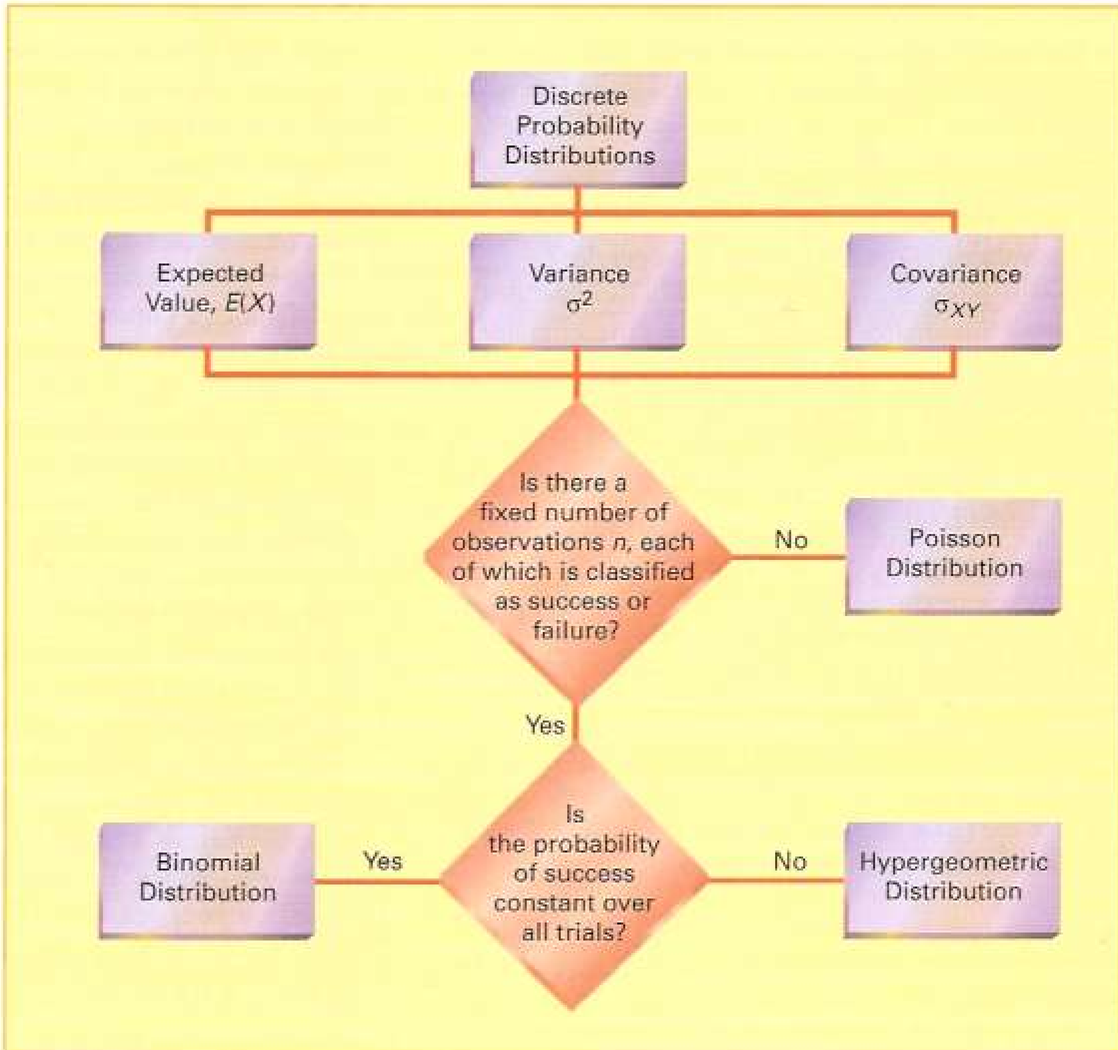

Discrete Probability Distributions
Expected Value, $E(X)$
Variance σ^2
Covariance σ_{XY}
Is there a fixed number of observations n, each of which is classified as success or failure?
No
Poisson Distribution
Yes
Is the probability of success constant over all trials?
Yes
Binomial Distribution
No
Hypergeometric Distribution

OR: Addition Rule
AND: Multiplication Rule

Independent Trials

(T1, 3:30)

