BinomialToolsCheatsheet

March 3, 2020

1 PROBABILITY TOOLS IN ENGINEERING

1.1 Discrete problems *(Working with quantities of things)*

1.1.1 BINOMIAL DISTRIBUTION (*The answer to* Yes/No *&* Good/Bad *questions,* for independent to each other objects)

The binomial distribution is frequently used to model the number of successes/fails in a n-long population(or sequence of trials), where p is the probability of success/fail. It can also show the number of trials needed to achieve a first success/failure.
The binomial distribution is a special case of Poisson distribution.

- Trials are **Independent** (the outcome of yes/no for each object, does not dependent on the other objects)

- Trials **with replacement** *(in each trial the examined object/component re-enters again the system, the system remains the same throught the process)*

Equation:

$P(k) = \binom{n}{k} * p^k q^{n-k} = \frac{n!}{k!(n-k)!} * p^k (1-p)^{n-k}$
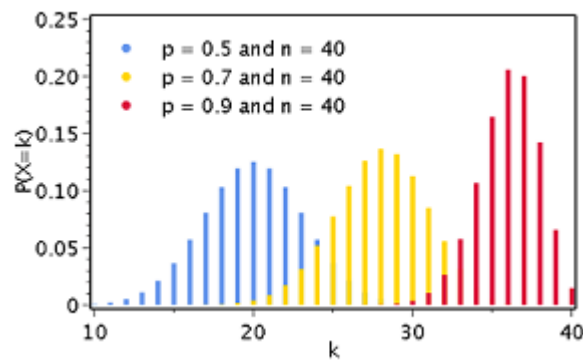- k: number of desired yes
- n: number of population or trials
- p: probability of yes to occur
- q =(1-p): probability of no to occur
- $\binom{n}{k}$: Binomial coefficient, Combinations of k objects out of n objects; Their order does not matter
- $p^k q^{n-k}$: Probability of any n-long sequence

Engineering Examples:

1. Probability of failures in an experiment, by examing a portion of results

A series of one hundred binary results is inspected by analysing four randomly selected results. Results are independent. If one of the four failed, the experiment is rejected. What is the probability that the experiment is accepted if it contains five failures?

Steps before solution:
- Results are clearly separated quantities (discrete).
- Each result is "independent".
So, the Binomial equation applies

Solution:
- the population of objects we want to examine is 4, so **n=4**.
- the experiment is accepted only if 0 failed results appear, so **k=0**.
- we know there are 5 in 100 failed tests; that means that if we take 1 random result it has a 5/100 probability to be a failed one, so **p=5%=5/100=0.05**.
- the question can be denoted as $P(k = 0) =?$.

By using the equation we have: $P(k = 0) = \frac{4!}{0!(4-0)!}0.05^0(1 - 0.05)^{4-0} = 0.814$, which is **the probability to accept the experiment, by examing 4 out of 100 results with a 5% fail probabilty**.

! What is interesting is that for the same experiment we can examine different questions.
eg. What is the probability that the experiment is *not accepted* if it contains five failures?
- the experiment is not accepted if 1 or 2 or 3 or 4 failed results appear, so **k= 1 or 2 or 3 or 4**.
- "or" in probabilities means addition, so **P(k=[1,2,3,4])=P(k>1)=P(k=1)+P(k=2)+P(k=3)+P(k=4)**.

By using the equation we have: $P(k = 1) = \frac{4!}{1!(4-1)!}0.05^1(1 - 0.05)^{4-1} = 0.171$.
By using the equation for each k we have: $P(k > 1) = 0.171 + 0.014 + 0.0004 + 0.0000006 = 0.185$, which is the probability to not accept the experiment, by examing 4 out of 100 results with a 5% fail probabilty.

Funny fact: adding the two solutions 0.814+0.185 we have 0.999999 which is 100%. That means that there is 100% chance that the experiment will be accepted or not.

2. Probability that a customer will purchase defected products

It is known that, in a certain manufacturing process, 1% of the products are defective. If the a customer purchases 50 of these products selected at random, what is the probability that he receives 2 or less defective products?

