Chapter 3

Discrete and Continuous Random Variables

Discrete Random Variables:

Binomial Distribution.

A Binomial Random Variable is a repeated Bernoulli trial.

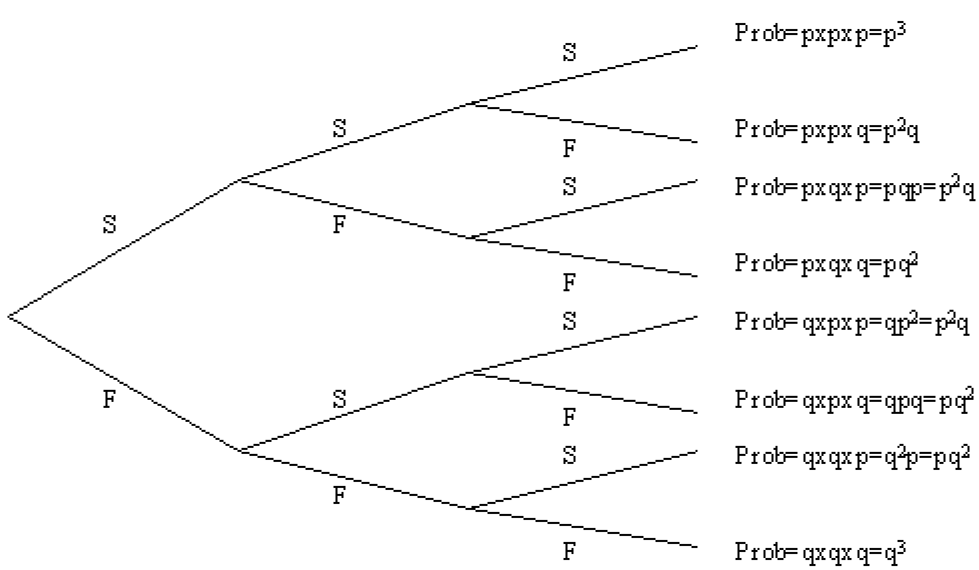
For a single experiment to be a Bernoulli trial there are three criteria that need to be met:

1. Each trial results in one of two possible outcomes, denoted success (S) or failure (F).

2. The probability of success (S) remains constant from trial-to-trial and is denoted by p. Write q = 1−p for the constant probability of failure (F).

3. The trials are independent.
Then to have a Binomial random variable we have 4 properties in total, the three above, and the following:
4) The experiment consists of n repeated trials.

We can use a tree diagram to see all the possible outcomes associated with an experiment of this type. The tree below is for a Binomial random variable with three trials, hence the three branches.



Source: http://www.icse.xyz/discuss/binomial/derive.html

Using the tree, we can see there are only 4 outcomes if we do not care about the ordering. They are:

(S,S,S), (S,S,F), (S,F,F), (F,F,F)

Counting from the tree we see that these outcomes occur with the following rates:

(S,S,S) occurs only one way, with probability p^3;

(S,S,F) occurs 3 ways [(S,S,F), (S,F,S), (F,S,S)], with probability p^2q^1;

(S,F,F) occurs 3 ways [(S,F,F), (F,S,F), (F,F,S)], with probability p^1q^2;

(F,F,F) occurs only one way, with probability q^3.

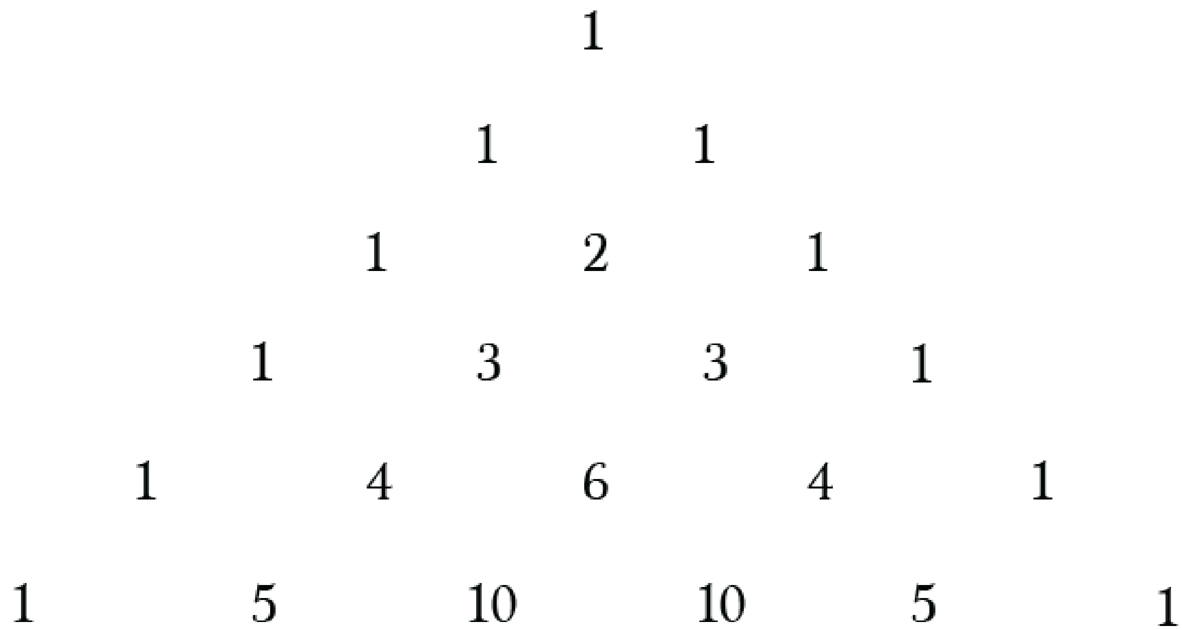
These combinations

1, 3, 3, 1 are derived using ${}_nC_r = \binom{n}{r}$

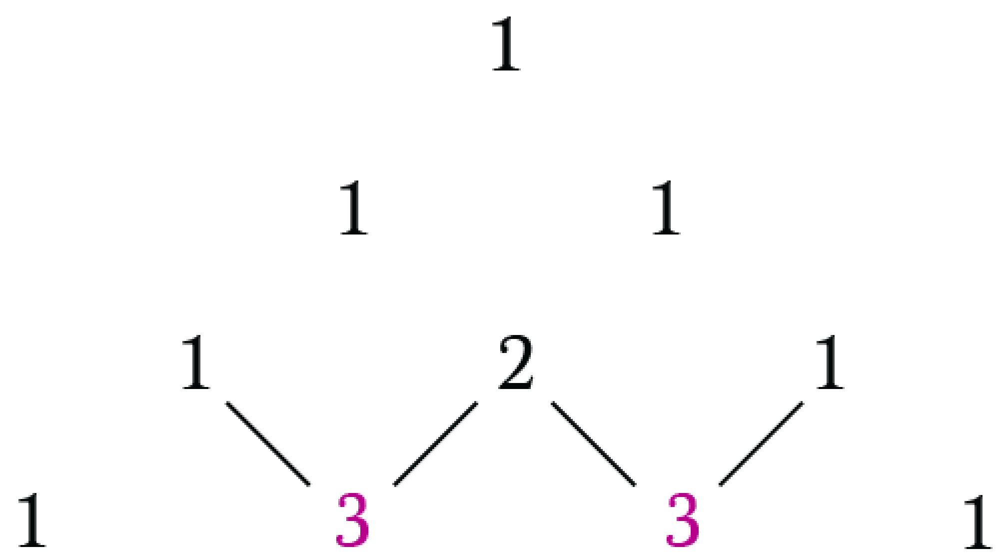
We can use pascals triangle to quickly derive all Binomial coefficients by noting the following:

$$\binom{0}{0}$$
$$\binom{1}{0} \quad \binom{1}{1}$$
$$\binom{2}{0} \quad \binom{2}{1} \quad \binom{2}{2}$$
$$\binom{3}{0} \quad \binom{3}{1} \quad \binom{3}{2} \quad \binom{3}{3}$$
$$\binom{4}{0} \quad \binom{4}{1} \quad \binom{4}{2} \quad \binom{4}{3} \quad \binom{4}{4}$$
$$\binom{5}{0} \quad \binom{5}{1} \quad \binom{5}{2} \quad \binom{5}{3} \quad \binom{5}{4} \quad \binom{5}{5}$$

These coefficients are the following:



And we can see that the rows in the triangle are easily calculated by adding the previous rows elements together as follows:



Source: https://medium.com/i-math/top-10-secrets-of-pascals-triangle-6012ba9c5e23

So, for the example with 3 trials, we can see that the formula for a Binomial random variable, has 1, 3, 3, 1 combinations of outcomes.

The formula for a Binomial Random Variable is given by:

- Binomial random variable: $X \sim \text{Bin}(n,p)$:

$$b(x; n, p) = \binom{n}{k} p^k (1-p)^{n-k}$$

Mean: np; Variance $np(1\text{-}p)$

Example:

A manufacturer of a type of RFID chip similar to BiChip that can be used to track financial information including both fiat (currency created by governments) and cryptocurrencies states that 98% of all installations their chip works as designed and expected.

What is the probability that the chip is successfully implanted in the next 4 out of 5 implantation surgeries?

Using the formula for the Binomial we have:

b(4; 5, 0.98) = $\binom{5}{4}(0.98)^4(0.02)^1$= 5*0.9223*0.02 = 0.0922

What is the probability that none of the next three are unsuccessful?

b(0;3, 0.98) = $\binom{3}{0}(0.98)^0(0.02)^3$ = 1*1*0.000008 = 0.000008

What is the average number of chips that fail in every shipment of 1000 chips?

Using the formula for the mean, with n= 1000 and q = 0.02

1000*0.02 = 20

Multinomial Distribution

Now consider the case where we not just two outcomes, but more than two outcomes. We no longer have a series of repeated Bernoulli trials. We have the following assumptions:

1. Each trial results in k possible outcomes, denoted E_1, E_2, ..., E_k.

2. The probability of each event k remains constant from trial-to-trial and is denoted by p_1, p_2, ..., p_k.

3. The trials are independent.
4. The experiment consists of n repeated trials.

The formula for the Mulitnomial is given by:

- Mulitnomial random variable: X~Mult(k) :

$$f(x_1, x_2, \dots, x_k; p_1, p_2, \dots, p_k, n) = \binom{n}{x_1, x_2, \dots, x_k} p_1^{x_1}, p_2^{x_2}, \dots, p_k^{x_k} = \frac{n!}{x_1! x_2! \dots x_k!},$$

with

$$\sum_{i=1}^{k} x_i = n \text{ and } \sum_{i=1}^{k} p_i = 1$$

Example:

Using the previous example of a chip similar to a BiChip, we are now interested in how long the chip will last once implanted. Previous studies have shown that the distribution of failure times for the chip were:

0 < 12 months	0.3
12 < 36 month	0.5
36 + months	0.2

Among the next 8 chips that are implanted what are the probabilities that:

2 will last less than 12 months
5 will last between 12 and 36 months
1 will last more than 36 months

Using the formula for a Multinomial we get:

$$f(2,5,1;\ 0.3,\ 0.5,\ 0.2) = \frac{8!}{2!5!1!}(0.03)^2(0.05)^5(0.02)^1 = 0.0945$$

Hypergeometric Distribution

Now we consider the case where the trials are not independent, and therefore the probabilities do not stay constant trial to trial. In the case where we have two outcomes we now have the following assumptions:

1. Each trial results in one of two possible outcomes, denoted success (S) or failure (F).

2. The probability of success (S) and therefore for failure (F) does not remains constant from trial-to-trial.

3. The trials are not independent, they are conditional on previous outcomes. (Sampling without replacement).

4. The experiment consists of n repeated trials.

To see how we form the Hypergeometric distribution, we have the case where is a lot (of some product, object or thing) of N (total) units where a are defective. Then we are going to select n units from the lot without replacement. On the first draw, we can see that the probability of selecting a defective unit is a/N. However, on the second draw we can see that because we are not replacing the 1^{st} sampled unit back, the probability of selecting a defective unit in the second draw is now either (a-1)/(N-1) if we selected a defective first or a/(N-1) if the first unit was not defective.