The following examples are of a set of 3 die rolls.
We know die rolls are INDEPENDENT.

$$P(\text{Three "1's" are rolled}) = \frac{1}{6}\times\frac{1}{6}\times\frac{1}{6} = \frac{1}{216}$$

$$P(\text{All Outcomes } \le 2) = \frac{1}{3}\times\frac{1}{3}\times\frac{1}{3}$$

This is the probability the a 1 or a 2 is rolled out of 6 possible outcomes. Thus for each trial

$$P(\text{1 or 2 rolled}) = P(\text{1 rolled}) + P(\text{2 rolled}) = \frac{1}{6}+\frac{1}{6} = \frac{2}{6} = \frac{1}{3}$$

$$P(\text{Not All "6's"}) = 1 - P(\text{All "6's"}) = 1 - \frac{1}{6}\times\frac{1}{6}\times\frac{1}{6} = \frac{216}{216} - \frac{1}{216} = 99.53\%$$

These are **MUTUALLY EXCLUSIVE** and **COLLECTIVELY EXHASUTIVE** which means they are **COMPLEMENTARY**.

$$P(\text{No "6's"}) = \frac{5}{6}\times\frac{5}{6}\times\frac{5}{6} = \frac{125}{216} = 58\%$$

This is the probability of anything but a 6 being rolled and doing so 3 times.

$$P(\text{Exactly Two "6's" rolled in 3 trials}) = \frac{1}{6}\times\frac{1}{6}\times\frac{5}{6} = \frac{5}{216} = 2.3\%$$

What is the probability of rolling 6 exactly twice? Must consider all possible outcome combinations. First examine the individual events.

$P(\bar{6}, 6, 6) = \frac{1}{6}\times\frac{1}{6}\times\frac{5}{6}$ $\quad P(6, \bar{6}, 6) = \frac{1}{6}\times\frac{1}{6}\times\frac{5}{6}$ $\quad P(6, 6, \bar{6}) = \frac{1}{6}\times\frac{1}{6}\times\frac{5}{6}$

These are **MUTUALLY EXCLUSIVE** and **COLLECTIVELY EXHASUTIVE** which means we can use the addition rule ("or").

$$P(\bar{6}, 6, 6) + P(6, \bar{6}, 6) + P(6, 6, \bar{6}) = \frac{1}{6}\times\frac{1}{6}\times\frac{5}{6} + \frac{1}{6}\times\frac{1}{6}\times\frac{5}{6} + \frac{1}{6}\times\frac{1}{6}\times\frac{5}{6} = 6.94\%$$

Say we change the question, what is the probability of rolling exactly 3 6's in 10 rolls?

$$P(6, 6, 6, \bar{6}, \bar{6}, \bar{6}, \bar{6}, \bar{6}, \bar{6}, \bar{6}) = \frac{1}{6}\times\frac{1}{6}\times\frac{1}{6}\times\frac{5}{6}\times\frac{5}{6}\times\frac{5}{6}\times\frac{5}{6}\times\frac{5}{6}\times\frac{5}{6}\times\frac{5}{6}$$

Now we must find the probability of all the other roll combinations and take the sum as above. The other combinations will have the same probability. **Is there a better way?**

Binomial Probability Distribution

When to use:

- '*n*' identical trials
  - e.g.: 15 tosses of a coin; ten light bulbs taken from a warehouse
- Two mutually exclusive outcomes on each trials
  - e.g.: Head or tail in each toss of a coin; defective or not defective light bulb

Can have more than two outcomes as long as they can be categorized as success and not success.

- Trials are independent
  - The outcome of one trial does not affect the outcome of the other
- Constant probability for each trial
  - e.g.: Probability of getting a tail is the same each time we toss the coin
    - Two sampling methods
      - Infinite population without replacement
      - Finite population with replacement

Examples:

- A manufacturing plant labels items as either defective or acceptable
- A firm bidding for contracts will either get a contract or not
- A marketing research firm receives survey responses of "yes I will buy" or "no I will not"
- New job applicants either accept the offer or reject it

Binomial is a **discrete** distribution. The Binomial Distribution is used when the discrete random variable of interest is the number of successes obtained in a sample of n observations.

EXAM Exam is more about recognizing the proper probability distribution to use for a given type of data. "Is this a good problem to use binomial on?"

Binomial Distribution Formula

$$P(X) = \frac{n!}{X!(n-X)!} p^X (1-p)^{n-X}$$

P(X) = probability of **X** successes in **n** trials, with probability of success **p** on each trial

X = number of 'successes' in sample, (X = 0, 1, 2, ..., n)

n = sample size (number of trials or observations)

p = probability of "success"

Example: Flip a coin four times, let x = # heads:
n = 4
p = 0.5
1 - p = (1 - .5) = .5
X = 0, 1, 2, 3, 4

- The number of combinations of selecting X objects out of n objects is

$$\binom{n}{X} = \frac{n!}{X!(n-X)!}$$

What is the probability of one success in five observations if the probability of success is .1?

X = 1, n = 5, and p = .1

$$P(X=1) = \frac{n!}{X!(n-X)!} p^X (1-p)^{n-X}$$

$$= \frac{5!}{1!(5-1)!} (.1)^1 (1-.1)^{5-1}$$

$$= (5)(.1)(.9)^4$$

$$= .32805$$

Note: P(X => 3) = P(X=3) + P(X=4), where n=4.
Also: P(X < 3) = P(X=0) + P(X=1) + P(X=2), where n=4.

Excel's Binomial Function

=**BINOMDIST**(no. of successes, no. of trials, prob. of success, cumulative?)

Example

=BINOMDIST(2,8,0.5, FALSE) (=0.11)

"Probability of tossing (exactly) two heads within 8 trials"

=BINOMDIST(2,8,0.5, TRUE) (=0.14)

"Probability of tossing two heads or less within 8 trials"

For the 10 rolls with exactly 3 sixes probability: =Binomdist(3, 10, 1/6, False) = 15.5%

For P(3 outcomes >= 5 in 20 trials) the probability of success for each trial would be

$$p = \frac{\text{two possible successes}}{\text{six total possible outcomes}} = \frac{2}{6}$$

Coin toss, p = .5: P(Exactly 10 heads in 20 tosses) = Binomdist(10, 20, .5, False)
(trials are independent)

The **True/False** in the **Binomdist()** function indicates weather or not the **cumulative** probability should be returned.
Example:
P(10 Heads or Less in 20 tosses) = P(10H in 20 tosses) + P(9H in 20 tosses) + P(8H in 20 tosses) + P(7H in 20 tosses) + … + P(1H in 20 tosses) + P(0H in 20 tosses)

To find this probability in Excel we would specify **TRUE** *in the* **binomdist()** *function.*

P(10 Heads or Less in 20 tosses) = Binomdist(10, 20, .5, True) = 58.8%

Also note:

P(11 or more heads in 20 tosses) = 1 – P(10 or less heads in 20 tosses)

Binomial Setting

Examples

- Number of times newspaper arrives on time (i.e., before 7:30 AM) in a week/month
- Number of times I roll “5” on a die in 20 rolls
- Number of times I toss heads within 20 trials
- Students pick random number between 1 and 10. Number of students who picked “7”
- Number of people who will vote “Republican” in a group of 20
- Number of left-handed people in a group of 40

Example: Probability of receiving the newspaper before 7:30 is known to be 1/3.
What is P(exactly 5 newspapers before 7:30 in a week) ?
Use the binomial distribution, each weekday is an independent event.

P(5 successes out of 7 events) = Binomdist(5, 7, 1/3, False) = 3.84%

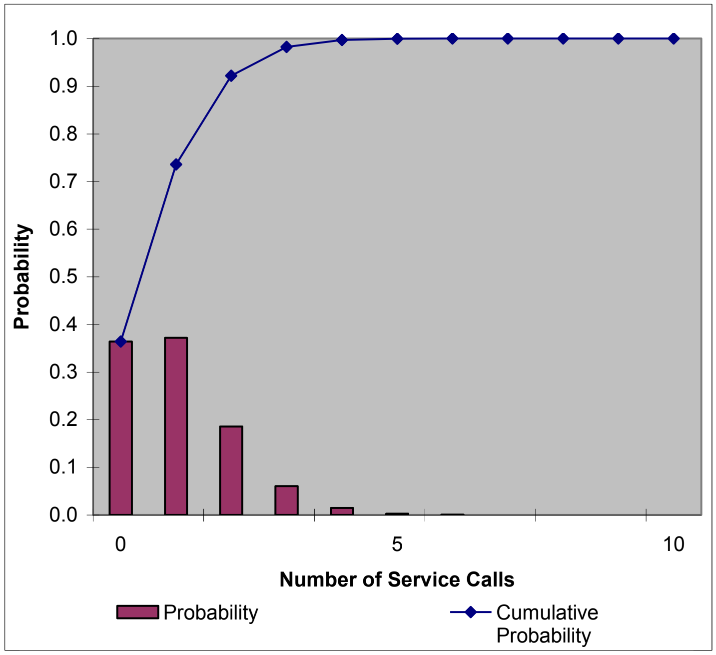
Simple Event: Always check if outcomes are equally likely.

Ex: P(10 students out of 40 choose the number 7) = Binomdist(10, 40, .1, False) = .359%

Service Center Staffing

Numer of Repair Calls per Day	Probability	Cumulative Probability
0	0.3642	0.36417
1	0.3716	0.73577
2	0.1858	0.92157
3	0.0607	0.98224
4	0.0145	0.99679
5	0.0027	0.99952
6	0.0004	0.99994
7	5.3633E-05	0.99999
8	5.8832E-06	1
9	5.603E-07	1
10	4.6882E-08	1
11	3.4792E-09	1
12	2.3076E-10	1
13	1.3766E-11	1
14	7.4249E-13	1
15	3.6367E-14	1
16	1.6235E-15	1
17	6.6266E-17	1
18	2.4793E-18	1
19	8.5219E-20	1
20	2.6957E-21	1
21	7.8592E-23	1
22	2.1143E-24	1
23	5.2528E-26	1
24	1.206E-27	1
25	2.5597E-29	1

Assumptions
- 50 computers sold **(plot is only showing 10)**
- Prob. customer calls for service: p= 0.02
- Want < 5% that there is no engineer



Assume repair man always takes a full day. p = 0.02 is the probability a customers calling for a repair man on any given day. There are a total of 50 customers so having 50 repair men would assure always having one available to send. But with p=.02, the probability that all 50 customers will call for repair in the same day is (.02)^50= 1.1259E-85 , very small.

So we are able to compromise and want the number of repair men necessary so the probability of a customer calling and one is not available is < 5%. What is that number of repair men? This is a **CUMULATIVE BINOMIAL** Distribution problem. In these discrete trial cases cumulative is just the sum of the probabilities up to the point in question. We are asking an "OR" question so we use the addition rule.

The key to Cumulative Probability questions is to realize you are answering the question **"What is the probability I have this number of successes OR LESS".**

In this example we see that it is virtually certain we will have 6 service calls ***OR LESS*** in a day.

In this case we are asking P(0 service calls in a day) "OR" P(1 service call in a day) "OR" P(2 service calls in a day) "OR" P(3 service calls in a day). We have to **SUM** these probabilities because having 3 service calls in a day means you've already realized the probabilities of having 0, 1, or 2 service calls in that day. This is why the **CUMULATIVE** probability **INCREASES** as the number of successes increase.

In this problem we are asked how many successes can we sum and still have 95% confidence that we will have enough repair men? We can restate the problem in terms of the inverse, P(< 5% chance of no repair man available).

P(0 repair calls) = 36.4% [63.6% chance no repair man]

P(1 repair call **OR LESS**) = P(0 repair calls) + P(1 repair calls) = 73.5%
[26.5% chance no repair man]
P(2 repair calls **OR LESS**) = P(0 repair calls) + P(1 repair calls) + P(2 repair calls) = 92.1%
[7.9% chance no repair man]
P(3 repair calls **OR LESS**) = P(0 repair calls) + P(1 repair calls) + P(2 repair calls) + P(3 repair calls) = 98.2%
[1.8% chance no repair man]

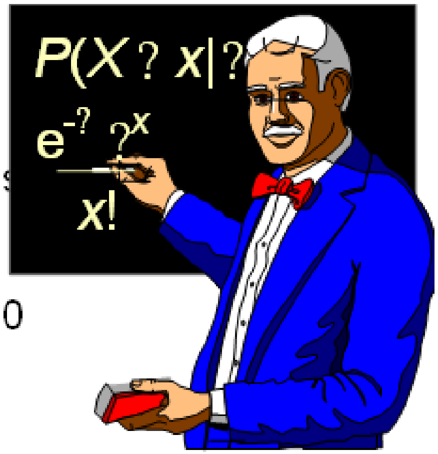
In this case we have to go all the way up to 3 repair men before we have a less than 5% chance of not being able to respond.

EXAM When is it Binomial and when is it Poisson? A good rule of thumb to follow, if there is a maximum number of successes it is a Binomial problem. If the number of successes is open ended it is a Poisson problem.

Binomial has an upper limit by definition. For instance we say, a coin is flipped 5 times…

Poisson Distribution

- Poisson Process:
  - Discrete events in an "interval"
    - ? The probability of One Success in an interval is stable
    - ? The probability of More than One Success in this interval is 0
  - The probability of success is independent from interval to interval
  - e.g.: number of customers arriving in 15 minutes
  - e.g.: number of defects per case of light bulbs



We use the Poisson Distribution in discrete counting processes where we are counting the number of successes. We can compute the probability of some number of successes in a given interval if the average or expected number of successes per interval is known. (lambda) We must be considering a continuous interval which can be shrunk to zero. As the interval is shrunk to zero the probability of a success in that interval must also go to zero. The number of events in one interval is INDEPENDENT from the number of events in another interval.

- Apply the Poisson Distribution when:
  - You wish to count the number of times an event occurs in a given area of opportunity
  - The probability that an event occurs in one area of opportunity is the same for all areas of opportunity
  - The number of events that occur in one area of opportunity is independent of the number of events that occur in the other areas of opportunity
  - The probability that two or more events occur in an area of opportunity approaches zero as the area of opportunity becomes smaller
  - The average number of events per unit is λ (lambda)

$$P(X) = \frac{e^{-\lambda}\lambda^{x}}{X!}$$

where:

X = number of successes per unit

λ = expected number of successes per unit

e = base of the natural logarithm system (2.718

P(X) = the probability of X successes given a knowledge of lambda.

Excel's Poisson Function

=**POISSON**(no. of occurences, mean, cumulative?)
(SUCCESSES)

Example

=POISSON(5,2,FALSE) False gives exactly (=0.036)

"Probablity that (exactly) five customers arrive wihtin an hour when the overall average is two"

=POISSON(5,2,TRUE) True gives "or less" (Cumulative) (=0.983)

"Probability that five or less customers arrive wihtin an hour when the overall average is two"

P(6 or more in 1 hour) = 1 – P(5 or less in 1 hour) = 1 – Poisson(5,2,True)

	λ								
x	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
0	0.9048	0.8187	0.7408	0.6703	0.6065	0.5488	0.4966	0.4493	0.4066
1	0.0905	0.1637	0.2222	0.2681	0.3033	0.3293	0.3476	0.3595	0.3659
2	0.0045	0.0164	0.0333	0.0536	0.0758	0.0988	0.1217	0.1438	0.1647
3	0.0002	0.0011	0.0033	0.0072	0.0126	0.0198	0.0284	0.0383	0.0494
4	0.0000	0.0001	0.0003	0.0007	0.0016	0.0030	0.0050	0.0077	0.0111
5	0.0000	0.0000	0.0000	0.0001	0.0002	0.0004	0.0007	0.0012	0.0020
6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0003
7	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Example: Find P(X = 2) if λ = .50

$$P(X = 2) = \frac{e^{-\lambda}\lambda^X}{X!} = \frac{e^{-0.50}(0.50)^2}{2!} = .0758$$

The number of **successes** is an **INTEGER** because we are counting discrete events.
Lambda is a **REAL** number and is given in **PER INTERVAL** units.

Be careful to compare the units of lambda to the units of the interval. For example P(exactly 10 customers in two hours when lambda = 2 customers / hour) = Poisson(10, 4, False), **lambda = 2 per hour and the interval is 2 hours.**

Poisson Setting

Examples

- Number of accidents at an intersection in 6 months
- Number of people entering a bank in a 30-minute interval
- Number of kids ringing the doorbell in 30 minutes for Halloween
- Number of times a Microsoft machine crashes within 24 hours
- Number of sewing flaws per (100) garment(s)

Poisson Example:

Customers arrive during lunch hour, of interest, the number of customers arriving each minute.
Evaluate the Poisson distributions applicability:
Event: Customer arriving
Area of Opportunity: 1 minute interval
Independence: we assume that the probability of a customer arriving in a given 1 minute interval is the same as the probability for all other one minute intervals and the arrival of one customer in a given interval has no effect on the probability of any other customer in any other interval.
Probability: the probability of a two or more customers arriving in an interval approaches zero as the length of the interval approaches zero.

It is known that on average 3 customers arrive per minute during the lunch hour: $\lambda = 3$
What is the probability of 2 customers arriving in a minute?

$$P(X=2) = \frac{e^{-3.0}(3.0)^2}{2!} = \frac{9}{(2.71828)^3(2)} = 0.2240$$

What is the probability of more than two customers will arrive?

$$P(X>2) = P(X=3) + P(X=4) + \cdots + P(X=\infty)$$

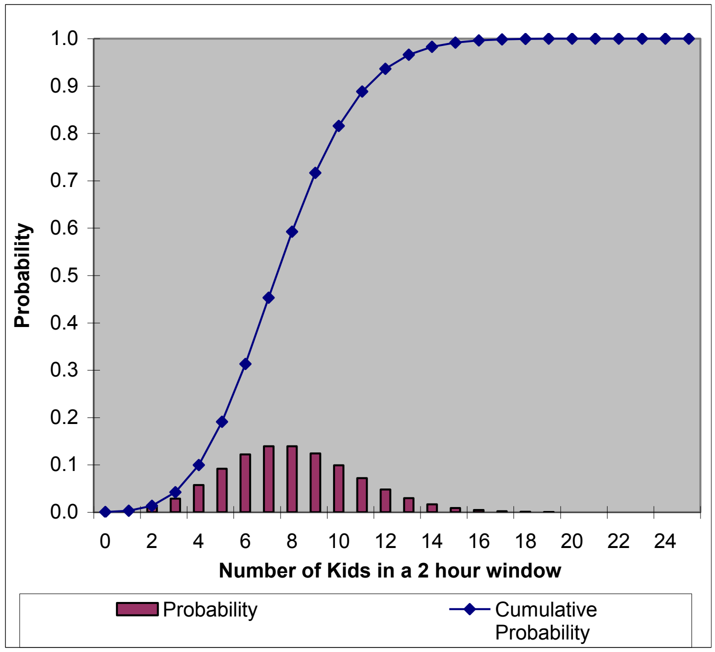
$$P(X>2) = 1 - P(X \le 2) = 1 - [P(X=0) + P(X=1) + P(X=2)]$$

$$P(X>2) = 1 - \left[\frac{e^{-3.0}(3.0)^0}{0!} + \frac{e^{-3.0}(3.0)^1}{1!} + \frac{e^{-3.0}(3.0)^2}{2!}\right]$$
$$= 1 - [0.0498 + 0.1494 + 0.2240] = 1 - 0.4232 = 0.5768$$

Halloween

Numer of kids in a 2 hour window	Probability	Cumulative Probability
0	0.000335	0.000335
1	0.002684	0.003019
2	0.010735	0.013754
3	0.028626	0.042380
4	0.057252	0.099632
5	0.091604	0.191236
6	0.122138	0.313374
7	0.139587	0.452961
8	0.139587	0.592547
9	0.124077	0.716624
10	0.099262	0.815886
11	0.072190	0.888076
12	0.048127	0.936203
13	0.029616	0.965819
14	0.016924	0.982743
15	0.009026	0.991769
16	0.004513	0.996282
17	0.002124	0.998406
18	0.000944	0.999350
19	0.000397	0.999747
20	0.000159	0.999906
21	0.000061	0.999967
22	0.000022	0.999989
23	0.000008	0.999996
24	0.000003	0.999999
25	0.000001	1.000000

Assumptions
On average 4 kids per hour (lambda). Arrivals are independent, not groups.
Note: Lambda units do not match problem units!
- Want < 1% or less chance of running out of candy.



Compare the units between event interval and lambda, they do not match! Must tell Excel lambda is 8. Want a 1% ***or less*** probability of running out of candy. "**Or Less**" tells us the problem is **CUMULATIVE**.