scores between 55 and 75?

(b) How many students would have to take the examination so that the probability that their average grade is between 60 and 70 is at least 80% ?

Solution: (a) Probability a student scores between 55 and 75

= P(55<X< 75)

By Chebyshev's inequality we have

$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2}$

P{55 < X < 75} $\geq 1 - \frac{1}{k^2}$

Hence

$\mu - k\sigma = 55 \Rightarrow 65 - k \times 5 = 55 \Rightarrow 5k = 10 \Rightarrow k = 2$

So, $P\ (55 < X < 75)\ \geq 1 - \frac{1}{2^2} = 1 - \frac{1}{4} = \frac{3}{4} = 0.75$

and so the probability that this week's production will be between 40 and 60 is at least 0.75.

(b) $\bar{X} = sample\ Mean$

$\mu_{\bar{X}} = \mu = 65, \quad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{n}}$

$P(\mu_{\bar{X}} - k\sigma_{\bar{X}} < \bar{X} < \mu_{\bar{X}} + k\sigma_{\bar{X}})\ \geq 1 - \frac{1}{k^2}$ Eq a

$\Rightarrow P(65 - k \times \frac{5}{\sqrt{n}} < \bar{X} < 65 + k \times \frac{5}{\sqrt{n}}) \geq 1 - \frac{1}{k^2}$

$\Rightarrow 1 - \frac{1}{k^2} = 0.8 \Rightarrow \frac{1}{k^2} = 0.2 \Rightarrow k^2 = 5 \Rightarrow k = \sqrt{5}$

Now $P(60 < \bar{X} < 70)\ \geq 0.8$ Eq b

Comparing equations a and b, we get

$65 - k \times \frac{5}{\sqrt{n}} = 60$

$\Rightarrow 65 - \sqrt{5} \times \frac{5}{\sqrt{n}} = 60$

$\Rightarrow \sqrt{5} \times \frac{5}{\sqrt{n}} = 5$

$\Rightarrow n = 5$

6. Let X~ Exponential (λ). Using Chebyshev's inequality find an upper bound for

P (|X−EX| ≥b), where b>0.

Solution:

We have E(X) = $\frac{1}{\lambda}$ and VarX= $\frac{1}{\lambda^2}$.

Using Chebyshev's inequality, we have

P (|X−EX| ≥b) ≤ $\frac{\sigma^2}{b^2}$

$$\Rightarrow \; P(|X - EX| \geq b) \leq \frac{1}{\lambda^2 b^2}$$

BOOLE'S INEQUALITY

In probability theory, **Boole's inequality**, also known as the union bound, says that for any finite or countable set of events, the probability that at least one of the events happens is no greater than the sum of the probabilities of the individual events. **Boole's inequality** is named after George **Boole**.

Boole's Inequality provides an upper bound on the chance of a union.
That is, the chance that at least one of the events occurs can be no larger than the sum of the chances.

Use of Boole's Inequality:

1. In weather forecasting
For example, if the weather forecast says that the chance of rain on Saturday is 40% and the chance of rain on Sunday is 10%, then the chance that it rains at some point during those two days is at least 40% and at most 50%.
To find the chance exactly, you would need the chance that it rains on both days, which you don't have. Assuming independence doesn't seem like a good idea in this setting. So you cannot compute an exact answer, and must be satisfied with bounds.
Though bounds aren't exact answers or even approximations, they can be very useful.
2. To find system reliability
3. To find maximum market portfolio

4. To find extreme values in Actuarial Science

Let $\{C_k\}$ be an arbitrary sequence of events, then

$(i)\ P(\bigcup_{k=1}^{n} C_k) \ \leq \sum_{k=1}^{n} P(C_k)$

$(ii)\ P(\bigcup_{k=1}^{\infty} C_k) \ \leq \sum_{k=1}^{\infty} P(C_k)$

(iii) P $(\bigcap_{k=1}^{n} C_k) \ \geq \ \sum_{k=1}^{n} P(C_k) - (n-1)$

Proof:

(i) We will prove the Boole's inequality by using the method of induction:

When n = 1, the inequality is P$(C_1) = P(C_1)$

We can write this as P$(C_1) \leq P(C_1)$

When n = 2, the inequality is

$P(C_1 \cup C_2) = P(C_1) + P(C_2) - P(C_1 \cap C_2)$

[From Addition law of probability]

$\Rightarrow \ P(C_1 \cup C_2) \leq P(C_1) + P(C_2)$

[Since $P(C_1 \cap C_2) is\ a\ positive\ value$]

Boole's inequality

This is another proof of Boole's inequality, one that is done using a proof technique called proof by induction. For your quiz on October 22, you may use the proof by induction, the textbook proof, or any other proof that is valid. Any valid proof that is written 100% correctly will merit full credit for your first quiz score.

Theorem (Boole's Inequality; Theorem 1.3.8 of Hogg, McKean, Craig). *Let $\{C_n\}$ be an arbitrary sequence of events. Then*

$$P\left(\bigcup_{n=1}^{\infty} C_n\right) \leq \sum_{n=1}^{\infty} P(C_n).$$

Proof. First, we will establish that the assertion holds true for finite unions and summations; in other words, we will prove that the statement

$$P\left(\bigcup_{k=1}^{n} C_k\right) \leq \sum_{k=1}^{n} P(C_k).$$

for any positive integer n holds true. For this, we will proceed by induction. For the base step, we will prove that the statement for $n = 1$ holds true. A union of one set C_1 only is the set C_1 itself. Furthermore, any summation of one number $P(C_1)$ is the number $P(C_1)$ itself. So the assertion for one set C_1 becomes

$$P(C_1) \leq P(C_1),$$

which is clearly true. This completes the proof of the base step. For the induction step, we assume that the statement for $n = m$ holds true; in other words, we assume that the statement

$$P\left(\bigcup_{k=1}^{m} C_k\right) \leq \sum_{k=1}^{m} P(C_k)$$

holds true. With this assumption, we will prove that the statement for $n = m + 1$ holds true; in other words, we will prove the statement

$$P\left(\bigcup_{k=1}^{m+1} C_k\right) \leq \sum_{k=1}^{m+1} P(C_k).$$

Indeed, using the inclusion-exclusion principle, the first axiom of probability, and what we assumed in this induction step, we have

$$\begin{aligned}
P\left(\bigcup_{k=1}^{m+1} C_k\right) &= P\left(\bigcup_{k=1}^{m} C_k \cup C_{m+1}\right) \\
&= P\left(\bigcup_{k=1}^{m} C_k\right) + P(C_{m+1}) - P\left(\bigcup_{k=1}^{m} C_k \cap C_{m+1}\right) \\
&\leq P\left(\bigcup_{k=1}^{m} C_k\right) + P(C_{m+1}) - 0 \\
&= P\left(\bigcup_{k=1}^{m} C_k\right) + P(C_{m+1}) \\
&\leq \sum_{k=1}^{m} P(C_k) + P(C_{m+1}) \\
&= \sum_{k=1}^{m+1} P(C_k).
\end{aligned}$$

This completes our proof by induction, and establishes that the statement

$$P\left(\bigcup_{k=1}^{n} C_k\right) \leq \sum_{k=1}^{n} P(C_k).$$

(ii)

$$P\left(\bigcup_{k=1}^{\infty} C_k\right) = P\left(\lim_{n\to\infty}\bigcup_{k=1}^{n} C_k\right)$$

$$= \lim_{n\to\infty} P\left(\bigcup_{k=1}^{n} C_k\right)$$

$$\leq \lim_{n\to\infty}\sum_{k=1}^{n} P(C_k)$$

$$= \sum_{k=1}^{\infty} P(C_k),$$

Proof: (iii) P $(\bigcap_{k=1}^{n} C_k) \geq \sum_{k=1}^{n} P(C_k) - (n-1)$

If we apply Boole's inequality to C_k^c, we have,

$P(\bigcup_{k=1}^{n} C_k^c) \leq (\sum_{k=1}^{n} P(C_k^c)$

and using the fact that $\cup\, C_k^c = (\cap\, C_k)^c$ *and* $P(C_k^c) = 1 - P(C_k)$, *we get,*

$P(\bigcup_{k=1}^{n} C_k^c)$ = P $(\bigcap_{k=1}^{n} C_k)^c$ = 1- P $(\bigcap_{k=1}^{n} C_k)$

⇒ 1- P $(\bigcap_{k=1}^{n} C_k) \leq (\sum_{k=1}^{n} P(C_k^c)$

⇒ 1- P $(\bigcap_{k=1}^{n} C_k) \leq \sum_{k=1}^{n}(1 - P(C_k))$

⇒ 1- P $(\bigcap_{k=1}^{n} C_k) \leq n - \sum_{k=1}^{n} P(C_k)$

⇒ P $(\bigcap_{k=1}^{n} C_k) \geq \sum_{k=1}^{n} P(C_k) - (n-1)$

Examples:

1. There are 10,000 transister chips with Pr{transister failure} ≈ 1/1,000,000.

Chip fails if any transister fails.

Pr{Chip fails}

= Pr{ ∪ [ith transister fails]} ≤ Σ Pr{ith transister fails}

≈ (10,000) × $\frac{1}{1000000}$= 0.01

2. Your friend tells you that he had four job interviews last week. He says that based on how the interviews went, he thinks he has a 20% chance of receiving an offer from each of the companies he interviewed with. Nevertheless, since he interviewed with four companies, he is 90% sure that he will receive at least one offer. Is he right?

Let C_k be the event that your friend receives an offer from the k th company, k=1,2,3,4. Then, by the union bound (Boole' Inequality)

$$(i)\ P\left(\bigcup_{k=1}^{n} C_k\right) \leq \sum_{k=1}^{n} P(C_k)$$

$$= 0.2+ 0.2+0.2+0.2 =0.8$$

Thus the probability of receiving at least one offer is less than or equal to 80%.

3 Let A1,A2,...,An be any events. Define the indicator random variables X1, X2,...,Xn as

$$Xi = 1 \text{ if } A_i \text{ occurs}$$

$$= 0 \text{ otherwise}$$

We define X=X1+X2+X3+...+Xn, prove $P(X \geq 1) = P(\bigcup_{i=1}^{n} A_i)$

Solution:

E(X) =EX1+EX2+EX3+...+EXn

(by linearity of expectation)

=P(A1)+P(A2)+...+P(An)

$\Rightarrow E(X) = \sum_{i=1}^{n} P(A_i)$

Using Markov's inequality, we get $P\{(X) \geq a\} \leq \frac{E(X)}{a}$

Choosing a=1, we get, $P\{(X) \geq 1\} \leq \frac{E(X)}{1}$

$\Rightarrow P(X \geq 1) \leq E(X)$

$\Rightarrow P(X \geq 1) \leq \sum_{i=1}^{n} P(A_i)$ Eq. 1

From Boole's Inequality, we get

$P(\cup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i)$ Eq. 2

Combining Equations 1 and 2, we get,

$P(X \geq 1) = P(\cup_{i=1}^{n} A_i)$

CAUCHY SCHWARTZ'S INEQUALITY

What is the Cauchy-Schwarz Inequality?

The Cauchy-Schwarz Inequality (also called Cauchy's Inequality, the Cauchy-Bunyakovsky-Schwarz Inequality and Schwarz's Inequality) is useful for bounding expected values that are difficult to calculate. It allows you to split E[X Y] into an upper bound with two parts, one for each random variable.

The formula is:

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

Given that X and Y have finite variances.

