DISCRETE PROB. DISTRIBUTIONS

<u>Defn</u>: discrete uniform distribution

If the random variable X has k equally likely outcomes then the discrete uniform distribution is given by

$$f(x;k) = \frac{1}{k}$$

Notation: $f(x;k)$ indicates k is a parameter. In this case the parameter k is the total number of possible outcomes of the random variable X

The Binomial distribution

<u>Defn</u>: Bernoulli Trial. If an experiment has two possible outcomes such as success or failure it can be called a Bernoulli Trial.

<u>Defn</u>: Bernoulli Process:

1. The experiment consists of n repeated trials.
2. Each trial is a Bernoulli trial
3. The probability of success, denoted by p, remains constant from trial to trial.
4. The repeated trials are independent.

<u>Note</u>: Requirement 3. means that experiments where we sample with replacement are candidates for a Bernoulli process, but those where we sample without replacement can't be Bernoulli processes because p changes in that case.

Example: A manufacturing process results in %5 defective products. If we select 3 products from this manufacturing process and inspect them, what are the probabilities for the different numbers of defective products?

Soln: * Each product selection is a Bernoulli trial because it has two outcomes: defective (D) and non-defective (N)

* The whole experiment is a Bernoulli process because:
  1. There are 3 repeated trials
  2. Each trial is a Bernoulli trial
  3. The probability of defective $p = 0.05$ for each trial.
  4. The 3 trials are independent from each other. The fact that the first product is D or N has no bearing on the second.

Outcome	x
NNN	0
NDN	1
NND	1
DNN	1
NDD	2
DND	2
DDN	2
DDD	3

Lets compute $P(X=0)$:

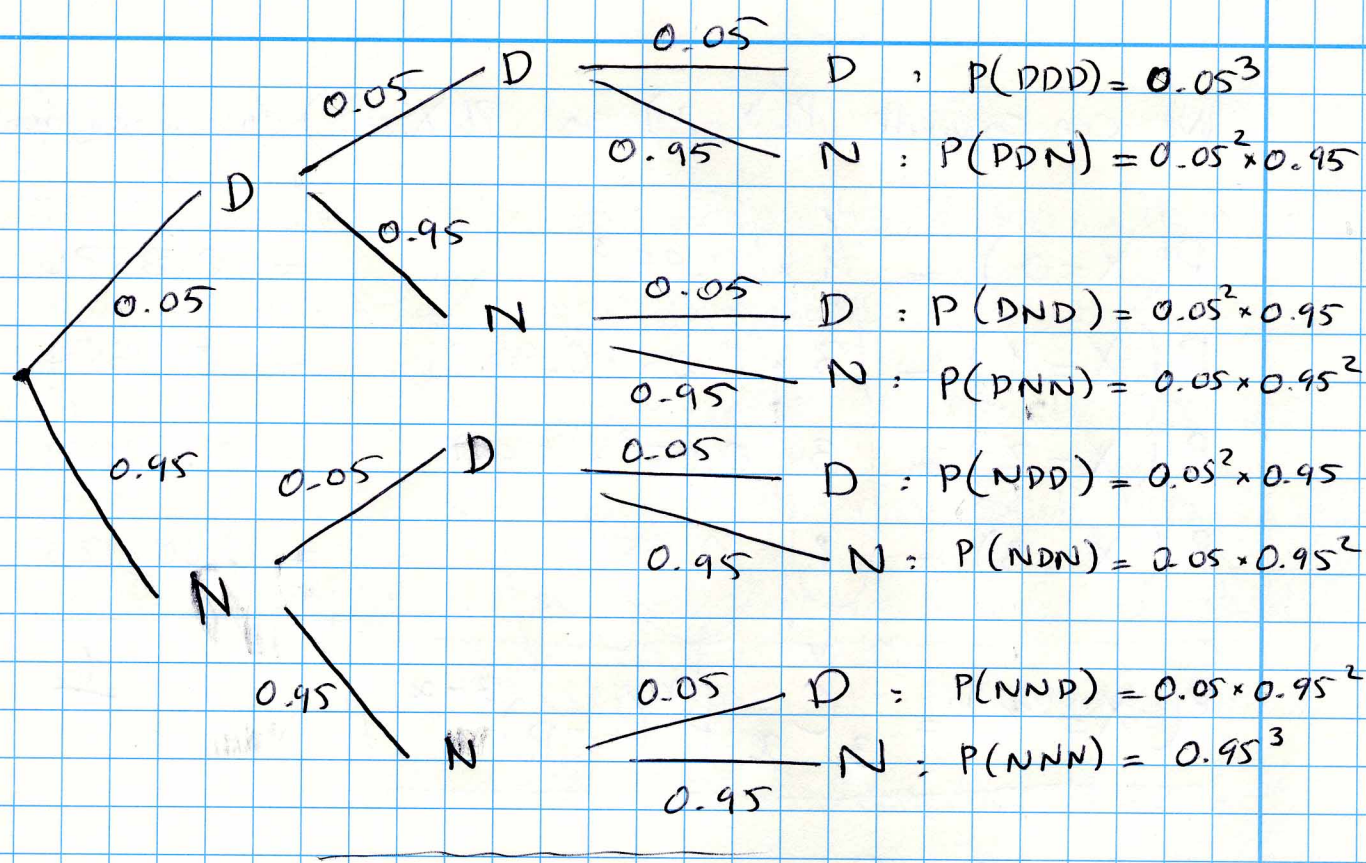
- $X=0$ happens only for NNN
- We'll use the multiplicative rule and since the trials are independent

$$P(NNN) = \underbrace{(1-0.05)}^{3} = 0.95^3$$

Each trial has 2 outcomes

$P(D) = 0.05$

so $P(N) = 1 - 0.05 = 0.95$



Lets compute $P(X=1)$:

* $X=1$ happens for atcomes DNN, NDN, NND

- $P(DNN) = 0.05 \times 0.95 \times 0.95 = 0.05 \times 0.95^2$

$P(NDN) = 0.95 \times 0.05 \times 0.95 = 0.05 \times 0.95^2$

$+ \; P(NND) = 0.95 \times 0.95 \times 0.05 = 0.05 \times 0.95^2$

$+$

$P(X=1) = 3 \times 0.05 \times 0.95^2$

<u>Important things to notice</u>:

a) $P(DNN) = P(NDN) = P(NND)$

b) $P(X=x)$ depends only on

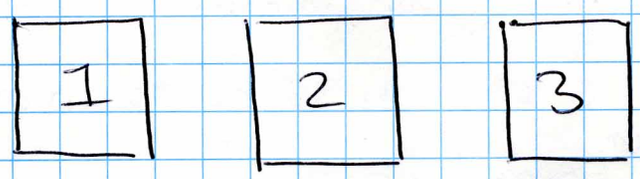
1. The number of defectives x
2. The probability of each trial resulting in a defective $p = 0.05$

We can compute $P(X=2)$ and $P(X=3)$ in a similar manner

$$P(X=0) = 1 \times 0.95^3 = 0.8574$$
$$P(X=1) = 3 \times 0.05 \times 0.95^2 = 0.1354$$
$$P(X=2) = 3 \times 0.05^2 \times 0.95 = 0.0071$$
$$P(X=3) = 1 \times 0.05^3 = 0.000125$$
$$+ \qquad 1$$

$$\boxed{P(X=x) = {}_3C_x \, p^x (1-p)^{3-x}}$$

Think of ${}_3C_x$ as the number of ways to choose x spots out of 3 spots to place the x defectives.



Let these 3 cards denote the 3 spots. Lets place all 3 in a bag, then choose 2 without replacement. We place the defectives in those 2 spots. Now the outcomes are

1 2] → results in DDN
2 1
1 3] → results in DND
3 1
2 3] → results in NDD
3 2

hence selection order doesn't matter

$$\frac{3 \times 2}{2} = {}_3C_2 \text{ ways.}$$

<u>Defn</u> : Binomial distribution.

The number X of successes in a Bernoulli process consisting of n Bernoulli trials is called a <u>binomial random variable</u>. If each trial can result in a <u>success</u> with probability p and a <u>failure</u> with probability $q = 1-p$ then the probability distribution of the binomial random variable X is called the <u>binomial distribution</u> and is given as

$$\boxed{P(X = x) = b(x; n, p) = {}_nC_x \, p^x q^{n-x}}$$

<u>Notes</u> : * The n, p in $b(x; n, p)$ are the 2 parameters. n is the number of trials. p is the probability of success for each trial.

* q is always $1-p$
* The book uses the notation $\binom{n}{x}$ for ${}_nC_x$
* For our previous example $n = 3$, $p = 0.05$

  so $P(X = x) = b(x; 3, 0.05) = {}_3C_x \, 0.05^x \, 0.95^{3-x}$

* For our previous example

  Success = Defective

  Failure = Non-defective

<u>Example continued</u>: For the same manufacturing process with $p=0.05$ of generating a defective product at each trial, what is the probability that a quality control engineer who takes a sample of 100 products will find

a) 10 defectives

b) 10 or more defectives

<u>Soln</u>: This is a Bernoulli process. So $P(X=x)$ is a binomial distribution with $n=100, p=0.05$

a) $$P(X=10) = {}_{100}C_{10} \times 0.05^{10} \times (1-0.05)^{100-10}$$

$$= {}_{100}C_{10} \times 0.05^{10} \times 0.95^{90}$$

$$\approx 0.0167$$

Hard to compute with a calculator. Easy to compute in MATLAB. Use the factorial command

$${}_{100}C_{10} = \frac{100!}{90!\,10!} = \underbrace{\frac{\text{factorial}(100)}{\text{factorial}(90) \times \text{factorial}(10)}}_{\text{MATLAB}}$$

b) $P(X \geq 10) = \sum_{i=10}^{100} P(X = i)$

$$= \sum_{i=10}^{100} {}_{100}C_i \times 0.05^i \times 0.95^{100-i}$$

Again one can easily compute this in MATLAB with a for loop.

```
for i = 10 : 100
    result = result + (factorial(100)/
    (factorial(i)*factorial(100-i)) *
        0.05^i * 0.95^(100-i)
end
```

The answer comes out as $P(X \geq 10) = 0.0282$

Remember these are exactly the numbers given for the example first day of class.

$$\boxed{\begin{array}{l}\text{The mean and variance of the binomial distribution } b(x; n, p) \text{ are} \\ \qquad \mu = np \quad \text{and} \quad \sigma^2 = npq\end{array}}$$

We will be able to prove these later when we learn about the means and variances of linear combinations of random variables.

Binomial distribution example

A pharmceutical company has developed a drug for a certain disease that cures 80% of the patients. If a group of 50 patients suffering from this disease are treated with this drug:

a) What is the expected number of patients that will be cured?

b) What is the variance?

c) What is the probability that 30 of the 50 will be cured?

d) What is the probability that more than 30 will be cured?

Solution: X: the number of patients cured at of 50 is a binomial random variable with $n=50$, $p=0.8$

a) The mean of a binomial r.v. is given by np,

so $\mu = E[X] = np = 50 \times 0.8 = 40$

b) The variance of a binomial r.v. is given by npq,

so $\sigma^2 = 50 \times 0.8 \times (1-0.8) = 8$

c) $P(X=30) = b(30; n=50, p=0.8) = {}_{50}C_{30}\, 0.8^{30}\, 0.2^{20} = 6.1 \times 10^{-4}$

d) $P(X>30) = \sum_{i=31}^{50} {}_{50}C_i\, 0.8^i\, 0.2^{50-i} = 0.9991$

In a more realistic scenario, the pharmaceutical company would be asking the following:

We think we have a drug that cures 80% of all patients with a certain disease. We tested the drug on a sample of 50 patients and observed that 35 were cured. Can we claim that the drug indeed cures with $p=0.8$?

soln. : We now have one of the two tools necessary to answer this question:

$$P(X=35) = b(35; n=50, p=0.8) = 0.0299$$

But what can we do with this to either prove or disprove the claim that $p=0.8$? We will learn more about this when we study confidence intervals and hypothesis testing.