Therefore, we note that there are $\binom{a}{x}$ ways to choose a defective, and there are $\binom{N-a}{n-x}$ ways to choose the non-defective. And there are $\binom{N}{n}$ ways to choose n objects from N total. Which is what we use to derive the hypergeometric distribution.

It can also be shown that when n/N is very small, the binomial is a good approximation to the hypergeometric.

The formula for the Hypergeometic is give by:

- Hypergeometric random variable: X~Hyp(k) :

$$h(x; N, n, k) = P(X = k) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

Mean: n(k/N)

Example:

Using the previous example with the chip like the BiChip, we are given a lot of 100 chips where there are 25 that are defective. What is the probability that in a sample of 10 chosen at random, that 2 are defective?

Using the formula for a Hypergeometric we get:

h(2; 10,25, 100) = $\frac{\binom{25}{2}\binom{75}{8}}{\binom{100}{10}}$ = 0.292

Using a Binomial approximate the same probability. Using the formula for a Binomial we get:

b(2; 10, 0.25) = $\binom{10}{2}(0.25)^2(0.72)^8$= 0.282

Negative Binomial Distribution:

If we are now interested not in a fixed number of Bernoulli trials, but instead we want to wait until a fixed number of successes occurs we use a Negative Binomial Distribution. We have the following assumptions:

1. Each trial results in one of two possible outcomes, denoted success (S) or failure (F).

2. The probability of success (S) remains constant from trial-to-trial and is denoted by p. Write q = 1−p for the constant probability of failure (F).

3. The trials are independent.

4. The experiment consists of x repeated trials until the kth success.

This means we are seeking number of successes in a sequence of independent and identically distributed Bernoulli trials before a specified (non-random) number of successes occurs. This definition is where x is the total number of trials needed to get k successes.

And the formula for the Negative Binomial Distribution is given by:

- Negative binomial random variable: X~Neg(k,p) :

$$b^*(x;k,p) = \binom{x-1}{k-1} p^k(1-p)^{x-k}, x = k, k+1, k+2, \ldots$$

Mean: kp/(1-p)

Example:

We have 7 runs where we are comparing two products A and B. If in previous runs product A is better than B 55% of the time, what is the probability that A is better overall than B in the 7 comparison runs.

First, we define overall as A>B in at least 4 runs. This means we have either that the 4^{th} success happens on the 4^{th}, 5^{th}, 6^{th}, or 7^{th} run.

Next, we use the formula to find each of the individual probabilities using the negative binomial:

b*(4;4,0.55) + b*(5; 4,0.55)+ b*(6;4,0.55) + b*(7;4,0.55) =
0.0915+0.01647+0.1853+0.1668 = 0.6083

Geometric Distribution:

If we consider a series of Bernoulli trials, and if we are interested in which trial the first success will occur, we use the Geometric distribution. We have the following assumptions:

1. Each trial results in one of two possible outcomes, denoted success (S) or failure (F).

2. The probability of success (S) remains constant from trial-to-trial and is denoted by p. Write q = 1−p for the constant probability of failure (F).

3. The trials are independent.

4. The experiment consists of k repeated trials until the 1st success.

This means we are seeking the probability distribution of the number of Bernoulli trials needed to get one success. And the formula for the Geometric Distribution is given by:

- Geometric random variable: $X \sim \text{Geo}(p)$:

$$P(X = k) = (1 - p)^{k-1}p$$

  Mean: $1/p$

Poisson Distribution:

Consider the case where we have an outcome (often failures or errors) occurring in a time interval or in a region in space. This is called a Poisson process. Let us have the following assumptions:

1. The number of events occurring in non-overlapping intervals are independent
2. The probability of exactly one event in a short time interval is ≈ λ * h, where lambda is the average rate of the event
3. The probability of 2 or more events in a short interval h is essentially zero.

If we consider the case where we partition the unit interval into n disjoint subintervals of length 1/n. By condition (2) the probability of one change occurring in one small subinterval is approximately λ × 1/n. By condition (3) the probability of two or more changes is essentially equal to 0. By condition (1) we have a sequence of n Bernoulli trials with probability p = λ × 1/n.

Therefore, we have:

$f(x) = P(X = x) \approx \frac{n!}{x!(n-x)!}\left(\frac{\lambda}{n}\right)^{x}\left(1-\frac{\lambda}{n}\right)^{n-z}$

Letting n → ∞ we get

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

Then, the actual probability distribution is given by a binomial distribution where we have the number of trials is sufficiently larger than the number of successes in those trials. Which is why we use the limit. The formula for a Poisson random variable is given by:

- Poisson random variable: $X \sim \text{Poi}(\lambda)$:

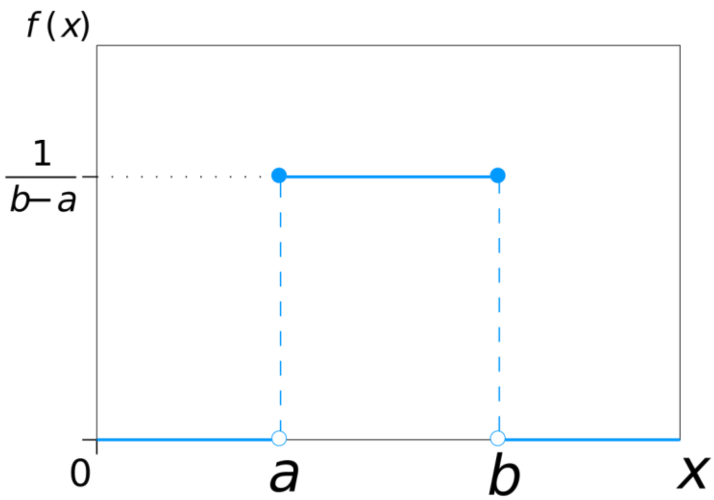
$$p(x;\lambda t) = \frac{e^{-\lambda t}(\lambda t)^{x}}{x!}, x = 0,1,2,\ldots$$

  Mean: λt; Variance λt

Continuous Random Variables

A continuous random variable is a random variable where the data can take infinitely many values.

Uniform Distribution

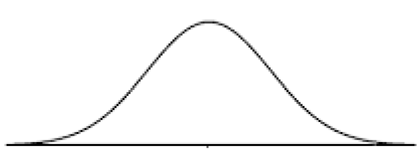


$$F(x; A, B) = \begin{cases} \frac{1}{(B-A)}, & A \leq x \leq B \\ 0, & \text{elsewhere.} \end{cases}$$

Mean = (A+B)/2 and Var = $(B-A)^2/12$

Normal Distribution

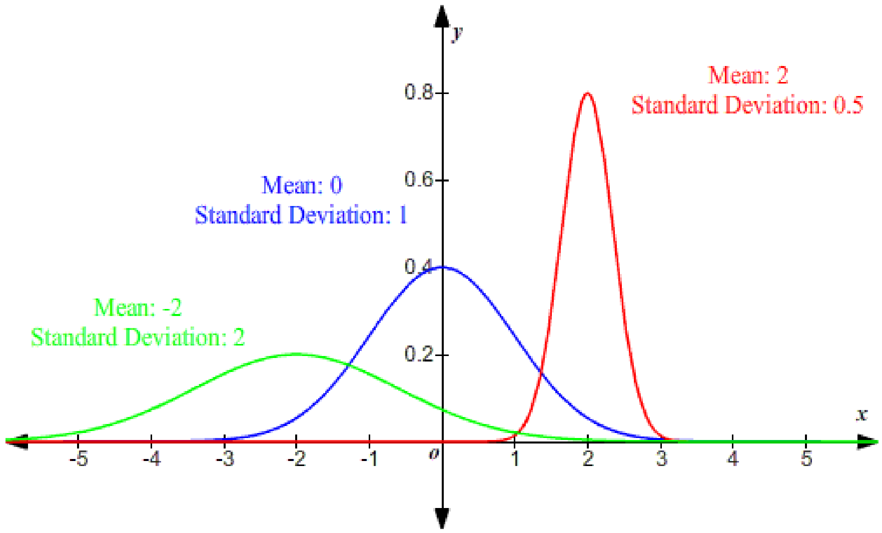
The most important continuous distribution is the Normal Distribution.



$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{for } -\infty < x < \infty$$

Where $X \sim N(\mu, \sigma^2)$

We can see that the mean and variance alter the shape of the distribution

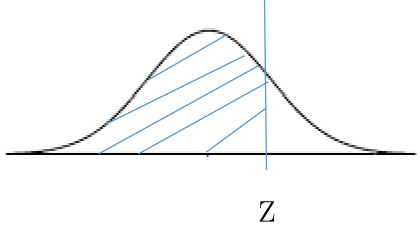


We have the following properties of a normal distribution:

1. The mode, which is the maximum point on the horizontal axis occurs at the mean.
2. The distribution is symmetric through the mean and so the median also occurs at the mean.
3. The curve has points of inflection at one standard deviation from the mean (plus or minus).
   a. It is concave downward if $\mu - \sigma < X < \mu + \sigma$
   b. It is concave upward otherwise.
4. The normal curve approaches the horizontal axis asymptotically as we proceed in either direction away from the mean.
5. The total area under the curve is equal to 1.
6. There is no closed form for this distribution. To get values for a Normal random variable we must standardize. We convert this probability to one involving the N(0, 1) distribution by (i) Subtracting the mean μ (ii) Dividing by the standard deviation σ. Subtracting the mean re-centers the distribution on zero. Dividing by the standard deviation re-scales the distribution so it has standard deviation 1.
   a. Standardization: If X is a normal random variable with mean μ and variance σ^2, then

$$Z = \frac{X - \mu}{\sigma} \text{ has a standard normal distribution}$$

The standard normal tables (A.3 in text) let us read off probabilities of the form $P(Z < z)$.



We can make linear combinations of Normal random variables and they are also Normal random variables.

- If X and Y are two independent normal random variables such that

$$X \sim N(\mu_1, \sigma_1^2) \text{ and } Y \sim N(\mu_2, \sigma_2^2)$$

  then

$$X - Y \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$$

- If X and Y are two independent normal random variables such that

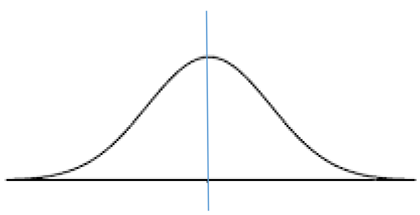
$$X \sim N(\mu_1, \sigma_1^2) \text{ and } Y \sim N(\mu_2, \sigma_2^2)$$

  then

$$X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$
$$aX \sim N(a\mu_1, a^2\sigma_1^2)$$
$$aX + bY \sim N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$$

Example:

Suppose $Z \sim N(0, 1)$, what is $P(Z < 0)$?



We can see that N(0,1) is symmetric around the mean and therefore the $P(Z < 0) = 0.5$

What is P(Z < 1.0)

Using the tables, the column z is the first 2 digits and the header row is the third digit.

z	**0.0**	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	5040	5080	5120	5160	5199	5239	5279	5319	5359
0.1	0.5398	5438	5478	5517	5557	5596	5636	5675	5714	5753
0.2	0.5793	5832	5871	5910	5948	5987	6026	6064	6103	6141
0.3	0.6179	6217	6255	6293	6331	6368	6406	6443	6480	6517
0.4	0.6554	6591	6628	6664	6700	6736	6772	6808	6844	6879
0.5	0.6915	6950	6985	7019	7054	7088	7123	7157	7190	7224
0.6	0.7257	7291	7324	7357	7389	7422	7454	7486	7517	7549
0.7	0.7580	7611	7642	7673	7704	7734	7764	7794	7823	7852
0.8	0.7881	7910	7939	7967	7995	8023	8051	8078	8106	8133
0.9	0.8159	8186	8212	8238	8264	8289	8315	8340	8365	8389
1.0	**0.8413**	8438	8461	8485	8508	8531	8554	8577	8599	8621
1.1	0.8643	8665	8686	8708	8729	8749	8770	8790	8810	8830

From this table we can identify that $P(Z < 1.0) = 0.8413$

If Z ~ N(0, 1) what is $P(Z > 0.92)$? Using the fact that the tables give the area to the left of the Z value we can see that $P(Z > 0.92) = 1 - 0.8212 = 0.1788$

If Z ~ N(0, 1) what is $P(-1.96 < Z < 1.96)$?