Steps before solution:
- Products are clearly separated quantities (discrete).
- Each product is "independent".
So, the Binomial equation applies

Solution:
- the population of objects we want to examine is 50, so **n=50**.
- to be successful, the defected products must be 2 or less, so **k= 0 or 1 or 2**.
- we know there is 1% chance that a product is defected, so **p=1%=0.01**.

- the question can be denoted as $P(k = [0, 1, 2]) = P(k \leq 2) =?$.
- "or" in probabilities means addition, so **P(k=[0,1,2])=P(k≤ 2)=P(k=0)+P(k=1)+P(k=2)**.

By using the equation we have: $P(k = 0) = \frac{50!}{0!(50-0)!} 0.01^0 (1 - 0.01)^{50-0} = 0.605$.

By using the equation for each k we have: $P(k \leq 2) = 0.605 + 0.305 + 0.075 = 0.985$, which is **the probability to send the shipment with 2 defected products or less, if there is a 1% fail probabilty during manufacturing**.

3. Probability of defects in Quality control of products, by examing a portion of products

An engineer is testing whether or not 90% of the DVD players produced by his company conform to specifications and randomly selects a batch of 12 DVD players from each day's production. The day's production is acceptable provided no more than 1 DVD player fails to meet specifications. Otherwise, the entire day's production has to be tested.

1. What is the probability that the engineer incorrectly passes a day's production as acceptable if only 80% of the day's DVD players actually conform to specification?

2. What is the probability that the engineer unnecessarily requires the entire day's production to be tested if in fact 90% of the DVD players conform to specifications?

Question 1
Steps before solution:
- Products are clearly separated quantities (discrete).
- Each product is "independent".
So, the Binomial equation applies

Solution a:
- the population of objects we want to examine is 12, so **n=12**.
- to incorrectly pass the products, the defected products must be more than 2, so **k= 2 or 3 or 4 ... or 12**.
- we know there is 80% chance that a product is ok, so **p=80%=0.8**.
- the question can be denoted as $P(k = [2, 3, 4, ..., 12]) = P(k > 1) =?$.
- "or" in probabilities means addition, so
P(k>1)=P(k=2)+P(k=3)+P(k=4)+P(k=5)+...+P(k=10)+P(k=11)+P(k=12).

This approach is not practical as we have to calculate a lot of numbers.
The chance that we will be correct or wrong is 100%. So an easier way to examine this problem is:
- **P(correct or wrong) =100%=1 =P(correct)+P(wrong) =>**
- **P(wrong) = 1 - P(correct) =>**
- So, **P(k=[2, 3, 4,..., 12]) = 1 - P(k=[0, 1])**.
- By using this relation, we only need to calculate the chance that we are correct *(Solution b)*.

Solution b:
- the population of objects we want to examine is 12, so **n=12**.
- to be successful, the defected products must be 1 or less, so **k= 0 or 1**.
- we know there is 80% chance that a product is ok, so the chance for a defect is **p** = 1 - 80% = **0.2**.
- the question can be denoted as $P(k \leq 1) =?$.

- "or" in probabilities means addition, so **P(k≤ 1)=P(k=0)+P(k=1)**.

By using the equation we have: $P(k = 0) = \frac{50!}{0!(50-0)!}0.01^0(1 - 0.01)^{50-0} = 0.28$.

By using the equation for each k we have: $P(k \leq 1) = 0.28 + 0.376 = 0.656$, which is **the probability to correctly pass the products with 1 defected product or less, if there is a 20% fail probabilty during production**.

By using the previously mentioned relation **P(wrong)=1-P(correct)**, we calculate that **P(wrong)=1-0.656=0.344**.

1.1.2 Independent trials / experiments *(planning formulas)*

1. **Probability of an incorrect test/trial.**

   - Trials/tests are **Independent** (the outcome of yes/no for each trial, does not dependent on the other trials)
   - Test **Relability** Experiment **Validity** (A *correct Test* doesn't mean *correct conlcusion*; The metrics used for validating a hypothesis/parameter should be thoroughly examined)

$P(\text{Test Incorrect}) + P(\text{Test Correct}) = 1, P()>0$

$P(\text{Test Incorrect}) = P(\text{Human Error} |or| \text{System Error}) = P(\text{Human Error}) + P(\text{System Error})$ (Human error, System error : independent to each other)