What this is basically saying is that for two random variables, X and Y, the expected value of the square of them multiplied together **$E(XY)^2$** will always be less than or equal to the expected value of the product of the squares of each. **$E(X^2)E(Y^2)$.**

Applications: The Cauchy-Schwarz inequality is arguably the inequality with the widest number of applications. As well as probability and statistics, the inequality is used in many other branches of mathematics, including:

- Classical Real and Complex Analysis,
- Hilbert spaces theory,
- Numerical analysis,
- Qualitative theory of differential equations.
- linear algebra
- Probability theory
- Vector algebra

Statements:

(i) $\left(\sum_{i=1}^{n} X_i^2\right)\left(\sum_{i=1}^{n} Y_i^2\right) \geq \left(\sum_{i=1}^{n} X_i Y_i\right)^2$

(ii) $E(X^2)E(Y^2) \geq \left(E(XY)\right)^2$

(iii) $|\rho(X,Y)| \leq 1$

[where $\rho(X,Y) is\ the\ correlation\ coefficient\ between\ X\ and\ Y$]

(ii)

Look at the following two expectations:

$$E[(aX+bY)^2] = a^2E[X^2] + b^2E[Y^2] + 2abE[XY] \geq 0$$

$$E[(aX-bY)^2] = a^2E[X^2] + b^2E[Y^2] - 2abE[XY] \geq 0$$

Now let $a^2 = E[Y^2]$ and $b^2 = E[X^2]$. This gives

$$2abE[XY] \geq -2a^2b^2$$

$$2abE[XY] \leq 2a^2b^2$$

Dividing by $2ab$ results in

$$-\sqrt{E[X^2]}\sqrt{E[Y^2]} \leq E[XY] \leq \sqrt{E[X^2]}\sqrt{E[Y^2]}$$

which is equivalent to

$$|E[XY]| \leq \sqrt{E[X^2]}\sqrt{E[Y^2]}$$

$$\Rightarrow Squarring\ both\ sides\ we\ get, E(X^2)E(Y^2) \geq \left(E(XY)\right)^2$$

(iii) Normalize X and Y as follows. Let μ_x, μ_y, σ_x and σ_Y be their means and standard deviations, respectively. Then

$$\frac{X-\mu_X}{\sigma_X} \quad \text{and} \quad \frac{Y-\mu_Y}{\sigma_Y}$$

each have zero mean and unit standard deviation. (In particular this will mean, below, that their second moments are 1.) We can create a new pair of random variables

$$\left(\frac{X-\mu_X}{\sigma_X} \pm \frac{Y-\mu_Y}{\sigma_Y}\right)^2.$$

Since each takes non-negative values, the means are non-negative as well [xxx xref forward to where this is proved …it seems obvious but actually requires proof]:

$$E\left[\left(\frac{X-\mu_X}{\sigma_X} \pm \frac{Y-\mu_Y}{\sigma_Y}\right)^2\right] \geq 0.$$

FOILing out we have

$$E\left[\left(\frac{X-\mu_X}{\sigma_X}\right)^2 \pm \frac{2(X-\mu_X)(Y-\mu_Y)}{\sigma_X\sigma_Y} + \left(\frac{Y-\mu_Y}{\sigma_Y}\right)^2\right] \geq 0.$$

Using the linearity of expectation and recalling that the normalized variables have second moments equal to 1, we have

$$\begin{aligned}
2 \pm 2E\left[\frac{(X-\mu_X)(Y-\mu_Y)}{\sigma_X\sigma_Y}\right] &\geq 0 \\
-1 \leq E\left[\frac{(X-\mu_X)(Y-\mu_Y)}{\sigma_X\sigma_Y}\right] &\leq 1 \\
-\sigma_X\sigma_Y \leq E\left[(X-\mu_X)(Y-\mu_Y)\right] &\leq \sigma_X\sigma_Y \\
-\sigma_X\sigma_Y \leq \mathrm{Cov}(X,Y) &\leq \sigma_X\sigma_Y \\
|\mathrm{Cov}(X,Y)| &\leq \sigma_X\sigma_Y.
\end{aligned}$$

$$\Rightarrow \left|\frac{Cov(X,Y)}{\sigma_x\sigma_y}\right| \leq 1 \Rightarrow |\rho(X,Y)| \leq 1$$

Let X and Y be two random variables with $EX = 1, Var(X) = 4$, and $EY = 2, Var(Y) = 1$. Find the maximum possible value for $E[XY]$.

Solution

Using $\rho(X,Y) \le 1$ and $\rho(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$, we conclude

$$\frac{EXY - EXEY}{\sigma_X \sigma_Y} \le 1.$$

Thus

$$\begin{aligned} EXY &\le \sigma_X \sigma_Y + EXEY \\ &= 2 \times 1 + 2 \times 1 \\ &= 4. \end{aligned}$$

In fact, we can achieve $EXY = 4$, if we choose $Y = aX + b$.

$$Y = aX + b \quad \Rightarrow \quad \begin{cases} 2 = a + b \\ 1 = (a^2)(4) \end{cases}$$

Solving for a and b, we obtain

$$a = \frac{1}{2}, \quad b = \frac{3}{2}.$$

Note that if you use the Cauchy-Schwarz inequality directly, you obtain:

$$\begin{aligned} |EXY|^2 &\le EX^2 \cdot EY^2 \\ &= 5 \times 5. \end{aligned}$$

Thus

$$EXY \le 5.$$

But $EXY = 5$ cannot be achieved because equality in the Cauchy-Schwarz is obtained only when $Y = \alpha X$. But here this is not possible.

The Cauchy Schwarz inequality

$$\left(\sum_{i=1}^{n} a_i^2\right).\left(\sum_{i=1}^{n} b_i^2\right) \geq \left(\sum_{i=1}^{n} a_i b_i\right)^2$$

taking $a_i = |x_i - \bar{x}|$ and $b_i = 1/n$,

$$\left(\sum_{i=1}^{n} |x_i - \bar{x}|^2\right).\left(\sum_{i=1}^{n} \frac{1}{n^2}\right) \geq \left(\sum_{i=1}^{n} |x_i - \bar{x}|.\frac{1}{n}\right)^2$$

$$\left(\sum_{i=1}^{n} (x_i - \bar{x})^2\right).\left(n.\frac{1}{n^2}\right) \geq \left(\frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}\right)^2$$

$$\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n} \geq \left(\frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}\right)^2$$

$$\sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}} \geq \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$$

$$S.D \geq M.D$$

WEAK LAW OF LARGE NUMBERS:

The Weak Law of Large Numbers, also known as Bernoulli's theorem, states that if you have a sample of independent and identically distributed random variables, as the sample size grows larger, the sample mean will tend toward the population mean.

Definition . Let f(x) be a density with mean μ and variance σ^2.
For i.i.d. random variables X1,X2,...,Xn, drawn from f(x), the sample mean, denoted by $\overline{X}$, is defined as

$$\overline{X} = (X1+X2+...+Xn)/n.$$

Note that since the Xi's are random variables, the sample mean $\overline{X}$, is also a random variable.

In particular, we have

$$E[\overline{X}] = (EX1+EX2+...+EXn)/n \qquad \text{(by linearity of expectation)}$$

$$= \frac{nE(X)}{n} \qquad \text{(since } E(X_i) = E(X) \text{ for all i)}$$

$$= E(X)$$

Also, the variance of $\overline{X}$ is given by

$$Var(\overline{X}) = Var(X1+X2+...+Xn)/n^2 \qquad \text{(since } Var(aX)=a^2Var(X))$$

$$= \frac{Var(X_1)+Var(X_2)+var(X_3)+\cdots....+Var(X_n)}{n^2}$$

(since the Xi's are independent)

$$= \frac{nVar(X_i)}{n^2} \qquad \text{(since Var(Xi)=Var(X))}$$

$$= \frac{Var(X)}{n}$$

Now let us state and prove the weak law of large numbers (WLLN).

Let X1 , X2 , ... , Xn be i.i.d. random variables with a finite expected value $E(Xi) = \mu < \infty$. Then, for any $\epsilon>0$,

$$\lim_{n\to\infty} P[|\overline{X} - \mu| \geq \epsilon] = 0$$

We shall prove the result only under the additional assumption that the random variables have a finite variance σ^2.

Now, as $E(\overline{X}) = \mu$ and Var $(\overline{X}\,)\ \frac{Var(X)}{n} = \frac{\sigma^2}{n}$.

it follows from Chebyshev’s inequality that

$$P[|\overline{X} - \mu| \geq \epsilon] \leq \frac{\sigma^2}{n\epsilon^2}$$

$$\Rightarrow P[|\overline{X} - \mu| \geq \epsilon] \leq \frac{\sigma^2/\epsilon^2}{n}$$

Which tends to 0 as n tends to infinity.

Weak law of large numbers Let f(x) be a density with mean μ and variance σ^2. Let $\overline{X}$ be the sample mean of a random sample of size n from f(x). Let ϵ and δ be any two specified numbers satisfying $\epsilon > 0$ and $0 < \delta < 1$. If n is any integer greater than $\frac{\sigma^2}{\epsilon^2 \delta}$, then

Then $P[|\overline{X} - \mu| < \epsilon] \geq 1 - \delta$

Proof: From Markov's Inequality we get, $P\{(X) \geq a\} \leq \frac{E(X)}{a}$

$$\Rightarrow P\{g(X) \geq a\} \leq \frac{Eg(X)}{a}$$

$$\Rightarrow P\{g(X) < a\} \geq 1 - \frac{Eg(X)}{a}$$

Let $g(X) = (\overline{X} - \mu)^2$ and $a = \epsilon^2$

Then, $P[|\overline{X} - \mu| < \epsilon] = P[(\overline{X} - \mu)^2 < \epsilon^2] \geq 1 - \frac{E(\overline{X}-\mu)^2}{\epsilon^2}$

$$\Rightarrow P[(\overline{X} - \mu)^2 < \epsilon^2] \geq 1 - \frac{\frac{\sigma^2}{n}}{\epsilon^2}$$

$$\Rightarrow P[(\overline{X} - \mu)^2 < \epsilon^2] \geq 1 - \delta$$

For $\delta > \frac{\sigma^2}{n\epsilon^2}$ or $n > \frac{\sigma^2}{\delta\epsilon^2}$

$$\Rightarrow P[|\overline{X} - \mu| < \epsilon] \geq 1 - \delta$$

Example: Suppose that some distribution with an unknown mean has variance equal to 1. How large a sample must be taken in order that the probability wi1l be at least .95 that the sample mean $\overline{X}$ will lie within 0.5 of the population mean?

Solution: We have $\sigma^2 = 1$, $\epsilon = .5$, and $\delta = .05$;

Therefore $n > \frac{\sigma^2}{\delta\epsilon^2} = \frac{1}{0.05(0.5)^2} = 80$

Example: How large a sample must be taken in order that you are 99 percent certain that is within the sample mean $\overline{X}$ will lie within $.5\sigma$ of the population mean?

Solution: We have $\sigma^2 = 1$, $\epsilon = .5\sigma$, and $\delta = .0.01$;

Therefore $n > \frac{\sigma^2}{\delta\epsilon^2} = \frac{1}{0.01(0.5\sigma)^2} = \frac{1}{0.01(0.5)^2} = 400$

Example: A certain brand of lightbulb has lifetime that is exponentially distributed with mean A hours, A unknown. I try to estimate A by letting n lightbulbs run independently, and recording & averaging their lifetimes. How large should n be, so that I can be at least 90% sure that the estimate I get is within 5% of the actual average A?

Solution: Here $\overline{X} = (X1+...+Xn)/n$ with Xi ~exponential (λ) (λ unknown),

$\mu = \frac{1}{\lambda} = A$, $\sigma^2 = \frac{1}{\lambda^2} = A^2$, $\varepsilon = 0.05\,A$, $1 - \delta = 0.9$; $\delta = 0.1$

$n > \frac{\sigma^2}{\delta\epsilon^2} = \frac{A^2}{0.1(0.05A)^2} = 4000$

Want n large enough so that this probability is at most.1, so n=400
large enough.

Central-limit theorem Let f(x) be a probability distribution function with mean μ and finite variance σ^2.

Let $\overline{X}$ be the sample mean of a random sample of size n from f(x). Let the random variable Z be defined by

$$Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Then, the distribution of Z approaches the standard normal distribution as n approaches infinity.