Using the tables and the symmetry we can see that $P(-1.96 < Z < 1.96) = 0.95$.

If X ~ N(3500, 5002) and we want to calculate $P(X < 3100)$?

First, we standardize X to be Z. Then we use the tables to find the probability.

Z = (3100 – 3500)/500 = -0.8 and $P(Z < -0.8) = 1 - P(Z < 0.8)$

$= 1 - 0.7881 = 0.2119$

Suppose two robots A and B have been trained to navigate the campus in a race.

X = Time of run for robot A, X ~ N(80, 102)

Y = Time of run for robot B, Y ~ N(78, 132)

On any given day what is the probability that A runs the university race faster than B?

Let D = X − Y be the difference in times of A and B

If A is faster than rat B then D < 0 so we want to know P(D < 0)?

Using the identity $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$ and $X - Y \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$
We see D = X − Y ~ N(80 − 78, 102 + 132) = N(2, 269).

P(D < 0) = P(Z < (0-2)/16.40) = P(Z < −0.122 = 0.45142

Normal approximations to the Binomial and Poisson

Binomial

Under certain conditions we can use the Normal distribution to approximate the Binomial distribution. For large values of n, the distributions of the count X and the sample proportion p are approximately normal.

Then we have the following that will hold:

If X ~ Bin(n, p) then $\mu = np$ and $\sigma^2 = npq$ where q = 1 − p. For large n and p (that is not too small or too large) we have:

X ~ N(np, npq) n > 10 and p ≈ 1/2

OR

n > 30 and p moving away from ½

However, we cannot simply just use a typical normal standardization. We must consider the fact that we are using a continuous distribution to approximate a discrete distribution. This is done using a continuity correction.

$$Z = \frac{X + 0.5 - np}{\sqrt{npq}}$$

Example:

Suppose $X \sim Bin(12, 0.5)$ what is $P(4 \leq X \leq 7)$?

For this distribution we have $\mu = np = 6$ and $\sigma^2 = npq = 3$. We can use a $N(6, 3)$ distribution as an approximation.

$P(4 \leq X \leq 7)$ transforms to $P(3.5 < X < 7.5)$ then we have

$P(3.5 < X < 7.5) = P((3.5 - 6)/\sqrt{3} < Z < (7.5 - 6)/\sqrt{3}) = P(-1.443 < Z < 0.866)$ where $Z \sim N(0, 1) = 0.732$

Poisson

Remember that a Poisson random variable could be thought of as the limiting distribution of a series of Binomial random variables. So as the number of trials increases the Binomial is a good approximation for the Poisson. Also, we can also use the Normal distribution to approximate a Poisson distribution under certain conditions. These conditions have to do with the mean of the distribution.

If $X \sim P(\lambda)$ then we have $\mu = \lambda$ and $\sigma^2 = \lambda$

For large λ (say $\lambda > 20$; though it is better if λ is 1000)

$X \sim N(\lambda, \lambda)$

As with the Binomial approximation we need to use a continuity correction.

$$Z = \frac{X + 0.5 - \lambda}{\sqrt{\lambda}}$$

Example:

A radioactive source emits particles at an average rate of 25 particles per second. What is the probability that in 1 second the count is less than 27 particles?

X = No. of particles emitted in 1s, is $X \sim P(25)$

So, we can use a $N(25, 25)$ as an approximate distribution.

Again, we need to make a continuity correction So $P(X < 27)$ transforms to $P(X < 26.5)$.

$P(X < 26.5) = P((X - 25)/5 < (26.5 - 25)/5) = P(Z < 0.3)$ where $Z \sim N(0, 1)$

$= 0.6179$

Exponential Random Variable and Erlang Random Variables:

- Exponential random variable: $X \sim \text{Exp}(\lambda)$:

$$f(x; \lambda) = \lambda e^{-\lambda x}$$

  Mean: 1/ λ

- Erlang random variable: $X \sim \text{Erl}(k,\lambda)$:

$$f(x; k, \lambda) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!} \text{ for } x, \lambda \geq 0$$

  Mean: k/λ

Additional notes will be added for the lecture on Monday June 5th after the lecture.