$P(HumanError) = \frac{h}{n}, P(SystemError) = \frac{s}{n}$

   - h: Number of total Human errors
   - s: Number of total System errors
   - n: Number of total population or trials/tests
   - h + s ≤ n
   - ph: Probability of Human errors
   - ps: Probability of System errors
   - p: probability to occur
   - q=(1-p) : probability

Formula:

```
[11]: h = 5
      s = 20
      n = 50
      print("P(Human Error)=", h/n, "%")
      print("P(System Error)=", s/n, "%")
      print("P(Test Incorrect)=", (h/n)+(s/n), "%")
```

```
P(Human Error)= 0.1 %
P(System Error)= 0.4 %
P(Test Incorrect)= 0.5 %
```

```
[12]: ph = 0.1
      ps = 0.3
      print("P(Human Error)=", ph, "%")
```

```
print("P(System Error)=", ps, "%")
print("P(Test Incorrect)=", ph+ps, "%")
```

```
P(Human Error)= 0.1 %
P(System Error)= 0.3 %
P(Test Incorrect)= 0.4 %
```

2. **Probability of correct test/trial in a row.**

- Trials/tests are **Independent** (the outcome of yes/no for each trial, does not dependent on the other trials)
- Test **Relability** Experiment **Validity** (A *correct Test* doesn't mean *correct conlcusion*; The metrics used for validating a hypothesis/parameter should be thoroughly examined)

$P(\text{Test Incorrect}) + P(\text{Test Correct}) = 1, P() > 0$

$P(\text{Test Incorrect}) = P(\text{Human Error }|\text{or}| \text{ System Error}) = P(\text{Human Error}) + P(\text{System Error})$ (Human error, System error : independent to each other)

$P(k \geq n) = p * \sum_{k=n}^{\infty}(1-p)^{k-1} = (1-p)^{n-1}$

- k: Correct trials/tests in a row
- n: Wanted number of correct trials/tests
- p: probability of correct to occur
- q=(1-p) : probability of incorrect to occur

```
[30]: p = 0.01
      n = 10
      print("P(Test correct) = ", p, "%")
      print("P(Tests correct for ", n, " trials) = ", (1-p)**(n-1), "%")
```

```
P(Test correct) =  0.01 %
P(Tests correct for  10  trials) =  0.9135172474836408 %
```

3. **Number of tests/trials needed for specific confidence level.**

- Trials/tests are **Independent** (the outcome of yes/no for each trial, does not dependent on the other trials)
- Test **Relability** Experiment **Validity** (A *correct Test* doesn't mean *correct conlcusion*; The metrics used for validating a hypothesis/parameter should be thoroughly examined)

$P(\text{Test Incorrect}) + P(\text{Test Correct}) = 1, P() > 0$

$P(\text{Test Incorrect}) = P(\text{Human Error }|\text{or}| \text{ System Error}) = P(\text{Human Error}) + P(\text{System Error})$ (Human error, System error : independent to each other)

$n \geq \frac{log(a)}{log(q)}$

$N = n * k$

- n: number of trials/tests necessary for 1-a confidence level result
- a: Probability of wrong conclusion (Type error I)
- 1-a: Confidence level
- p: Probability of correct test to occur

5

- q = 1-p : Probability of incorrect test to occur
- k: Tests needed to validate a Hypothesis (is not included in this chapter)
- N: number of total tests needed to verify a conclusion in the experiment

```python
import math
a = 0.05
p = 0.1
print("P(Test correct) = ", 1-q, "%")
print("Confidence level = ", 1-a, "%")
print("Number of tests/trials for ", 1-a, " confidence level result = ", math.
 ↪log(a)/math.log(1-p))
```

```
P(Test correct) =  0.09999999999999998 %
Confidence level =  0.95 %
Number of tests/trials for  0.95  confidence level result =  28.43315880574342
```

```python
import math
a = 0.05
p = 0.7
k = 10
result = k * math.log(a)/math.log(1-p)
print("Number of tests/trials for ", 1-a, " confidence level result = ", math.
 ↪log(a)/math.log(1-p))
print("Total number of trials/tests to verify experiment Hypothesis = ", result)
```

```
Number of tests/trials for  0.95  confidence level result =  2.4882059318866436
Total number of trials/tests to verify experiment Hypothesis =
24.882059318866435
```

[ ]: