Tail Sum Formula

Let X be a random variable.

$$E(X) = \underbrace{\sum_{\text{all } x} x \cdot \underbrace{P(X=x)}}\; = \underbrace{\sum_{\text{all } x} \underbrace{P(X \geq x)}}$$

Choose whichever you think is easier to solve :)

random Var (distribution

normal variable.

$P(X \geq x)$ what does this mean?

Let "X be a random variable"

"X represents a number in an event"

Rolling a die → let X be the number we roll.

$P(X = 1)$ "The probability that 1 is the number we roll"

$P(X \geq x)$

$P(X \geq 5) = P(X=5) + P(X=6)$

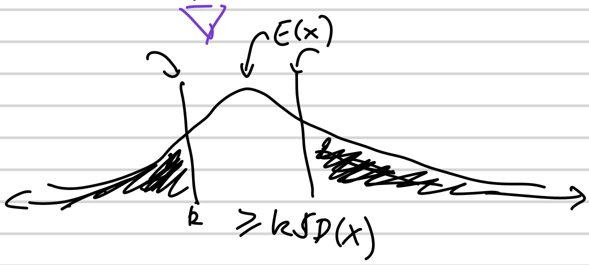
= "The probability that the number we roll is greater than or equal to 5"

Markov & Chebyshev Inequalities

Markov: $P(X \geq a) \leq \dfrac{E(X)}{a}$ $\quad (X \geq 0,\ a > 0)$

Chebychev (Chebyshev) : $P(|X - E(X)| \geq k\,SD(X)) \leq \dfrac{1}{k^2}$ $\quad k \neq 0$

area $\leq \dfrac{E(X)}{a}$



Exercise 63 In Toronto, the average income is roughly $\$65,000$. According to a report by Fong and Partners, you need an annual salary of $\$135,000$ to be considered "middle class".

(1) Find an upper bound for the percentage of individuals who are considered at least middle class, *i.e.*, their income is over 135,000.

(2) According to statcan, the standard of deviation of incomes for 25 – 29 year olds, is $\boxed{\$30,000.}$ Can you find a better upper bound for the percentage of individuals (aged 25 – 29) who are considered at least middle class?

(Hint: you need to use two different inequalities)

Let X be income.

$E(X) = 65$

① $P(X \geq 135) \leq \frac{E(X)}{135} = \frac{65}{135} = 0.48148$ $\boxed{48\%}$

② $SD(X) = 30$ $\boxed{P(|X - E(X)| \geq k\,SD(X))} \leq \frac{1}{k^2}$

$P(X \geq 135)$ $P(|X - 65| \geq k \cdot SD(X))$

$= P(X - 65 \geq 135 - 65)$

$= P(X - 65 \geq 70)$ $X - 65 \geq 70$

$\cancel{\leq P\left(\frac{X-65}{SD(X)} \geq \frac{70}{SD(X)}\right)}$ $SD(X) = 30$ $X \geq 65$ $|X-65| \geq 70$

$= P\left(X - 65 \geq \frac{70\,SD(X)}{SD(X)}\right)$ $SD(X) = 30$ $\cancel{X \leq 65}$ $|X-65| \geq 70$

$= P\left(X - 65 \geq \frac{70}{30} SD(X)\right)$ $-(X-65)$ $\boxed{65 - X \geq 70}$

$X \leq -5$

$= P\left(|X - 65| \geq \underset{k}{\frac{7}{3}} SD(X)\right)$

$\leq \frac{1}{\left(\frac{7}{3}\right)^2}$

$= \frac{1}{\frac{49}{9}}$

$= \frac{9}{49}$

$= 0.18367$

$P(X \geq 135) \leq 0.18367$

18 %

Moments, Variants, Central Limit theorem.

i^{th} - Moments $\rightarrow$ $E(f(X)) = \sum_{\text{all } x} f(x)\, P(X=x)$

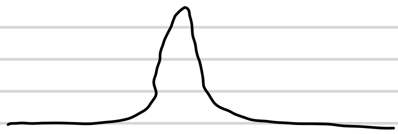
$\rightarrow f(X) = X^i$

2nd Moment $E(X^2) = \sum_{\text{all } x} x^2 P(X=x)$

1st Moment $E(X) = \sum_{\text{all } x} x\, P(X=x)$

Variance: $Var(X) = E(X^2) - E(X)^2$

High Variance



Low Variance

Central limit theorem

$S_n = X_1 + X_2 + \cdots + X_n$ (ind. & same distribution) Equal in distribution

$P(X_i = x) = P(X_j = x),\ \forall i, j, x$

$$P\left(a \le \frac{S_n - E(S_n)}{SD(S_n)} \le b\right) \approx \Phi(b) - \Phi(a)$$

$$\frac{a - \frac{1}{2} - \mu}{\sigma}$$ E SD

$$\boxed{\begin{aligned} E(S_n) &= n\,E(X_i) \\ SD(S_n) &= \sqrt{n}\,SD(X_i) \end{aligned}}$$

Square root law

X_i, X_j They're two different random variables We assume they are equal in distribution

$$\boxed{P(X_i = x) = P(X_j = x)\ \forall x}$$

Rolling a 6-sided die
Pulling a # out of a hat w/ 6 numbers

Exercise 66 Let D_i be a random digit between 0 and 9. Let X_i be the last digit of D_i^2 (so if $D_i = 8$ then $D_i^2 = 64$ and $X_i = 4$). Let $\bar{X}_n = \frac{X_1+\ldots+X_n}{n}$ be the average of a large number n where each X_i is determined from randomly chosen D_i.

Probably should use central lim. Theorem

(1) Predict the value of $\bar{X}_n$ for large n.

(2) If you just had to predict the last digit of $\bar{X}_{100}$ what digit should you chose to maximize your chance of being right, and what is the probability?

D_i	0	1	2	3	4	5	6	7	8	9
D_i^2	0	1	4	9	16	25	36	49	64	81
X_i	0	1	4	9	6	5	6	9	4	1

$P(X_i=1) = \frac{2}{10}$ $\quad P(X_i=0) = \frac{1}{10}$ $\quad P(X_i=2) = 0$

$P(X_i=3) = 0$ $\quad P(X_i=4) = \frac{2}{10}$ $\quad P(X_i=5) = \frac{1}{10}$

$P(X_i=6) = \frac{2}{10}$ $\quad P(X_i=7) = 0$ $\quad P(X_i=8) = 0$

$P(X_i=9) = \frac{2}{10}$

$$\frac{1}{10} + \frac{2}{10} + 0 + 0 + \frac{2}{10} + \frac{1}{10} + \frac{2}{10} + 0 + 0 + \frac{2}{10} = \frac{10}{10} = 1$$

✓ $E(X_i) = \sum_{\text{all}} x\,P(X_i=x) = \cancel{0 \cdot \frac{1}{10}} + 1\cdot\frac{2}{10} + \cancel{2\cdot 0} + \cancel{3\cdot 0} + 4\cdot\frac{2}{10} + 5\cdot\frac{1}{10}$

$+ 6\cdot\frac{2}{10} + \cancel{7\cdot 0} + \cancel{8\cdot 0} + 9\cdot\frac{2}{10}$

$= 4.5$

$E(X_i^2) = \sum_{\text{all } x} x^2 P(X_i=x) = 1^2\cdot\frac{2}{10} + 4^2\cdot\frac{2}{10} + 5^2\cdot\frac{1}{10} + 6^2\cdot\frac{2}{10}\cdot 9^2\cdot\frac{2}{10}$

$= 293/10$

$Var(X_i) = E(X_i^2) - E(X_i)^2 = \left(\frac{293}{10}\right) - (4.5)^2 = 9.05$

$SD(X_i) = \sqrt{Var(X_i)} = \sqrt{9.05} = 3.008322$

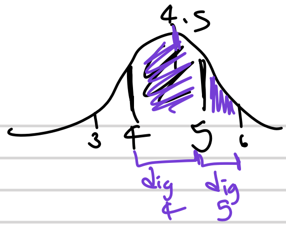
(1) $E(\bar{X}_n) = E\left(\frac{X_1+X_2+\cdots+X_n}{n}\right) = \frac{1}{n} E(X_1 + \cdots + X_n)$

$= \frac{1}{\cancel{n}} \cancel{n} \cdot E(X_i)$

$= E(X_i)$

$= 4.5$

$E(\overline{X}_n) = 4.5$



$\boxed{\text{mode} \quad \lfloor np + p \rfloor}$

only binomial distrib^n

2 digits are either 4 or 5.

$\lfloor E(x) + p \rfloor \quad p < 1/2$ (Assume)

$\lfloor 4.5 + p \rfloor = \boxed{4}$

$P(\overline{X}_n = ?)$

does not exist, but we assume $< 1/2$

$\boxed{P(4 \le \overline{X}_n < 5)}$ $\quad P(5 \le \overline{X}_n < 6)$ $\quad P\left(a \le \boxed{\frac{S_n - E(S_n)}{SD(S_n)}} \le b\right)$

$S_n = X_1 + \cdots + X_n$

$= P(4 \le \frac{S_n}{n} < 5)$ $\quad \overline{X}_n = \frac{X_1 + \cdots + X_n}{n} = \frac{S_n}{n}$

$n = 100$

$= P(4 \le \frac{S_{100}}{100} < 5)$

$\boxed{SD(S_n) = \sqrt{n}\, SD(X_i) \quad E(S_n) = n\, E(X_i)}$

$= P(4 \cdot 100 \le S_{100} < 5 \cdot 100)$

Square root law

$= P(400 - E(S_n) \le S_{100} - E(S_n) < 500 - E(S_n))$

$= P\left(\frac{400 - E(S_n)}{SD(S_n)} \le \frac{S_{100} - E(S_n)}{SD(S_n)} < \frac{500 - E(S_n)}{SD(S_n)}\right)$

$E(X_i) = 4.5$

$SD(X_i) = 3.008322$

$= P\left(\frac{400 - 450}{30.08322} \le \sim < \frac{500 - 450}{30.08322}\right)$

$\Rightarrow E(S_{100}) = \overset{n}{100} \cdot \overset{E(X_i)}{4.5} = 450$

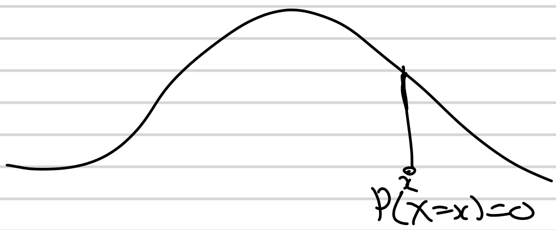
$= P(-1.662334 \le \sim < 1.66234)$

$\Rightarrow SD(S_{100}) = \overset{\sqrt{n}}{\sqrt{100}} \cdot \overset{SD(X_i)}{3.008322} = 30.08322$

$\overset{CLT}{\approx} \Phi(1.66234) - \Phi(-1.66234)$

$= 0.903$

$\boxed{90.3\%}$



Exercise 65 Suppose you were walking down the street and you found $\$100,000$ just lying there with no one around. You pick it up and decide to keep it because, we're in a word problem so why not? You then decide to invest that money like a responsible adult. There are four different stock options, but you don't know which will give you a profit and which will take your money. You know one of them gives you a profit of 200 dollars for every 1000 you invest, another one gives 100 for every 1000 you invest, a third gives you nothing for every 1000 you invest and the last one makes you lose 100 for every 1000 you invest.

(1) Suppose you randomly choose to invest all $100,000$ into one stock. What is the probability that you will earn (at least) $8,000$ dollars?

(2) Suppose you decide to invest 1000 dollars, 100 times where each time you randomly chose a stock to invest in. What is the probability that you will earn (at least) $8,000$ dollars? (Use the central limit theorem)

Exercise 64 According to the book, in roulette, a "house special" is a bet on the numbers 0, 00, 1, 2 and 3 out of the 38 total possible numbers. So you have a $\frac{5}{38}$ chance of winning and for every dollar you bet, you get seven dollars back if you win (and nothing if you lose). If you make the "house special" bet 300 times where you bet \$1 each time, what is the chance that you have at least as much money as you started with? (In other words, what's the probability you end with greater than or equal to \$300.) Use normal approximation.

Exercise 62 Suppose that we have three different events: A, B, and C. Let $P(A) = \frac{1}{5}$, $P(B) = \frac{1}{4}$ and $P(C) = \frac{1}{3}$. Let X be the total number of events A, B and C that occur.

(1) What is the expected value of X?

(2) If the events are disjoint, what is the variance?

(3) If the events are independent, what is the variance?